research articles Journal of proteome research

# Differential Detergent Fractionation for Non-electrophoretic Eukaryote Cell Proteomics

**Fiona M. McCarthy,† Shane C. Burgess,*,† Bart H. J. van den Berg, Marek D. Koter, and G. Todd Pharr**

*College of Veterinary Medicine, PO Box 6100, Mississippi State University, Mississippi 39762*

**Received August 26, 2004**

Differential detergent fractionation (DDF), which relies on detergents to sequentially extract proteins from eukaryotic cells, has been used to increase proteome coverage of 2D-PAGE. Here, we used DDF extraction in conjunction with the nonelectrophoretic proteomics method of liquid chromatography and electrospray ionization tandem mass spectrometry. We demonstrate that DDF can be used with 2D-LC ESI $MS^2$ for comprehensive cellular proteomics, including a large proportion of membrane proteins. Compared to some published methods designed to isolate membrane proteins specifically, DDF extraction yields comprehensive proteomes which include twice as many membrane proteins. Two-thirds of these membrane proteins have more than one trans-membrane domain. Since DDF separates proteins based upon their physicochemistry and subcellular localization, this method also provides data useful for functional genome annotation. As more genome sequences are completed, methods which can aid in functional annotation will become increasingly important.

**Keywords:** membrane • gene ontology • functional annotation • CHUNK-1 • Sidekick-1 • AVTR · CPE-FGFR • GABA



## Introduction

The aim of expression proteomics is to obtain comprehensive cellular protein expression profiles. However, whole cell proteomics is challenging. First, the dynamic range of proteins within eukaryotic cells varies by as much as $10^8$;[1] second, physicochemical properties of cellular proteins are very different. Traditionally, two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE) systems have been used for expression proteomics.[2] However, many hydrophobic proteins are poorly soluble in electrophoresis buffers and are under-represented in gels.[3] Nonelectrophoretic proteomics, based on liquid chromatography and electrospray ionization tandem mass spectrometry (LC ESI $MS^2$), may have advantages on expression proteomics and be complimentary to gel-based methods.[4] Poor membrane-protein solubility may be circumvented with LC ESI $MS^2$ because separation can be based on the physiochemical properties of more soluble peptides, derived from protease-digestion, rather than the intact poorly soluble precursor proteins themselves. Membrane proteomics is important because the plasma membrane is a cell's primary interface for interacting with its environment; membrane proteins direct processes involved in cell adhesion, cell−cell communication, proliferation and differentiation. In addition, organelle membrane proteins are essential for intracellular signaling. The importance of membrane proteins is clear from genome analyses. One-third of all currently described genes code for membrane proteins.[5,6] One-half of the mass of the plasma membrane is protein, and this proportion is even greater in internal organelle membranes.[7]

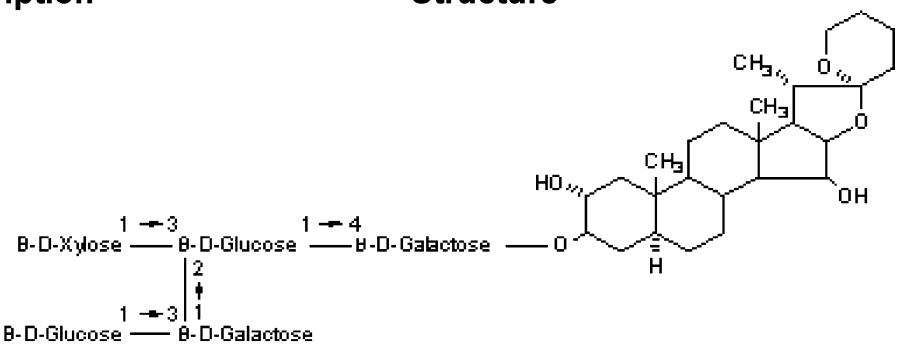
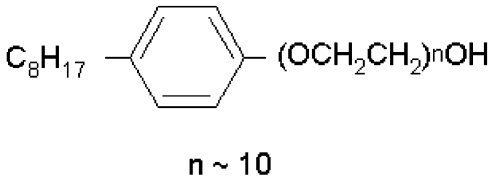
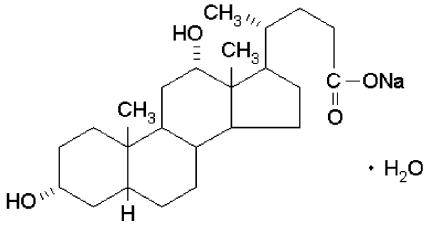
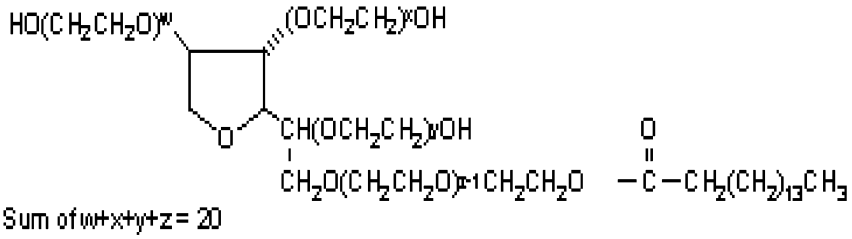
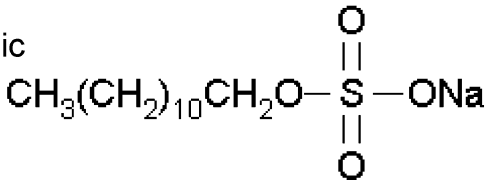
A separate issue is that identification and annotation of open reading frames lags behind genome sequencing. Methods that aid in the elucidation of protein function, structure, localization, and expression kinetics would greatly assist functional genome annotation.[8,9] Protein separation based on subcellular fractionation defines a protein's cellular location and can be used to help infer its function. Differential detergent fractionation (DDF) has been used as a rapid, practical method of subfractionating eukaryotic cells prior to 2D-PAGE to increase the visualized proteome.[10] DDF uses a series of different detergents to sequentially extract cellular proteins while keeping the cell architecture intact. We use a chicken B-lymphocyte system to demonstrate how DDF combined with LC ESI $MS^2$ can be used to obtain functional information about proteins. We have adapted DDF for use immediately prior to LC ESI $MS^2$ expression proteomics. Proteins are first isolated using DDF and then are digested into smaller, more soluble, peptides.

Using DDF in conjunction with 2D LC ESI $MS^2$, we achieved comprehensive cellular proteomes including a large proportion of membrane proteins. Few membrane proteins were found in more than one DD fraction and presence in multiple fractions creates a "DDF profile" that represents a protein's subcellular localization. Specific DDF profiles were generated; each contained proteins with specific characteristics. We used the specific DDF profiles of proteins, for which large amounts of data exists, to suggest functions for poorly described proteins with the same DDF pattern. Using DDF, almost twice as many membrane proteins were identified when compared to metha-

* To whom correspondence should be addressed. E-mail: burgess@cvm.msstate.edu.
† Both authors contributed equally to this work.


