research articles Journal of proteome research

Comprehensive Proteomic Analysis of Human Cervical−Vaginal Fluid

Surendra Dasari,[‡] Leonardo Pereira,[†] Ashok P. Reddy,[||] John-Edward A. Michaels,[||] Xinfang Lu,[‡] Thomas Jacob,[||] Archana Thomas,[||] Matthew Rodland,[‡] Charles T. Roberts, Jr.,[‡] Michael G. Gravett,[§] and Srinivasa R. Nagalla*,[‡,||]

Departments of Pediatrics and Obstetrics and Gynecology, Oregon Health and Science University, Portland, Oregon 97239, Department of Obstetrics and Gynecology, University of Washington, Seattle, Washington 98101, and ProteoGenix, Inc., Portland, Oregon 97213

Received October 13, 2006

Cervical−vaginal fluid (CVF) is a potential rich source of biomarkers for enhancing our understanding of human parturition and pathologic conditions affecting pregnancy. In this study, we performed a comprehensive survey of the CVF proteome in pregnancy utilizing multidimensional liquid chromatography (2D-LC) coupled with mass spectrometry and gel-electrophoresis-based protein separation and identification. In total, 150 unique proteins were identified using multiple protein identification algorithms. Metabolism (32%) and immune response-related (22%) proteins are the major functional categories represented in the CVF proteome. A comparison of the CVF, serum, and amniotic fluid proteomes showed that 77 proteins are unique to CVF, while 56 and 17 CVF proteins also occur in serum and amniotic fluid, respectively. This data set provides a foundation for evaluation of these proteins as potential CVF biomarkers for noninvasive diagnosis of pregnancy-related disorders.

Keywords: vaginal fluid • pregnancy • proteome • prematurity

Introduction

Cervical−vaginal fluid (CVF) is a complex biological fluid consisting of water, electrolytes, low-molecular-weight organic compounds (glucose, amino acids, and lipids), cells (leukocytes, lymphocytes, and epithelial cells), and a multitude of proteins and proteolytic enzymes that are predominantly synthesized by the endocervix.[1] CVF also contains secretions from vaginal cells, which include mucins, defensins, complement factors, immunoglobins, lactoferrin, and collectins.[1] CVF flows over and lubricates the entire female reproductive tract, including the vaginal, cervical, and uterine areas. CVF forms the first line of defense against external pathogens, signals fertility, and aids insemination, pregnancy, and labor.[1,2] CVF also contains flora such as *Lactobacilli crispatus* and *Lactobacilli vaginalis*. Secretions from this flora impart a low pH to the CVF, which enhances its antipathogen activity.[1] Any imbalance in the vaginal flora or invasion of external flora results in bacterial vaginosis. In response to bacterial vaginosis, the secretion of several cytokines such as IL-1α, IL-1β, IL-10, IL-6, and TNF-α into the CVF by the cervical and vaginal endoepithelia changes.[3,4] Failure to curb bacterial vaginosis has been positively correlated with cervical cancer,[5] pelvic inflammatory disease,[6] endometritis,[7,8] and tubal infertility.[8] Bacterial vaginosis in pregnant women has been correlated with an increased risk of preterm labor and preterm birth.[9]

The cytokines and other defense molecules present in CVF also play an important role in infection, replication, and proliferation of sexually transmitted viruses such as HIV and Herpes Simplex Virus (HSV) in the vagina.[10−12] Analysis of the cationic polypeptide fraction of the CVF has identified 20 polypeptides that contribute to anti-HIV activity.[13] Previous studies have also identified a role for CVF in the trapping of HIV virions, thus, preventing infection.[14,15] Recent studies have detected a correlation between several immune-response molecules in CVF and the incidence of subclinical premature rupture of membranes (PROM), which leads to preterm birth.[16,17] During pregnancy, CVF could contain amniotic fluid (AF) derived from the uterus, due to either the disruption or parallel secretions of the chorionic-decidual interface. This “leakage” of AF into CVF provides the basis for the current noninvasive diagnosis for the presence of the fetal fibronectin, which has been used to predict preterm labor in women.[18]

CVF is an important potential diagnostic site to monitor maternal and fetal health in pregnant women because of its minimally invasive collection method compared to that for AF, that is, amniocentesis. A comprehensive catalog of proteins expressed in the CVF proteome could enable better insight into the potential role of various CVF proteins that could contribute to or reflect complications during pregnancy or vaginal pathologies. Current proteomics technology is capable of achieving this goal by permitting the identification of a significant fraction of the proteins and peptides that are present in a

* To whom correspondence should be addressed. E-mail, nagallas@ohsu.edu; tel, 503-494-1928; fax, 503-494-4821.
‡ Department of Pediatrics, Oregon Health and Science University.
† Department of Obstetrics and Gynecology, Oregon Health and Science University.
|| ProteoGenix, Inc.
§ University of Washington.

10.1021/pr0605419 CCC: $37.00 © 2007 American Chemical Society
Published on Web 03/21/2007

complex biological sample. The multidimensional Protein Identification Technology (2D-LC) technique has proven to be an effective tool to probe entire proteomes of small organisms[19] and proteomes of complex biological fluids, such as human plasma and saliva, which have a wider dynamic range of protein abundance.[20,21] These studies have demonstrated that multiple proteomic techniques (gel- and liquid-based separation approaches) and data-mining tools are complementary and, when used on the same set of samples, can produce a more comprehensive set of results.[22,23] To date, most CVF analyses have been based on either immunodetection of known proteins or limited proteomics methods utilizing gel-based electrophoresis techniques.[13] In this study, we have utilized multidimensional liquid chromatography coupled with mass spectrometry (2D-LC), gel-electrophoresis-based protein separation and identification, and multiple data-mining tools for a comprehensive analysis of the CVF proteome during normal pregnancy.

Materials and Methods

Sample Collection And Processing. This study was approved by the IRB committee at Oregon Health & Science University. All subjects were identified prospectively and gave informed written consent to participate in the study. Seven subjects representing the second trimester of pregnancy (16−21 weeks), at a mean gestational age (GA) of 18.5 weeks and a standard deviation of ±2.3 weeks were recruited. All subjects were asymptomatic and recruited from our outpatient office between 16 and 21 weeks gestation as they presented for prenatal care, provided they met the inclusion criteria. Main inclusion criteria were singleton pregnancy, intact amniotic membranes, no evidence of cervical−vaginal infection, and a fetus without major congenital anomaly or aneuploidy. All 7 of these subjects delivered at term, and none of them developed preeclampsia or gestational diabetes. CVF samples were collected by placing 2 sterile G-in, Dacron-tipped plastic applicators (Solon, Skowhegan, ME) into the posterior vaginal fornix and rotating them for 15 s during sterile speculum examination. Following collection, protein was extracted into phosphate-buffered saline with a protease inhibitor cocktail (Roche Diagnostics, Alameda, CA). Samples were spun after extraction to remove any debris and cellular material, and the supernatant was stored at −70 °C. Two pooled samples (GA 16−18 and 19−21 weeks) were prepared ($n = 3$ for each pool) by combining GA-matched samples. A total of 530 μg of protein from each pooled sample was acetone-precipitated and dissolved in 10 mM Tris, pH 8.5, for 2D-LC analysis. A total of 100 μg of the remaining individual sample was used for one-dimensional gel electrophoresis (1D-GE).

