ZEBRAFISH
Volume 6, Number 3, 2009

DOI: 10.1089/zeb.2009.0591

Original Article

Large-Scale Proteome Profile of the Zebrafish (*Danio rerio*) Gill for Physiological and Biomarker Discovery Studies

Andrea G. De Souza,[1,*] Tyson J. MacCormack,[2,*] Nan Wang,[1] Liang Li,[1] and Greg G. Goss[2]

Abstract

Zebrafish are an important model in vertebrate genetics, developmental biology, physiology, and toxicology. In this study, we established the first large-scale proteome profile of a teleost fish tissue using a shotgun method based on two-dimensional liquid chromatography–electrospray ionization tandem mass spectrometry. Proteome coverage was significantly improved with the application of a sequential protein solubilization method for protein fractionation and a precursor ion exclusion method for improving peptide and protein identification efficiency. Five thousand seven hundred sixteen proteins were identified with an estimated false-positive matching rate of 1.34%, and the proteome exhibited excellent coverage of important biochemical pathways relevant to the function of the gill in respiration, ion and acid–base homeostasis, and energy metabolism. Numerous established and potential biomarkers of stress, disease, and environmental contamination were also expressed in the gill. Annotation information was completely lacking for >30% of the detected proteins, highlighting the need for advancements in bioinformatics analysis techniques to complement this research. Nevertheless, the results provide important insights into the physiological function of the gill as well as its role as an environmental interface. We discuss the significance of these findings in the context of exploratory physiological and toxicological studies.

Introduction

Establishing the complete proteome of an organism provides a strong base from which hypotheses concerning physiological adaptations or responses to stress can be designed. Proteomic profiling is particularly useful for identifying new biomarker proteins, as it describes the biological characteristics underlying a particular phenotype.[1–4] It can also be applied to study the effects of emerging contaminants whose bioactive mechanisms may be complex and not easily predicted using standard toxicological methods.[5] Due to the complexity of the physiological stress response, single proteins are rarely specific or sensitive enough to be used reliably as biomarkers.[4] By studying entire biochemical pathways, proteomics can delineate characteristic patterns of change specific to a stimulus. In clinical settings, using a panel of biomarker proteins increases the specificity and accuracy of disease diagnosis relative to using an individual protein alone.[6,7]

While genomic studies provide valuable information, the transcriptome does not account for the posttranscriptional and posttranslational regulation of protein expression. In many cases, there is poor correspondence between changes in transcript level and protein expression.[8,9] The proteome accounts for these complex regulatory processes and potentially provides a much more accurate snap-shot of an organism's physiological status. However, proteome profiling is much more challenging than transcriptome analysis, and to date, technological restrictions have severely limited the number and type of proteins analyzed. Two-dimensional (2D) gel electrophoresis protein separation combined with tandem mass spectrometry (MS/MS) is the traditional method used to assess protein expression and proportional abundance changes between samples.[10,11] The cumbersome nature of this method reduces the number of proteins addressed, with most studies identifying <100 proteins.[1,12,13] In addition, membrane proteins tend to be under-represented in gel-based analyses because hydrophobic proteins may precipitate during isoelectric focusing and are not transferred to the subsequent sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis separation.[14,15]

Departments of [1]Chemistry and [2]Biological Sciences, University of Alberta, Edmonton, Alberta, Canada.
*These two authors contributed equally to this work.

Two-dimensional liquid chromatography–electrospray ionization MS/MS (LC-ESI MS/MS) shotgun methods typically identify more proteins than gel-based techniques and provide a more favorable separation environment for hydrophobic proteins.[15] Wang *et al.*[16] used LC-ESI MS/MS to identify over 1200 proteins from the cytosolic fraction of the zebrafish liver, increasing proteome coverage by more than an order of magnitude over previous studies of whole fish tissues.[17–21] Recently, close to 1400 proteins were identified in zebrafish embryos using similar techniques.[22] These studies illustrate the power and sensitivity of LC-ESI MS/MS shotgun methods and show that a comprehensive proteome is within reach utilizing a combination of existing techniques.

The zebrafish (*Danio rerio*) is an excellent candidate for proteomic studies as it is a well-characterized vertebrate model widely used in a variety of disciplines, including genetics, developmental biology, and physiology.[23–25] The genome of the zebrafish is mostly sequenced and partially annotated, which facilitates proteomic analyses by allowing for the identification and characterization of proteins using existing databases. Fish are a valuable model system for investigating the impact of toxins on aquatic environments, and the zebrafish is gaining popularity as a model for this discipline.[26–31] Fish require a large surface area for gas and ion exchange with their environment, and these functions are largely accomplished by the gills.[32] The gills can make up >60% of the total surface area of the animal[33] and provide a direct route for the uptake of contaminants.[34]

The goal of the current study was to develop and apply a 2D LC-ESI MS/MS shotgun technique for the analysis of the zebrafish gill and determine the proteome coverage achievable using current state-of-the-art technology. This work provides a baseline proteome of the zebrafish gill for use in physiological, developmental, and toxicological studies. The analysis of the proteome focused on proteins relevant to the normal function of the gill and on biomarkers of toxicant and stress exposure. Using our method, we were able to identify and partially characterize the most complete proteome of a fish tissue published to date. The data provided excellent coverage of predicted biochemical pathway components and identified numerous established and potential biomarkers. In this context, we also highlight a number of bioinformatics challenges that must be resolved to facilitate the widespread use of high-throughput methods for quantitative proteomic profiling.

Materials and Methods

Chemicals and reagents

All chemicals and reagents were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Tricaine methanesulfonate was purchased from Syndel Laboratories Ltd. (Vancouver, BC, Canada). The total protein extraction kit was purchased from Biochain Institute Inc. (Hayward, CA) and the bicinchoninic acid (BCA) assay kit was purchased from Pierce Biotechnology (Rockford, IL). Dithiothreitol (DTT) and iodoacetamide (IAA) were purchased from BioRad (Mississauga, ON, Canada). Sequencing-grade modified trypsin and LC-MS grade acetonitrile, acetone, water, and methanol (MeOH) were purchased from Fisher Scientific Canada (Edmonton, AB, Canada).

Gill tissue collection and sample preparation

A total of 24 zebrafish (12 male, 12 female, body mass 200–400 mg) were taken from the Zebrafish Breeding Facility at the University of Alberta Biosciences Aquatics Facility. All fish were reared and maintained at 28°C. Nacre strain (A/B background) zebrafish were used in this study.

To reduce the content of blood and plasma proteins in the gill tissue, the circulatory system was flushed with ice-cold phosphate-buffered saline (PBS) containing 50 IU/mL Na-heparin. Perfusion cannulae consisted of a length of PE10 tubing with an 8-mm-long tip fabricated from a 30 G needle. The cutting edge of the needle was removed and the tip was polished. Fish were anaesthetized in a 1.0 g/L solution of tricaine methanesulfonate and positioned ventral-side up on a moistened sponge under a dissecting microscope. The body cavity was then opened and quickly rinsed with ice-cold PBS. The heart was exposed by blunt dissection and one ventricular wall was carefully pierced with a sharp 30 G needle. The cannula was inserted and held in place by clamping the ventricle wall around it using blunt forceps. The heart was subsequently perfused at 300 μL/min with ice-cold PBS containing 50 IU/mL Na-heparin until gill tissue was visibly cleared of blood (1–5 min).