10.1021/pr049842d CCC: $30.25 
Published on Web 01/25/2005

**Table 1.** Properties of Detergents Used in DDF

| Detergent name, description (supplier, product #) | Structure |
|---|---|
| Digitonin, mild non-ionic (Sigma, D-141) |  |
| Triton-X-100, non-ionic (Sigma, T-8787) |  |
| Deoxycholate (DOC), anionic (Sigma, D-6750) |  |
| Tween-40, non-ionic (Sigma, P-1504) |  |
| Sodium Dodecyl Sulfate (SDS), anionic (Sigma, L-4509) |  |

nol extraction.[11] We show that DDF is a simple method of prefractionation for LC ESI $MS^2$ expression proteomics, facilitates functional genome annotation, allows analysis of proteins with multiple transmembrane domains (TMD), and isolates a larger proportion of membrane proteins than other commonly used isolation methods.[11,12]

## Materials and Methods

**Lymphocyte Isolation.** Chicken Bursa of Fabricius B-cells were used to prepare a pure preparation of $10^8$ cells for DDF. Bursas were collected from five 21-day-old Ross 508 chickens and bursal B-cells isolated as previously described.[13] For the comparison of different extraction methods, chicken peripheral blood lymphocytes (PBL) were used. PBL were isolated from 15 mL of chicken blood using Histopaque-1077 (Sigma, St Louis, MO). Cells were counted using a haemocytometer and three samples, each containing $6 \times 10^7$ cells, were used.

**Protein Extraction by DDF.** DDF sequentially extracts proteins from intact cells using a series of detergents while preserving cell architecture. Table 1 shows the properties of each detergent used in DDF. Low retention tubes (Thermo-Electron Corporation; San Jose, CA) were used throughout. The first detergent, digitonin, interacts with cholesterol to form pores in the cell membrane and extract soluble proteins from the cytosol.[14] Triton X-100 solubilizes membrane and organelle proteins and a combination of deoxycholate (DOC) and Tween

40 extract the soluble nuclear fraction. Nuclear matrix proteins and the more insoluble proteins are solubilized using SDS. DDF was used to isolate proteins from bursal B-cells. DDF was done as previously described[10] but seven additional digitonin washes were used. Cytosolic proteins were extracted by sequential 30 min incubation in cytosolic DDF buffer containing digitonin and the membrane fraction collected by incubating the cells in 10% Triton X-100 DDF buffer for 30 min and then removing the soluble protein. Nuclear DDF buffer containing deoxycholate was added to the remaining samples and the cell pellets frozen to disrupt the nucleus prior to homogenization. Nuclear proteins were collected from the resulting soluble fraction and the sample aspirated through an 18 gauge needle and treated (37 °C, 1 h) with a mixture of DNase I (50U; Invitrogen, Carlsbad CA) and RNase A (50 mg; Sigma-Aldrich, St Loius, MO) to digest intact nucleic acids. The remaining pellet was solubilized in DDF buffer containing 5% SDS. To confirm that the majority of the cytosolic proteins were extracted the protein samples were analyzed by SDS-PAGE. A 10% volume sample of each of the fractions was separated on a 10% denaturing acrylamide gel and the proteins visualized by Coomassie blue.

**Membrane Protein Isolation Using SDS Only.** Cells were centrifuged (520 × g, 10 min, 4 °C) and the pellet resuspended in 5% SDS (w/v), 10 mM sodium phosphate buffer, pH 7.4. The sample was freeze−thawed several times to enhance protein solubilization. The solution was then passed through an 18 gauge needle 5 times to shear genomic DNA then treated with DNase I and RNase A as above.

**Membrane Protein Isolation Using Carbonate Fractionation−Methanol Extraction.** Membrane proteins were also isolated using a methanol extraction procedure.[11] Briefly, PBLs were incubated in ice cold 0.1 M carbonate buffer, pH 11.0 to remove soluble proteins.[15] Integral membrane proteins were then extracted by addition of 50 mM ammonium bicarbonate and 60% methanol to disrupt the phospholipid bilayer. The resulting protein mixture was digested for MS by the direct addition of sequencing grade modified porcine trypsin (Promega; Madison, WI) at a 1:50 (w/w) ratio of trypsin:protein.

**Trypsin Digestion and 2D-LC ESI MS².** For each of the above procedures, equal amounts of protein were precipitated with 25% tricholoroacetic acid to remove salts and detergents. Protein pellets were resuspended in 0.1 M ammonium bicarbonate, 5% HPLC grade acetonitrile (ACN), reduced (5mM DTT, 65 °C, 5 min), alkylated (10 mM iodoacetamide, 30 °C, 30 min) and then trypsin digested until there was no visible pellet (1:50 w/w 37 °C, 16 h). Peptides were desalted using a peptide microtrap (Michrom BioResources, Inc., Auburn, CA) and eluted using a 0.1% triflouroacetic acid, 95% ACN solution. Desalted peptides were dried in a vacuum centrifuge and resuspended in 20 $\mu$L of 0.1% formic acid. LC analysis was accomplished by strong cation exchange followed by reverse phase liquid chromatography coupled directly in line with ESI ion trap MS. Samples were loaded into a liquid chromatography gradient ion exchange system containing a Thermo Separations P4000 quaternary gradient pump (ThermoElectron Corporation; San Jose, CA) coupled with a 0.32 × 100 mm BioBasic strong cation exchange column. A flow rate of 3$\mu$L/min was used for both SCX and RP columns. A salt gradient was applied in steps of 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 57, 64, 71, 79, 90, 110, 300, and 700 mM ammonium acetate in 5% ACN, 0.1% formic acid and the resultant peptides loaded directly into the sample loop of a 0.18 × 100 mm BioBasic C18 reverse phase liquid chromatography column of a Proteome X workstation (ThermoElectron). The reverse phase gradient used 0.1% formic acid in ACN and increased the ACN concentration in a linear gradient from 5% to 30% in 30 min and then 30% to 65% in 9 min followed by 95% for 5 min and 5% for 15 min.

The spectrum collection time was 59 min for every strong cation exchange step. The LCQ Deca ion trap mass spectrometer was configured to optimize the duty cycle length with the quality of data acquired by alternating between a single full MS scan followed by three tandem MS scans on the three most intense precursor masses (as determined by Xcalibur mass spectrometer software in real time) from the full scan. The collision energy was normalized to 35%. Dynamic mass exclusion windows were set at 2 min, and all of the spectra were measured with an overall mass/charge ($m/z$) ratio range of 200−2000.

