Analytical Biochemistry 394 (2009) 177–185

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Proteome analysis of a single zebrafish embryo using three different digestion strategies coupled with liquid chromatography–tandem mass spectrometry

Ying Lin [a], Yan Chen [b], Xiaoxu Yang [b], Dong Xu [b], Songping Liang [a,b,*]

[a] College of Life Science, Peking University, Beijing 100871, People's Republic of China
[b] College of Life Science, Hunan Normal University, Changsha, Hunan 410081, People's Republic of China

ARTICLE INFO

Article history:
Received 4 March 2009
Available online 28 July 2009

Keywords:
Zebrafish
Single embryo
Urea
Sodium deoxycholate
Performic acid
LC–MS/MS

ABSTRACT

Zebrafish is a powerful model to analyze vertebrate embryogenesis and organ development. Although a number of genes have been identified to specify embryonic development processes, only a few large-scale proteomic analyses have been reported in regard to these events to date. Here the total proteins of a single embryo were analyzed by urea-, sodium deoxycholate (SDC)-, and performic acid (PA)-assisted trypsin digestion strategies coupled to capillary liquid chromatography–tandem mass spectrometry (CapLC–MS/MS) identification. In total, 509 and 210 proteins were detected from the embryos at 72 and 120 hours postfertilization (hpf), respectively, with a false identification rate of less than 1%. Approximately 95% of those proteins could be observed by combining the urea- and SDC-assisted digestion strategies, suggesting that these two methods are more effective than the PA-assisted method. Compared with 0.5% SDC, 1% SDC was more effective to identify proteins in zebrafish embryos. In addition, removal of the predominant yolk proteins could significantly improve protein identification efficiency. Our study represents the first overview of the protein expression profile of a single zebrafish embryo at 72 or 120 hpf. More important, this single individual proteome methodology could be applied to multiple development stages of wide-type or mutant embryos, providing a simple and powerful way to further our understanding of embryonic development.

© 2009 Published by Elsevier Inc.

Introduction

Zebrafish (*Danio rerio*) is currently a common and powerful model offering numerous opportunities to understand vertebrate embryogenesis and organ development due to its many advantageous innate qualities (e.g., optical clarity of the embryo, rapid embryonic development, large clutch sizes, ease of manipulation, permeability to small molecules) [1–6]. The immense significance of using zebrafish as a model for the human disease and drug discovery studies has been receiving increasing attention [7–10]. The zebrafish genome has been sequenced and interrogated with regard to gene transcription patterns during various stages of development, and many genes with potential functions in various developmental processes have been identified [4,9,11]. The importance of engaging proteomics in zebrafish study to derive biological value at the protein level was highlighted by a previous review [12]. Surprisingly, large-scale proteomic analysis has rarely been reported [13–18]. The few available proteome studies, however, clearly show that the comparison of protein levels or isoform shifts during different zebrafish development stages can provide crucial information on translational and posttranslational regulation that is undetectable by gene expression analysis alone [16,19].

Most zebrafish proteomic studies have employed gel-based protein separation coupled to mass spectrometry (MS)[1] identification. For example, Tay and coworkers identified 55 unique proteins from 108 protein spots in gel by MS [13], and 348 unique proteins at 72 hours postfertilization (hpf) and 317 unique proteins at 120 hpf were identified using two-dimensional gel electrophoresis (2-D PAGE) and matrix-assisted laser desorption/ionization–tandem time-of-flight (MALDI–TOF/TOF) MS [14]. Although the 2-D PAGE approach has led to important findings, it has obvious limitations. First, low-abundant proteins might not be detected given that the amount of sample that can be loaded on a 2-D gel is limited and

* Corresponding author. Present address: College of Life Sciences, Hunan Normal University, Changsha, Hunan 410081, People's Republic of China. Fax: +86 731 886 1304.
E-mail address: liangsp@hunnu.edu.cn (S. Liang).

[1] *Abbreviations used:* MS, mass spectrometry; hpf, hours postfertilization; 2-D PAGE, two-dimensional gel electrophoresis; MALDI–TOF/TOF, matrix-assisted laser desorption/ionization–tandem time-of-flight; 1-D, one-dimensional; LC, liquid chromatography; MS/MS, tandem MS; PA, performic acid; SDC, sodium deoxycholate; CapLC, capillary LC; DTT, dithiothreitol; IAA, iodoacetamide; TFA, trifluoroacetic acid; PMSF, phenylmethylsulfonyl fluoride; HCCA, α-cyano-4-hydroxycinnamic acid; SDS, sodium dodecyl sulfate; HPLC, high-performance LC; MeOH, methanol; ACN, acetonitrile; TPCK, L-(tosylamido 2-phenyl)ethyl chloromethyl ketone; KF, Potassium fluoride; IPI, International Protein Index; GO, gene ontology; GRAVY, grand average hydropathy; ER, endoplasmic reticulum; MW, molecular weight.

0003-2697/$ - see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.ab.2009.07.034

the cellular concentrations of proteins vary widely [19,20]. Second, proteins with certain properties, such as high hydrophobicity, are difficult or even impossible to be resolved on a 2-D gel [20,21]. Third, the high abundance of yolk proteins in the early zebrafish embryos interferes with the application of 2-D PAGE. To avoid these obstacles, shotgun proteomics provides a good alternative approach in which the protein mixture is first digested into more soluble fragments and then the resulting peptide mixture is separated by various means and analyzed by MS [22–24].

Shotgun proteomic analysis involves solution-based digestion that leads to proteolytic degradation of cell or tissue lysates or subfractions, followed by analysis of the resulting complex peptide mixtures by one-dimensional (1-D) or 2-D liquid chromatography (LC) coupled with tandem MS (MS/MS) [25–28]. Several studies have shown that this approach is capable of detecting proteins over a very wide dynamic range of concentrations [29–32]. The digestion efficiency of proteins directly affects the efficiency of protein identification. Incomplete digestion retards the identification of proteins. Effective digestion of proteins, particularly the hydrophobic integral membrane proteins, often requires enhancement of their denaturation and solubilization. Therefore, the denaturation and solubilization of proteins are not only the initial step but also the critical step. To effectively identify the proteins in a complex mixture, many different methods designed to improve digestion efficiency by enhancing the denaturation and solubilization of proteins have been reported, including thermal denaturation [33], chemical denaturation (detergents or organic solvents) [34–36], and microwave irradiation [37,38].

Urea is a chaotropic agent that can bind proteins and destroy native interactions. It actively participates in the unfolding process of proteins without interfering with peptide analysis by standard LC–MS/MS because it does not bind to ion exchange or reversed-phase resins and, thus, is readily removed before the peptides are eluted from the column [39]. With an amphipathic characteristic, sodium dodecyl sulfate (SDC) is able to disrupt hydrophobic interactions to improve the unfolding and solubilization process of proteins, exposing more cleavage sites to tryptic digestion [35,40]. Performic acid (PA) completely oxidizes methionine to methionine sulfone and cysteine to cysteic acid, increasing the hydrophilicity of the oxidized proteins, especially for the transmembrane proteins, and thereby increasing the accessibility for proteolytic attack [41]. Our previous work has demonstrated that sodium deoxycholate (SDC) and PA could significantly improve the efficiency of the identification of membrane proteins [35,40]. In the current study, we carried out a proteome analysis of a single zebrafish embryo based on urea-, SDC-, and PA-assisted trypsin digestion strategies coupled to capillary LC (CapLC)–MS/MS identification. Our study represents the first overview of the protein expression profiles of single zebrafish embryo at two different developmental stages, 72 and 120 hpf, and comparing these two profiles would further our knowledge in zebrafish embryogenesis.