A full workflow is outlined in Figure 1. After perfusion, whole gill baskets were excised and placed in a flat-bottomed microcentrifuge tube on dry ice, and proteins extracted according to the protocol provided by the manufacturer (Biochain Institute Inc.). A modification to the protocol involved the addition of 2 mM phenylmethylsulfonyl fluoride (dissolved in EtOH) to the extraction buffer immediately before solubilization. Each gill was homogenized on ice using a disposable pestle, rotated on an end-over-end rotator (4°C, 20 min), and centrifuged (20800 rcf, 4°C, 20 min); the supernatant was collected and stored at −80°C. A BCA assay was conducted to determine the gill protein concentration using bovine serum albumin as the standard.

Acetone precipitation and in-solution digestion

The individual gill protein extracts were pooled to generate one sample. Proteins were reduced with 900 mM DTT for 1 h at 37°C, cooled at room temperature, and alkylated with 2.2 molar equivalents of 450 mM IAA for 1 h at room temperature in the dark. The concentrations of DTT and IAA were chosen to reduce sample dilution upon reduction and alkylation. For every 5 μg of protein, 0.225 μmol of DTT and 0.500 μmol of IAA were used to reduce and alkylate the disulfide bonds, respectively. Proteins were then precipitated with four times the volume of acetone (precooled at −80°C), kept at −20°C overnight, and centrifuged (20800 rcf, 4°C, 10 min); the supernatant was removed and discarded.

The protein pellet was partially solubilized in 100 mM NH_4HCO_3, sonicated in an ice bath (1 min), vortexed (1 h, 4°C), and centrifuged (20800 rcf, 4°C, 10 min). The supernatant was collected and diluted to 50 mM NH_4HCO_3. Solubilization of the protein in the remaining pellet proceeded using the following solvents sequentially: 60% MeOH, 8 M urea, and 1% SDS. A BCA assay was conducted to determine the protein concentration in each protein sample. The urea and SDS samples were diluted to 1 M and 0.05%, respectively, before digestion with trypsin. The trypsin-to-protein ratio was 1:25 (w/w), and digestion occurred at 37°C overnight.

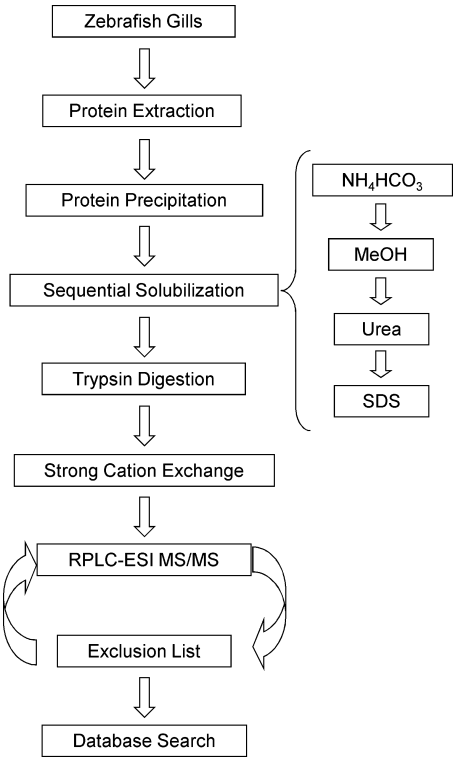

FIG. 1. Workflow for proteome analysis of zebrafish gill.

Cation exchange chromatography

Each of the four peptide samples was further separated on a PolySULFOETHYL A column (PolyLC, Columbia, MD, 5 μm, 300 Å, 2.1×250 mm) by strong cation exchange (SCX) on an Agilent 1100 HPLC system (Palo Alto, CA). Mobile Phase A consisted of 10 mM KH_2PO_4 (pH 2.76) and Mobile Phase B consisted of 10 mM KH_2PO_4 and 0.5 M KCl (pH 2.76). The elution gradient was 0% B for 7 min, 0–6% B from 7 to 8 min, 6–28% B from 8 to 36 min, 28–40% B from 36 to 44 min, 40–60% B from 44 to 49 min, 60–100% B from 49 to 53 min, 100% B from 53 to 58 min, 100–0% B from 58 to 60 min, and held at 0% B for 10 min at a flow rate of 0.200 mL/min. Fractions were collected at 1 min intervals from 16 to 70 min.

Peptide desalting and quantitation

Before LC-ESI MS/MS, the samples collected from the SCX were subjected to salt removal on a Polaris C18 A column (Palo Alto, CA, 3 μm, 300 Å, 4.6×50 mm) on an Agilent 1100 HPLC system with a ultraviolet detector. In addition to salt removal, simultaneous peptide quantitation was performed based on the area of the peptide peak at 214 nm.[35] Mobile Phase A consisted of 0.1% trifluoroacetic acid in water and Mobile Phase B consisted of 0.1% trifluoroacetic acid in acetonitrile. The elution gradient was 2.5% B for 5.50 min, 2.5–85% B from 5.50 to 5.51 min, 85% B from 5.51 to 15.50 min, 85–2.5% B from 15.50 to 15.51 min, and held at 2.5% B from 15.51 to 30.00 min at a flow rate of 1.000 mL/min. Fraction collection occurred at 7.50 min for a total of 1.10 min. Adjacent fractions from the SCX fractionation were pooled if the total amount of peptides in individual fractions was <1 μg.

LC-ESI Q-Tof MS and MS/MS analysis

The desalted SCX peptide samples were analyzed using a Q-Tof Premier™ mass spectrometer equipped with a nanoACQUITY® UPLC system (Waters, Milford, MA). Approximately 1 μg of sample was loaded onto an Atlantis dC18 column (Waters, 3 μm, 100 Å, 75 μm×150 mm). Mobile Phase A consisted of 0.1% formic acid in water and Mobile Phase B consisted of 0.1% formic acid in acetonitrile. The elution gradient was 2–7% B from 0 to 2 min, 7–20% B from 2 to 85 min, 20–30% B from 85 to 110 min, 30–45% B from 110 to 115 min, 45–90% B from 115 to 120 min, held at 90% B for 5 min, and 90–2% B for 5 min at 250 nL/min. A precursor ion exclusion (PIE) strategy in LC MS/MS was used where ions whose acquired MS/MS spectra resulted in positive peptide identification were excluded from MS/MS acquisition in the subsequent run of the adjacent SCX sample.[36] This strategy improves the overall peptide identification efficiency, as well as enables the sequencing of low abundance peptides.