**Protein Identification.** Mass spectra and tandem mass spectra were used to search subsets of the nonredundant protein database (nrpd) downloaded from the National Center for Biotechnology Institute (NCBI; 06/14/04) using TurboSEQUEST (Bioworks Browser 3.2; ThermoElectron). For most of this work, we restricted our analysis to an avian subset of the nrpd (AVIAN DB; search terms: chicken, gallus, cornix, aves, turkey, and ostrich NOT plant, yeast, bacteria, virus). However, the draft chicken genome is poorly annotated and very few internal cellular proteins were present in the AVIAN DB. We therefore used a nonavian-vertebrate subset of the NR protein database (NAV DB; search terms: mammal, Homo, Rattus, Mus, fish, and excluding the terms used to create the AVIAN DB) to identify more internal cellular proteins. Gene ontology (GO) lists were downloaded from AMIGO (http://www.godatabase.org/cgi-bin/Amigo/go.cgi) for the following cellular compartments: nucleus (GO:0005634), endoplasmic reticulum (GO:0005783), Golgi apparatus (GO:0005794) and mitochondrion (GO:0005739). Trypsin digestion was applied in silico to AVIAN DB and NAV DB and including mass changes due to cysteine carbamidomethylation and methionine oxidation. The peptide (MS precursor ion) mass tolerance was set to 1.5 Da and the fragment ion (MS²) mass tolerance was set to 1.0 Da. Peptide matches were considered genuine if they were ≥ 6 amino acids and consistent with described X correlation and ΔCn values.[16]

**Protein Analysis.** Using TurboSEQUEST peak areas of the MS² fragment ions were calculated and summed for each identified protein. The sum of the areas under the curve for each peptide was calculated to obtain a relative semiquantitation value for each protein. When a protein was detected in more than one detergent fraction, the quantitation values were normalized with respect to the largest value to determine the relative amounts of protein found in each sample.

Each protein was classified with respect to its cellular component and biological process using GO annotation.[17] When no GO annotation was available, proteins were annotated manually based on literature searches and closely related homologues (determined by BLASTP).[18] EMBL Bioinformatic Harvester [19] was used to determine GO annotation terms for homologues and these terms were ascribed to the avian proteins. Avian proteins that had no close homologue were analyzed using tools provided by the ExPASy Proteomics Server,[20] PSORT II,[21] and SOSUI [22] and DAS [23] TMD prediction tools.

**Chicken Plasma Analysis.** Immunoglobulins (Igs) were depleted from 150 $\mu$L of pooled plasma sample from 14 Ross x Ross 6-week-old broiler chickens by affinity adsorption chromatography using protein A/G (USB Corporation). The plasma was diluted with an equal amount of phosphate buffered saline

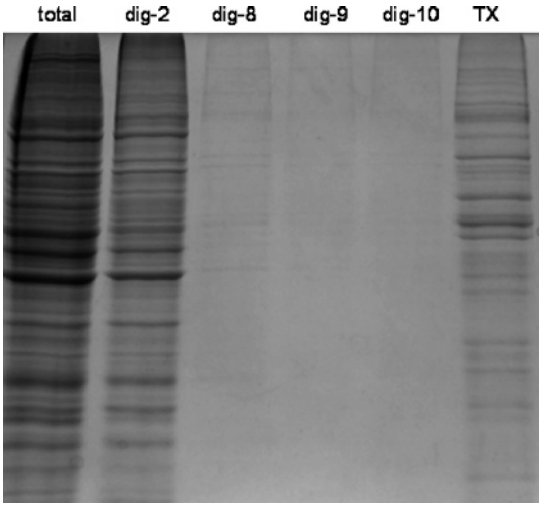



**Figure 1.** Silver stained SDS-PAGE of bursal B-cell DD fractions. Proteins extracted by sequential digitonin washes are compared to total cell proteins and proteins extracted using Triton X-100. Total = total protein; dig = digitonin wash; TX = Triton X-100 wash.

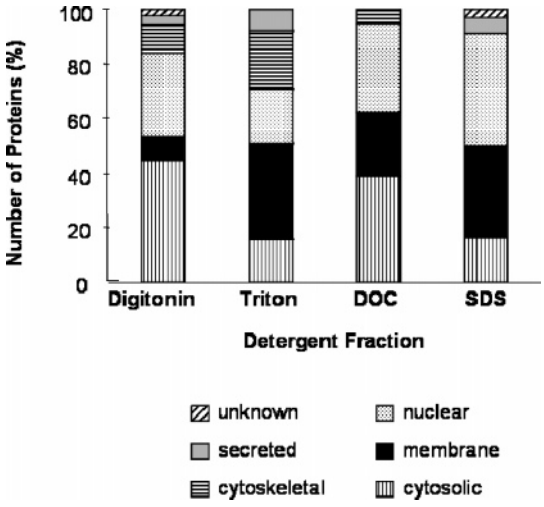



**Figure 2.** Cellular categorization of proteins identified from different bursal B-cell DD fractions. Proteins from each detergent extraction are classified based upon GO cellular component.

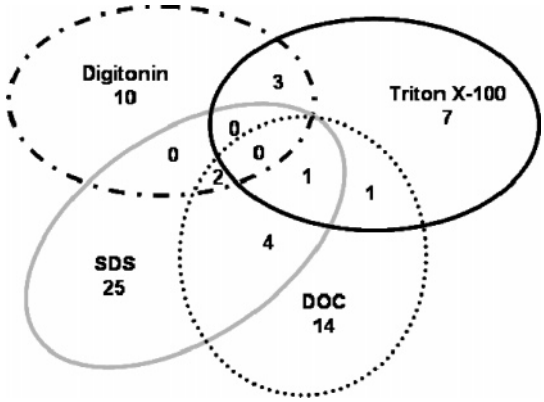



**Figure 3.** Distribution of membrane proteins identified from different bursal B-cell DD fractions. Very few membrane proteins are found in more than one detergent fraction.

(PBS;150 mM NaCl, 10 mM sodium phosphate, pH 8.0) and incubated (30 min on ice). Immunoglobulin-depleted plasma was separated from the protein A/G beads by centrifugation (200 × *g*). The beads were washed three times with 5 volumes of PBS, and the washes were pooled with the immunoglobulin-depleted plasma. The diluted immunoglobulin-depleted plasma sample was then precipitated using trichloroacetic acid and resuspended in 10 mM HCO3 NH4, 5% ACN, pH 7.5. The samples were then digested with modified porcine trypsin and analyzed using 2D-LC ESI $MS^2$ as described above except the salt gradient was in steps of 10, 23, 37, 51, 70, 99, 300, and 700 mM ammonium acetate (Sigma-Aldrich) in 5% ACN, 0.1% formic acid.

## Results

**Different Proteins are Isolated from Different Detergent Fractions.** Cells were depleted of abundant cytosolic proteins by repeated washes in digitonin buffer. To confirm that we were depleting abundant cytosolic proteins, equivalent amounts of digitonin and Triton X-100 washes were compared using 10% SDS-PAGE. By the eighth wash most cytosolic proteins were removed (Figure 1).

The $MS^2$ data from each of the DDF samples were analyzed using the AVIAN DB subset and the resulting proteins, classified with respect to their subcellular location, are shown in Figure 2. We found a 4-fold increase in the membrane proteins identified from the Triton X-100 fraction, when compared with the digitonin fraction (Figure 2).