Multidimensional Liquid Chromatography (2D-LC). A total of 530 μg of protein from each pooled sample was dried and dissolved in 100 μL of digestion buffer containing 8 M urea, 1 M Tris base, 100 mM methylamine, and 10 mM $CaCl_2$ (pH 8.5). Samples were reduced and alkylated by first incubating at 50 °C in 12.5 μL of 0.9 M dithiothreitol (DTT) for 15 min and then in 25 μL of 1.0 M iodoacetamide in dark at room temperature for another 15 min. An additional 12.5 μL of 0.9 M DTT, along with 210 μL of water and 1N NaOH, was added to the solution to adjust its pH to 8.5. Samples were digested with 40 μL of 1 mg/mL trypsin (Promega) stock solution overnight at 37 °C. Digestion was stopped with 40 μL of formic acid and desalted using C18 SepPak Plus cartridges. Digests (1 mL) were injected onto a polysulfoethyl strong cation-exchange (SCX) column (2.1-mm i.d. × 100 mm, 5-μm particle size and 300-Å pore size (The Nest Group, Southborough, MA)) and fractionated using an HPLC equipped with a UV detector and a fraction collector. Solvent A was 10 mM potassium phosphate (pH 3) with 25% acetonitrile (ACN), and solvent B was 10 mM potassium phosphate (pH 3) and 350 mM KCl with 25% ACN. A 95-min gradient at a flow rate of 200 μL/min was employed for fractionation of peptides. In total, 80 fractions were collected, evaporated, and resuspended in 100 μL of 0.1% TFA for desalting using a 96-well Vydac C18 silica spin plate (The Nest Group, Southborough, MA). Fractions were eluted in 80% ACN/0.1% formic acid (FA), evaporated, and resuspended in 20 μL of 5% FA, and 5 μL of each fraction was analyzed on a Q-Tof-2 mass spectrometer connected to a CapLC (Waters, Milford, MA).

One-Dimensional Gel Electrophoresis (1D-GE) Analysis. A total of 100 μg of protein from an individual sample was reduced with iodoacetamide and resolved on a Tris-tricine, 10−20% gradient SDS-PAGE gel. The gel was stained with Coomassie blue R-250. Each lane was sliced into 25 individual bands, destained, and digested in-gel with trypsin for 24 h at 37 °C. The peptides were extracted in ammonium bicarbonate and then filtered with a 0.22-μm MultiScreen filter plate (Millipore, Billerica, MA). Filtered solutions were dried, reconstituted in 5% formic acid, and analyzed twice (technical replicates) on a Q-Tof-2 mass spectrometer equipped with a CapLC (Waters, Inc., Milford, MA).

Mass Spectrometry. 2D-LC fractions and gel digests were further separated using a Nanoease C18 75-μm i.d. × 15-cm fused-silica capillary column (Waters Inc., Milford, MA) and a 95-min water/ACN gradient. The mass spectrometer was calibrated using Glu1Fibrinopeptide B. An MS−MS/MS survey method was used to acquire spectra. Masses from m/z 400 to 1500 were scanned for MS survey and masses from m/z 50 to 1900 for MS/MS. In total, 27 397 MS/MS spectra were acquired from the 2D-LC fractions. Raw MS/MS spectra were preprocessed with ProteinLynx Global Server v.2.1 software (Waters, Inc., Milford, MA).

Protein and Peptide Identification. Supporting Information Figure 1 shows the protein and peptide identification workflow. Raw MS/MS spectra from either 2D-LC samples or 1D-GE samples were further processed by de-isotoping and centroiding the raw data. Preprocessed MS/MS spectra from different fractions of the sample were pooled for further analysis. Peptides present in the sample were identified by matching pooled MS/MS spectra to a combined protein database containing known contaminants and forward and reverse entries from the Swiss-Prot database (version 46.6) selected for human species. Peptide identification searches were performed using three independent search engines: TurboSequest (ThermoFinnigan, Waltham, MA), X! Tandem,[24] and OpenSea.[25,26] Sequest and X! Tandem are database search engines that match experimental spectra to theoretical spectra generated from a theoretical enzymatic digest of the protein database. OpenSea is a *de novo* sequence-based search engine that performs an error-tolerant matching between inexact *de novo* sequences and protein sequences in the database. Peaks software (Bioinformatics Solutions, Ontario, CA) was used to provide *de novo* sequences to the OpenSea search engine. The reduction and alkylation step of the sample processing introduces a fixed cabamidomethylation modification on all cysteine residues in the proteins. Hence, all the programs were configured to use the modified cysteine mass (160.03 Da) as standard mass for all cysteine residues. The *de novo* sequencer and all search engines were

configured to use monoisotopic masses to calculate parent and fragment ion masses. Peaks software was configured to use 0.2 and 0.1 Da for parent ion and fragment ion mass tolerances, respectively. The top five candidate *de novo* sequences reported by Peaks software for each MS/MS spectra were fed to OpenSea for error-tolerant database matching. OpenSea was configured to use 0.25 Da as the fragment ion mass tolerance. For Sequest searches, a parent ion mass tolerance of 2.0 Da was used to calculate parent ion mass. X! Tandem was configured to use mass tolerances of 0.5 and 0.25 Da for parent and fragment ions, respectively. To expedite Sequest searches, it was not configured to search for any variable modifications. In turn, on the basis of our prior experience, X! Tandem and OpenSea were configured to search for variable modifications (i.e., oxidation of methionine; pyroglutamic acid formation at the N terminus; carbamylation of the N terminus; dehydration of internal serine, threonine, aspartic acid, and glutamic acid residues; and deamidation intermediates of glutamine and asparagine) that could have been present in the MS/MS spectra either as artifacts from sample processing or peptide fragmentation mechanisms. Peptide identifications from individual search engines were combined into protein identifications using probabilistic protein identification algorithms implemented in Scaffold (v1.3.2, Proteome Software, Portland, OR).

