A Survey of the Impact of Deyolking on Biological Processes Covered by Shotgun Proteomic Analyses of Zebrafish Embryos

Fatima Rahlouni,¹ Szabolcs Szarka,^{1,*} Vladimir Shulaev,² and Laszlo Prokai¹

Abstract

Deyolking, the removal of the most abundant protein from the zebrafish (*Danio rerio*) embryo, is a common technique for in-depth exploration of proteome-level changes *in vivo* due to various environmental stressors or pharmacological impacts during embryonic stage of development. However, the effect of this procedure on the remaining proteome has not been fully studied. Here, we report a label-free shotgun proteomics survey on proteome coverage and biological processes that are enriched and depleted as a result of deyolking. Enriched proteins are involved in cellular energetics and development pathways, specifically implicating enrichment related to mitochondrial function. Although few proteins were removed completely by deyolking, depleted molecular pathways were associated with calcium signaling and signaling events implicating immune system response.

Introduction

VITELLOGENIN, THE ZEBRAFISH yolk protein, is the most abundant class of protein in the embryo.¹ It is the nutritional source for the growing embryo, playing a critical role in both systemic and regulated pathways in development.^{1–3} However, the depth of proteome analysis with zebrafish embryos using proteomics applications is limited because of the high abundance of vitellogenins. Shotgun proteomic analyses, a mass spectrometry-based technique that allows the global identification of proteins in a sample,⁴ are therefore limited by the overwhelming presence of vitellogenin. Proteolytic peptides of this yolk protein could potentially suppress the ionization of other, less abundant proteolytic peptides of nonyolk proteins.^{5,6} When using conventional data-dependent acquisitions, the redundant selection of precursor ions from high-abundance peptides to sequence by tandem mass spectrometry (MS/MS) also decreases proteome coverage provided through shotgun proteomics.⁷

To overcome this obstacle to in-depth proteomic analyses of zebrafish embryos, a deyolking procedure¹ is done on the embryos to remove as much of the vitellogenin as possible to reduce sample complexity before mass spectrometric analysis thereby increasing the chances of detecting low-abundance proteins. This technique has been previously shown to improve data analysis and increase the number of proteins identified.

However, the impact of deyolking on proteins enriched and depleted by the procedure, including biological processes represented by these proteins, has not been surveyed.^{1,8}

The problem of the suppression of nonyolk peptides is most pronounced in the analysis of early embryos, when the yolk is at its largest in comparison to the growing embryo. As the embryo develops in later stages, the yolk shrinks in size as it is consumed by the embryo, and the problem is reduced to the point that deyolking can cause more problems to the analysis; that is, the risks outweigh the rewards. Here, we present a bioinformatic analysis of the effect of deyolking on the zebrafish embryo proteome based on a shotgun proteomic method and label-free quantification to illustrate that deyolking does not have a uniform effect on the detection or abundance of nonyolk proteins.

Our previous results demonstrated the suitability of a shotgun approach relying on data-dependent liquid chromatography–tandem mass spectrometry (LC-MS/MS) for initial, “screening-driven” proteomic applications using label-free proteomics.^{9–11} Specifically, comparison of results obtained by two-dimensional separation involving meticulous prefractionation through strong cation-exchange chromatography followed by reversed-phase LC-MS/MS with those obtained from shotgun LC-MS/MS analyses demonstrated that most molecular functions and biological processes were represented using either methodology.¹²

¹Department of Pharmacology and Neuroscience, University of North Texas Health Science Center, Fort Worth, Texas.

²Department of Biological Sciences, University of North Texas, Denton, Texas.

*Current affiliation: LGC, Fordham, Cambridgeshire CB7 5WW, United Kingdom.

To avoid extensive and laborious prefractionation along with its disadvantages, which includes potential sample loss with each additional sample preparation step introduced and overlap of peptides and proteins in adjoining fractions, long LC gradient runs as an alternative to prefractionation also have been used successfully.^{13–15} In addition, many laboratories take advantage of label-free methods for quantitative proteomics.¹⁶ Recently, the MS/MS total ion current (“MS² TIC”) method has been successfully introduced into proteomics workflow to permit untargeted quantification.¹⁷ MS² TIC measurements can be employed reportedly over a large dynamic range,¹⁷ which is ideally suited for a label-free evaluation of the commonly used deyolking procedure of zebrafish embryos¹ in the context of shotgun proteomics studies. Therefore, we implemented this approach to survey possible gains and losses in covering biological processes associated with the enrichment and unintentional depletion of nonyolk proteins.

Materials and Methods

Chemicals

High performance liquid chromatography (HPLC) grade solvents were all obtained from Fisher Scientific (Atlanta, GA). Sequencing grade trypsin was from Applied Biosystems (Foster City, CA). All other chemicals were acquired through Sigma-Aldrich (St. Louis, MO), unless otherwise stated.

Zebrafish maintenance

Animals were treated in compliance with the guidelines set by the National Institutes of Health and the Institutional Animal Care and Use Committee at the University of North Texas Health Science Center. Wild-type, reproductively mature zebrafish were ordered from Aquatica Biotech (Sun City Center, FL). Animals were maintained in natural mating trios (one male and two females) in our aquatic habitat under standard laboratory conditions¹⁸ at a constant temperature about 28.5°C and a 14:10-h light:dark cycle, and fed two to three times daily with commercial tropical fish flake food supplemented by brine shrimp (Petsolutions, Beavercreek, OH). Directly after spawning, fish eggs were collected and cleaned with fish water. Embryos were maintained and treated in embryo medium at about 28.5°C. At 24 h post-fertilization, zebrafish were sorted for viability. Since embryos received nourishment from their yolk ball, no additional maintenance was required.¹⁸

Deyolking of zebrafish embryos in vivo

At 5 dpf, 600 embryos were split into two groups where half were subjected to a deyolking process as described previously,¹ and the other half was not. Deyolked and yolk-intact specimens were washed with embryo medium followed by resuspension in a lysis buffer (8 M urea, 1% CHAPS, cOmplete Mini Protease Inhibitor Cocktail Tablet [Roche, Indianapolis, IN] in 0.1 M phosphate-buffered saline). The suspension was then snap frozen in liquid nitrogen, with 10 subsequent freeze-thaw cycles done to ensure sufficient lysis of the cells. Samples were centrifuged at 1000 g for 5 min at 4°C, the supernatant was collected, and protein concentration was measured via the BCA Protein Assay Kit

(Piercenet, Rockford, IL). The lysates were stored at –80°C until in-solution digestion was performed.