Although DDF reportedly extracts proteins based on their subcellular location,[24,25] we found membrane proteins were isolated using each of the detergent buffers (Figure 2). However, very few are extracted in more than one fraction (Figure 3). A total of 67 membrane proteins were identified using the AVIAN DB, and only 16.4% were found in more than one detergent fraction. The properties of these proteins are shown in Table 2, along with the total numbers of peptides identified and the percent protein coverage by those peptides. Protein function (GO biological process) and subcellular localization (GO cellular component) are also included. We examined the structure and localization of these membrane proteins in more detail to determine if a protein's DDF pattern related to its physiochemical properties and expression pattern within the cell.

**Membrane Proteins are Fractionated by DDF on the Basis of Their Number of TMD and Subcellular Localization.** Plasma membrane proteins are isolated by different detergents based upon the number of TMD they contain. The average TMD score for proteins identified from the digitonin fraction is 1.7, while this score increases to 2.9 for the Triton X-100 fraction, 3.0 for the DOC-Tween fraction, and 3.4 for SDS isolated membrane proteins. We found that proteins annotated as membrane-associated tended to be isolated with the digitonin buffer, for example, annexin A2 is a membrane associated protein that localizes in the lamina beneath the plasma membrane and this protein was identified from the digitonin fraction. Triton X-100 typically isolates plasma membrane receptors with 1−3 TMD while DOC isolates integrins and cadherins, as well as some membrane proteins that are predicted to have up to 7 TMD (e.g., acetylcholine receptor). The largest TMD proteins identified from DDF were in the SDS fraction and these include ion transporters which are expressed in the plasma membrane as well as proteins from organelles, such as SERCA3.

In addition to the plasma membrane, many cell organelles have lipid membranes. By including proteins identified from the NAV DB, we obtained lists of proteins GO-annotated to

**Table 2.** Membrane Proteins Identified by DDF

| GI # | Protein | #TMD | #Peptides Identified | %peptide coverage | DDF | GO:BP | GO:CC |
|---|---|---|---|---|---|---|---|
| 113128 | Muscarinic acetylcholine receptor M4 | 7 | 8 | 36.4 | T | G-protein coupled receptor protein signaling pathway | integral to membrane |
| 113949 | Annexin A2 | 0 | 9 | 37.5 | D | development | plasma membrane; soluble fraction |
| 114380 | Sodium/potassium-transporting ATPase alpha-3 chain | 7 | 23 | 51.1 | S | cation transport | integral to membrane |
| 119359 | Endoplasmin | 1 | 35 | 70.0 | T/TD | protein folding | plasma membrane; endoplasmic reticulum |
| 124950 | Integrin alpha-8 | 1 | 13 | 26.5 | S | cell-matrix adhesion | integral to membrane |
| 124958 | Integrin alpha-V (Vitronectin receptor alpha subunit) | 2 | 16 | 31.4 | S | cell-matrix adhesion | integral to membrane |
| 127513 | IgM | 1 | 18 | 85.7 | D/T | immune response | membrane fraction |
| 231487 | Neuronal acetylcholine receptor | 6 | 6 | 20.7 | TD | ion transport | integral to membrane |
| 345659 | Connexin-56 | 4 | 12 | 60.0 | TD | cell communication | integral to membrane |
| 462150 | GABA receptor | 5 | 7 | 26.9 | TD/S | gamma-aminobutyric acid signaling; chloride transport | integral to membrane |
| 542694 | Gamma-aminobutyric acid A receptor gamma 1 chain | 5 | 8 | 28.6 | S | gamma-aminobutyric acid signaling; chloride transport | integral to membrane |
| 627334 | Protein-tyrosine-phosphatase (EC 3.1.3.48), receptor | 1 | 28 | 38.9 | S | cell surface receptor linked signal transduction | integral to membrane |
| 633796 | CPE-FGFR | 1 | 5 | 62.5 | D/T | protein amino acid phosphorylation | membrane |
| 1042020 | GABA type C receptor rho 2 (RACK1) | 5 | 3 | 16.7 | TD/S | ion transport | integral to membrane |
| 1070476 | Epidermal growth factor receptor | 2 | 37 | 59.7 | S | transmembrane receptor protein tyrosine kinase signaling | integral to membrane |
| 1184241 | Bu-1 | 1 | 8 | 57.1 | TD | function unknown | membrane |
| 1478410 | AVT receptor | 7 | 1 | 100.0 | D/T | G-protein coupled receptor protein signaling pathway | integral to membrane |
| 1536804 | Immunoglobulin lambda chain | 0 | 1 | 20.0 | T | membrane associated | immune response |
| 1621232 | HEMCAM | 3 | 12 | 36.4 | T | cell adhesion | membrane |
| 1706627 | Ephrin type-A receptor 5 | 1 | 30 | 60.0 | D | transmembrane receptor protein tyrosine kinase signaling | integral to membrane |
| 1708864 | LDL receptor-related protein 1 (LRP) | 1 | 66 | 26.2 | D/TD/S | endocytosis | integral to membrane; coated pit |
| 2134296 | C-Delta-1 | 2 | 15 | 39.5 | S | cell communication | membrane |
| 2134427 | Transforming growth factor beta receptor type II | 1 | 11 | 32.4 | TD | transmembrane receptor protein Ser/Thr kinase signaling | integral to membrane |
| 2313049 | C-Serrate-2 | 1 | 22 | 34.9 | TD | cell communication | membrane |
| 2498677 | Neuropilin-1 | 2 | 16 | 34.0 | TD | cell adhesion | integral to membrane |
| 3023428 | Cadherin-10 | 1 | 9 | 19.1 | TD | homophilic cell adhesion | integral to membrane; gap junction |
| 3243095 | Putative chemokine receptor | 7 | 5 | 29.4 | T | G-protein coupled receptor protein signaling pathway | integral to membrane |
| 4323233 | Cadherin | 3 | 2 | 22.2 | S | homophilic cell adhesion | integral to membrane |
| 4884676 | Receptor tyrosine kinase | 5 | 25 | 36.8 | TD | transmembrane receptor protein tyrosine kinase signaling | membrane |
| 5725448 | CHUNK-1 | 3 | 11 | 45.8 | D/TD/S | function unknown | membrane |
| 5921704 | Voltage-dependent L-type calcium channel alpha-1S | 12 | 7 | 58.3 | S | calcium ion transport | integral to membrane |
| 6448447 | Mitochondrial carrier homolog 2 (Mtch2) | 3 | 8 | 50.0 | T | transport | inner mitochondrial membrane |
| 6708125 | Cone potassium-dependent sodium-calcium exchanger | 11 | 6 | 24.0 | S | ion transport | integral to membrane |
| 8133188 | Voltage-dependent anion channel (Porin-2) | 12 | 12 | 75.0 | TD/S | anion transport | mitochondrial outer membrane |
| 8928354 | Sortilin-related receptor | 1 | 24 | 24.7 | D | lipid transport; enmdocytosis | membrane |
| 10241574 | Teneurin-2 | 1 | 37 | 24.3 | D | cell adhesion | integral to membrane |
| 11245468 | Type I inositol triphosphate receptor | 6 | 13 | 61.9 | S | calcium ion transport | membrane |
| 12229805 | Ephrin type-B receptor 3 | 1 | 22 | 41.5 | TD | transmembrane receptor protein tyrosine kinase signaling | integral to membrane |
| 13161045 | Integrin alpha 3A subunit | 1 | 5 | 10.9 | S | cell-matrix adhesion | integral to membrane |
| 13431348 | Cadherin-6 | 1 | 9 | 20.5 | S | homophilic cell adhesion | integral to membrane |
| 14043015 | Death receptor 6 | 2 | 22 | 59.5 | S | apoptosis | integral to membrane |
| 14210853 | OL-protocadherin | 1 | 12 | 24.0 | TD | homophilic cell adhesion | membrane |
| 15076564 | ChB1 | 2 | 7 | 28.0 | S | function unknown | membrane |
| 17366584 | Cadherin-7 | 2 | 8 | 19.5 | S | homophilic cell adhesion | integral to membrane |
| 18203648 | SERCA3 | 10 | 46 | 78.0 | T/TD/S | calcium ion transport | integral to membrane |
| 18266871 | Neuronal acetylcholine receptor | 10 | 6 | 22.2 | S | ion transport | integral to membrane |
| 20140635 | Transferrin receptor protein 1 | 1 | 19 | 42.2 | S | endocytosis | integral to membrane |
| 22775582 | ATP/ADP antiporter | 6 | 22 | 73.3 | TD/S | mitochondrial inner membrane | mitochondrial transport |
| 23194254 | Sidekick-1 | 1 | 26 | 26.8 | D | receptor activity | membrane |
| 29242837 | MHC class II antigen | 1 | 6 | 66.7 | T | antigen presentation, exogenous antigen | integral to membrane |
| 29372484 | Bombesin-like peptide receptor subtype 3.5 | 6 | 6 | 35.3 | S | G-protein coupled receptor protein signaling | integral to membrane |
| 37962275 | Leptin receptor | 2 | 14 | 24.6 | S | cell surface receptor linked signal transduction | integral to membrane |
| 38045635 | Integrin beta-5 | 2 | 29 | 54.7 | TD | cell-matrix adhesion | integral to membrane |
| 38372283 | Neurofascin | 1 | 23 | 30.3 | TD | cell adhesion | integral to membrane |
| 45382353 | Interferon alpha/beta receptor 1 | 2 | 8 | 24.2 | D | receptor activity | integral to membrane |
| 45382793 | CD40-homologue | 2 | 3 | 23.1 | D | immune response ; apoptosis | membrane |
| 45383039 | Vanilloid receptor-like protein | 5 | 18 | 40.0 | S | cation transport | integral to membrane |
| 45384050 | Neuron-glia cell adhesion molecule (Ng-CAM) | 1 | 9 | 13.6 | S | cell adhesion | integral to membrane |
| 47604958 | CD8 alpha chain | 1 | 5 | 41.7 | S | immune response | integral to membrane |
| 50737406 | PREDICTED: similar to desmoglein 4 | 1 | 16 | 27.6 | D | homophilic cell adhesion | integral to membrane; |
| 50737409 | PREDICTED: similar to Desmoglein 2 | 1 | 15 | 26.8 | S | homophilic cell adhesion | integral to membrane; intercellular junction |
| 50749252 | PREDICTED: similar to ADAM 8 | 1 | 12 | 46.2 | T | proteolysis and peptidolysis | integral to membrane |
| 50749938 | PREDICTED: similar to ADAM 9 | 1 | 16 | 33.3 | TD | proteolysis and peptidolysis; protein kinase cascade | integral to membrane |
| 50750371 | PREDICTED: similar to integrin alpha-4 subunit | 1 | 17 | 24.6 | D | cell-matrix adhesion | integral to membrane |
| 50761482 | PREDICTED: similar to gp130-like monocyte receptor | 2 | 8 | 21.6 | TD | transmembrane receptor protein tyrosine kinase signaling | integral to membrane |
| 50787137 | PREDICTED: similar to teneurin-4 | 2 | 5 | 38.5 | D | cell adhesion receptor activity; immune response | integral to plasma membrane; extracellular |
| 52138677 | R-cadherin | 1 | 11 | 27.5 | S | homophilic cell adhesion | integral to membrane |