Protein identification from MS/MS data

Raw LC MS/MS data were lock-mass corrected, deisotoped, and converted to peak list files by ProteinLynx Global Server 2.2.5 (Waters). Peptide sequences were identified by automated database searches of peak list files generated from each LC-ESI MS/MS run using the Mascot search program (Matrix Science, London, United Kingdom). Mascot searches were performed individually for the SCX fractions of the tryptic digest generated from each of the solvents used to dissolve the proteome sample. These individual search results were then merged for each solubilization method. A final merge of all results from four solvents generated the final protein list.

For each LC-ESI MS/MS run, database searches were restricted to *D. rerio* in the NCBInr database (searched in August 2007 with a database containing 4815286 sequences and 1665828716 residues). The following search parameters were selected for all database searches: enzyme, trypsin; maximum missed cleavages, 1; fixed modifications, carbamidomethyl (C); variable modifications, N-acetyl (protein), oxidation (M), pyro-glu (N-term E), pyro-glu (N-term Q); peptide tolerance, ±30 ppm; fragment mass tolerance, ±0.2 Da; peptide charge, 1+, 2+, 3+; instrument type, ESI-QUAD-TOF modified to include immonium ions and a-series ions as possible fragmentations. Two additional variable modifications were selected for the urea-solubilized samples: carbamyl (K) and carbamyl (N-term). The search results, including protein names, accession IDs, molecular mass, unique peptide sequences, ion score, Mascot threshold score for identity, calculated molecular mass of the peptide, and the difference (error) between the experimental and calculated masses, were extracted to Excel files using in-house software. The identified peptides with scores lower than the Mascot threshold score for identity at a confidence level of 95% were subsequently removed from the protein list. Redundant peptides identified for different proteins were deleted, and redundant proteins identified under the same gene name but different accession IDs were also deleted. The final unique protein or peptide list was generated by merging all the protein or peptide lists according to the following rules: only unique proteins (under unique gene names) and peptides with the highest scores were retained; each peptide was associated to one unique

protein; only the first hit within each identified protein group was kept as a representative protein. All redundant peptides with lower identification scores were deleted, in addition to redundant proteins with either lower scores or lower number of peptides.

The target-decoy search strategy was applied to assess the false-positive peptide matching rate in our analysis by searching the MS/MS spectra against the forward and reversed zebrafish proteome sequences.[37,38] This approach involved re-searching the matched spectra from the Mascot database search using the forward or correct proteome sequences against the reversed proteome sequence or decoy. The decoy peptide matches with scores above the threshold scores at the 95% confidence level were then compared to those in the forward sequence search. If the score of an MS/MS spectrum matched with a decoy peptide was equal to or higher than that of the same spectrum matched with a correct peptide, a false-positive match was registered. The false-positive matching rate was calculated using the following equation:

$$2 \times n_{\text{reversed}} / [n_{\text{forward}} + n_{\text{reversed}}],$$

where n_{reversed} and n_{forward} are the number of matches from the reversed (decoy) and forward (correct) sequence, respectively.

Protein annotation

Figure 2 illustrates the general workflow used for the analysis of the final list of identified proteins generated by the methods described above. Proteins identified with only an alphanumerical code and the descriptors "hypothetical," "novel," and/or "predicted" were further analyzed using the Universal Protein Resource[39] ID mapping and retrieve tools to identify alternate, biologically relevant protein names for subsequent categorization and pathway analysis.

Gene Ontology (GO) terms[40] associated with the identified proteins were retrieved using The Database for Annotation, Visualization, and Integrated Discovery (DAVID) functional annotation tool.[41] For proteins that DAVID was unable to retrieve any associated GO terms, GI accession numbers were mapped to GO terms using the UniProt ID mapping tool. When necessary, GO codes were translated using the AmiGO tool on the GO website.

In some cases, proteins were categorized according to their participation in specific biochemical pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database,[42] the DAVID functional annotation tool, and manual analysis based on GO terms and primary literature searches. For pathway coverage data, the gill proteome was manually analyzed against KEGG pathways predicted for *D. rerio*.

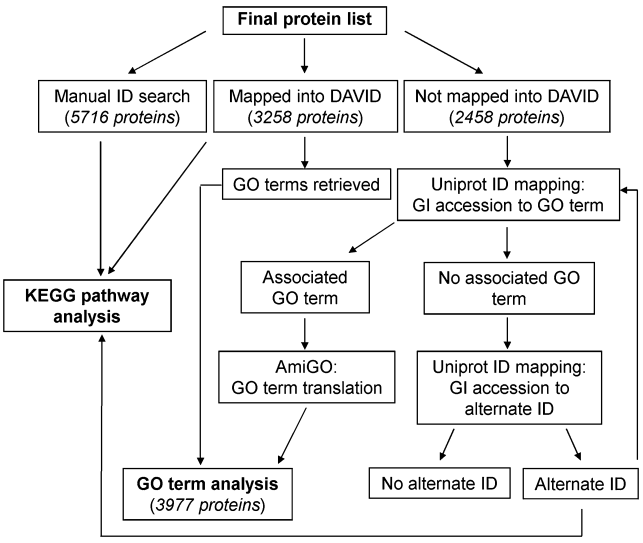

FIG. 2. Workflow for bioinformatics analysis of zebrafish gill.

Results

Using the advanced LC-ESI-MS/MS method developed in our lab, we have identified 5716 proteins expressed in the zebrafish gill (see Supplemental Table S1, available online at www.liebertonline.com). Raw data are available upon request from the authors; to increase data accessibility and provide interactive data processing of the raw and metadata generated from one or more organs of the zebrafish, efforts are currently under way to establish an integrated, publicly available database of the zebrafish proteome. The false-positive discovery rate was estimated at 1.34% using a previously described method.[37,38] Although 49% of the proteins identified were from single-peptide matches, the confidence of protein identification based on single-peptide matches is still high, as judged by the low false-positive matching rate determined from the dataset. It should be noted that, compared to a low-resolution mass analyzer such as an ion trap for recording collision-induced dissociation mass spectra, the Q-Tof instrument generates MS/MS spectra with a high resolution and a high mass measurement accuracy, which contributes to a low false-positive matching rate.

To improve proteome coverage, we employed several novel approaches, including sequential solubilization of proteins in the sample for simplifying the proteome before 2D-LC MS/MS analysis. Tissue samples contain a complex mixture of proteins with varying solubilities. Exploiting a number of solvents with different properties after acetone precipitation can fractionate the proteome sample into subproteomes and allow a more thorough analysis by 2D-LC MS/MS. In a previous study from our lab,[16] similar methods were used to expand proteome coverage to over 1200 proteins in just the cytosolic fraction of the zebrafish liver. Figure 3 shows the distribution of the number of proteins identified in each of the four fractions generated by the sequential solubilization protocol. Eight hundred three (14%), 104 (2%), 1297 (23%), and 860 (15%) proteins were exclusively found in the NH_4HCO_3, MeOH, urea, and SDS fractions, respectively. This data clearly show that when dealing with a complex sample, sequential solubilization significantly increases proteome coverage compared to a single solubilization protocol. Previously published proteomic analyses on fish tissue typically identified 10–100 proteins. Only two studies have identified more than 1000 proteins (1384 proteins[22] and 1204 proteins[16]) and both used 2D LC-ESI MS/MS. Our work clearly demonstrates the potential of this technique for more comprehensive coverage.