In-solution digestion

Approximately 500 µg of protein from each deyolked and nondeyolked samples was reduced, alkylated, and digested with trypsin as previously reported.¹⁹ Following tryptic digestion, the enzymatic reaction was terminated by acidifying the samples to pH <2.0 with acetic acid and the digests were desalted using C18 Sep-Pak solid-phase extraction cartridges (Waters, Milford, MA). The desalted fish embryonic tryptic digests were further dried with a SpeedVac and subsequently reconstituted in 50 µL of 5% (v/v) acetonitrile in water containing 0.1% (v/v) acetic acid and aliquots of 5 µL were used for LC-MS/MS analyses.

Data-dependent LC-MS/MS

The samples were analyzed in triplicate using a hybrid linear ion trap–Orbitrap mass spectrometer (LTQ-Orbitrap Velos Pro; Thermo Scientific, Waltham, MA) equipped with a nano-electrospray ionization source and operated with Xcalibur (version 2.2 SP1) and LTQ Tune Plus (version 2.7) data acquisition software. Online reversed-phase HPLC was performed with Waters nano-ACQUITY UPLC (Waters) system.

A sample (5 µL) was automatically loaded onto the IntegraFrit™ sample trap (2.5 cm × 75 µm; New Objective, Woburn, MA), for preconcentration and desalting, at a flow rate of 1.5 µL/min in a loading solvent containing 0.1% (v/v) acetic acid and 5% (v/v) acetonitrile in 94.9% (v/v) water before injection onto a reverse-phase column (NAN75-15-03-C18-PM; 75 µm i.d. × 15 cm; LC Packings, Sunnyvale, CA) packed with C₁₈ beads (3 µm, 100 Å pore size; PepMap) for 5 min. Mobile-phase buffer A consisted of 0.1% (v/v) acetic acid and 99.9% (v/v) water, and mobile-phase buffer B consisted of 0.1% (v/v) acetic acid and 99.9% (v/v) acetonitrile.

Following salt removal and injection onto the analytical column, peptides were separated using the following gradient conditions: (1) 5 min in 95% solvent A for equilibration; (2) linear gradient to 40% solvent B over 240 min and holding at 40% solvent B for isocratic elution for 5 min; (3) increasing the gradient to 90% solvent B and maintaining for 5 min; and finally (4) 95% solvent A in the next 20 min. The flow rate through the column was 250 nL/min. Peptides eluted through a Picotip emitter (internal diameter 10 ± 1 µm; New Objective) were directly sampled by the nano-electrospray source of the mass spectrometer.

Spray voltage and capillary temperature during the gradient run were maintained at 1.5 kV and 200°C. Conventional data-dependent mode of acquisition was utilized in which an accurate m/z survey scan was performed in the Orbitrap cell followed by parallel MS/MS linear ion trap analysis of the top five most intense precursor ions. Orbitrap full-scan mass spectra were acquired at 60,000 mass resolving power (m/z 400) from m/z 350 to 1600 using the automatic gain control mode of ion trapping. Peptide fragmentation was performed by collision-induced dissociation (CID) in the linear ion trap using a 2.0-Th isolation width and 35% normalized collision energy with helium as the target gas. The precursor ion that had been selected for CID was dynamically excluded from further MS/MS analysis for 180 s.

TABLE 1. SIGNIFICANTLY ENRICHED PROTEINS IDENTIFIED BY COMBINED MASCOT AND SEQUEST DATABASE SEARCHES AND COMBINED SPECTRAL COUNTING AND MS² TIC FEATURES GENERATED IN SCAFFOLD 4 SOFTWARE COMPARING DEYOLKED AND NONDEYOLKED CONTROL ZEBRAFISH EMBRYOS