Key: D, digitonin; DDF, differential detergent fraction; GI #, Genbank accession number; GO:BP, Gene Ontology biological process; GO:CC, Gene Ontology cellular compartment; S, sodium dodecylsulfate; T, triton X-100; TD, tween 40 deoxycholate.

localize to specific subcellular structures. We analyzed 22 mitochondrial proteins, 13 ER proteins, and 34 proteins from the nucleus and show that organelles also have specific fractionation patterns. Proteins specific to the ER, for example calreticulin and SERCA1, are isolated from both Triton X-100 and DOC. Triton X-100 buffer specifically isolates proteins from the mitochondria. Of the 22 mitochondrial proteins we identified, 20 were isolated in the Triton X-100 fraction. The exceptions are Porin-2 (gi:8133188) and ATP/ADP antiporter (gi:22775582). Both Porin-2 and ATP/ADP antiporter are identified in more than one DDF.

**Membrane Proteins Found in More Than One Fraction Have Multiple Subcellular Locations.** Several membrane proteins were identified in more than one detergent fraction. For each of these proteins, the relative amount from each detergent fraction was quantified (Figure 4). Also included in Figure 4 are several proteins with well documented cellular distributions, which serve as markers for the different subcellular compartments. The most striking example of this is hsp70, a ubiquitous chaperone protein which was isolated from all four detergent fractions. Since DDF compartmentalizes well-characterized proteins based on subcellular location, we can use this technique not only to isolate proteins but also as a tool to define protein location and infer some function. As examples, our data shows that both the arginine vasotocin receptor (AVTR, gi:1478410) and the chick pigment epithelium derived-fibroblast growth factor receptor (CPE-FGFR gi:633796) proteins were identified from both digitonin and Triton X-100 fractions. These two proteins have the same DDF profile as IgM, suggesting that they exist as both a soluble and membrane bound form. Proteomic analysis of chicken plasma confirmed that both AVTR and CPE-FGFR exist as soluble forms.