Materials and methods

Materials

Trypsin (proteomic sequencing grade), dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), phenylmethylsulfonyl fluoride (PMSF), SDC, and α-cyano-4-hydroxycinnamic acid (HCCA) were purchased from Sigma–Aldrich (St. Louis, MO, USA). Acrylamide, bisacrylamide, thiourea, urea, glycine, Tris, and sodium dodecyl sulfate (SDS) were obtained from Amresco (Solon, OH, USA). High-performance LC (HPLC)-grade formic acid, acetone, methanol (MeOH), and acetonitrile (ACN) were obtained from Fisher Scientific Canada (Edmonton, Canada). L-(Tosylamido 2-phenyl)ethyl chloromethyl ketone (TPCK)-modified porcine trypsin was purchased from Promega (Madison, WI, USA). Water was obtained from a Milli-Q Plus purification system (Millipore, Bedford, MA, USA). All other reagents were domestic products of the highest grade available.

Embryo maintenance and sample preparation

Zebrafish maintenance and embryo collection were carried out as described previously [42]. Briefly, embryos were grown until 72 or 120 hpf and collected. To evaluate the effect of the deyolking process, the embryo at 72 hpf was deyolked with deyolking buffer (½ Ginzburg Fish Ringer) without calcium [43,44]. Subsequently, a single embryo was immersed in liquid nitrogen and then lysed in 10 μl of buffer containing 8 M urea, 25 mM NH_4HCO_3, 5 mM Na_3PO_4, 1 mM potassium fluoride (KF), 1 mM sodium orthovanadate, and 1 mM PMSF at pH 8.2. This sample was then sonicated with a KQ 5200DE sonicator (400 W, 40 kHz, Kunshan Sonicator, Kunshan, China) in an icewater bath using four 10-s bursts, with a 1-min rest between bursts, followed by centrifuging at 14,000*g* for 30 min at 4 °C to pellet cellular debris. Protein concentration was determined by a BCA protein assay kit according to the manual (Beyotime Biotechnology, Jiangsa, China). The samples were stored at −80 °C for later analysis.

PA-assisted trypsin digestion

A 6% PA solution was prepared with 200 μl of H_2O_2, 60 μl of HCOOH, and 740 μl of water and heated at 50 °C for 3 min. A 10-μl sample was first dissolved in formic acid and then stepwise diluted with water and the 6% PA solution. The final concentration of both H_2O_2 and formic acid was 3% (v/v). The mixture was incubated on ice for 4 h, followed by the addition of 2 volumes of cold Milli-Q water, and then lyophilized. After the oxidized sample was dissolved in 40 μl of 50 mM NH_4HCO_3, TPCK-modified porcine trypsin (Promega) was added to a final trypsin/protein ratio of 1:40 (w/w), and finally the mixture was incubated for 14 h at 37 °C.

Urea-assisted trypsin digestion

The reduction of disulfide bonds and thiol group alkylation were carried out as described previously [45]. Briefly, a 10-μl sample was reduced with DTT (2 mM final concentration) for 1 h at 56 °C. Free thiol groups were blocked by adding a double volume of IAA (4 mM final concentration) and incubating for 1 h at room temperature in the dark. After urea was added to a final concentration of 2 M, tryptic digestion was performed immediately for 14 h at 37 °C using a 1:40 (w/w) trypsin/protein ratio in a 40-μl reaction volume. The resulting solution was lyophilized.

SDC-assisted trypsin digestion

A 10-μl sample was reduced by 2 mM DTT at 56 °C for 1 h and alkylated by 4 mM IAA at room temperature in the dark for 1 h. Tryptic digestion (1:40 [w/w] enzyme/protein ratio) was performed separately at 37 °C for 14 h in 40 μl of 50 mM NH_4HCO_3 solution (pH 8.0) containing 0.5 or 1.0% SDC. To quench the enzymatic reactions and remove SDC prior to mass spectrometric analysis, the solutions were acidified by adding TFA (0.1% final concentration) and centrifuged at 15,000*g* for 15 min.

CapLC–MS/MS analysis and protein identification

The resultant digest was completely dried by a SpeedVac (Thermo Savant, New York, NY, USA) and dissolved in 30 μl of 0.1% formic acid. Then it was analyzed by an Agilent 1200 capillary system

(Agilent, Waldbronn, Germany) coupled to an ion trap mass spectrometer (HCTultra, Bruker Daltonics, Bremen, Germany) with an electrospray ionization source. Before separation, digest was desalted and concentrated with 0.1% formic acid at a flow rate of 20 μl/min on a short C18 precolumn (Zorbax SB, 500 μm i.d., 3.5 cm, Agilent) connected in the front of an analytical capillary C18 PepMap column. When separating peptides on the capillary column (180 μm i.d., 15 cm, LC Packings, Sunnyvale, CA, USA), the flow rate of elution solution was 3 μl/min, which was generated by a cap-flow splitter cartridge (3/500) from an initial pump flow rate of 500 μl/min. The injection volume for each sample was 30 μl in all experiments. For peptide elution, the following solvents were used: solvent A (98% H_2O, 1.9% ACN, and 0.1% formic acid) and solvent B (95% ACN, 4.9% H_2O, and 0.1% formic acid). The eluting gradients used to separate tryptic digests were 5–35% solvent B in 147 min, 35–80% solvent B in 15 min, followed by 80% solvent B in 10 min and then to 5% solvent B in 10 min. The mass spectrometer was operated in the positive ion mode at a 4000-V capillary voltage. The nebulizer pressure was 10 psi. The flow rate of drying gas was 5 μl/min. The temperature of drying gas was 300 °C. The full MS scan mode was standard enhanced (m/z 350–1600). The five most abundant ions detected in each MS scan were selected for collision-induced dissociation (CID) with 1.0 V collision energy using the data-dependent MS/MS mode over the m/z range of 100–2000. System control and data collection were made by Esquire Control software (version 6.0, Bruker Daltonics).

Data analysis and bioinformatics

Raw spectrum data were processed and Mascot-compatible files were created using Data Analysis 3.4 software (Bruker Daltonics) with the following parameters: compound threshold, 10,000; maximum number of compounds, 100,000; retention time window, 1.0 min. Searches were performed using Mascot software (version 2.1, Matrix Science, London, UK), and the International Protein Index (IPI) zebrafish database (version 3.41, 44,226 protein sequences, http://www.ebi.ac.uk) was used for peptide and protein identification [46]. Search parameters were set as follows: enzyme, trypsin; allowance of up to one missed cleavage; mass tolerance, 2.0 Da; MS/MS mass tolerance, 0.8 Da; fixed modification parameter, carbamidomethylation (C); variable modification parameters, oxidation (at Met). For the PA-treated samples, cysteic acid of cysteine residues and oxidization of methionine to methionine sulfone were set as fixed modifications and H,W oxidation was set as a variable modification. Proteins were generally identified on the basis