<i>Protein name</i>	<i>Accession number</i>	<i>Molecular weight (kDa)</i>	<i>Student's t-test or G-test^a (p < 0.05)</i>	<i>Yolk average: total MS² TIC or spectral counts^b</i>	<i>Devolked average: total MS² TIC or spectral counts^b</i>
Aconitase 2, mitochondrial ^c	F8W4M7_DANRE	86	0.002	Not identified without deyolking	1.70E+04 ± 2.34E+03
Actin, alpha 1, skeletal muscle ^d	Q9I8V1_DANRE	42	0.008	1.92E+06 ± 2.52E+05	4.20E+06 ± 3.86E+05
Actinodin 2 ^c	A9IRX1_DANRE	57	0.037	Not identified without deyolking	1.59E+04 ± 5.16E+03
Adenosine monophosphate deaminase 1 (Isoform M) ^{c,e}	Q6P3G5_DANRE	83	0.045	Not identified without deyolking	5.43E+04 ± 1.88E+04
Adenylate kinase isoenzyme 1 ^{c,e}	Q68EH2_DANRE	21	0.030	4.69E+04 ± 1.48E+04	1.02E+05 ± 8.06E+03
ATP synthase subunit alpha ^d	Q08BA1_DANRE	60	<0.001	8.32E+04 ± 4.65E+03	3.71E+05 ± 1.02E+04
ATP synthase subunit beta ^d	FIR2V7_DANRE	56	0.001	Not identified without deyolking	3.75E+05 ± 4.71E+04
ATPase, H ⁺ transporting, lysosomal, V1 subunit Ab ^{c,e}	E7FCD8_DANRE	68	0.005	Not identified without deyolking	2.94E+05 ± 5.26E+04
BetaA2-2-crystallin ^{c,e}	Q45FX9_DANRE	24	0.015	Not identified without deyolking	4.97E+03 ± 1.21E+03
Brain-subtype creatine kinase ^{c,e}	Q8AY63_DANRE	43	0.012	Not identified without deyolking	1.82E+05 ± 4.15E+04
Calsequestrin ^{c,e}	FIQNI9_DANRE	32	0.018	Not identified without deyolking	8.68E+04 ± 2.24E+04
Clathrin heavy chain ^{c,e}	FIQXT6_DANRE	192	0.012	Not identified without deyolking	4.91E+04 ± 1.12E+04
Creatine kinase, mitochondrial 2 (Sarcomeric) ^{c,e}	FIR966_DANRE	46	0.036	Not identified without deyolking	2.02E+04 ± 6.49E+03
Creatine kinase, muscle a ^d	Q6PCS6_DANRE	43	0.007	Not identified without deyolking	6.30E+05 ± 1.22E+05
Creatine kinase, muscle b ^d	A2BHA3_DANRE	43	0.004	Not identified without deyolking	2.53E+05 ± 4.26E+04
Crystallin, beta B1 ^c	Q7T306_DANRE	29	0.041	Not identified without deyolking	5.63E+04 ± 1.89E+04
Dachsous cadherin-related 1a ^{c,e}	FIQ256_DANRE	352	0.004	2.26E+04 ± 2.26E+04	1.64E+05 ± 5.81E+03
Elongation factor 1-gamma ^c	X1WDW2_DANRE	50	0.001	Not identified without deyolking	5.85E+04 ± 7.10E+03
Epiphykan ^c	EFIG_DANRE	75	<0.001	Not identified without deyolking	8.00E+04 ± 4.24E+03
Fast skeletal muscle myosin light polypeptide 3 ^c	B0S5X0_DANRE	17	0.014	7.23E+05 ± 1.19E+05	1.48E+06 ± 1.34E+05
Fructose-bisphosphate aldolase ^c	Q9I8U7_DANRE	40	0.025	Not identified without deyolking	5.98E+03 ± 1.70E+03
GDP dissociation inhibitor 2 ^{c,e}	Q803Q7_DANRE	51	0.007	Not identified without deyolking	6.23E+04 ± 1.23E+04
Glyceraldehyde-3-phosphate dehydrogenase ^d	Q6TNT9_DANRE	36	0.001	Not identified without deyolking	4.31E+05 ± 4.81E+04
G-protein-coupled receptor 98 ^{c,e}	FIR3D3_DANRE	674	0.019	3.56E+03 ± 3.56E+03	2.50E+04 ± 4.30E+03
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 ^{c,e}	G3P2_DANRE	37	0.034	Not identified without deyolking	1.32E+05 ± 4.16E+04
Heat shock 60kD protein 1 (Chaperonin) ^c	FIQEZ1_DANRE	61	0.046	Not identified without deyolking	1.90E+04 ± 6.61E+03
Heterochromatin protein 1, binding protein 3 ^c	GBB1_DANRE	65	0.017	Not identified without deyolking	6.55E+04 ± 1.65E+04
Heterogeneous nuclear ribonucleoprotein A0, like ^{c,e}	Q803B0_DANRE	32a	0.006	Not identified without deyolking	8.92E+04 ± 1.70E+04
Heterogeneous nuclear ribonucleoprotein L ^{c,e}	Q7ZU48_DANRE	59	0.017	Not identified without deyolking	9.16E+04 ± 2.31E+04
Histone H3.2 ^c	Q7ZW09_DANRE	15	0.010	6.46E+04 ± 3.93E+04	2.49E+05 ± 9.72E+03
Hydroxysteroid (17-beta) dehydrogenase 10 ^c	H32_DANRE	27	0.017	Not identified without deyolking	4.23E+04 ± 1.07E+04
Malate dehydrogenase ^d	X1WC76_DANRE	35	<0.001	3.66E+04 ± 2.70E+04	5.85E+05 ± 8.91E+03
Malic enzyme ^{c,e}	Q5XJS8_DANRE	67	0.009	Not identified without deyolking	1.81E+04 ± 3.83E+03
Myosin light chain, phosphorylatable, fast skeletal muscle a ^c	Q7T334_DANRE	19	0.006	2.08E+05 ± 2.71E+04	4.76E+05 ± 4.29E+04

(continued)

TABLE 1. (CONTINUED)