Protein identifications that had at least one unique, highly confident (probability ≥ 0.9) peptide identification were considered likely to be present in the sample. A protein was accepted into the comprehensive list without manual validation if it was confidently identified in at least one of the samples with three highly confident unique peptide hits. Proteins that did not meet this filtering criterion were manually validated. Manual validation was performed using all criteria listed in reference 21, enhanced fragmentation C-terminal to aspartic acid,[27] and the presence of low-mass immonium ions (proline, valine, isoleucine, leucine, histidine, phenylalanine, and tyrosine) whenever these residues were present in the peptide sequence.

Results

Human CVF was analyzed using two different proteomics techniques: 2D-LC and 1D-GE. Two pooled samples were trypsinized and subjected to SCX fractionation, resulting in a total of 40 fractions. An individual sample was fractionated using 1D-GE, and the resulting bands were subject to in-gel trypsin digestion. In total, 27 397 MS/MS spectra were collected by analyzing all fractions on a LC−ESI-qTOF mass spectrometer. All MS/MS spectra were searched using Sequest, X! Tandem, and OpenSea. Peptide identifications from all programs were assembled into protein identifications using Scaffold.

In total, 831 proteins at the single-peptide identification level were identified when the lowest possible peptide identification probability thresholds (0.2) were employed. Thirty percent of the identified proteins were false-positive identifications (reverse database entries). Several protein isoforms and proteins that were subsets of other proteins were present in the list. Furthermore, the low-scoring (peptide identification probability < 0.9) peptide identifications did not manifest the necessary characteristics to pass the manual validation criterion listed in the Materials and Methods. A large proportion (54%) of the protein hits were also single-peptide identifications. Since single-peptide protein identifications are more likely to be false-positives and, therefore, insufficient for protein quantitation and inferring pathobiological function, a peptide identification probability of 0.9 was established as a minimal criterion to consider only highly confident peptide and protein identifications. Degenerate protein identifications were grouped together and reported as one entry, and any proteins that were subsets of other proteins were removed from the analysis.

A total of 206 unique proteins from all experiments was mapped to 55% of the experimental MS/MS spectra after applying the filtering described above. Three and 15% of the identified proteins in the list are false-positive identifications and single-peptide identifications, respectively. A total of 177 proteins remained after removing contaminants such as keratins, trypsin, and bovine casein. A total of 105 proteins that had at least three unique peptide hits in at least one of the experiments was accepted without further manual validation. The remaining protein identifications were manually validated using the criteria listed in the Materials and Methods. An additional 45 proteins passed manual validation; 29 of them had at least two unique peptide hits, and 16 had a single peptide hit. This increased the number of proteins that were identified with at least two distinct peptide hits to 134, and with at least one distinct peptide hit to 150.

To ensure the reliability of protein identifications, we performed all searches with a combined database that was constructed with reverse entries of the database appended at the end of the forward sequences. The number of reverse database entries that passed all criteria for protein identification was considered to reflect the reliability of the protein identification criteria outlined in the Materials and Methods. Since none of the reverse entries met these criteria, the reliability of the protein identifications was estimated to be 100%.

MS/MS spectral counting is generally considered to be a sensitive and semiquantitative method for measuring protein abundances.[28] However, homologous proteins pose a greater problem for accurate MS/MS spectral count representation due to their high sequence similarity. To avoid either inflation or deflation of MS/MS spectral counts of homologous proteins, a final level of filtering was performed to combine MS/MS spectral counts of protein homologues that share greater than 50% sequence homology. For example, squamous cell carcinoma 1 and 2 antigens share greater than 90% sequence homology. Although we have identified peptide hits that suggest the presence of both proteins in the sample, their MS/MS spectral counts were combined and represented as a single entry. Proteins that were combined under this criterion were IGHA1 and IGHA2; IGHG1, IGHG2, and IGHG4; SCCA1 and SCAA2; and SPR2A, SPR2B, and SPR2D. MS/MS spectral counts of peptides common to proteins that do not share high sequence homology were pulled toward the protein that was considered most likely (greater number of peptide hits) to be present in the sample. Finally, a combined MS/MS spectral count for each protein was established by combining the respective MS/MS spectral counts of the protein in all experiments. The combined MS/MS spectral count was normalized by the total number (12 827) of MS/MS spectra that were matched to noncontaminant proteins at a single-peptide probability threshold of 0.9 in all experiments. The normalized spectral counts are not strictly quantitative, but they can be used to gauge the relative abundance of the proteins present in a sample with respect to each other.

The final 134 proteins that had at least two unique peptide hits and passed manual validation are listed in Table 1 by their decreasing order of normalized MS/MS spectral counts. The