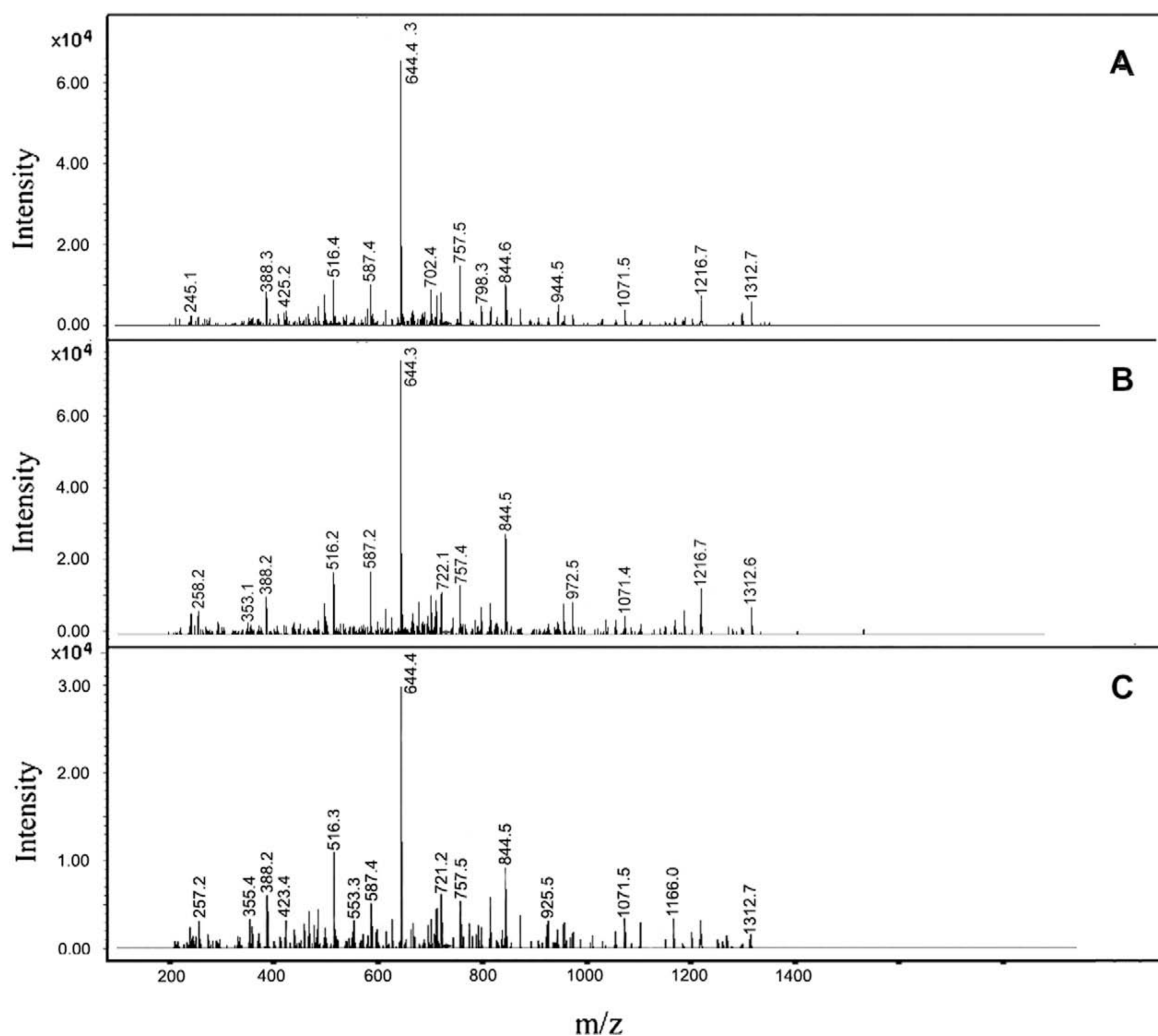

Fig. 1. LC–MS/MS spectra of the tryptic peptide IEDEQSLGAQLQK (precursor m/z 729.9, 2+) identified in the samples prepared by three different digestion methods: (A) urea-assisted tryptic digestion; (B) SDC-assisted tryptic digestion; (C) PA-assisted digestion.

of two or more peptides whose ion scores exceeded the threshold of $P < 0.05$, indicating identification at the 95% confidence level. If proteins were identified by a single peptide, the spectrum was inspected manually. For a protein to be confirmed, the assignment needed to be based on four or more y- or b-series ions (e.g., y4, y5, y6, y7). A Perl Script was written in-house to parse Mascot output files (html files) into XML files suitable for subsequent data analysis. For each of the digestion strategies, five biological replicate samples were analyzed. False-positive rates were calculated as described previously [47]. In brief, the exported mass spectra were searched using Mascot against a composite database consisting of the zebrafish IPI and a duplicate of the same database in which the amino acid sequence of each protein entry was reversed. Protein hits up to an accumulated false-positive rate of 1% were considered as true-positive protein identifications. All identified proteins had IPI database accession numbers, and many of these proteins had assigned Gene Ontology (GO) numbers [46] that were used to retrieve the protein information in the database. The grand average hydropathy (GRAVY) value for identified proteins was calculated using ProtParam software [48] available at http://cn.expasy.org.

Results

Comparison of three trypsin digestion strategies

In this experiment, a single zebrafish embryo at 72 hpf was first deyolked with deyolking buffer without calcium and was then lysed and sonicated. After the samples were solubilized separately and digested in 2 M urea, 3% PA, or 1% SDC, the proteins were identified by CapLC–MS/MS (model HCT; Bruker). The results of these protocols were subjected to Mascot search, and the identification efficiency was evaluated based primarily on the total number of the identified proteins.

Fig. 1 shows three MS/MS spectra of the same tryptic peptide (IEDEQSLGAQLQK, precursor m/z 729.9, 2+) from the three digests of zebrafish embryos at 72 hpf. Fig. 1A and C show the spectra of the peptides from the urea and PA digestion strategies, respectively, and Fig. 1B shows the spectrum of the peptide from the SDC method. It was found that, compared with Fig. 1A and C, the peptide detected in SDC digest has a larger number of y-fragment ions with an excellent signal/noise ratio (Fig. 1B). This result indicates that the SDC method is able to efficiently digest proteins and facilitate protein identification from zebrafish embryos, allowing large-scale proteome analysis of zebrafish larvae.

Five replicates of single zebrafish embryo proteome were subjected to each of the three digestion methods. We combined data from these replicates and selected only those proteins with a false identification rate of less than 1% (the detailed data analysis method is described in Materials and methods and in Ref. [47]). In total, proteome analysis using the three digestion protocols led to the identification of 509 proteins at 72 hpf (see Supplemental Tables 1 and 2 in the supplementary material). The total numbers of nonredundant proteins identified based on digestion in 2 M urea, 1% SDC, and 3% PA were 379, 378, and 181, respectively. Among these proteins, 121 (or 23.8%) proteins were found in all three digests, whereas 86 (or 16.9%), 89 (or 17.5%), and 26 (or 5.1%) proteins were observed uniquely in the urea-, SDC-, and PA-assisted digests, respectively (Fig. 2A). Of the 509 proteins, 483 (or 94.9%) could be identified using a combination of the urea- and SDC-assisted tryptic digestion strategies.