<i>Protein name</i>	<i>Accession number</i>	<i>Molecular weight (kDa)</i>	<i>Student's t-test or G-test^a (p < 0.05)</i>	<i>Yolk average: total MS² TIC or spectral counts^b</i>	<i>Devolked average: total MS² TIC or spectral counts^b</i>
Myosin, heavy polypeptide 1.2, skeletal muscle ^f	B8A561_DANRE	222	<0.001	Not identified without devolking	8.45E+01 ± 4.18E+01 ^f
NADH dehydrogenase (ubiquinone) alpha subcomplex, 5 ^{c,e}	Q5BJA2_DANRE	13	0.003	Not identified without devolking	4.92E+04 ± 7.63E+03
Nucleoside diphosphate kinase ^{c,e}	Q7SXL4_DANRE	17	0.025	4.41E+04 ± 1.57E+04	1.59E+05 ± 2.92E+04
Periostin, osteoblast specific factor b ^{c,e}	E7FBL9_DANRE	86	0.030	1.01E+04 ± 1.01E+04	5.16E+04 ± 7.47E+03
	F1QHK3_DANRE				
	Q75U66_DANRE				
Phosphorylase ^{c,e}	Q503C7_DANRE	97	0.034	9.02E+03 ± 9.02E+03	1.64E+05 ± 4.79E+04
Pyrophosphatase (inorganic) 1b ^{c,e}	B7ZD39_DANRE	33	0.015	Not identified without devolking	9.36E+04 ± 2.29E+04
Pyruvate kinase ^d	F1QEN1_DANRE	67a	<0.001	Not identified without devolking	2.74E+05 ± 8.59E+03
	F1QSE0_DANRE				
	F1RKB3_DANRE				
Ribosomal protein S5 ^{c,e}	Q6PC80_DANRE	23	0.011	Not identified without devolking	1.76E+05 ± 3.94E+04
Synaptosomal-associated protein 91 ^{c,e}	E7F4Z2_DANRE	94	0.049	Not identified without devolking	5.59E+04 ± 2.00E+04
	E9QBM8_DANRE				
Synaptotagmin binding, cytoplasmic RNA interacting protein ^e	F1QY96_DANRE	77	0.011	5.81E+04 ± 1.46E+04	2.31E+05 ± 3.56E+04
Titin A ^d	Q6NUZ3_DANRE				
Titin B ^d	F1R7N8_DANRE	3487	0.009	1.86E+04 ± 1.86E+04	1.84E+05 ± 2.95E+04
Transforming growth factor, beta-induced ^{c,e}	F1Q6U3_DANRE	3026	0.026	2.79E+04 ± 2.47E+03	1.60E+05 ± 3.83E+04
Trypsine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide ^{c,e}	Q503K1_DANRE	74	0.022	Not identified without devolking	1.03E+05 ± 2.84E+04
Tubulin, alpha 1c ^f	Q803M8_DANRE	28	<0.001	Not identified without devolking	1.72E+04 ± 1.35E+03
Ubiquinol-cytochrome c reductase core protein I ^{c,e}	B8A518_DANRE	50	0.027	2.43E+00 ± 0.72E+00 ^f	9.96E+00 ± 0.24E+00 ^f
Uncharacterized ionotropic glutamate receptor family protein si:ch211-251b21.1 ^c	F1QUE3_DANRE	55	0.004	Not identified without devolking	1.99E+04 ± 3.44E+03
Uncharacterized actin family protein zgc:86725 ^d	B8ILR6_DANRE	50	<0.001	Not identified without devolking	1.69E+05 ± 1.43E+04
	F1RCB6_DANRE	42	0.001	1.60E+06 ± 7.82E+04	3.27E+06 ± 1.84E+05

Minimum protein identification confidence: 99%, minimum peptide identification confidence: 95% requiring at least 2 peptides per protein. G-test or Student's *t*-test was accepted at *p* < 0.05 with at least a 1.5-fold change.

^aStudent's *t*-test for total MS² TIC, unless significant enrichment was detected only with spectral counting for which the G-test is reported.

^bSpectral counts, if significant enrichment detected only with spectral counting—otherwise, total MS² TIC. Total MS² TIC and spectral counts are given as average ± standard error.

^cSignificant enrichment detected only using MS² TIC criteria.

^dSignificant enrichment detected with both spectral counting and using MS² TIC criteria.

^eIdentification made by SEQUEST alone.

^fSignificant enrichment detected only with spectral counting.

MS² TIC, MS/MS total ion current.

Database compilation, label-free quantification, and signaling pathway analysis

MS/MS data generated by data-dependent acquisitions were extracted by BioWorks version 3.3 and searched against a composite UniProt zebrafish protein sequence database (release 2014_07, 41267 entries) using the Mascot and SEQUEST search algorithms within Proteome Discoverer (version 1.4; Thermo Scientific). The Proteome Discoverer application reverses all protein sequences to achieve a decoy database to search and, using the Percolator node, false discovery rates (FDRs) are calculated. Mascot and SEQUEST were searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 25.0 ppm assuming the digestion enzyme trypsin with the possibility of one missed cleavage. Carbamidomethylation of cysteine was specified as a fixed modification while oxidation of methionine and deamidation of asparagine and glutamine were specified as variable modifications in the database search.

The software program Scaffold (version 4.4.1; Proteome Software, Inc., Portland, OR) was employed to validate MS/MS-based peptide and protein identifications. Peptide information was accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm.²⁰ Protein identifications, where protein probabilities were assigned by the Protein Prophet,²¹ were accepted if they could be established at greater than 99.0% probability and contained at least two identified peptides.

Scaffold 4 readily extracts the MS² TIC value for each identified peptide and calculates the total TIC for each protein, and provides the spectral count. To test for significant changes in protein expression between treatments, the Student's *t*-test was performed on the normalized TIC values and accepted at $p < 0.05$ requiring at least a 1.5-fold change, and the *G*-test was performed on the normalized spectral count values and accepted at $p < 0.05$ with at least a 1.5-fold change. FDR was calculated for the significantly enriched and depleted proteins as number of decoy identifications per total protein identifications.

Additionally, Ingenuity Pathway Analysis® (IPA®; Qiagen, Redwood City, CA) was utilized to derive annotations along with potential protein interaction networks from the associated proteins in the zebrafish embryos. IPA processes and maps identifiers according to their HomoloGene to the orthologs found in IPA's data bank where curated content is specific to human, mouse, and rat.

Data sharing

The data have been deposited to the ProteomeXchange²² with identifier PXD002124.