Table 1. Human CVF Proteome

Swiss-Prot acc. no.[a]	protein description	pI[b]	MW[c]	function[d]	normalized spectral count[e]	AF/Serum[f]
P02768	Serum albumin precursor	5.43	39.30	Transport	18.84	A, S
P01857, P01859, P01861	Ig gamma-1 chain C region, Ig gamma-2 chain C region, Ig gamma-4 chain C region	8.46, 7.66, 7.18	36.08, 35.9, 35.9	Immune Response	10.35	A, S
Q9UBC9	Small proline-rich protein 3	8.86	18.10	Cell Differentiation	8.6	A
P29508, P48594	Squamous cell carcinoma antigen 1, Squamous cell carcinoma antigen 2	6.35, 5.86	44.5, 44.8	Metabolism	6.3	
P06702	Calgranulin B	5.90	85.60	Immune Response	5.82	A,S
P07355	Annexin A2	4.69	55.30	Cell Differentiation	5.33	A, S
P04083	Annexin A1	6.96	11.10	Immune Response	4.28	A
Q01469	Fatty acid-binding protein, epidermal	6.82	15.00	Metabolism	3.97	
P01834	Ig kappa chain C region	5.58	11.60	Immune Response	3.79	A, S
P02787	Serotransferrin precursor	5.22	51.20	Transport	2.82	A, S
P05109	Calgranulin A	5.98	22.80	Immune Response	1.78	A, S
Q9HC84	Mucin-5B precursor	6.24	587.60	Transport	1.4	A
P04080	Cystatin B	8.39	39.30	Enzyme Regulator	1.09	A
P07476	Involucrin	7.56	38.40	Cell Differentiation	1.08	A
P01040	Cystatin A	5.38	11.00	Enzyme Regulator	1.04	A, S
P35321	Cornifin A	8.85	9.90	Cell Differentiation	1.03	A
Q09666	Neuroblast differentiation associated protein AHNAK	6.29	312.30	Cell Differentiation	1	
P01842	Ig lambda chain C regions	6.91	11.20	Immune Response	0.97	A, S
P30740	Leukocyte elastase inhibitor	5.90	42.70	Enzyme Regulator	0.94	
P05164	Myeloperoxidase precursor	6.51	10.80	Immune Response	0.84	
P02788	Lactotransferrin precursor	6.70	75.10	Immune Response	0.82	A, S
P80188	Neutrophil gelatinase-associated lipocalin precursor	9.02	20.50	Transport	0.81	A, S
P01009	Alpha-1-antitrypsin precursor	5.37	44.30	Immune Response	0.76	A, S
P61626	Lysozyme C precursor	9.28	14.70	Metabolism	0.76	A
P01876, P01877	Ig alpha-1 chain C region, Ig alpha−2 chain C region	6.08, 5.71	37.6, 36.5	Immune Response	0.73	A, S
P04792	Heat-shock protein beta-1	8.58	35.90	Metabolism	0.73	
P10599	Thioredoxin	5.44	23.20	Metabolism	0.72	A
P62988	Ubiquitin	6.56	8.60	Metabolism	0.65	
P01833	Polymeric-immunoglobulin receptor precursor	5.59	83.30	Signal Transduction	0.47	A
P12429	Annexin A3	5.37	70.90	Enzyme Regulator	0.44	
Q8TDL5	Long palate, lung and nasal epithelium protein 1	6.69	50.20	Function Not Assigned	0.44	
P00450	Ceruloplasmin precursor	5.41	120.00	Transport	0.43	A, S
P35326, P35325, P22532	Small proline-rich protein 2A, Small proline-rich protein 2B, Small proline-rich protein 2D	8.81, 8.81, 8.77	8, 7.97, 7.9	Cell Differentiation	0.41	
P60709	Actin, cytoplasmic 1	5.29	41.60	Cell Organization	0.38	A, S
P01011	Alpha-1-antichymotrypsin precursor	5.33	47.60	Immune Response	0.36	A, S
P22528	Cornifin B	8.85	9.90	Cell Differentiation	0.34	
P00738	Haptoglobin precursor	6.13	45.20	Metabolism	0.3	A, S
P62328	Thymosin beta-4	5.02	4.90	Cell Organization	0.3	A
P18510	Interleukin-1 receptor antagonist protein precursor	5.51	123.60	Immune Response	0.27	
P01024	Complement C3 precursor	6.02	187.00	Immune Response	0.24	A, S
P07737	Profilin-1	4.62	68.40	Cell Organization	0.22	A
P02790	Hemopexin precursor	8.50	78.10	Transport	0.21	A, S
P14780	Matrix metalloproteinase-9 precursor	5.18	9.00	Metabolism	0.21	
P04406	Glyceraldehyde-3-phosphate dehydrogenase, liver	9.30	52.10	Metabolism	0.2	S
P15924	Desmoplakin	5.95	69.20	Cell Differentiation	0.2	S
P08107	Heat shock 70 kDa protein 1	4.94	84.50	Metabolism	0.19	
Q9NQ38	Serine protease inhibitor Kazal-type 5 precursor	8.50	120.70	Immune Response	0.19	
P12724	Eosinophil cationic protein precursor	5.63	36.20	Metabolism	0.18	
P04279	Semenogelin-1 precursor	6.64	38.60	Cell Differentiation	0.17	S
O60437	Periplakin	5.44	204.50	Function Not Assigned	0.16	
P09211	Glutathione *S*-transferase P	5.06	53.50	Metabolism	0.16	
P02749	Beta-2-glycoprotein I precursor	8.37	36.20	Immune Response	0.15	A, S
P07108	Acyl-CoA-binding protein	6.99	47.00	Transport	0.15	
P59665	Neutrophil defensin 1 precursor	6.54	10.20	Immune Response	0.15	S
O60235	Transmembrane protease, serine 11D precursor	8.69	46.20	Metabolism	0.13	
P03973	Antileukoproteinase 1 precursor	6.43	49.30	Enzyme Regulator	0.13	A
P04075	Fructose-bisphosphate aldolase A	6.95	59.60	Enzyme Regulator	0.13	
P14923	Junction plakoglobin	5.69	78.40	Transport	0.13	S
P62805	Histone H4	11.36	11.20	Cell Organization	0.12	
P62937	Peptidyl-prolyl cis-trans isomerase A	7.82	17.90	Metabolism	0.12	A
Q02383	Semenogelin-2 precursor	9.04	62.90	Cell Differentiation	0.12	S
P02774	Vitamin D-binding protein precursor	5.67	66.40	Transport	0.11	A, S
P07858	Cathepsin B precursor	8.47	14.90	Metabolism	0.11	
P24158	Myeloblastin precursor	7.79	24.20	Metabolism	0.11	

Table 1. (Continued)