To further optimize the concentration of SDC to increase the efficiencies of proteolysis and protein identification, we used 0.5 and 1% SDC to digest single zebrafish larva at the 120-hpf stage. As shown in Fig. 2B, 210 proteins with a false rate of less than 1%

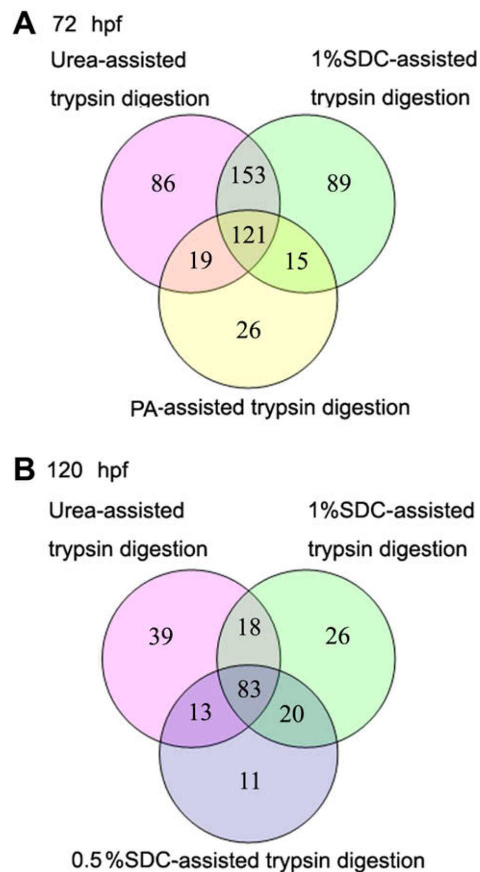

Fig. 2. Comparison of protein identification results from three digestion methods. (A) In total, 509 proteins were identified in the single zebrafish embryo proteome at 72 hpf (379, 378, and 181 proteins from the urea-, SDC-, and PA-assisted digests, respectively). (B) In total, 210 proteins were identified at 120 hpf (153, 147, and 127 proteins from the urea-, 0.5% SDC-, and 1% SDC-assisted digests, respectively).

were identified by 2 M urea, 0.5% SDC, and 1% SDC digestion strategies in total (see Supplemental Table 3). Among those proteins, 127 and 147 proteins were identified in the 0.5 and 1% SDC digests, respectively, with 24 (or 14.0%) and 44 (or 25.7%) of these proteins being observed exclusively in their corresponding digest (Fig. 2B). These findings suggest that 1% SDC is more effective than 0.5% SDC in the protein identification from zebrafish larvae.

Effects of deyolking

A major obstacle in applying proteomics to analyze zebrafish embryos is the high proportion of yolk proteins in early zebrafish larvae [43]. To evaluate the effects of deyolking, we compared the protein profile of whole (or undeyolked) embryo with that of a deyolked one under the same experimental conditions. As shown in Fig. 3A, 336 and 201 proteins were identified in the deyolked and undeyolked embryos, respectively, when urea-assisted digestion was employed. In total, 179 unique proteins were identified exclusively in the deyolked larva, whereas only 44 unique proteins were detected distinctively in the undeyolked larva. The subcellular localization of the proteins identified in the deyolked and undeyolked embryos is shown in Fig. 3B. Obviously, the deyolking method enabled us to identify more proteins, especially those with currently unknown localization (114 in deyolked and 54 in undeyolked), nuclear proteins (66 in deyolked and 34 in undeyolked), and ribosome proteins (19 in deyolked and 7 in undeyolked). Fig. 3C represents the biological process classification of the proteins identified in the deyolked and undeyolked embryos at

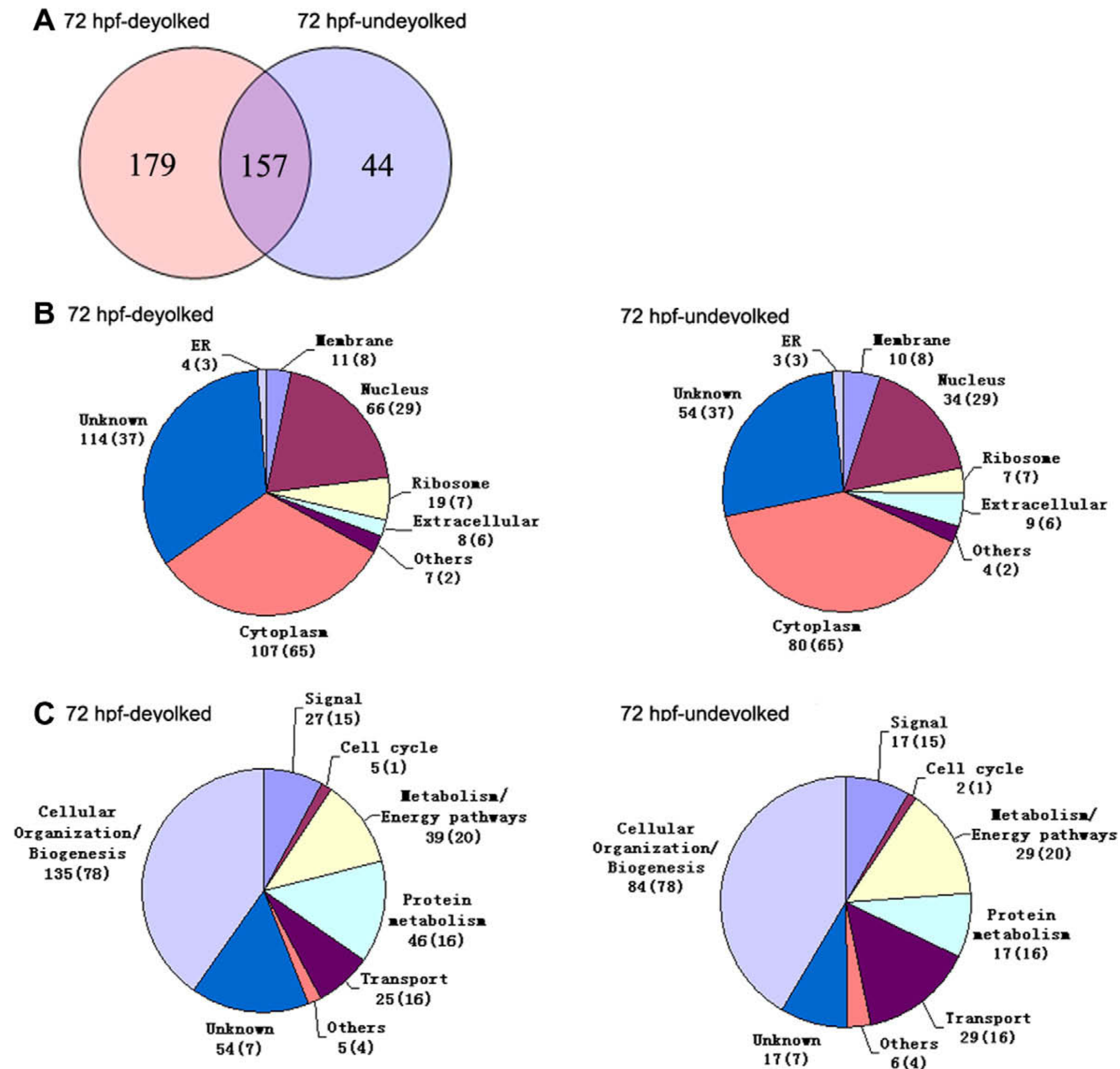

Fig. 3. Comparison of protein identification results between the deyolked sample and the untreated one at 72 hpf in the urea method: (A) overlap between the identified proteins; (B) subcellular localizations; (C) biological processes of the proteins. The numbers under the name of each group indicate the proteins numbers identified from the specific sample. The numbers of overlapped proteins are given in parentheses.