Results

Through combing analyses of yolk-intact and deyolked zebrafish embryos and using the complementary Mascot and SEQUEST database search algorithms jointly, our study identified 159 proteins (6 unique proteins were found in the yolk-intact samples, 107 proteins were identified from both yolk-intact and deyolked samples, and 46 unique proteins were discovered in the deyolked samples). These protein identifications passed rigorous validation criteria both by Peptide Prophet²⁰ and Protein Prophet²¹ with at least two

unique tryptic peptides identified for each protein. The number of identified proteins through our reversed-phase nanoflow LC-MS/MS with a 4-h gradient compared well to that reported upon using the laborious fractionation-based multidimensional protein identification technology (Mud-PIT, also a shotgun approach) with doubly- and triply-charged precursors for data-dependent MS/MS sequencing (242 identified proteins) after deyolking.²³

We combined spectral counting and MS² TIC to detect quantitative differences as previously reported.¹⁷ Spectral counting, which is not as sensitive as MS² TIC, alone gave 26 enriched or depleted proteins from this dataset, whereas MS² TIC had nearly three times as many. The use of MS² TIC resulted in substantial improvement in identifications over spectral counting, where only three proteins (B8A561_DANRE, B8A518_DANRE, and B3DFP9_DANRE) significant with spectral counting were not found significant by MS² TIC, one of which (B8A561_DANRE) was also found to be identified only after deyolking. Using rigorous criteria for the evaluation of significant protein expression differences, 69 confidently identified proteins were markedly affected by the deyolking process (Tables 1 and 2).

FDR for the identification of the significantly enriched and depleted proteins was $< 2\%$ at the protein level with only one decoy (F1Q8S1_DANRE-DECOY) matching our set criteria. The majority of the proteins were observed to be enriched by deyolking, and the bulk of the proteins that were depleted were indeed vitellogenins. Nearly 20% of all proteins identified in this study were revealed only in the deyolked sample, which represents gain in coverage of the zebrafish embryo proteome from the deyolking procedure. Of the depleted proteins, only two nonyolk proteins (about 3% of the proteins significantly affected by deyolking) were removed to the extent that they eluded confident identification by shotgun proteomics.

We used IPA to find functional interactions between the enriched and depleted proteins identified in our study (Fig. 1). Of the 69 proteins significantly affected by deyolking, 52 were mapped to three networks combined in Figure 1A. The enriched proteins were also mapped to two networks that were associated with (1) skeletal and muscular system development and function, amino acid metabolism, small molecule biochemistry; and (2) cellular compromise, cellular function and maintenance, auditory and vestibular system development and function. A summary of the pathways and biological functions represented by the enriched proteins is shown in Table 3.

Of the two nonyolk proteins that were completely removed, both were keratins (E7FCX7_DANRE and Q1LXJ9_DANRE, which were identified by peptides that are uniquely attributable to zebrafish and similar fish species) present in the combined network. Overall, the combined network represented by depleted proteins was fairly small but the associated molecular pathways pointed to calcium signaling and signaling events implicating immune system response (Fig. 1C and Table 4).

Discussion

The widespread use of the zebrafish as model organism in proteomics, and more specifically the deyolking of the embryos to improve proteome coverage, underscores the need

TABLE 2. SIGNIFICANTLY DEPLETED PROTEINS IDENTIFIED BY COMBINED MASCOT AND SEQUEST DATABASE SEARCHES AND COMBINED SPECTRAL COUNTING AND MS² TIC FEATURES GENERATED IN SCAFFOLD 4 SOFTWARE COMPARING DEYOLKED AND NONDEYOLKED CONTROL ZEBRAFISH EMBRYOS

<i>Protein name</i>	<i>Accession number</i>	<i>Molecular weight (kDa)</i>	<i>Student's t-test or G-test^a (p < 0.05)</i>	<i>Yolk average: total MS² TIC or spectral counts^b</i>	<i>Devolked average: total MS² TIC or spectral counts^b</i>
AHNAK nucleoprotein ^{c,d}	F1QZ50_DANRE FIR1J9_DANRE	685	0.001	1.40E+05 ± 1.33E+04	4.03E+03 ± 4.03E+03
Apolipoprotein A-II ^e	B3DFP9_DANRE	16a	0.026	7.27E+00 ± 2.57E+00 ^b	1.16E+00 ± 0.16E+00 ^b
Apolipoprotein Bb, tandem duplicate 1 ^f	F1QQ87_DANRE Q5TZ29_DANRE	413	0.002	5.00E+05 ± 5.90E+04	8.77E+04 ± 9.96E+03
ATPase, Ca ⁺⁺ transporting, cardiac muscle, fast twitch 1 ^f	Q5U3A4_DANRE	110	0.008	9.65E+05 ± 1.33E+05	2.60E+05 ± 5.70E+04
Crystallin, gamma M2d17 ^d	Q64Z70_DANRE A6H8Q4_DANRE	21	0.008	1.29E+05 ± 8.07E+03	5.17E+04 ± 1.37E+04
Keratin 4 ^f	B0S6N1_DANRE	54	0.002	1.69E+06 ± 2.28E+05	8.20E+04 ± 5.21E+04
Keratin type 1 c19e ^d	F1QK60_DANRE	50	0.049	1.27E+05 ± 4.56E+04	Completely removed by devolking
Myosin, heavy chain b ^{c,f}	Q1LXJ9_DANRE F1QVX3_DANRE X1WF87_DANRE	223	0.041	1.70E+06 ± 1.27E+05	6.43E+05 ± 3.32E+05
Myosin, heavy polypeptide 1.1, skeletal muscle ^{c,f}	B8A568_DANRE	222	0.013	7.03E+06 ± 2.77E+05	5.62E+06 ± 1.83E+05
Myosin, heavy polypeptide 1.3, skeletal muscle ^{c,d}	B8A569_DANRE	222	0.009	6.94E+06 ± 3.11E+05	5.29E+06 ± 1.45E+05
Myosin, heavy polypeptide 2, fast muscle specific ^d	Q6IQX1_DANRE	222	0.013	5.25E+06 ± 1.32E+05	4.45E+06 ± 1.35E+05
Type I cyokeratin, enveloping layer, like ^{c,f}	E7FCX7_DANRE E9QDY3_DANRE	46	0.028	2.02E+05 ± 5.99E+04	Completely removed by devolking
Uncharacterized histone 2 family protein si:ch211-113a14.11 ^f	Q1LXJ7_DANRE Q9PWD8_DANRE R4GE02_DANRE	27	0.036	7.40E+06 ± 5.81E+05	5.24E+06 ± 3.86E+05
Vitellogenin 1 ^f	Q1LWN2_DANRE	149	<0.001	1.77E+06 ± 6.66E+04	1.35E+04 ± 6.93E+03
Vitellogenin 2 ^f	Q1MTC4_DANRE	180	<0.001	7.11E+05 ± 2.14E+04	Completely removed by devolking
Vitellogenin 5 ^f	FIR2S5_DANRE	149	<0.001	1.18E+06 ± 7.16E+04	Completely removed by devolking
Vitellogenin 7 ^f	FIR2T3_DANRE Q1MTC6_DANRE	149	0.002	1.64E+06 ± 2.12E+05	7.65E+03 ± 7.65E+03