Swiss-Prot acc. no.[a]	protein description	p*I*[b]	MW[c]	function[d]	normalized spectral count[e]	AF/Serum[f]
P00441	Superoxide dismutase	5.70	15.80	Cell Differentiation	0.1	
P02763	Alpha-1-acid glycoprotein 1 precursor	5.00	21.50	Immune Response	0.1	A, S
P02765	Alpha-2-HS-glycoprotein precursor	5.00	21.50	Signal Transduction	0.1	A, S
P04040	Catalase	5.55	54.30	Metabolism	0.1	S
P13796	L-plastin	6.42	95.10	Function Not Assigned	0.1	A, S
P54108	Cysteine-rich secretory protein-3 precursor	8.11	25.50	Immune Response	0.1	
O43707	Alpha-actinin 4	5.27	104.80	Cell Organization	0.09	
P06733	Alpha enolase	5.71	13.20	Metabolism	0.09	
P11142	Heat shock cognate 71 kDa protein	5.01	70.40	Metabolism	0.09	S
P18206	Vinculin	6.44	331.60	Transport	0.09	S
P26038	Moesin	6.09	67.60	Cell Organization	0.09	
P27482	Calmodulin-related protein NB-1	4.30	16.70	Immune Response	0.09	
P32926	Desmoglein-3 precursor	4.76	101.70	Transport	0.09	
P67936	Tropomyosin alpha 4 chain	4.67	28.40	Function Not Assigned	0.09	S
Q02487	Desmocollin-2 precursor	4.80	84.70	Transport	0.09	
Q9UGL9	NICE-1 protein	9.13	9.70	Function Not Assigned	0.09	
P00558	Phosphoglycerate kinase 1	8.30	44.50	Metabolism	0.08	
P01625	Ig kappa chain V-IV region Len	7.92	12.63	Immune Response	0.08	A, S
P01871	Ig mu chain C region	6.35	49.50	Immune Response	0.08	A
P16402	Histone H1.3	11.02	22.20	Cell Organization	0.08	
P63104	14-3-3 protein zeta/delta	4.73	27.70	Metabolism	0.08	A
P02679	Fibrinogen gamma chain precursor	5.24	48.50	Cell Proliferation	0.07	A, S
P08311	Cathepsin G precursor	9.89	25.50	Metabolism	0.07	
P60174	Triosephosphate isomerase	6.51	26.50	Metabolism	0.07	
P80723	Brain acid soluble protein 1	4.64	22.50	Function Not Assigned	0.07	
P01617	Ig kappa chain V-II region TEW	5.69	12.30	Immune Response	0.06	S
P01620	Ig kappa chain V-III region SIE	8.70	11.80	Immune Response	0.06	A, S
P05387	60S acidic ribosomal protein P2	4.26	11.50	Metabolism	0.06	
P11021	78 kDa glucose-regulated protein precursor	4.82	11.60	Metabolism	0.06	S
P29373	Retinoic acid-binding protein II, cellular	5.43	15.60	Metabolism	0.06	
O75223	Protein C7orf24	5.07	21.00	Function Not Assigned	0.05	
P07900	Heat shock protein HSP 90-alpha	5.88	37.80	Transport	0.05	
P13987	CD59 glycoprotein precursor	5.20	70.20	Signal Transduction	0.05	A
P31151	S100 calcium-binding protein A7	6.26	11.31	Cell Differentiation	0.05	
P31947	14-3-3 protein sigma	4.68	27.80	Cell Proliferation	0.05	S
P37837	Transaldolase	6.36	37.50	Metabolism	0.05	
P47929	Galectin-7	7.00	14.90	Transport	0.05	S
Q16610	Extracellular matrix protein 1 precursor	6.19	58.80	Signal Transduction	0.05	A, S
Q99880	Histone H2B.c	10.32	13.81	Cell Organization	0.05	S
O95171	Sciellin	9.38	75.30	Cell Differentiation	0.04	
P01028	Complement C4 precursor	6.66	192.70	Immune Response	0.04	A, S
P01042	Kininogen precursor	6.34	71.90	Immune Response	0.04	A, S
P04004	Vitronectin precursor	9.11	11.70	Immune Response	0.04	A, S
P07237	Protein disulfide-isomerase precursor	6.11	9.90	Metabolism	0.04	S
P08603	Complement factor H precursor	11.37	26.70	Immune Response	0.04	A, S
P14618	Pyruvate kinase, isozymes M1/M2	7.95	57.70	Metabolism	0.04	
P20670	Histone H2A.o	10.90	13.95	Cell Organization	0.04	S
P20810	Calpastatin	4.99	76.50	Enzyme Regulator	0.04	
P22735	Protein-glutamine gamma-glutamyltransferase K	5.68	89.70	Metabolism	0.04	
Q06830	Peroxiredoxin 1	8.27	22.10	Cell Differentiation	0.04	
Q13835	Plakophilin 1	9.29	82.80	Signal Transduction	0.04	
P00747	Plasminogen precursor	7.04	90.50	Metabolism	0.03	A, S
P05386	60S acidic ribosomal protein P1	9.19	83.80	Metabolism	0.03	
P08670	Vimentin	6.14	136.90	Function Not Assigned	0.03	
P28799	Granulins precursor	6.43	63.50	Cell Proliferation	0.03	
P30086	Phosphatidylethanolamine-binding protein	7.43	20.90	Enzyme Regulator	0.03	
P35237	Placental thrombin inhibitor	5.18	42.60	Enzyme Regulator	0.03	
P01591	Immunoglobulin J chain	4.62	15.60	Immune Response	0.02	S
P02647	Apolipoprotein A-I precursor	5.56	30.80	Metabolism	0.02	A, S
P02675	Fibrinogen beta chain precursor	8.54	55.90	Cell Proliferation	0.02	A, S
P13639	Elongation factor 2	10.72	15.60	Metabolism	0.02	
P18669	Phosphoglycerate mutase 1	5.46	17.10	Metabolism	0.02	A
Q9UBX7	Kallikrein 11 precursor	9.23	31.03	Metabolism	0.02	
Q9UKR3	Kallikrein 13 precursor	8.79	28.90	Metabolism	0.02	

[a] Proteins having at least two peptide identifications found in human CVF are listed with their Swiss-Prot/TrEmbl accession number and description. Homologous protein identifications are grouped together as a single entry. [b,c] Theoretical p*I*s (*b*) and monoisotopic molecular weights (*c*) were calculated using the CalPI/MW tool[58] on the Swiss-Prot Web site. [d] Functional annotation was performed using the DAVID database[29] bioinformatics resource. [e] Combined spectral counts for each of the protein identifications from both 1D-GE and 2D-LC experiments were normalized by the total number (12 827) of MS/MS spectra that were matched (noncontaminant proteins) at a single-peptide probability threshold of 0.9 in all samples. The proteins in the table are ordered by decreasing normalized spectral counts. [f] Proteins that were also seen in either AF (A) and/or serum (S) are marked accordingly (see text for discussion).