72 hpf. The proteins involved in protein metabolism, signal, cell cycle, and cellular organization/biogenesis are relatively enriched in the deyolked sample (Fig. 3C). In addition, SDC- and PA-assistant digests were also employed to analyze zebrafish larvae at the 72-hpf stage, and similar phenomena were observed (see Supplemental Fig. 1). These results indicate that removal of the predominant yolk proteins benefits the identification of zebrafish embryonic proteins.

Physicochemical characterization

The physicochemical properties (predicted p*I*, molecular weight [MW], and GRAVY) of the proteins identified by three different methods were analyzed. As shown in Fig. 4A and B, there was no significant difference in the average p*I* and the average MW among the three methods. GRAVY is a commonly used parameter to gauge the hydropathy of proteins or peptides. It is generally accepted that proteins with a negative GRAVY value are classified as hydrophilic proteins, whereas those with a positive value are classified as hydrophobic proteins [49]. In our experiments, we categorized all of the identified proteins into three groups according to their GRAVY index: the proteins with GRAVY index lower than −0.5 were considered as strongly hydrophilic, those with GRAVY index between −0.5 and 0 were considered as mildly hydrophilic, and those with GRAVY index higher than 0 were considered as hydrophobic. The GRAVY index distribution for the proteins found in each digest is summarized in Fig. 4C. Hydrophilic proteins were found to be the major part in each digestion (Fig. 4C). In total, 343 (or 90.5%), 335 (or 88.6%), and 159 (or 87.8%) proteins were identified in urea-, SDC-, and PA-assisted digests, respectively. The relatively smaller number of the proteins identified as hydrophobic probably results from the isolation procedure. Most hydrophobic proteins locate in membranes. In general, to analyze these proteins by proteomic methods, the membrane components should be isolated by specific processes. However, in our protocol, we did not employ such steps. Therefore, we would not expect to find a large number of hydrophobic proteins in our experiments.

Differential protein expression during development

Zebrafish embryonic development between the hatching period (72 hpf) and the larval stage (120 hpf) is characterized by rapid maturation of primal organs [14]. Our approach yielded 256 unique proteins at 72 hpf and 210 unique proteins at 120 hpf from

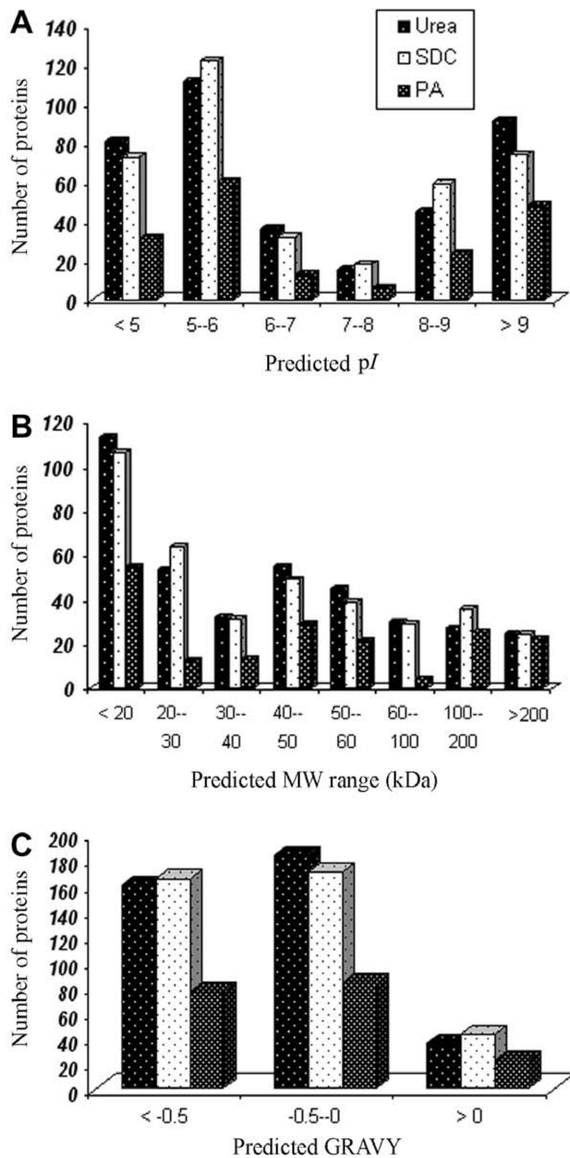

Fig. 4. Comparison of proteins identified by three different methods based on physicochemical parameters from the 72-hpf embryos: (A) predicted p*I*; (B) predicted MW; (C) predicted GRAVY.

undeyolked embryos under the same experimental conditions (see Fig. 5A and Supplemental Table 4). Interestingly, a large fraction of the proteins, 120 of 330 (or 36.4% of the total identified proteins), were identified exclusively at 72 hpf (Fig. 5A), including myeloid leukemia-associated SET translocation protein (Setb) and brain-type fatty acid-binding protein (Fabp7), which play an important role in the early nervous system development [50–52]. On the other hand, 73 proteins (or 22.4% of the total identified proteins) were identified exclusively at 120 hpf, including parvalbumin 4, fast muscle troponin I, and TnC. The former is a calcium-binding protein and is correlated with the maturation of locomotor activity in developing striated muscle [53]. The latter two are involved in muscle development [54,55].

The cellular distribution of the proteins identified at the two stages is shown in Fig. 5B. Compared with 120 hpf larva, more proteins with clear subcellular localizations (e.g., cytoplasm, extracellular, ribosome, nucleus, membrane, endoplasmic reticulum [ER]) were identified from 72 hpf larva, 188 to 142 proteins (Fig. 5B). Furthermore, we classified the identified proteins according to biological processes. As shown in Fig. 5C, the proteins showed a similar distribution, with 40 to 44% proteins annotated to cellular organization/biogenesis, 13 to 15% annotated to metabolism/energy pathways, 12 to 14% annotated to transport, approximately 10% annotated to signal, 7 to 9% annotated to protein metabolism, approximately 2% annotated to cell cycle, 1 to 2% annotated to others, and 8 to 10% annotated to unknown functions. However, we also noticed that the proteins involved in metabolism/energy pathways and protein metabolism were relatively more abundant at 72 hpf than at 120 hpf.

Discussion

In a typical shotgun proteomic analysis, proteins are digested with proteases such as trypsin into a complex peptide mixture and then analyzed by MS. Digestion is the key step; therefore, effective digestion could significantly increase the efficiency of protein identification, particularly for hydrophobic integral membrane proteins. In this work, we compared the effects of three different digestion strategies on the protein identification of a single zebrafish embryo at 72 hpf. As a result, of the 509 proteins identified from the embryo at 72 hpf, 378 (or 74.3%) proteins were identified after SDC-assisted trypsin digestion. Another 105 (or 20.6%) proteins were identified after urea-assisted trypsin digestion. The use of PA resulted in the identification of only the remaining 26 (or 5.1%) proteins. These results demonstrate that, compared with the other two methods, the SDC-assisted digestion protocol provides the largest proteome coverage of the zebrafish embryo, a meaningful mixture consisting of various proteins with a wide range of concentrations and with different susceptibilities to tryptic proteolysis, whereas the PA-assisted digestion protocol is obviously unsuitable for the analysis of zebrafish embryo proteins. In addition, compared with 0.5% SDC, 1% SDC could identify even more unique proteins through assisting protein digestion. This result is consistent with our previous report that SDC can be used in a high concentration in solubilizing proteins because it is well compatible with trypsin and dilution is unnecessary prior to trypsin digestion [40].