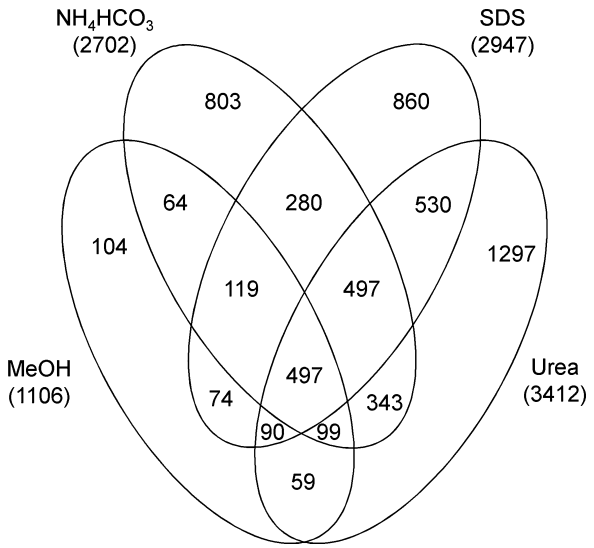

FIG. 3. A comparison of the protein identification results from the four protein solubilization techniques. The numbers in parentheses indicate the total number of proteins identified in each fraction.

Sequential solubilization is a powerful technique to fractionate the protein sample, but steps must be taken during the subsequent analysis to ensure that protein identification is accurate. Urea and SDS are effective for solubilizing hydrophobic proteins,[43,44] but these solutes can introduce problems in shotgun methods based on proteome digestion and LC MS/MS analysis.[45] Urea efficiently solubilizes membrane proteins, but the formation of cyanic acid can be problematic because it reacts with primary amines. This side reaction can be accounted for in database searches by selecting the variable modifications of carbamylation at lysines and N-termini of proteins. SDS molecules can strongly bind to some peptides and suppress peptide signals in MS analysis. To reduce SDS interference, we used SCX before reversed phase separation and introduction of the sample into the mass spectrometer. Using these additional sample preparation techniques, we were able to increase proteome coverage.

Proteome coverage was also improved through the use of an optimal PIE technique developed in our lab.[36] Although the initial protein sample was fractionated via solubilization, SCX, and RP before analysis, it is still extremely complex and not all peptides are able to be analyzed by MS/MS. The PIE method involves generating a list of positively identified peptides from one SCX fraction and then excluding these ions from MS/MS spectral acquisition in the subsequent SCX fractions. In other studies, the application of the PIE method improved the likelihood of scanning less abundant peptides and increased proteome coverage by 45% or more.[36]

Protein identification was based on the sequence match of one or more peptides. Protein modifications are difficult to detect, as sequence coverage by peptides is usually low (in this work, only a few protein isoforms were detected where different regions of peptide sequences were identified). The zebrafish genome contains approximately 17330 genes that can potentially code for proteins.[46] The identification of 5716 unique proteins therefore represents 36% of the total potential proteome expressed in the gill alone. Given that only a single tissue was investigated under steady-state conditions, we feel it is reasonable to suggest that this represents a significant proportion of the total number of proteins present in the gill at the time of sampling.

A significant issue encountered in the subsequent proteome analysis was that 44% of the proteins identified in the gill were described only as hypothetical, predicted, novel, or by an alphanumeric code (e.g., Zebrafish Genome Collection [ZGC] proteins). The lack of descriptive names made it difficult to classify these proteins in a meaningful biological context, particularly in the absence of associated GO terms (see below). To resolve this issue and improve the biological analysis of the proteome, we manually searched the available databases to identify more relevant names for the 2139 hypothetical proteins. Alternate names were retrieved for 1105 of these proteins with the majority of these described only as novel proteins or by ZGC numbers. Similar searches were performed for novel and ZGC protein names, and a number of more biologically relevant descriptors were retrieved for use in the pathway analysis described below.

Some degree of annotation information was retrieved for approximately 3977 proteins. The DAVID functional annotation tool[41] was able to retrieve GO terms for roughly 57% of the identified gill proteins, with the remaining annotations found using the UniProt ID mapping tool. The paucity of information for the other approximately 1739 proteins highlights the urgent need for advancements in annotation and bioinformatics techniques to facilitate large-scale proteomic analyses of this type.

Despite this gap in knowledge, the proteome exhibited excellent coverage of a number of biochemical pathways relevant to the function of the gill (Table 1) as well as numerous established biomarkers (Table 2). We utilized the KEGG database to examine 12 pathways important to the physiological role of the zebrafish gill, and coverage ranged from 16% to 88% (Table 1). In several cases, we identified multiple proteins matching a single component of a particular pathway. An example of the extent of coverage we were able to achieve is given for glycolysis (Fig. 4), where we identified multiple proteins for many steps in the pathway.

Discussion

A major drawback of transcriptomic studies is the potentially weak correspondence between transcript levels and protein expression.[9] The expression of some proteins can vary more than 20-fold in the absence of changes in mRNA levels, and similar changes in transcript levels can occur with no effect on protein expression.[47] The expression and function of proteins determines the biology of an organism; therefore, developing a method for analyzing a complete proteome is particularly valuable for physiological studies.

The extensive level of proteome coverage we have established in this study should greatly facilitate subsequent physiological, toxicological, and comparative studies of the fish gill. Unlike genomic studies, we clearly identify which proteins are expressed in the zebrafish gill under control conditions and provide important baseline information. With this database, we can now design specific hypotheses to investigate the function of single proteins, protein–protein interactions, or whole biochemical pathways. To emphasize the utility and relevance of proteomic profiling to these types of

Table 1. Coverage of Selected Biochemical Pathways by Proteins Expressed in the Zebrafish Gill

Pathway	# *predicted*	# *identified*	% *coverage*
Citrate cycle (TCA cycle)	16	14	88
Proteasome	25	22	88
Ribosome (small subunit)	27	23	85
Ribosome (large subunit)	28	22	79
Pyruvate metabolism	16	12	75
Glycolysis	23	17	74
Metabolism of xenobiotics by cytochrome p450	11	7	64
Nitrogen metabolism: reduction and fixation	8	5	63
Fatty acid elongation in mitochondria	5	3	60
Starch and sucrose metabolism	10	6	60
Urea cycle and metabolism of amino groups	15	9	60
Fatty acid metabolism	14	8	57
Oxidative phosphorylation	79	37	47
Ubiquitin-mediated proteolysis	76	12	16

Pathway components were predicted by the Kyoto Encyclopedia of Genes and Genomes database.
TCA, trichloroacetic acid.

studies, the proteome was analyzed based upon established physiological roles of the fish gill and the biochemical pathways supporting these functions.