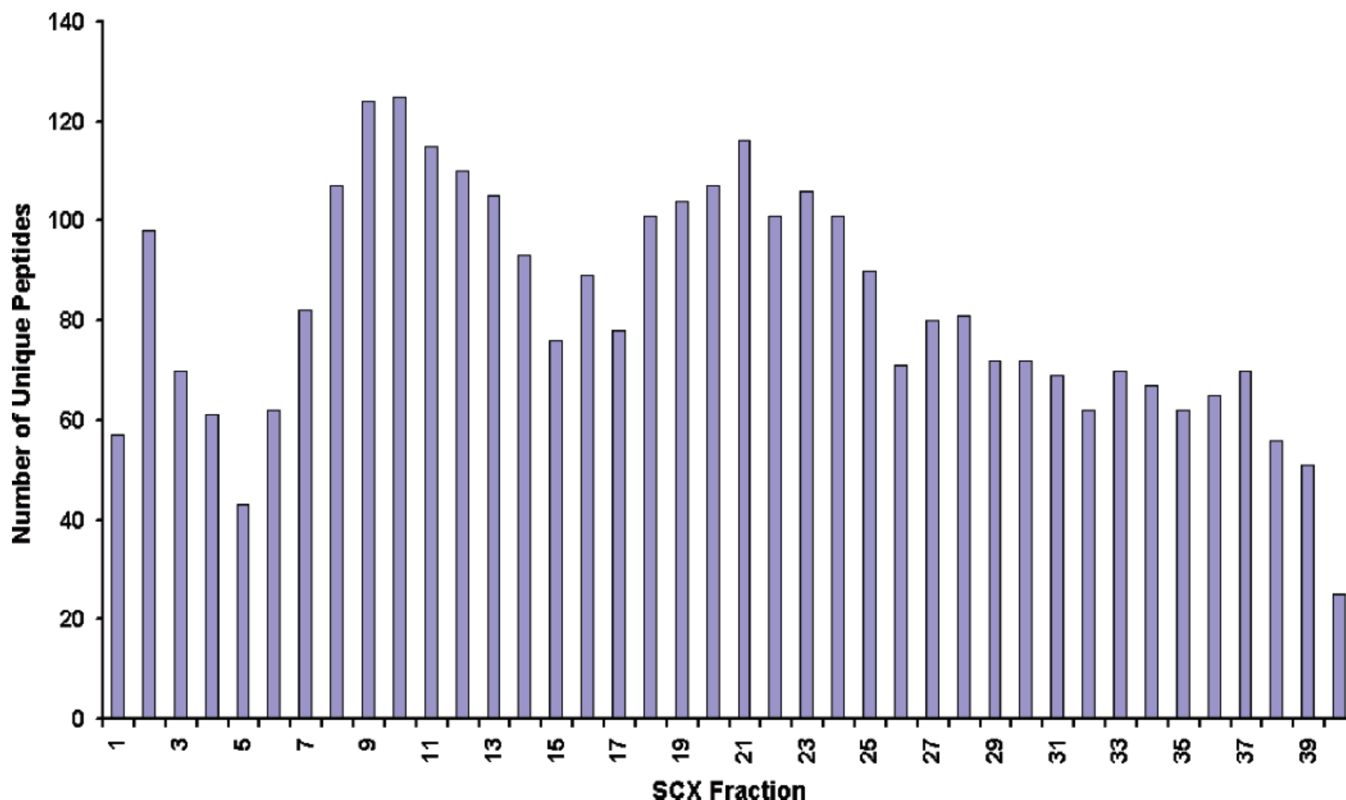

Figure 1. Distribution of the number of unique peptides identified in SCX fractions. The total number of unique peptides identified per fraction in human CVF 2D-LC samples shows the advantage of the technique over traditional gel-based electrophoresis techniques.

16 proteins that had a single peptide hit and passed the manual validation are listed in Supporting Information Table 1 by their decreasing order of combined MS/MS spectral counts. Proteins listed in the tables are functionally annotated based on the classification from the Database for Annotation, Visualization and Integrated Discovery (DAVID).[29]

The CVF proteins found in this study were cross-referenced with the HUPO plasma proteome[20,30] and the AF proteome.[31−33] The HUPO plasma proteome was further curated by converting the IPI database protein accessions into Swiss-Prot/TrEmbl protein accessions wherever possible and removing common contaminants like keratin. Protein isoforms reported in the HUPO plasma proteome were collapsed into a single protein entry due to the lack of direct MS/MS spectral evidence that could resolve different isoforms. The curated HUPO plasma proteome (526 proteins) was combined with the 195 proteins listed in Anderson et al.[20] to make a nonredundant, highly confident HUPO plasma proteome (data not shown). CVF protein identifications were compared to the curated HUPO plasma proteome and the AF proteome based on their Swiss-Prot/TrEmbl protein annotation and are marked accordingly (A, found in amniotic fluid; S, found in serum) in the last column of the corresponding tables.

Discussion

The 2D-LC technique is known to provide enhanced fractionation compared to traditional gel-based electrophoresis methods. Figure 1 shows the number of unique peptides identified per SCX fraction from the 2D-LC fractionation. Clearly, the enhanced fractionation of the technique, when coupled with RP-HPLC, contributed to the identification of a greater number of unique peptides per SCX fraction and, overall, a large number of protein identifications in the sample.

Recent studies have shown that a MS/MS dataset can be thoroughly characterized using multiple search engines to identify the peptides in that data set.[22] When different search engines are used to identify peptides in a data set, they identify different sets of MS/MS spectra due to the difference in heuristics that are encoded in the corresponding search engines. Thus, a combination of different search engine results on the same data set gives a more comprehensive list of peptide identifications. In this study, we have used three different search engines to identify the peptides present in the samples: Sequest, X! Tandem, and OpenSea. Using this combinatorial approach, we were able to identify 59% of the acquired MS/MS spectra in one of the 2D-LC experiments. The breakdown of percentages of spectral identifications (above the score cutoffs of the corresponding programs) between the three programs (Figure 2a) shows that only 38% of spectra were identified by all three programs, whereas 22% of spectra were identified uniquely by only one of the programs. Interestingly, 15% of the spectra were identified solely by the OpenSea search engine. This is due to the ability of OpenSea to identify spectra with missing fragment ions and unexpected sequence modifications. The total number of candidate proteins identified in the sample was also increased due to the combinatorial search technique. Among a total of 118 candidate protein identifications as shown in Figure 2b, 66% were identified by all three programs, whereas 13% were identified uniquely by only one of the programs. Thus, the combinatorial search technique employed in this study identified more peptide and candidate protein identifications from the data sets.

The composition of maternal CVF changes with gestational age and vaginal health. The overlap of proteins identified in two biological replicates from 2D-LC experiments is shown in Figure 3. Sixty-nine percent of the proteins were identified in

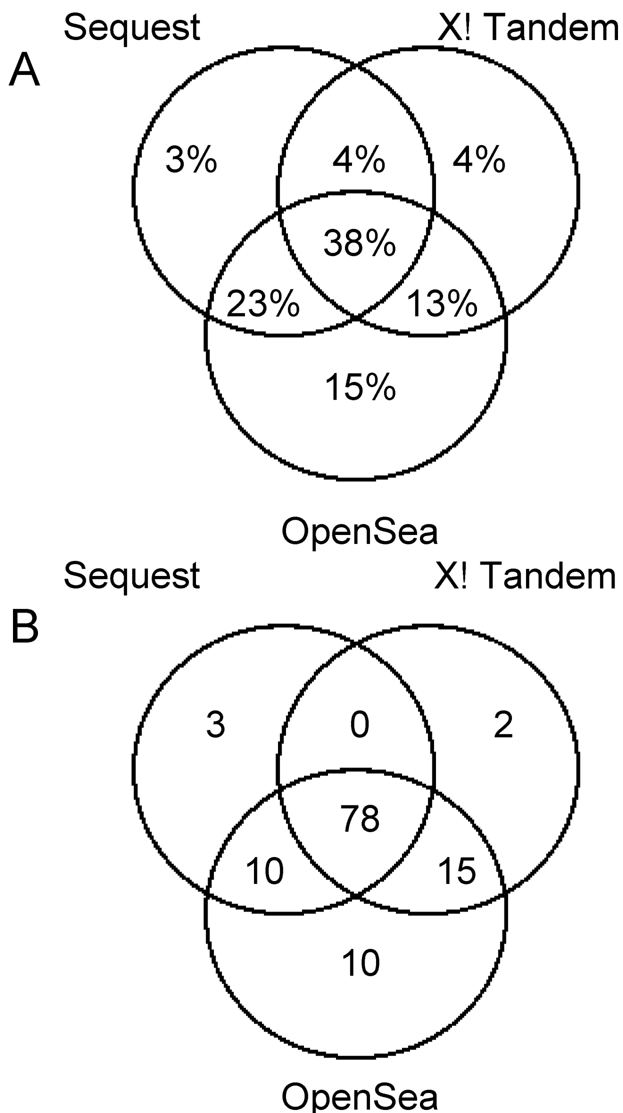

Figure 2. Spectral and protein identification overlap between different search engines. (A) A total of 9507 MS/MS spectra from a human CVF 2D-LC experiment was searched with Sequest, X! Tandem, and OpenSea search engines. When three independent search engines are used, a total of 5601 (59%) MS/MS spectra in the sample was matched to proteins at a two-unique-peptide identification threshold. The distribution of percentage of identified spectra (above threshold) between search engines shows that using multiple independent search engines identified more MS/MS spectra in a sample. (B) A total of 118 candidate proteins was identified when all three search engines were used to analyze MS/MS spectra from a single human CVF 2D-LC sample. The distribution of a number of candidate protein hits among three search engines shows that using multiple independent search engines identified more proteins.