The presence of the high proportion of yolk proteins in early zebrafish embryo retards the application of proteomics [43]. To date, only three groups have reported the proteomic analysis of zebrafish larva at the 72- or 120-hpf stage without deyolking steps [15,56,57]. All of these groups employed gel-based protein separation, and one group showed the protein spots pattern only in gel without identification [56]. Therefore, it remains unclear whether those previous proteome analyses without deyolking provide adequate information regarding the proteins involved in embryo development. Our data showed that, in the urea method, 336 unique proteins were identified in the deyolked embryo, among which 179 (or 47.1% of the total identified proteins from both embryos) proteins could not be identified in the undeyolked embryo at the same stage. Consistently, the deyolking process also enabled us to identify more proteins in the SDC and PA methods. These results indicate that the removal of the predominant yolk proteins is essential before the embryonic proteins can be identified.

In the current study, a larger number of unique proteins were identified at 72 hpf (256 proteins) than at 120 hpf (210 proteins) from undeyolked wide-type larva. We analyzed the physicochemical and biological characteristics of the identified proteins, which are summarized in Supplemental Table 4. When the proteins were sorted according to biological process, we found a similar distribution of functional categories at both developmental stages (Fig. 5C). In total, 45 proteins related to metabolism and energy production were observed at both stages, whereas 18 and 7 proteins were found uniquely at 72 and 120 hpf, respectively (Fig. 5C). The pro-

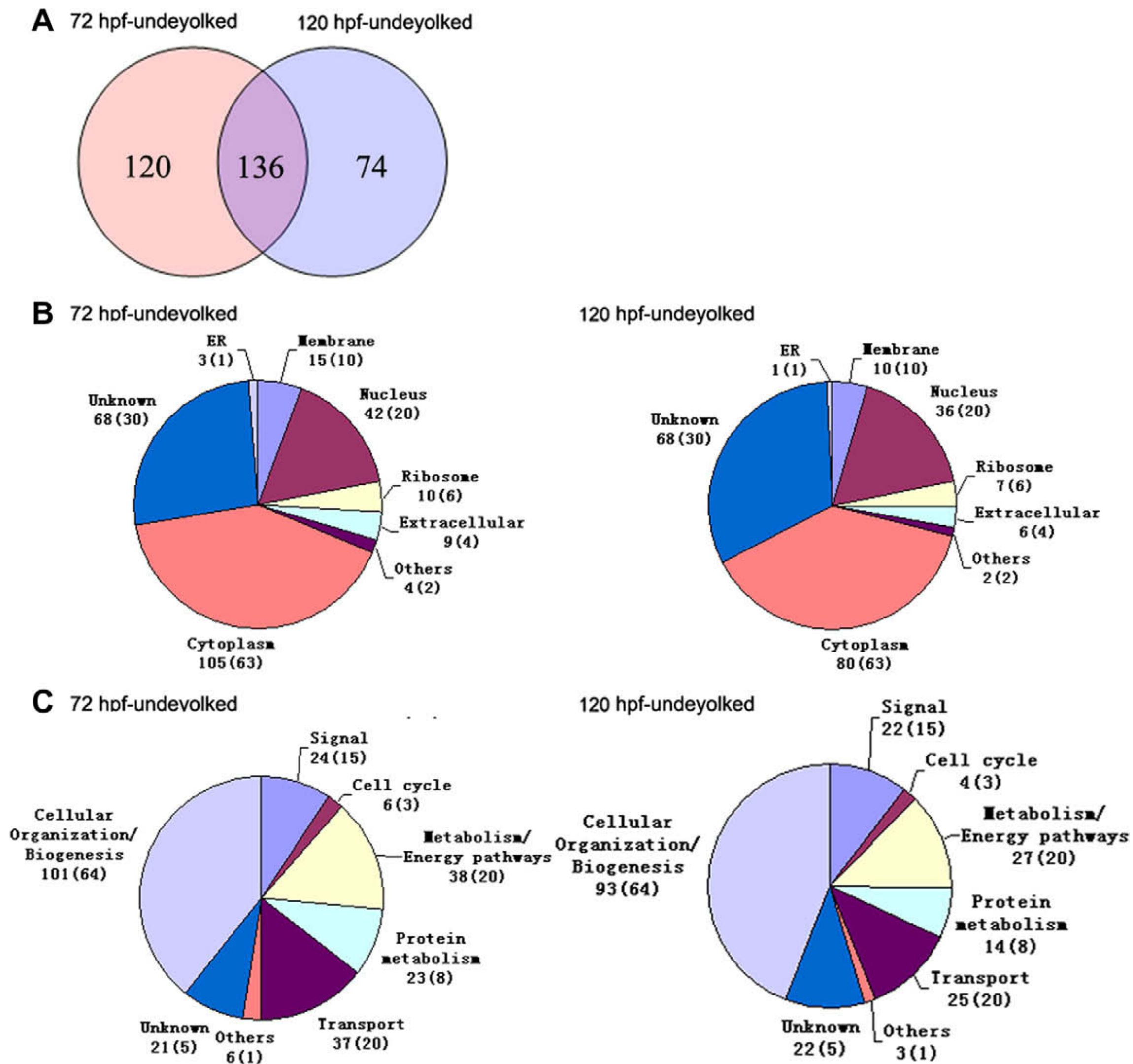

Fig. 5. Comparison of identified proteins from the single zebrafish proteome at 72 and 120 hpf: (A) overlap between 72 and 120 hpf; (B and C) subcellular localizations (B) and biological processes (C) of the proteins at 72 and 120 hpf. The numbers under the name of each group indicate the proteins numbers identified from the specific larval stages. The numbers of overlapped proteins between the two examined stages are given in parentheses.

teins involved in protein metabolism were relatively more abundant at 72 hpf (23 proteins) than at 120 hpf (14 proteins). These results may reflect the developmental differences between the two larval stages. In other words, this is consistent with the faster synthesis of cellular proteins during organismal growth at 72 hpf compared with that at 120 hpf [14].

The proteins with annotations identified at both stages included structural proteins (e.g., myosin, actin, cytokeratin, tubulin), 40S ribosomal proteins (e.g., 40S ribosomal protein S28, 40S ribosomal protein S8), cell cycle proteins (e.g., ef1a, zgc:162932), and all of the known forms of the yolk proteins (e.g., vitellogenin 1–7, Vg1). Many proteins associated with the tissue and organ structures were represented at both stages, including troponin C, myosin, heavy polypeptide 1/2, and actin alpha 1/2, which are specific to skeletal muscle; creatine kinase, which is specific to brain; and beta B1-crystallin, which is specific to lens. Apart from the proteins identified at both stages, 120 and 74 proteins were observed uniquely at 72 and 120 hpf, respectively (see Fig. 5A and Supplemental Table 4). These differentially expressed proteins can then be used to distinguish these two stages. For example, matrilin-4, apolipoprotein eb precursor, and brain-type fatty acid-binding protein were identified only at 72 hpf, whereas cdc48, parvalbumin 4, and translationally controlled tumor protein were unique to 120 hpf. In addition, some differentially expressed proteins reflect differences in differentiation states of several tissues. For example, parvalbumin 4, fast muscle troponin I, and TnC, three proteins involved in muscle maturation [53–55], were observed distinctively at 120 hpf, indicating that skeletal muscles become more and more robust during the developmental period from 72 to 120 hpf. H-FABP, a protein involved in cardiomyocyte growth and differentiation in neonatal hearts [58], was detected only at 120 hpf. Moreover, Setb and Fabp7, two proteins involved in early neurogenesis [50,52], were identified exclusively at 72 hpf. All of the above observations are consistent with the early organogenesis and body patterning known to occur at 72 hpf and with the late organogenesis and late tissue differentiation known to occur between 72 and 120 hpf [59].