Minimum protein identification confidence: 99%, minimum peptide identification confidence: 95% requiring at least two peptides per protein. G-test or Student's t-test was accepted at $p < 0.05$ with at least a 1.5-fold change.

^aStudent's t-test for total MS² TIC, unless significant enrichment was detected only with spectral counting for which the G-test is reported.

^bSpectral counts, if significant enrichment detected only with spectral counting—otherwise, total MS² TIC. Total MS² TIC and spectral counts are given as average ± standard error.

^cIdentification made by SEQUEST alone.

^dSignificant enrichment detected only using MS² TIC criteria.

^eSignificant enrichment detected only with spectral counting.

^fSignificant enrichment detected with both spectral counting and using MS² TIC criteria.

FIG. 1. Networks constructed through Ingenuity Pathway Analysis from nonyolk proteins affected by the deyolking procedure. **(A)** From all (69) enriched and depleted proteins, we mapped 52. Top functions included (1) amino acid metabolism, small molecule biochemistry, skeletal and muscular system development and function; (2) cellular compromise, cellular function and maintenance, energy production; and (3) nucleic acid metabolism, small molecule biochemistry, cell signaling. **(B)** From all (52) enriched proteins, we mapped 43. Top functions included (1) skeletal and muscular system development and function, amino acid metabolism, small molecule biochemistry; and (2) cellular compromise, cellular function and maintenance, auditory and vestibular system development and function. **(C)** From all (13) depleted proteins, excluding the vitellogenins, we mapped 9. Top functions included cell-mediated immune response, cellular movement, hematological system development and function. *Single lines* indicate protein-protein interactions from the network diagram and *arrows* specify proteins/compounds that regulate another protein. The intensity of *green* and *red* molecule colors represents the degree of down- or upregulation, respectively. *Solid* or *dashed lines* show direct or indirect interactions, respectively. Color images available online at www.liebertpub.com/zeb

TABLE 3. IPA® ANALYSIS: SUMMARY OF TOP PATHWAYS AND BIOLOGICAL FUNCTIONS REPRESENTED BY PROTEINS SIGNIFICANTLY ENRICHED UPON DEYOLKING ZEBRAFISH EMBRYOS

	p-Value
Canonical pathways	
Creatine-phosphate biosynthesis	2.82E-08
Gluconeogenesis I	1.58E-07
Protein kinase A signaling	9.03E-06
Mitochondrial dysfunction	2.06E-05
RhoGDI signaling	2.18E-05
Molecular and cellular functions	
Amino acid metabolism	1.37E-02–7.07E-09
Small molecule biochemistry	4.24E-02–7.07E-09
Nucleic acid metabolism	4.24E-02–6.14E-08
Cellular assembly and organization	4.43E-02–3.76E-05
Cellular development	4.32E-02–3.76E-05
Physiological system development and function	
Skeletal and muscular system development and function	4.43E-02–3.06E-09
Embryonic development	4.06E-02–3.76E-05
Organ development	4.32E-02–3.76E-05
Organismal development	4.06E-02–3.76E-05
Tissue development	4.62E-02–3.76E-05

for understanding the information gained from this procedure. This study presents, to our knowledge, the first proteomics evaluation of this recognized technique on the biological processes affected by deyolking. We used a shotgun, label-free approach on unfractionated samples to minimize additional laborious preparation steps or time-consuming methodologies.

TABLE 4. IPA® ANALYSIS: SUMMARY OF TOP PATHWAYS AND BIOLOGICAL PROCESSES REPRESENTED BY PROTEINS (EXCLUDING VITELLOGENINS) SIGNIFICANTLY DEPLETED UPON DEYOLKING ZEBRAFISH EMBRYOS

	p-Value
Canonical pathways	
Calcium signaling	7.50E-04
Calcium transport I	2.21E-03
Calcium-induced T lymphocyte apoptosis	1.57E-02
TR/RXR activation	2.07E-02
Nitric oxide signaling in the cardiovascular system	2.44E-02
Molecular and cellular functions	
Cell morphology	2.10E-02–2.46E-04
Cellular assembly and organization	1.81E-02–4.92E-04
Cellular compromise	6.63E-03–4.92E-04
Cellular development	3.90E-02–4.92E-04
Cellular growth and proliferation	3.90E-02–4.92E-04
Physiological system development and function	
Connective tissue development and function	2.10E-02–2.46E-04
Digestive system development and function	4.92E-04–2.46E-04
Endocrine system development and function	9.84E-04–2.46E-04
Organ morphology	8.83E-03–2.46E-04
Organismal development	3.47E-02–2.46E-04

We focus our discussion mainly on significantly enriched proteins, more specifically on those proteins and associated biological information that were only identified after deyolking, because it is considered the principal advantage of performing this procedure on zebrafish embryos concerning shotgun proteomic studies. However, attention to potential consequences of the procedure due to unintentional removal of nonyolk proteins during deyolking may also be necessary, as our study has revealed protein depletion by the process for the first time.