both biological replicates, whereas 31% of the proteins were identified solely in one of the replicates. This was not unexpected, since GA-dependent changes in the cervix play a significant biological role in the process of labor and delivery. The random sampling of low-abundance proteins by the mass spectrometer might also have contributed to the above-mentioned difference. Among 65 proteins that were identified by 1D-GE technical replicates, 69% were identified in both replicates, whereas 31% were identified uniquely in one of the replicates. This underscores the importance of having biological

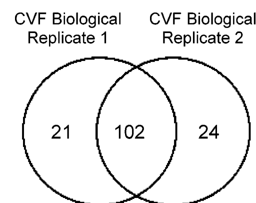

Figure 3. Protein identification distribution in biological replicates. The Venn diagram shows the distribution of protein identifications between CVF biological replicates that were analyzed using 2D-LC. Among a total of 147 identified proteins, 102 proteins were present in both samples, and 45 proteins were present in one sample or the other.

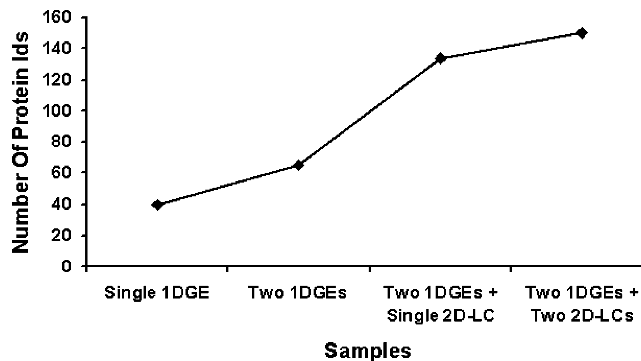

Figure 4. Protein identification distribution in different analytical approaches. The protein identification rate increased by addition of biological replicates and experimental replicates to the analysis. Among a total of 150 proteins, 62 proteins were detected by both 1D-GE and 2D-LC techniques, 85 were detected by 2D-LC alone, and 3 (all of them having at least 2 unique peptide identifications) were detected by 1D-GE alone.

and technical replicates when characterizing proteomes. The overall increase in the number of protein identifications with the addition of experiments to the analysis is summarized in Figure 4. A total of 40 proteins was identified by our protein identification criteria when using a single 1D-GE experiment. An increase of 15, 69, and 16 protein identifications was observed when a single 1D-GE technical replicate, the 2D-LC experiment, and its corresponding biological replicate were added to the analysis, respectively. This is the first comprehensive proteomics study that has employed a variety of analysis programs, technical replicates, biological replicates, and experimental methods to characterize the human CVF proteome.

The combinatorial proteomics approach applied in this study characterized the proteomic composition of CVF during pregnancy by uncovering a large number of proteins that were not previously known to be present in CVF. Table 1 and Supporting Information Table 1 list a comprehensive set of proteins present in CVF that are involved in homeostasis of the reproductive area and fetal protection. The tryptic peptide profile of the proteins listed in Table 1 is shown in Figure 5. Over 89% of the proteins had at least two unique peptide identifications. The peptide profile also shows that CVF contains a variety of proteins that have a wide range of tryptic peptide yields. Figure 6 shows the functional classification of the CVF proteome during pregnancy. The major functional groups in CVF are immune and defense-related molecules (such as calgranulins A and B) and metabolic molecules (ranging from proteases such