**DDF Defines Multiple Classes of Nuclear Proteins.** Detergent fractionation of nuclear proteins (NPs) has two levels of complexity. First, many of the proteins that function in the nucleus at some stage may be present in the cytoplasm (e.g.,

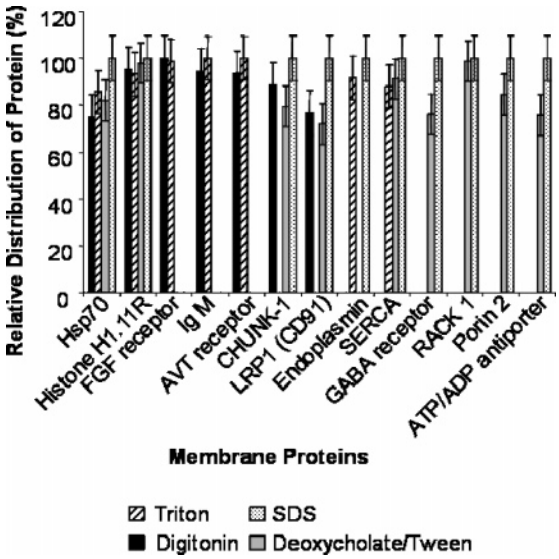



**Figure 4.** DDF profiles for membrane proteins extracted by more than one detergent buffer. Hsp70 and histone H1.11R are included as examples of proteins that were identified from each of the detergent fractions.

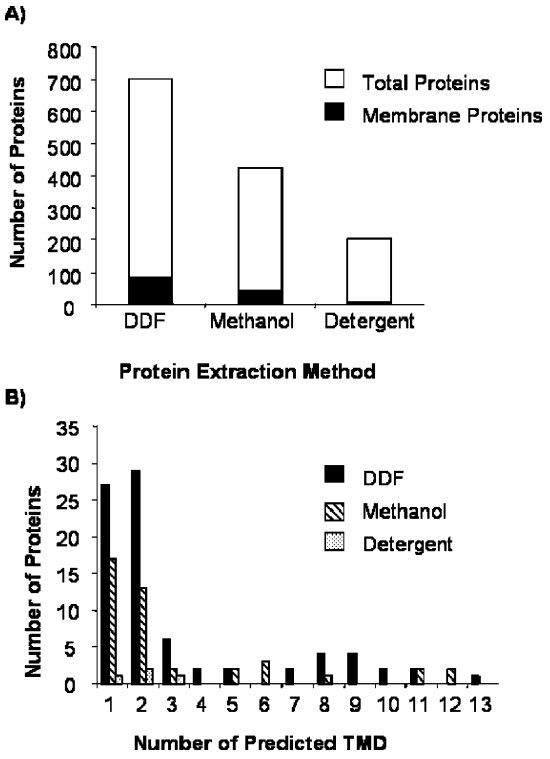



**Figure 5.** Comparison of DDF and other membrane extraction methods. Proteins were identified from equal numbers of PBLs extracted using either DDF, detergent or methanol. (A) Total number of proteins and membrane proteins for each extraction method. (B) Membrane protein distribution by predicted TMDs for each of the extraction methods.

NF$\kappa$B) and can thus be found in all the fractions. In the second case are NPs whose location is reported to be solely in the nucleus. We defined three classes of these nuclear-restricted NPs based on the detergent fractions in which they were isolated: (i) only in the DOC fraction, (ii) only in the SDS fraction, and (iii) in both DOC and SDS. Proteins isolated in DOC only are predominately transcription factors, while those preferentially isolated by SDS are bound to nucleic acids such as small nuclear ribonucleoproteins that form part of the spliceosome complex. Histones are closely associated with DNA, but although they are more abundant in DOC and SDS fractions, they were also extracted by digitonin and Triton X-100.

**DDF Compares Favorably with other Methods for Isolating Membrane Proteins.** We tested the ability of two different extraction methods to isolate membrane proteins compared to DDF. Proteins were extracted from PBL because they can be more easily collected than bursacytes. The most proteins were identified from DDF, and this sample also contained the highest proportion of membrane proteins (Figure 5a). When membrane proteins are analyzed by number of TMD, DDF provided the most comprehensive coverage. (Figure 5b).

## Discussion

It is almost axiomatic that a completely sequenced and well-annotated genome is fundamental for doing proteomics. However, many sequenced genomes are released with only preliminary annotation and functionally annotating these genomes is becoming an increasingly important issue. Even the human and mouse genomes require much more functional annotation. The work we present here demonstrates the converse, that proteomics can also be used to functionally annotate genomes. We purposely used relatively poorly functionally annotated chicken genome as a model.

Another proteomics challenge is to compile comprehensive proteomes when cell proteins are expressed over a very wide dynamic range. This problem is 2-fold: First, many proteins are genuinely present at very low levels within the cell; second, many proteins are difficult to detect because, although they may be abundant, they may not be easy to isolate. One approach to identifying less abundant proteins from a complex mixture in nonelectrophoretic proteomics is to pre-fractionate the complex sample prior to MS analysis, most commonly using LC. DDF is another deconvolution step prior to LC. Furthermore, by using different detergents, DDF efficiently isolates hydrophobic membrane proteins. Taken together, this means that DDF helps to compile comprehensive proteome coverage. But proteomics requires more than just a list of cell proteins; we demonstrate here that DDF is not only able to isolate proteins efficiently but that by isolating proteins using DDF we can get information on where the proteins are expressed within the cell.

DDF is efficient at extracting membrane proteins because it first removes the more abundant cytosolic proteins and it relies on the properties of more than one detergent to extract proteins from the cell. Proteins that are not soluble in one detergent may be solubilized by another. As shown in Figure 1, after eight digitonin washes there are significantly fewer proteins extracted and amounts of the most abundant soluble proteins are absent or greatly reduced. Detergents have been reported to interfere with ESI.[26,27] These studies analyzed proteins directly from 2D-PAGE, a procedure which uses considerably more detergent than that used in DDF. Furthermore, DDF detergents are removed after the proteins are digested into more soluble peptides. Our results show that the method we used must remove enough of the detergent so that we still get appropriate peptide ionization.

DDF yields electrophoretically distinct protein samples which, when analyzed using 2-D PAGE, represent subcellular compartments.[10,24,25] We show that this partitioning is not as distinct when fractions are analyzed using LC ESI $MS^2$; mem-

brane proteins were identified from each detergent fraction (Figure 2). Although we expected membrane proteins to partition in the Triton X-100 fraction − and indeed this fraction shows a 4-fold increase in membrane proteins when compared with the digitonin fraction—the proportion of membrane proteins identified in subsequent fractions is also significantly higher than those identified from the digitonin fraction. Since one of the advantages of using DDF is that we would be able to infer functional information about a protein based on its subcellular location, we analyzed this partitioning effect of DDF further.

While membrane proteins are isolated from each detergent fraction, few (16.4%) are isolated from more than one fraction; each detergent isolates a specific subset of membrane proteins (Figure 3). Membrane proteins isolated solely from the digitonin fraction typically contain only a single TMD or are only peripherally associated with the plasma membrane. For example, annexin A2 is located in the lamina near the plasma membrane and it appears that when digitonin interacts with cholesterol it can also extract proteins loosely associated with the plasma membrane. The general trend for isolation of membrane proteins is that each sequential detergent step extracted proteins with an increasing number of TMD. This is probably a direct consequence of the increasing strength of detergents used in the sequential washes. Digitonin inserts into the phospholipid bilayer to create pores; Triton X-100 strips proteins from membranes; and SDS dissociates protein complexes and solubilizes most proteins. Notably when membrane proteins with a low number of TMD (<5) were isolated solely from the DOC and SDS fraction they were integrins, cadherins and cell adhesion molecules. Since these molecules are involved with cell adhesion, they form strong interactions with other cells or extracellular matrix proteins.