Respiration and ion and acid–base homeostasis

The gill is the predominant site of oxygen uptake in fish and it contains a population of cells specialized for carrying out this function.[32] Approximately 14 proteins with GO terms related to oxygen binding, oxygen transport, or reactive oxygen species metabolism were identified in the zebrafish gill proteome (see Supplemental Table S2). These consisted of various alpha and beta hemoglobin subunits and myoglobin. Although all gills were fully flushed before analysis, it is likely that the high sensitivity of our methods identified proteins originating from trace amounts of blood in the tissue.

Freshwater fish maintain a high internal osmolarity relative to their environment, and the gill has a substantial architecture of ion regulatory machinery to sustain that gradient.[48] Zebrafish naturally inhabit ion-poor environments, but they are tolerant to a range of salinities from essentially ion-free to brackish waters.[49] Greater than 60 ion transport-related proteins were identified in the gill proteome (Supplemental Table S2), encompassing the majority of proteins hypothesized to be responsible for ion regulation and acid–base homeostasis in freshwater fish.[50,51] These include numerous isoforms of

Table 2. Biomarker Proteins Identified in the Zebrafish Gill Proteome

Protein name	*Protein function*
Acyl-coenzyme A dehydrogenase family, member 8	2-Methylnaphthalene degradation
Alcohol dehydrogenase 5	2-Methylnaphthalene degradation, methane metabolism, metabolism of xenobiotics by cytochrome p450
Glutathione s-transferase M	Metabolism of xenobiotics by cytochrome p450
Glutathione s-transferase PI	Metabolism of xenobiotics by cytochrome p450
Cytochrome p450 2AD3	Metabolism of xenobiotics by cytochrome p450
Cytochrome p450 family 2, subfamily J, polypeptide 24	Metabolism of xenobiotics by cytochrome p450
Cytochrome p450 family 3, subfamily C, polypeptide 1	Metabolism of xenobiotics by cytochrome p450
Novel protein similar to cytochrome p450 family 2, subfamily j	Metabolism of xenobiotics by cytochrome p450
NADPH-cytochrome p450 oxidoreductase	Metabolism of xenobiotics by cytochrome p450
ZGC:66393 (UDP glucuronosyltransferase 1 family A, B)	Metabolism of xenobiotics by cytochrome p450
Predicted: epoxide hydrolase 1, microsomal, partial	Metabolism of xenobiotics by cytochrome p450
Hydroxysteroid (17-beta) dehydrogenase 4	Caprolactam degradation
Sirtuin 2 (silent mating type information regulation 2, homolog)	Caprolactam degradation
Enoyl coenzyme A hydratase, short chain, 1, mitochondrial	Caprolactam degradation, benzoate degradation via coA ligation
Acetyl-coA acetyltransferase 2	Benzoate degradation via coA ligation
Manganese superoxide dismutase	Response to oxidative stress
Predicted: similar to extracellular superoxide dismutase	Response to oxidative stress
Superoxide dismutase 1, soluble	Response to oxidative stress
Homogentisate 1,2-dioxygenase	Styrene degradation
Adenosine deaminase, RNA-specific	Atrazine degradation
Metallothionein 2	Response to methylmercury
Arsenate resistance protein 2	Response to arsenic
Predicted: similar to heat shock cognate 70 kDa protein	Response to cadmium ion, response to Xenobiotic stimulus, response to heat
Vitellogenin 1	Response to endocrine disrupting compound
Vitellogenin 2	Response to endocrine disrupting compound
Vitellogenin 3 precursor	Response to endocrine disrupting compound

NADPH, nicotinamide adenine dinucleotide phosphate; ZGC, Zebrafish Genome Collection; UDP, uridine diphosphate.

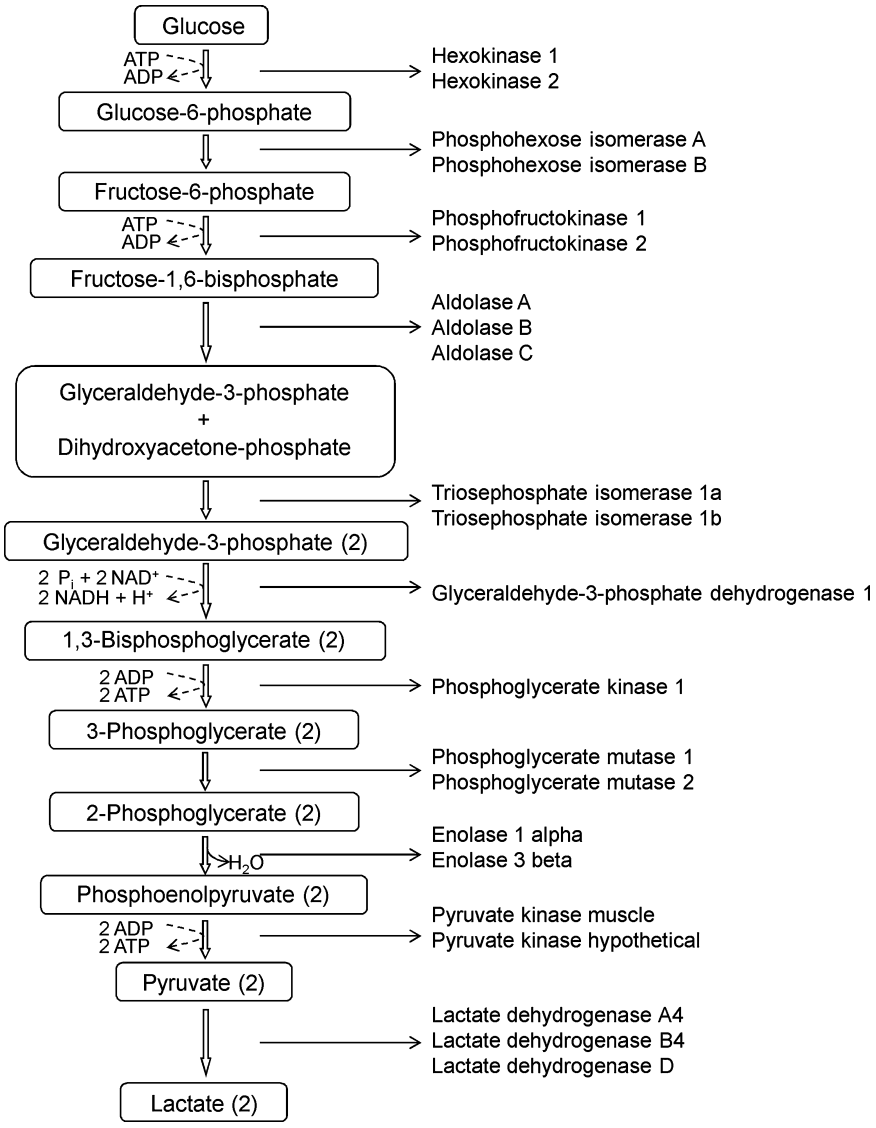

FIG. 4. Zebrafish gill proteome coverage of the enzymes involved in the metabolism of glucose to lactate via glycolysis and fermentation.