One concern with MS-based protein identifications from shotgun proteomics-based experiments is the need for complete and accurate databases, as well as appropriate search engines to obtain matches to database entries. Because the zebrafish genome has been fully sequenced and is available,²⁴ the databases are complete and accurate. There are many software options and search algorithms available to search acquired MS/MS spectra against a database, each of which making use of a different scoring model.^{25,26} These algorithms can be used jointly, as complementary scoring models, to increase the confidence of peptide identifications.^{27,28} All these algorithms implement strategies for estimation of FDR.²⁶

In our study, we used two widely accepted search engines with criteria recommended from previous work.^{10,11,17} In our hands, Mascot has conservative scoring criteria, which usually results in no false discovery, but definitely FDR of less than 1% at the protein level. However, when we combined Mascot searches with SEQUEST, we observed that there were additional identifications made only by the latter search engine, although at the expense of increasing the FDR. Nevertheless, FDR introduced by SEQUEST is controlled by the addition of another search algorithm,²⁹ in our case by Mascot's stringent modeling. While using multiple search algorithms improves peptide confidence and increases proteome coverage, the introduction of and/or increase in FDR may be a concern.

To improve detection of quantitative differences and increase proteome coverage, we used the combination of spectral counting and MS² TIC label-free quantification methods as previously reported.¹⁷ Spectral counting, an approach where the relative protein abundance is measured by the number of MS/MS spectra, is simple and easy to implement, but is limited in inaccurately representing quantitative measures like fold change. To increase dynamic range and improve sensitivity for low-abundance proteins, MS² TIC (the sum of the total ion current of the MS/MS spectra of all peptides for a protein) were used.¹⁷ However, MS² TIC is not as reproducible as spectral counting due to inherent variation in sampling over the chromatographic peak. On the other hand, two quantification approaches were indeed complementary and could be combined to increase the detection of proteins whose abundance was affected by the deyolking procedure.¹⁷

Our proteomic analysis has shown that performing the deyolking procedure not only depletes the embryonic proteome of vitellogenins, but also significantly enriches a number of other important proteins (Table 1). All the enriched proteins identified after deyolking are mapped to IPA networks where the cellular functions are mainly involved in metabolic processes and energy production, as well as cellular function and maintenance needed for maintenance/growth of the cells and development of the embryo including signaling pathways, such as RhoGDI signaling, that are important regulatory points in cell proliferation (Table 3 and Fig. 1B).

The represented physiological system functions from all enriched proteins is consistent with the embryonic stage of development, showing skeletal and muscular system development and function, as well as embryonic, organ, and tissue development. Of the 35 proteins that were identified only after de yolking, 10 were mitochondrial. The observed enrichment of mitochondrial proteins, especially those related to ATP production, was demonstrated by the implication of the energy producing pathways (creatine-phosphate biosynthesis, gluconeogenesis I, and protein kinase A signaling listed in Table 3).

Enrichment of membrane proteins also was a benefit of de yolking. For example, vacuolar H⁺-transporting ATPase (E7FCD8_DANRE represented in Fig. 1B by its mammalian ortholog *ATP6V1A*, a multisubunit enzyme that mediates acidification of intracellular organelles in eukaryotic cells³⁰), clathrin heavy chain (F1R966_DANRE for which the human ortholog *CLTC*, involved in numerous biological processes, represents the major polyhedral coat protein of coated pits and vesicles³¹), and the synaptosomal-associated 91-kDa protein (E7F4Z2_DANRE whose mammalian ortholog *SNAP91* is a membrane protein involved in clathrin binding³²) were identified only after yolk removal.

While the physiological role of *SNAP91* in zebrafish embryos has not been revealed, vacuolar H⁺-transporting ATPase has been found to localize mainly in the apical membrane of H⁺-pump-rich cells in the embryo's skin to pump out internal acid and function similar to acid-secreting intercalated cells of the mammalian kidneys.³³ Also, receptor-mediated endocytosis of yolk protein has been attributed to clathrin since its first observation.³¹

However, one also must be aware of unintentional removal of nonyolk proteins (Table 2). Although few such proteins were removed completely, significant depletion associated with calcium signaling pathways was revealed (Table 4). All the depleted proteins, excluding the vitellogenins, were mapped to a network where cell-mediated immune response, cellular movement, hematological system development and function were implicated (Fig. 1C). More specifically implicated are pathways such as calcium-induced T lymphocyte apoptosis and nitric oxide signaling in the cardiovascular system, which taken together with the mapped network implicate the development and function of the blood components. Depending on the context of experimental design, for example, proteomic studies focusing on immune-related diseases and pathways could potentially lose important information by analyzing only de yolked embryos.

In summary, our study may serve as a useful reference for future proteomic analyses of zebrafish embryos that consider the use of de yolking to increase proteome coverage. We have demonstrated that this procedure does not have a uniform effect on the detection or abundance of nonyolk proteins. Nevertheless, we identified several proteins by shotgun proteomics performed using both yolk-intact and de yolked zebrafish embryos that were covered with confidence only after the commonly used de yolking procedure. Enriched proteins represent the various cellular processes related to metabolic processes and signaling events necessary for growth and development.

We also have unveiled, for the first time, the potential impact that de yolking has concerning unintended removal of nonyolk proteins. Depleted proteins represent various

molecular pathways linked to calcium signaling and signaling events associated with immune system response and development. Altogether, we recommend using both yolk-intact and de yolked embryos to maximize proteome coverage in shotgun proteomics studies.

Acknowledgments

The Welch Foundation supported this study through an endowment (BK-0031, Chair in Biochemistry to L.P.). The authors further thank Shastazia White for her assistance with animal care and maintenance.

Disclosure Statement

No competing financial interests exist.