There are at least two important applications for our methodology. First, a large number of genes have been reported to regulate zebrafish development processes, but only a few large-scale proteomic analyses have been employed to investigate these events to date. The data here reflect only the proteins present at the two examined development stages. However, when more early embryonic stages were examined by this methodology, the pro-

teins contributing specifically to certain developmental events are expected to be resolved and then the zebrafish embryonic development processes might be reconstructed at the view of protein level. Second, the single embryo resolution enables us to compare the proteome between wild type and mutants. This will further our understanding of how the mutant protein impacts the embryonic proteome and how the mutant protein executes its biological functions, and knowing how the proteome changes in response to loss of a particular protein may provide insight into the molecular mechanism.

Conclusions

To our knowledge, this is the first report of a proteome study of a single zebrafish embryo by shotgun proteomics. The samples were subjected separately to three different trypsin digestions in urea, SDC, and PA. The resultant digests were analyzed individually by CapLC–MS/MS. In total, 509 unique proteins were identified at 72 hpf and 210 unique proteins were identified at 120 hpf. Approximately 95% of all the detected proteins were generated by combining the SDC- and urea-assisted trypsin digestions. Digestion in the presence of 1% SDC led to broader proteome coverage of a single zebrafish embryo than that in the presence of 0.5% SDC. The proteins involved in energy production and protein metabolism were relatively more abundant at 72 hpf than at 120 hpf, possibly reflecting the faster synthesis of cellular proteins for organismal growth at 72 hpf compared with that at 120 hpf [14]. This single individual proteome methodology, when applied to multiple embryonic stages or to compare the proteomes of individual wild-type and mutant embryos, is expected to further our understanding of embryonic development and how a particular protein impacts the protein profile.

Acknowledgments

We thank Xianchun Wang for critically reviewing the manuscript. This work was supported by National Basic Research Program of China (973 Program, Grant 2007CB914203) and the National Natural Science Foundation of China (Grant 30770437).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2009.07.034.

References

[1] C.B. Kimmel, Genetics and early development of zebrafish, Trends Genet. 5 (1989) 283–288.

[2] W. Driever, D. Stemple, A. Schier, L. Solnica-Krezel, Zebrafish: genetic tools for studying vertebrate development, Trends Genet. 10 (1994) 152–159.

[3] J.S. Eisen, Zebrafish make a big splash, Cell 87 (1996) 969–977.

[4] M.C. Fishman, Zebrafish genetics: the enigma of arrival, Proc. Natl. Acad. Sci. USA 96 (1999) 10554–10556.

[5] D.J. Grunwald, J.S. Eisen, Headwaters of the zebrafish: emergence of a new model vertebrate, Nat. Rev. Genet. 3 (2002) 717–724.

[6] L.I. Zon, R.T. Peterson, In vivo drug discovery in the zebrafish, Nat. Rev. Drug Discov. 4 (2005) 35–44.

[7] F.B. Pichler, S. Laurenson, L.C. Williams, A. Dodd, B.R. Copp, D.R. Love, Chemical discovery and global gene expression analysis in zebrafish, Nat. Biotechnol. 21 (2003) 879–883.

[8] G. Golling, A. Amsterdam, Z. Sun, M. Antonelli, E. Maldonado, W. Chen, S. Burgess, M. Haldi, K. Artzt, S. Farrington, S.Y. Lin, R.M. Nissen, N. Hopkins, Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development, Nat. Genet. 31 (2002) 135–140.

[9] P. Haffter, M. Granato, M. Brand, M.C. Mullins, M. Hammerschmidt, D.A. Kane, J. Odenthal, F.J. van Eeden, Y.J. Jiang, C.P. Heisenberg, R.N. Kelsh, M. Furutani-Seiki, E. Vogelsang, D. Beuchle, U. Schach, C. Fabian, C. Nusslein-Volhard, The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*, Development 123 (1996) 1–36.

[10] R.T. Peterson, S.Y. Shaw, T.A. Peterson, D.J. Milan, T.P. Zhong, S.L. Schreiber, C.A. MacRae, M.C. Fishman, Chemical suppression of a genetic mutation in a zebrafish model of aortic coarctation, Nat. Biotechnol. 22 (2004) 595–599.

[11] W.L. Stanford, J.B. Cohn, S.P. Cordes, Gene-trap mutagenesis: past, present, and beyond, Nat. Rev. Genet. 2 (2001) 756–768.

[12] D.R. Love, F.B. Pichler, A. Dodd, B.R. Copp, D.R. Greenwood, Technology for high-throughput screens: the present and future using zebrafish, Curr. Opin. Biotechnol. 15 (2004) 564–571.

[13] T.L. Tay, Q. Lin, T.K. Seow, K.H. Tan, C.L. Hew, Z. Gong, Proteomic analysis of protein profiles during early development of the zebrafish, *Danio rerio*, Proteomics 6 (2006) 3176–3188.

[14] M.B. Lucitt, T.S. Price, A. Pizarro, W. Wu, A.K. Yocum, C. Seiler, M.A. Pack, I.A. Blair, G.A. Fitzgerald, T. Grosser, Analysis of the zebrafish proteome during embryonic development, Mol. Cell. Proteomics 7 (2008) 981–994.

[15] S. Lemeer, M.W. Pinkse, S. Mohammed, B. van Breukelen, J. den Hertog, M. Slijper, A.J. Heck, Online automated in vivo zebrafish phosphoproteomics: from large-scale analysis down to a single embryo, J. Proteome Res. 7 (2008) 1555–1564.

[16] V. Link, L. Carvalho, I. Castanon, P. Stockinger, A. Shevchenko, C.P. Heisenberg, Identification of regulators of germ layer morphogenesis using proteomics in zebrafish, J. Cell Sci. 119 (2006) 2073–2083.

[17] N. Wang, L. Mackenzie, A.G. De Souza, H. Zhong, G. Goss, L. Li, Proteome profile of cytosolic component of zebrafish liver generated by LC–ESI MS/MS combined with trypsin digestion and microwave-assisted acid hydrolysis, J. Proteome Res. 6 (2007) 263–272.

[18] U. Gundel, D. Benndorf, M. von Bergen, R. Altenburger, E. Kuster, Vitellogenin cleavage products as indicators for toxic stress in zebrafish embryos: a proteomic approach, Proteomics 7 (2007) 4541–4554.

[19] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, Proc. Natl. Acad. Sci. USA 97 (2000) 9390–9395.

[20] M.R. Wilkins, E. Gasteiger, J.C. Sanchez, A. Bairoch, D.F. Hochstrasser, Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number, Electrophoresis 19 (1998) 1501–1505.

[21] V. Santoni, M. Molloy, T. Rabilloud, Membrane proteins and proteomics: un amour impossible?, Electrophoresis 21 (2000) 1054–1070

[22] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates 3rd, Direct analysis of protein complexes using mass spectrometry, Nat. Biotechnol. 17 (1999) 676–682.

[23] C.L. Lee, H.H. Hsiao, C.W. Lin, S.P. Wu, S.Y. Huang, C.Y. Wu, A.H. Wang, K.H. Khoo, Strategic shotgun proteomics approach for efficient construction of an expression map of targeted protein families in hepatoma cell lines, Proteomics 3 (2003) 2472–2486.