Na^+/K^+-ATPase and V-type H^+-ATPases, sodium bicarbonate cotransporters, anion exchangers, and several carbonic anhydrases. However, several proteins of interest such as the Na^+/H^+ exchanger and epithelial Na^+ channel were not identified in our analysis. A number of other Na^+, K^+, and Ca^{2+} channels typically found in muscle and nervous tissue were also identified. This is expected, given that gill tissue contains both vascular smooth muscle cells, neuro-epithelial cells, and mitochondrion-rich cells.[51–53]

Protein turnover

The gill plays a pivotal role in maintaining physiological homeostasis and, as such, it is expected that gill tissue would exhibit significant plasticity in the face of environmental changes.[54] The gills respond to changes in the internal and external environment by changing their complement of membrane transport proteins,[55] the levels of enzymes involved in energy metabolism,[56] and even the physical structure of the lamellae.[57] Protein synthesis rates in the fish gill are consistently high relative to those measured in other tissues.[54,58,59] It is not surprising, therefore, that we identified a large number of proteins related to protein synthesis and degradation (see Supplemental Table S3). The ribosome is the primary cellular component involved in protein synthesis, and our analysis identified 23 of 27 and 22 of 28 of the expected small and large ribosomal subunit proteins, respectively. The proteasome and lysosome are responsible for protein degradation[60] with short-lived proteins generally degraded by the proteasome and long-lived proteins degraded by the lysosome.[60,61] The proteasome also plays an important role in a variety of basic cellular processes, such as the regulation of cell cycle, division, development, and differentiation, and modulation of immune and inflammatory responses.[62,63] We identified 22 out of 25 of the protein subunits of the 26S proteasome. Information was not available for

expected lysosomal proteins for *D. rerio*. Only 12 of the expected 76 proteins involved in ubiquitin-mediated proteolysis were present in the identified zebrafish gill proteome (Supplemental Table S3). It is possible that a number of proteins in this pathway were not identified due to relatively low protein abundance in the sample.

Energy metabolism

High protein turnover rates and ion and acid–base regulation are energetically expensive processes and necessitate a relatively high organ-specific metabolic rate. In rainbow trout, gill NaCl uptake alone can account for 37% of tissue oxygen demand and up to 4% of the animal's total energy expenditure.[64] The gill is largely an aerobic tissue that exhibits a strong dependence upon glucose and lactate for ATP production.[65,66] The zebrafish gill proteome contained the full complement of glycolytic enzymes with several isoforms present for most proteins in the pathway (Fig. 4). A number of proteins critical to the regulation of carbohydrate metabolism were also found, including several subunits of AMP-activated protein kinase, an insulin-like growth factor 2 receptor, a glucagon-like peptide-2 receptor, pyruvate dehydrogenase phosphatase 2, as well as a predicted protein similar to glucose transporter X. In total, 222 proteins involved in the metabolism of carbohydrates, amino acids, and lipids were identified in the gill (see Supplemental Table S4).

Coverage is slightly less comprehensive for biochemical pathways such as oxidative phosphorylation (47%), whose constituents are largely localized to membranes. We have far greater coverage for pathways such as the citrate cycle (88%) or glycolysis (74%) that contain few membrane-associated proteins. Greater than 240 integral membrane proteins were identified from the zebrafish gill proteome. Although this number represents only 4% of the proteins identified, it is still far more than the numbers identified from fish tissue in any other study to date ($\sim$45 membrane proteins[22]; 78 membrane proteins[67]; see Supplemental Table S5). About 8%, 6%, 17%, and 19% of the integral membrane proteins were identified in the NH_4HCO_3, MeOH, urea, and SDS fractions, respectively. The remaining 53% were identified in more than one solvent, but always included urea or SDS as one of the solvents, further illustrating the importance of these two solvents in the solubilization of membrane proteins. The number of integral membrane proteins we identified is still likely a substantial underestimate of the true number present considering that annotation information is unavailable for >30% of the proteome.

Biomarkers

Comprehensive proteomic profiling is a powerful technique for biomarker discovery and it will be an important tool for defining the bioactivity of new compounds. A number of well-characterized biomarkers of environmental stress and toxicant exposure were expressed in the zebrafish gill (Table 2). Only five proteins were predicted to be involved in the metabolism of xenobiotics by cytochrome P450 in zebrafish, but we identified approximately 10 unique proteins associated with this GO term in the gill proteome. A number of proteins related to the degradation of specific toxins like 1- and 2-methylnapthalene, caprolactam, and atrazine were also expressed. Several proteins significant to metal toxicity were identified, including metallothionein 2.

In addition to the specific biomarkers referred to above, >40 heat shock proteins (HSPs), chaperonins, and related proteins were expressed in the gill (see Supplemental Table S6). These proteins generally interact with unfolded or denatured proteins and facilitate refolding, repair, or degradation processes.[68] HSPs are ubiquitously present in all organisms and their expression levels are sensitive to a variety of stresses, including heat, hypoxia, pathogens, and acid–base and osmotic disturbances. This well-conserved response has lead to considerable interest in the use of HSPs as biomarkers of stress in animals.[68] The complexity and species-to-species variability of the heat shock response, however, has prohibited the use of HSPs as specific stress indicators in fish.[69] The sensitivity of proteomic profiling methods may help to clarify the relationship between HSP expression patterns and other biomarkers in response to a particular stimulus. Exploiting the well-characterized zebrafish model system can also minimize the uncertainties of working with lesser studied fish species.

Conclusions

We present the first large-scale proteome profile of a teleost fish tissue. Using advanced sample preparation and data-dependent MS/MS analysis techniques developed in our lab, we have identified 5716 proteins expressed in the zebrafish gill. The data provide critical baseline information on tissue-specific protein expression for future studies on gill physiology and aquatic toxicology. Future work will proceed on two fronts: one focused on refining the current method for increased proteome coverage (e.g., identifying more membrane proteins by enriching membrane proteome fractions from tissue extracts) and the second on the development and application of quantitative profiling techniques. We are currently optimizing a 2MEGA isotope-labeling method for relative quantification of proteins from differently treated samples.[70–72] We envision that the combination of this labeling method and the current protocol should provide a means of carrying out large-scale quantitative proteome profiling studies. Analyzing such large-scale quantitative datasets will require advanced bioinformatics tools that are currently unavailable. With the progression of genomic sequencing of nonmodel systems and improved bioinformatics analyses, this technique should be applicable for physiological studies on any organism.

Acknowledgments

Funding was provided by the Alberta Ingenuity Centre for Water Research and the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs program (L.L.). T.J.M. was supported by postdoctoral fellowships from Natural Sciences and Engineering Research Council of Canada and the Izaak Walton Killam Foundation.