References

1. Link V, Shevchenko A, Heisenberg CP. Proteomics of early zebrafish embryos. *BMC Dev Biol* 2006;6:1.
2. Gundel U, Benndorf D, von Bergen M, Altenburger R, Kuster E. Vitellogenin cleavage products as indicators for toxic stress in zebra fish embryos: a proteomic approach. *Proteomics* 2007;7:4541–4554.
3. Rohde LA, Heisenberg CP. Zebrafish gastrulation: cell movements, signals, and mechanisms. *Int Rev Cytol* 2007;261:159–192.
4. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* 2013;113:2343–2394.
5. Vaidyanathan S, Kell DB, Goodacre R. Selective detection of proteins in mixtures using electrospray ionization mass spectrometry: influence of instrumental settings and implications for proteomics. *Anal Chem* 2004;76:5024–5032.
6. Lu W, Yin X, Liu X, Yan G, Yang P. Response of peptide intensity to concentration in ESI-MS-based proteome. *Sci China Chem* 2014;57:686–694.
7. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004;76:4193–4201.
8. Lin Y, Chen Y, Yang X, Xu D, Liang S. Proteome analysis of a single zebrafish embryo using three different digestion strategies coupled with liquid chromatography-tandem mass spectrometry. *Anal Biochem* 2009;394:177–185.
9. Prokai L, Stevens SM Jr., Rauniyar N, Nguyen V. Rapid label-free identification of estrogen-induced differential protein expression in vivo from mouse brain and uterine tissue. *J Proteome Res* 2009;8:3862–3871.
10. Nagaprashantha LD, Talamantes T, Singhal J, Guo J, Vatsyayan R, Rauniyar N, *et al.* Proteomic analysis of signaling network regulation in renal cell carcinomas with differential hypoxia-inducible factor-2alpha expression. *PLoS One* 2013;8:e71654.
11. Talamantes T, Ughy B, Domonkos I, Kis M, Gombos Z, Prokai L. Label-free LC-MS/MS identification of phosphatidylglycerol-regulated proteins in *Synechocystis* sp. PCC6803. *Proteomics* 2014;14:1053–1057.
12. Stevens SM Jr., Duncan RS, Koulen P, Prokai L. Proteomic analysis of mouse brain microsomes: identification and bioinformatic characterization of endoplasmic reticulum proteins in the mammalian central nervous system. *J Proteome Res* 2008;7:1046–1054.
13. Tu C, Li J, Bu Y, Hangauer D, Qu J. An ion-current-based, comprehensive and reproducible proteomic strategy for comparative characterization of the cellular responses to

- novel anti-cancer agents in a prostate cell model. *J Proteomics* 2012;77:187–201.
14. Thakur SS, Geiger T, Chatterjee B, Bandilla P, Frohlich F, Cox J, *et al.* Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol Cell Proteomics* 2011;10:M110.003699.
15. Pirmoradian M, Budamgunta H, Chingin K, Zhang B, Astorga-Wells J, Zubarev RA. Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol Cell Proteomics* 2013;12:3330–3338.
16. Nahnsen S, Bielow C, Reinert K, Kohlbacher O. Tools for label-free peptide quantification. *Mol Cell Proteomics* 2013;12:549–556.
17. Freund DM, Prenni JE. Improved detection of quantitative differences using a combination of spectral counting and MS/MS total ion current. *J Proteome Res* 2013;12:1996–2004.
18. Westerfield M. *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Danio rerio)*. University of Oregon Press, Eugene, 2007.
19. Guo J, Prokai-Tatrai K, Nguyen V, Rauniyar N, Ughy B, Prokai L. Protein targets for carbonylation by 4-hydroxy-2-nonenal in rat liver mitochondria. *J Proteomics* 2011;74:2370–2379.
20. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002;74:5383–5392.
21. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75:4646–4658.
22. Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 2014;30:223–226.
23. Nesatyy VJ, Groh K, Nestler H, Suter MJ. On the acquisition of +1 charge states during high-throughput proteomics: Implications on reproducibility, number and confidence of protein identifications. *J Proteomics* 2009;72:761–770.
24. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 2013;496:498–503.
25. Mueller LN, Brusniak MY, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res* 2008;7:51–61.
26. Sadygov RG, Cociorva D, Yates JR 3rd. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods* 2004;1:195–202.
27. Alves G, Wu WW, Wang G, Shen RF, Yu YK. Enhancing peptide identification confidence by combining search methods. *J Proteome Res* 2008;7:3102–3113.
28. Dagda RK, Sultana T, Lyons-Weiler J. Evaluation of the consensus of four peptide identification algorithms for tandem mass spectrometry based proteomics. *J Proteomics Bioinform* 2010;3:39–47.
29. Rohrbough JG, Breci L, Merchant N, Miller S, Haynes PA. Verification of single-peptide protein identifications by the application of complementary database search algorithms. *J Biomol Tech* 2006;17:327–332.
30. Nishi T, Forgac M. The vacuolar (H⁺)-ATPases—nature's most versatile proton pumps. *Nat Rev Mol Cell Biol* 2002;3:94–103.
31. Brodsky FM. Diversity of clathrin function: new tricks for an old protein. *Annu Rev Cell Dev Biol* 2012;28:309–336.
32. Swanton E, Sheehan J, Bishop N, High S, Woodman P. Formation and turnover of NSF- and SNAP-containing “fusion” complexes occur on undocked, clathrin-coated vesicle-derived membranes. *Mol Biol Cell* 1998;9:1633–1647.
33. Horng JL, Lin LY, Huang CJ, Katoh F, Kaneko T, Hwang PP. Knockdown of V-ATPase subunit A (atp6v1a) impairs acid secretion and ion balance in zebrafish (*Danio rerio*). *Am J Physiol Regul Integr Comp Physiol* 2007;292:R2068–R2076.

Address correspondence to:

Laszlo Prokai, PhD, DSc

Department of Pharmacology and Neuroscience
University of North Texas Health Science Center

3500 Camp Bowie Boulevard
Fort Worth, TX 76107-2699

E-mail: laszlo.prokai@unthsc.edu