[24] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature 422 (2003) 198–207.

[25] W. Yan, H. Lee, E.C. Yi, D. Reiss, P. Shannon, B.K. Kwieciszewski, C. Coito, X.J. Li, A. Keller, J. Eng, T. Galitski, D.R. Goodlett, R. Aebersold, M.G. Katze, System-based proteomic analysis of the interferon response in human liver cells, Genome Biol. 5 (2004) R54.

[26] K. Ou, T.K. Seow, R.C. Liang, S.E. Ong, M.C. Chung, Proteome analysis of a human heptocellular carcinoma cell line, HCC-M: an update, Electrophoresis 22 (2001) 2804–2811.

[27] K.D. Welch, B. Wen, D.R. Goodlett, E.C. Yi, H. Lee, T.P. Reilly, S.D. Nelson, L.R. Pohl, Proteomic identification of potential susceptibility factors in drug-induced liver disease, Chem. Res. Toxicol. 18 (2005) 924–933.

[28] C.C. Wu, M.J. MacCoss, K.E. Howell, D.E. Matthews, J.R. Yates 3rd, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis, Anal. Chem. 76 (2004) 4951–4959.

[29] S.P. Gygi, B. Rist, T.J. Griffin, J. Eng, R. Aebersold, Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags, J. Proteome Res. 1 (2002) 47–54.

[30] S. Julka, F. Regnier, Quantification in proteomics through stable isotope coding: a review, J. Proteome Res. 3 (2004) 350–363.

[31] S.E. Ong, M. Mann, Mass spectrometry-based proteomics turns quantitative, Nat. Chem. Biol. 1 (2005) 252–262.

[32] M.R. Roe, T.J. Griffin, Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes, Proteomics 6 (2006) 4678–4687.

[33] Z.Y. Park, D.H. Russell, Thermal denaturation: a useful technique in peptide mass mapping, Anal. Chem. 72 (2000) 2667–2670.

[34] P. Suder, A. Bierczynska, S. Konig, J. Silberring, Acid-labile surfactant assists in-solution digestion of proteins resistant to enzymatic attack, Rapid Commun. Mass Spectrom. 18 (2004) 822–824.

[35] J. Zhou, T. Zhou, R. Cao, Z. Liu, J. Shen, P. Chen, X. Wang, S. Liang, Evaluation of the application of sodium deoxycholate to proteomic analysis of rat hippocampal plasma membrane, J. Proteome Res. 5 (2006) 2547–2553.

[36] W.K. Russell, Z.Y. Park, D.H. Russell, Proteolysis in mixed organic–aqueous solvent systems: applications for peptide mass mapping using mass spectrometry, Anal. Chem. 73 (2001) 2682–2685.

[37] L. Hua, T.Y. Low, S.K. Sze, Microwave-assisted specific chemical digestion for rapid protein identification, Proteomics 6 (2006) 586–591.

[38] W. Sun, S. Gao, L. Wang, Y. Chen, S. Wu, X. Wang, D. Zheng, Y. Gao, Microwave-assisted protein preparation and enzymatic digestion in proteomics, Mol. Cell. Proteomics 5 (2006) 769–776.

[39] B.J. Bennion, V. Daggett, The molecular basis for the chemical denaturation of proteins by urea, Proc. Natl. Acad. Sci. USA 100 (2003) 5142–5147.

[40] Y. Lin, J. Zhou, D. Bi, P. Chen, X. Wang, S. Liang, Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins, Anal. Biochem. 377 (2008) 259–266.
[41] J.J. Pesavento, B.A. Garcia, J.A. Streeky, N.L. Kelleher, C.A. Mizzen, Mild performic acid oxidation enhances chromatographic and top down mass spectrometric analyses of histones, Mol. Cell. Proteomics 6 (2007) 1510–1526.
[42] M.C. Mullins, M. Hammerschmidt, P. Haffter, C. Nusslein-Volhard, Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate, Curr. Biol. 4 (1994) 189–202.
[43] V. Link, A. Shevchenko, C.P. Heisenberg, Proteomics of early zebrafish embryos, BMC Dev. Biol. 6 (2006) 1.
[44] M. Westerfield, The Zebrafish Book. A Guide for the Laboratory Use of the Zebrafish (*Danio rerio*), 4th ed., University of Oregon Press, Eugene, 2000.
[45] J. Blonder, M.B. Goshe, R.J. Moore, L. Pasa-Tolic, C.D. Masselon, M.S. Lipton, R.D. Smith, Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography–tandem mass spectrometry, J. Proteome Res. 1 (2002) 351–360.
[46] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis 20 (1999) 3551–3567.
[47] J.E. Elias, S.P. Gygi, Target–decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nat. Methods 4 (2007) 207–214.
[48] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology—The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29.
[49] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[50] J.R. Gerstner, Q.Z. Bremer, W.M. Vander Heyden, T.M. Lavaute, J.C. Yin, C.F. Landry, Brain fatty acid binding protein (Fabp7) is diurnally regulated in astrocytes and hippocampal granule cell precursors in adult rodent brain, PLoS ONE 3 (2008) e1631.
[51] D.A. Lim, Y.C. Huang, T. Swigut, A.L. Mirick, J.M. Garcia-Verdugo, J. Wysocka, P. Ernst, A. Alvarez-Buylla, Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells, Nature 458 (2009) 529–533.
[52] R.Z. Liu, E.M. Denovan-Wright, A. Degrave, C. Thisse, B. Thisse, J.M. Wright, Differential expression of duplicated genes for brain-type fatty acid-binding proteins (Fabp7a and Fabp7b) during early development of the CNS in zebrafish (*Danio rerio*), Gene Expr. Patterns 4 (2004) 379–387.
[53] M. Olive, I. Ferrer, Postnatal development of parvalbumin immunoreactivity in striated muscles of the rat, Anat. Embryol. (Berl.) 190 (1994) 301–305.
[54] R. Ruksana, K. Kuroda, H. Terami, T. Bando, S. Kitaoka, T. Takaya, Y. Sakube, H. Kagawa, Tissue expression of four troponin I genes and their molecular interactions with two troponin C isoforms in *Caenorhabditis elegans*, Genes Cells 10 (2005) 261–276.
[55] P.Y. Rescan, New insights into skeletal muscle development and growth in teleost fishes, J. Exp. Zool. B 310 (2008) 541–548.
[56] E.A. Shrader, T.R. Henry, M.S. Greeley Jr., B.P. Bradley, Proteomics in zebrafish exposed to endocrine disrupting chemicals, Ecotoxicology 12 (2003) 485–488.
[57] S. Lemeer, R. Ruijtenbeek, M.W. Pinkse, C. Jopling, A.J. Heck, J. den Hertog, M. Slijper, Endogenous phosphotyrosine signaling in zebrafish embryos, Mol. Cell. Proteomics 6 (2007) 2088–2099.
[58] M.K. Tang, P.M. Kindler, D.Q. Cai, P.H. Chow, M. Li, K.K. Lee, Heart-type fatty acid binding proteins are upregulated during terminal differentiation of mouse cardiomyocytes, as revealed by proteomic analysis, Cell Tissue Res. 316 (2004) 339–347.
[59] W. Driever, M.C. Fishman, The zebrafish: heritable disorders in transparent embryos, J. Clin. Invest. 97 (1996) 1788–1794.