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. Bosworth CA, Chou CW, Cole RB, Rees BB. Protein expression patterns in zebrafish skeletal muscle: initial

characterization and the effects of hypoxic exposure. Proteomics 2005;5:1362–1371.
3. Wei Y, Chan LL, Wang D, Zhang D, Wang J, Dai J. Proteomic analysis of hepatic protein profiles in rare minnow (*Gobiocypris rarus*) exposed to perfluorooctanoic acid. J Proteome Res 2008;7:1729–1739.
4. Polanski M, Anderson NL. A list of candidate cancer biomarkers for targeted proteomics. Biomarker Insights 2006;1:1–48.
5. MacCormack TJ, Goss GG. Identifying and predicting biological risks associated with manufactured nanoparticles in aquatic ecosystems. J Ind Ecol 2008;12:286–296.
6. Mor G, Visintin I, Lai Y, Zhao H, Schwartz P, Rutherford T, *et al*. Serum protein markers for early detection of ovarian cancer. Proc Natl Acad Sci 2005;102:7677–7682.
7. Xiao T, Ying W, Li L, Hu Z, Ma Y, Jiao L, *et al*. An approach to studying lung cancer-related proteins in human blood. Mol Cell Proteomics 2005;4:1480–1486.
8. Marchant GE. Toxicogenomics and toxic torts. Trends Biotechnol 2002;20:329–332.
9. Pandey A, Mann M. Proteomics to study genes and genomes. Nature 2000;405:837–846.
10. Chanson A, Sayd T, Rock E, Chambon C, Santé-Lhoutellier V, Potier de Courcy G, Brachet P. Proteomic analysis reveals changes in the liver protein pattern of rats exposed to dietary folate deficiency. J Nutr 2005;135:2524–2529.
11. Strey CW, Winters MS, Markiewski, MM, Lambris JD. Partial hepatectomy induced liver proteome changes in mice. Proteomics 2005;5:318–325.
12. Melle C, Ernst G, Schimmel B, Bleul A, Koscielny S, Wiesner A, *et al*. Biomarker discovery and identification in laser microdissected head and neck squamous cell carcinoma with ProteinChip technology, two-dimensional gel electrophoresis, tandem mass spectrometry, and immunohistochemistry. Mol Cell Proteomics 2003;2:443–452.
13. Tay TL, Lin Q, Seow TK, Tan KH, Hew CL, Gong Z. Proteomic analysis of protein profiles during early development of the zebrafish, *Danio rerio*. Proteomics 2006;6:3176–3188.
14. Macher BA, Yen TY. Proteins at membrane surfaces—a review of approaches. Mol Biosyst 2007;3:705–713.
15. Speers AE, Wu CC. Proteomics of integral membrane proteins—theory and application. Chem Rev 2007;107:3687–3714.
16. Wang N, MacKenzie L, De Souza AG, Zhong H, Goss GG, Li L. 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 2007;6:263–272.
17. Brunt J, Hansen R, Jamieson DJ, Austin B. Proteomic analysis of rainbow trout (*Oncorhynchus mykiss*, Walbaum) serum after administration of probiotics in diets. Vet Immunol Immunopathol 2008;121:199–205.
18. Ky CL, de Lorgeril J, Hirtz C, Sommerer N, Rossignol M, Bonhomme F. The effect of environmental salinity on the proteome of the sea bass (*Dicentrarchus labrax* L.). Anim Genet 2007;38:601–608.
19. Martin SAM, Mohanty BP, Cash P, Houlihan DF, Secombes CJ. Proteome analysis of the Atlantic salmon (*Salmo salar*) cell line SHK-1 following recombinant IFN-γ stimulation. Proteomics 2007;7:2275–2286.
20. Mezhoud K, Praseuth D, Puiseux-Dao S, François JC, Bernard C, Edery M. Global quantitative analysis of protein expression and phosphorylation status in the liver of the medaka fish (*Oryzias latipes*) exposed to microsystin-LR: I. Balneation study. Aquat Toxicol 2008;86:166–175.
21. Smith RW, Wang J, Bucking CP, Mothersill CE, Seymour CB. Evidence for a protective response by the gill proteome of rainbow trout exposed to X-ray induced bystander gills. Proteomics 2007;7:4171–4180.
22. Lucitt MB, Price TS, Pizarro A, Wu W, Yocum AK, Seiler C, *et al*. Analysis of the zebrafish proteome during embryonic development. Mol Cell Proteomics 2008;7:981–994.
23. Dahm R, Geisler R. Learning from small fry: the zebrafish as a genetic model organism for aquaculture fish species. Mar Biotechnol 2006;8:329–345.
24. Jonz MG, Nurse CA. Epithelial mitochondria-rich cells and associated innervation in adult and developing zebrafish. J Comp Neurol 2006;497:817–832.
25. Phelps HA, Neely MN. Evolution of the zebrafish model: from development to immunity and infectious disease. Zebrafish 2005;2:87–103.
26. Henry TR, Spitsbergen JM, Hornung MW, Abnet CC, Peterson RE. Early life stage toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in zebrafish (*Danio rerio*). Toxicol Appl Pharm 1997;142:56–68.
27. Jorgensen A, Andersen O, Bjerregaard P, Rasmussen LJ. Identification and characterisation of an androgen receptor from zebrafish *Danio rerio*. Comp Biochem Physiol 2007;146C:561–568.
28. Kausch U, Alberti M, Haindl S, Budczies J, Hock B. Biomarkers for exposure to estrogenic compounds: gene expression analysis in zebrafish (*Danio rerio*). Environ Toxicol 2008;23:15–24.
29. Ramírez OAB, García FP. Genotoxic damage in zebrafish (*Danio rerio*) by arsenic in waters from Zimapán, Hidalgo, Mexico. Mutagenesis 2005;20:291–295.
30. Schnurstein A, Braunbeck T. Tail moment versus tail length—application of an *in vitro* version of the comet assay in biomonitoring for genotoxicity in native surface waters using primary hepatocytes and gill cells from zebrafish (*Danio rerio*). Ecotoxicol Environ Safety 2001;49:187–196.
31. Wu SM, Zheng YD, Kuo CH. Expression of mt2 and smt-B upon cadmium exposure and cold shock in zebrafish (*Danio rerio*). Comp Biochem Physiol 2008;148C:184–193.
32. Perry SF. The chloride cell: structure and function in the gills of freshwater fishes. Annu Rev Physiol 1997;59:324–357.
33. Rombough P, Moroz B. The scaling and potential importance of cutaneous and branchial surfaces in respiratory gas exchange in larval and juvenile walleye *Stizostedion vitreum*. J Exp Biol 1997;200:2459–2468.
34. Wood C, Kelly SP, Zhou BS, Fletcher M, O'Donnell M, Eletti B, Pärt P. Cultured gill epithelia as models for the freshwater fish gill. Biochim Biophys Acta 2002;1566:72–83.
35. Wang N, Xie C, Young JB, Li L. Off-line two-dimensional liquid chromatography with maximized sample loading to reversed-phase liquid chromatography-electrospray ionization tandem mass spectrometry for shotgun proteome analysis. Anal Chem 2009;81:1049–1060.
36. Wang N, Li L. Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis. Anal Chem 2008;80:4696–4710.
37. Elias J, Gygi S. Target-decoy search strategy for increased confidence in large-scale protein identification by mass spectrometry. Nat Methods 2007;4:207–214.