As well as the plasma membrane proteins, there are many organelle membrane proteins which, based on published data, should be isolated in the Triton X-100 fraction.[10] We determined that the majority of mitochondrial proteins are isolated solely from the Triton X-100 fraction. The exception was mitochondrial proteins with multiple TMD (discussed further below). Unlike mitochondrial proteins, ER specific proteins are isolated from both the Triton X-100 fraction and the DOC fraction. This is intuitive as the ER is an extensive membranous network that can be closely associated with the nuclear membrane.

Proteins GO-annotated to localize to the nucleus also had a complex DDF pattern. Essentially these proteins are in one of three groups: proteins were either isolated solely in the DOC buffer or the SDS buffer or were isolated by both of the detergent buffers. The DOC buffer isolated predominantly soluble transcription factors while the SDS buffer isolated mostly small nuclear ribonucleoproteins and centromere-associated proteins. It is likely that proteins tightly associated with chromosomal DNA cannot be extracted until the SDS buffer is added (after treatment with DNase I and RNase A). Similarly, nuclear ribonucleoproteins that are specifically associated with the membrane-bound spliceosome complexes are also isolated using the SDS buffer. Not surprisingly proteins associated with the splicosome were identified in the SDS fraction. Some transcription factors were present in both the DOC and SDS fractions. This may relate to how intimately they were bound to the DNA.

We found that generally proteins were isolated from detergent fractions based on TMD number and subcellular localization. However, some proteins were isolated by more than one detergent buffer. For each membrane protein found in more than one detergent fraction, the relative amount detected in each detergent fraction was quantitated by calculating peak areas for the fragment ions. ESI MS results in a signal which is directly proportional to the concentration of the analyte.[28] While it is not possible to obtain absolute values for analyte concentration without spiking the sample with a known reference, relative values for a protein can be determined by measuring the difference in peak areas between different sample runs[29,30] even in complex peptide mixtures.[31] By determining the relative quantities of each protein, it is possible to calculate the proportion of protein extracted using each different detergent.

For each protein detected in more than one fraction, we analyzed the proportion of that protein in each fraction (Figure 4). Detection of protein in more than one fraction may be a "carry-over" effect. However, none of these proteins fit the profile we would expect if this were the case; we do not see the majority of the protein in an early detergent extraction followed by decreases in the subsequent fractions. Instead the distribution of a protein across the different fractions, its "DDF profile", is a "fingerprint" for its subcellular expression. For example, the ubiquitously expressed chaperone hsp70 is found in all four detergent fractions while IgM, which has both a soluble and a membrane form, can be identified from both the digitonin fraction and the Triton X-100 fraction. A typical DDF profile for a histone protein (histone H1.11R, gi:121901) shows that, although present in all detergent fractions, it is more abundant in the DOC and SDS fractions. We would expect histones to be identified from DOC and SDS fractions just as other proteins that are intimately associated with DNA. However, histone proteins are expressed in cells at very high levels [7] and it is likely that we are detecting histones from cells actively undergoing mitosis. We would expect digitonin and Triton X-100 to extract histones from mitotic cells since the histones are present in the cytosol as part of the chromatin.

To determine whether DDF profiles could predict subcellular localization, we compared three proteins with similar profiles: IgM, AVTR, and CPE-FGFR. IgM exists as both membrane bound and soluble forms; we speculated that AVTR and CPE-FGFR would similarly have membrane bound and soluble isoforms. CPE-FGFR is a novel chicken fibroblast growth factor receptor homologous to human FGFR3. FGFR3 is also expressed as a soluble form[32] suggesting CPE-FGFR is also likely to have a soluble isoform. Conversely, the AVTR is predicted to have 7 TMD.[33] AVTR has the highest amino acid identity with the membrane-bound mammalian vasopressin receptor 2. However, while AVTR and mammalian vasopressin receptor 2 share some common functions,[34−36] AVTR is also involved in avian oviposition.[37] We detected AVTR and CPE-FGFR in chicken plasma.

Two proteins, for which little structural or expression data exists, extend this theory further: Sidekick-1 and CHUNK-1. Sidekick-1 was first isolated from Drosophilia[38] and homologues to Sidekick-1 have subsequently been reported in human, mouse and chicken.[39] Transmembrane prediction programs[22,23] predict 1−3 TMD for Sidekick-1 (depending on which program is used). Analysis of intracellular signaling motifs using PSORTII indicates that Sidekick-1 is cytosolic. Since Sidekick-1 was isolated solely from the digitonin faction it is less likely that sidekick-1 has 3 TMD, although it may be a membrane associated protein or a single TMD protein expressed on the plasma membrane. Similarly, TM-prediction programs vary in

the number of TMD they predict for CHUNK-1. Protein structure prediction is based upon known motifs and proteins which have a similar homology, and their success rate improves iteratively as more proteins with a known structure are entered into the protein databases. CHUNK-1 has no homology with any protein in the nrpd, making protein prediction results for this protein less reliable. Combining these prediction results with what is known about CHUNK-1's DDF profile can only aid in the elucidation of CHUNK-1 structure and function. CHUNK-1 is a protein that has a very similar DDF profile to CD91 (Figure 4). CD91 is a type I membrane protein that is cleaved into two noncovalently associated protein chains after it is inserted into the plasma membrane and CHUNK-1 may be associated with the plasma membrane in a similar fashion.

Endoplasmin (gi:119359) and the calcium pump protein SERCA3 (gi:18203648) have DDF profiles typical for ER localization. ER proteins were isolated from both the Triton X-100 and DOC fractions. The transferrin receptor has been used as a marker for the endosomal compartment and we have isolated this protein from the SDS fraction. In addition, the GABA receptor shown in Figure 4 refers to the $\gamma$-4 subunit, a unique isoform expressed in chickens which has closest homology to the $\gamma$-2 subunit. The $\gamma$-2 subunit has been shown to cycle through perinuclear endosomes, localizing with the transferrin receptor. Since the $\gamma$-4 subunit is isolated from both DOC and SDS, we expect that it too may cycle through the perinuclear endosomal compartment.

Several other proteins that are isolated by both the DOC and SDS detergent buffers are shown in Figure 4. The first of these, RACK1 (gi:5174447), is a cytoskeletal scaffolding protein, so we would expect it to be isolated in the SDS fraction with other cytoskeletal proteins. It is also identified from the DOC fraction; this is expected because it is part of a complex that recruits several transcription factors and must also be localized to the nucleus. Porin-2 and ATP/ADP antiporter are both mitochondrial proteins. Unlike other mitochondrial proteins we identified, these proteins were not isolated in the Triton X-100 fraction. Porin-2 contains 12 TMD which span the outer mitochondrial membrane while the ATP/ADP antiporter contains 6 TMD and localizes to the inner mitochondrial membrane. Mitochondrial proteins with a large number of TMD are not extracted by Triton X-100 but by the subsequent detergents. This effect is more pronounced for proteins which localize to the inner mitochondrial membrane, as they may not be extracted until the outer membrane is solubilized by the detergents.