38. Gong Y, Wang N, Wu F, Cass CE, Damaraju S, Mackey JR, Li L. Proteome profile of human breast cancer tissue generated by LC-ESI-MS/MS combined with sequential protein precipitation and solubilization. J Proteome Res 2008;7:3583–3590.
39. The UniProt Consortium. The Universal Protein Resource (UniProt). Nucleic Acids Res 2008;36:D190–D195.
40. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al*. Gene ontology: tool for the unification of biology. Nat Genet 2000;25:25–29.
41. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003; 4:R60.
42. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, *et al*. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008;36:D480–D484.
43. Aivaliotis M, Corvey C, Tsirogianni I, Karas M, Tsiotis G. Membrane proteome analysis of the green-sulfur bacterium *Chlorobium tepidum*. Electrophoresis 2004;25:3468–3474.
44. Santoni V, Molloy M, Rabilloud T. Membrane proteins and proteomics: un amour impossible? Electrophoresis 2000;21: 1054–1070.
45. Weiner JH, Li L. Proteome of the *Escherichia coli* envelope and technological challenges in membrane proteome analysis. Biochim Biophys Acta Biomembr 2008;1778:1698–1713.
46. Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, *et al*. Ensemble 2008. Nucleic Acids Res 2008;36:D707–D714.
47. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19:1720–1730.
48. Goss GG, Perry SF, Fryer JN, Laurent P. Gill morphology and acid-base regulation in freshwater fishes. Comp Biochem Physiol 1998;119A:107–115.
49. Engeszer RE, Patterson LB, Rao AA, Parichy DM. Zebrafish in the wild: a review of natural history and new notes from the field. Zebrafish 2007;4:21–40.
50. Parks SK, Tresguerres M, Goss GG. Theoretical considerations underlying Na^+ uptake mechanisms in freshwater fishes. Comp Biochem Physiol C 2008;148:411–418.
51. Parks SK, Tresguerres M, Goss GG. Interactions between Na^+ channels and Na^+/HCO_3^- cotransporters in the freshwater fish gill MR cell: a model for transepithelial Na^+ uptake. Am J Physiol 2007;292:C935–C944.
52. Jonz MG, Nurse CA. New developments on gill innervation: insights from a model vertebrate. J Exp Biol 2008;211:2371–2378.
53. Taylor EW, Jordan D, Coote JH. Central control of cardiovascular and respiratory systems and their interactions in vertebrates. Physiol Rev 1999;79:855–916.
54. Lyndon AR, Houlihan DF. Gill protein turnover: costs of adaptation. Comp Biochem Physiol 1998;119A:27–34.
55. Tresguerres M, Parks SK, Katoh F, Goss GG. Microtubule-dependent relocation of branchial V-H+-ATPase to the basolateral membrane in the Pacific spiny dogfish (*Squalus acanthias*): a role in base secretion. J Exp Biol 2006;209:599–609.
56. Driedzic WR, Gesser H. Energy metabolism and contractility in ectothermic vertebrate hearts: hypoxia, acidosis and low temperature. Physiol Rev 1994;74:221–258.
57. Sollid J, De Angelis P, Gundersen K, Nilsson GE. Hypoxia induces adaptive and reversible gross morphological changes in crucian carp gills. J Exp Biol 2003;206:3667–3673.
58. Lewis JM, Costa I, Val AL, Almeida-Val VMF, Gamperl AK, Driedzic WR. Responses to hypoxia and recovery: repayment of oxygen debt is not associated with compensatory protein synthesis in the Amazonian cichlid, *Astronotus ocellatus*. J Exp Biol 2007;210:1935–1943.
59. Lewis JM, Driedzic WR. Tissue-specific changes in protein synthesis associated with seasonal metabolic depression and recovery in the north temperate labrid, *Tautogolabrus adspersus*. Am J Physiol 2007;293:R474–R481.
60. Demarchi F, Brancolini C. Altering protein turnover in tumor cells: new opportunities for anti-cancer therapies. Drug Resist Update 2005;8:359–368.
61. Ohsumi Y. Protein turnover. IUBMB Life 2006;58:363–369.
62. Richardson PG, Hideshima T, Mitsiades C, Anderson K. Proteasome inhibition in hematologic malignancies. Ann Med 2004;36:304–314.
63. Glickman MH, Ciechanover A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol Rev 2002;82:373–428.
64. Morgan JD, Iwama GK. Energy cost of NaCl transport in isolated gills of cutthroat trout. Am J Physiol 1999;277:R631–R639.
65. Mommsen TP. Biochemical characterization of the rainbow trout gill. J Comp Physiol 1984;154B:191–198.
66. Soengas JL, Barciela P, Aldegunde M, Andrés D. Gill carbohydrate metabolism of rainbow trout is modified during gradual adaptation to sea water. J Fish Biol 1995;46:845–856.
67. Nazarian J, Hathout Y, Vertes A, Hoffman EP. The proteome survey of an electricity-generating organ (*Torpedo californica* electric organ). Proteomics 2007;7:617–627.
68. Basu N, Todgham AE, Ackerman PA, Bibeau MR, Nakano K, Schulte PM, Iwama GK. Heat shock protein genes and their functional significance in fish. Gene 2002;295:173–183.
69. Iwama GK, Afonso LOB, Todgham A, Ackerman P, Nakano K. Are HSPs suitable for indicating stressed states in fish? J Exp Biol 2004;207:15–19.
70. Ji C, Guo N, Li L. Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis. J Proteome Res 2005;4:2099–2108.
71. Ji C, Lo A, Marcus S, Li L. Effect of 2MEGA labeling on membrane proteome analysis using LC-ESI QTOF MS. J Proteome Res 2006;5:2567–2576.
72. Ji C, Zhang N, Damaraju S, Damaraju VL, Carpenter P, Cass CE, Li L. A study of reproducibility of guanidination-dimethylation labeling and liquid chromatography matrix-assisted laser desorption ionization mass spectrometry for relative proteome quantification. Anal Chim Acta 2007;585: 219–226.

Address correspondence to:
Greg G. Goss, Ph.D.
Department of Biological Sciences
University of Alberta
CW405 Biological Sciences Building
Edmonton, AB T5G 2E9
Canada

E-mail: greg.goss@ualberta.ca

Liang Li, Ph.D.
Department of Chemistry
University of Alberta
Edmonton, AB T6G 2G2
Canada

E-mail: liang.li@ualberta.ca