Our data suggests that each DDF step is selective for different parts of the cell. We realize that, in common with all current proteomics techniques, we are only identifying a subset of all proteins present in the cell. However, all of the well characterized proteins in our data set were present in the appropriate fraction. Also previous DDF analysis by 2-D PAGE[10,24,25] supports our assertion that DDF is a good predictor of subcellular localization. Since DDF can aid in the localization of proteins, then it follows that it can also aid in GO assignments. The purpose of GO is to ensure consistent terminology for gene products across species. This allows us to compare evolutionarily conserved homologues and make predictions about their structure, function and expression. Because DDF can provide information about a protein's subcellular localization, it is another tool in the gene annotation process.

Finally, high salt washes combined with methanol extraction efficiently isolate membrane proteins.[11] Since many membrane proteins with multiple TMD were identified using DDF, we compared DDF with methanol extraction. Although similar percentages of proteins with multiple predicted TMD were identified using DDF and methanol extraction (66% and 60% respectively), twice as many membrane proteins were identified by DDF in total (Figure 5). Furthermore, DDF provided more comprehensive coverage in terms of numbers of TMD.

**Acknowledgment.** This research was supported by grants from Life Sciences and Biotechnology Institute, MSU. The authors thank A. Cooksey, M. Kidd, T. Pechan, and A. Shack for their help.

## References

(1) Anderson, N. L.; Anderson, N. G. *Electrophoresis* **1998**, *19*, 1853−1861.
(2) Honore, B., Ostergaard, M.; Vorum, H. *Bioessays* **2004**, *26*, 901−915.
(3) Pasquali, C., Fialka, I.; Huber, L. A. *Electrophoresis* **1997**, *18*, 2573−2581.
(4) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat. Biotechnol.* **2001**, *19*, 242−247.
(5) Wallin, E.; von Heijne, G. *Protein Sci.* **1998**, *7*, 1029−1038.
(6) Stevens, T. J.; Arkin, I. T. *Proteins* **2000**, *39*, 417−420.
(7) Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K.; Walter, P. *Molecular Biology of the Cell*; Garland Science: New York; 2002.
(8) Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M.; Yuan, Y. *J. Mol. Biol.* **1998**, *283*, 707−725.
(9) Ge, H., Walhout, A. J.; Vidal, M. *Trends Genet.* **2003**, *19*, 551−560.
(10) Ramsby, M. L.; Makowski, G. S. *Methods Mol. Biol.* **1999**, *112*, 53−66.
(11) Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D. *J. Proteome Res.* **2002**, *1*, 351−360.
(12) Lehner, I., Niehof, M.; Borlak, J. *Electrophoresis* **2003**, *24*, 1795−1808.
(13) Funk, P. E.; Thompson, C. B. *Cell. Immunol.* **1996**, *186*, 75−81.
(14) Fiskum, G.; Craig, S. W.; Decker, G. L.; Lehninger, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 3430−3434.
(15) Howell, K. E.; Palade, G. E. *J. Cell Biol.* **1982**, *92*, 822−832.
(16) Washburn, M. P.; Yates, J. R., 3rd *Curr. Opin. Microbiol.* **2000**, *3*, 292−297.
(17) Camon, E., Barrell, D., Lee, V., Dimmer, E.; Apweiler, R. *In Silico Biol.* **2004**, *4*, 5−6.
(18) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(19) Liebel, U., Kindler, B.; Pepperkok, R. *Bioinformatics* **2004**, *20*, 1962−1963.
(20) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I., Appel, R. D.; Bairoch, A. *Nucleic Acids Res.* **2003**, *31*, 3784−3788.
(21) Nakai, K.; Horton, P. *Trends Biochem. Sci.* **1999**, *24*, 34−36.
(22) Hirokawa, T.; Boon-Chieng, S.; Mitaku, S. *Bioinformatics* **1998**, *14*, 378−379.
(23) Cserzo, M.; Wallin, E.; Simon, I.; von Heijne, G.; Elofsson, A. *Protein Eng.* **1997**, *10*, 673−676.
(24) Ramsby, M. L.; Makowski, G. S.; Khairallah, E. A. *Electrophoresis* **1994**, *15*, 265−277.
(25) Abdolzade-Bavil, A.; Hayes, S.; Goretzki, L.; Kroger, M.; Anders, J.; Hendriks, R. *Proteomics* **2004**, *4*, 1397−1405.
(26) Ishihama, Y.; Katayama, H.; Asakawa, N. *Anal. Biochem.* **2000**, *287*, 45−54.
(27) Loo, R. R.; Dales, N.; Andrews, P. C. *Protein Sci.* **1994**, *3*, 1975−1983.
(28) Voyksner, R. D.; Lee, H. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1427−1437.
(29) Bondarenko, P. V.; Chelius, D.; Shaler, T. A. *Anal. Chem* **2002**, *74*, 4741−4749.
(30) Chelius, D.; Bondarenko, P. V. *J. Proteome Res.* **2002**, *1*, 317−323.
(31) Chelius, D.; Zhang, T.; Wang, G.; Shen, R. F. *Anal. Chem.* **2003**, *75*, 6658−6665.
(32) Sturla, L. M.; Merrick, A. E.; Burchill, S. A. *Br. J. Cancer* **2003**, *89*, 1276−1284.

(33) Tan, F. L.; Lolait, S. J.; Brownstein, M. J.; Saito, N.; MacLeod, V.; Baeyens, D. A.; Mayeux, P. R.; Jones, S. M.; Cornett, L. E. *Biol. Reprod.* **2000**, *62*, 8−15.
(34) Acher, R.; Chauvet, J. *Biochimie* **1988**, *70*, 1197−1207.
(35) Koike, T. I.; Pryor, L. R.; Neldon, H. L.; Venable, R. S. *Gen. Comp. Endocrinol.* **1977**, *33*, 359−364.
(36) Nouwen, E. J.; Decuypere, E.; Kuhn, E. R.; Michels, H.; Hall, T. R.; Chadwick, A. *J. Endocrinol.* **1984**, *102*, 345−351.
(37) Jaccoby, S.; Singh, A. B.; Cornett, L. E.; Koike, T. I. *Gen. Comp. Endocrinol.* **1997**, *106*, 327−337.
(38) Nguyen, D. N.; Liu, Y.; Litsky, M. L.; Reinke, R. *Development* **1997**, *124*, 3303−3312.
(39) Yamagata, M.; Weiner, J. A.; Sanes, J. R. *Cell* **2002**, *110*, 649−660.

PR049842D