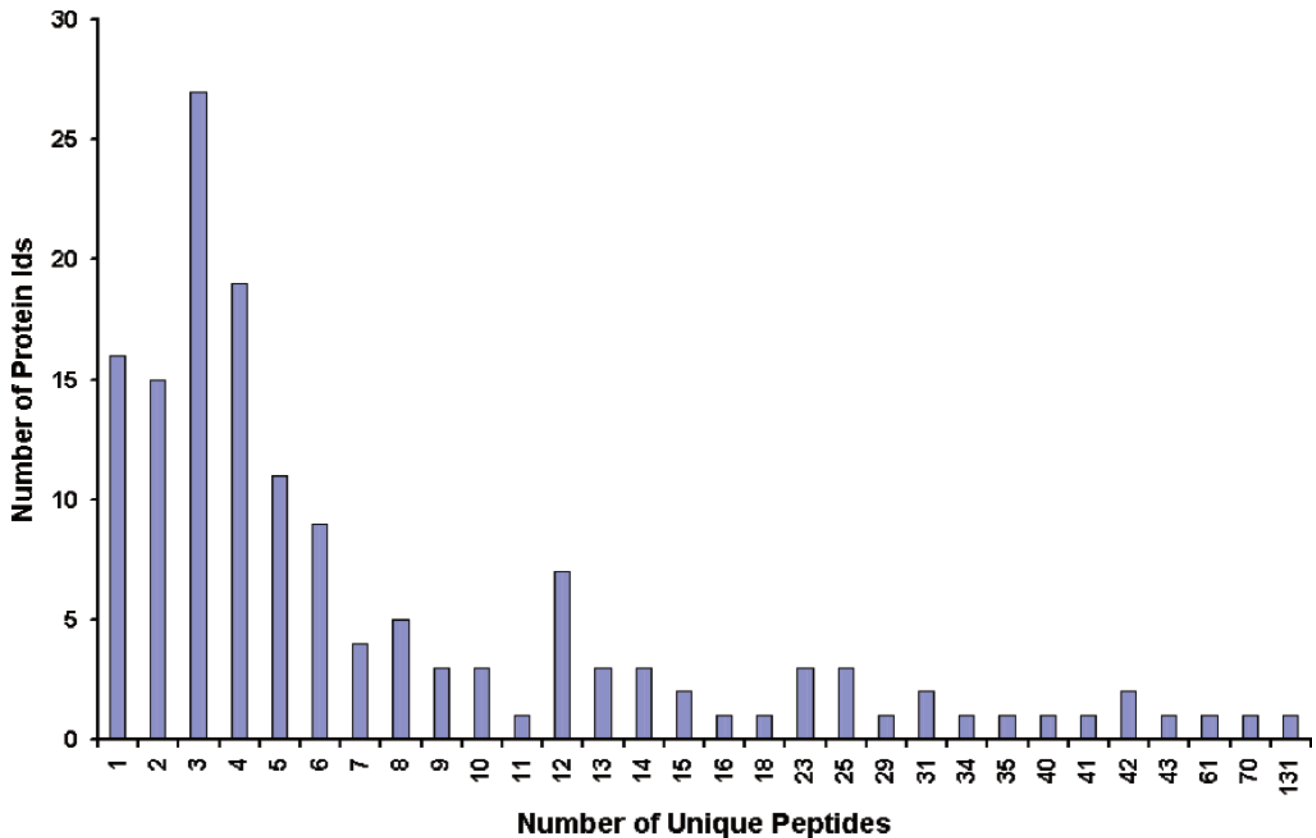

Figure 5. Tryptic peptide profile of proteins in human CVF. The tryptic peptide profile of the proteins identified in CVF shows that over 89% of the identifications had at least two unique peptide identifications; however, proteins with a wide range of tryptic peptide yields were identified in CVF.

as cathepsins B and G to chaperones, like heat shock protein 90-α (HSP 90-α).

The immunoresponse proteins found in this study fell into three categories: proinflammatory response molecules, anti-inflammatory response molecules, and antimicrobial molecules. Apart from commonly occurring immunogloblins, the most notable proinflammatory response molecules found in CVF are two calcium-binding proteins from the S100 family, calgranulins A and B. These proteins form a heterodimer mediated by Ca^{+2} ions and are commonly implicated in both acute-phase and chronic inflammation responses.[34] The relative abundance of these proteins in a control CVF sample, when compared to albumin (Table 1), suggests their vital role in fighting vaginal infections. It is also interesting to note that calgranulins A and B are also found in the intra-amniotic fluid during intra-amniotic infection,[35,36] which could lead to preterm labor and birth. Anti-inflammatory response molecules are vital during pregnancy to down-regulate the maternal immune response and prevent immune rejection of the fetus, or to avoid preeclampasia.[37,38] Several of the proteins we detected in CVF, most notably interleukin 1 receptor antagonist (IL1-ra) and heat shock protein 70 kDa (HSP70), belong to the group of molecules which, when expressed during pregnancy, help in down-regulation of the immune response. The secretion of HSP70 into the CVF during pregnancy complicated with vaginal infection induces the expression of IL1-ra.[39] Presumably, this is a mechanism to preserve a pregnancy on the immunoregulatory level, despite the untoward effects of infection.

Antimicrobial proteins play an important role in preventing infection of the vagina from bacterial and fungal pathogens. Confirming previous reports,[40] we have detected neutrophil defensin 1 (defensin family) and lactotransferrin in CVF, which are known to have antimicrobial properties and may protect the vagina from infections like *Neisseriae gonorrhoeae* and HSV.[40] Additionally, we have also detected several proteins from the histone family (H4, H2A, H2B, and H1.2). Traditionally, histones are considered to be intracellular proteins that are involved in chromatin arrangement inside the nucleus. However, recent studies have indicated that secreted neutrophil extracellular traps (NETs) contain histones,[41,42] and that secreted histones have a broad range of antimicrobial properties.[13,43−46] The detection of a wide range of pro- and anti-inflammatory response molecules, along with various antimicrobial molecules, suggests that the CVF has a complex milieu of innate immune response.

A major proportion (32%) of proteins found in this study are involved in various metabolic activities (Figure 6) like inflammatory regulation, protein degradation, and protease inhibition. Among the inflammatory regulation proteins we have observed are HSP90-A, bradykinin (kininogen 1 precursor), and kallikrein (kallikrein 11 and 13 precursors). HSP90-A has been recently reported to be involved in cell-mediated activation of the proinflammatory bradykinin-kallikrein complex.[47] Such cell-mediated immunity has been shown to be a key factor in defense against pathogens that infect the lower female genital tract.[48] The balance between proteases and protease inhibitors is critical for maintenance of healthy tissue, and imbalances often lead to serious cervical epithelial pathology. Among several proteases and antiproteases we observed in CVF, one of the interesting pairs is cathepsin B and α1 antitrypsin (A1AT).

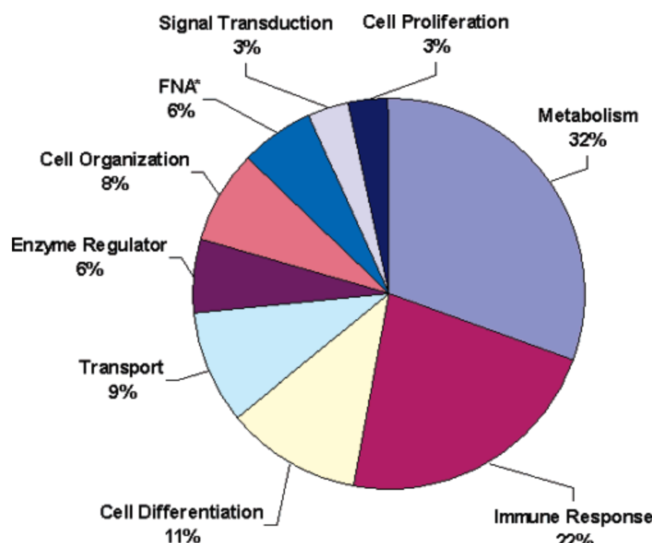

Figure 6. Functional classification of the human CVF proteome. Functional annotation of human CVF proteins was performed using functional classification tools available on the DAVID bioinformatics resource.[29] From a total of 150 identified proteins, 32% are involved in metabolism, 22% are involved in immune response, 11% are involved in cell differentiation, 9% are involved in transportation, 8% are involved in cell organization, 6% are involved in enzyme regulation, 3% are involved in signal transduction, and 3% in cell proliferation. No relevant functional annotations* (FNA; function not assigned) were found in the DAVID database for 6% of the identified proteins.

In cases of cervical carcinoma, the levels of cathepsin B in CVF are elevated, while the levels of A1AT are unchanged.[49−51] Thus, imbalance between protease and antiprotease expression in the cervix could be used to detect invasive cervical carcinomas. Detection of the above-mentioned metabolic proteins in CVF suggests that it contains enzymes that regulate a variety of functions ranging from regulation of inflammatory response to maintenance of cervical tissue health.

Apart from immune response and metabolic proteins, we also found proteins that aid in cell differentiation (11%), transport (9%), cell organization (8%), enzyme regulation (6%), signal transduction (3%), and cell proliferation (3%). A protein could have multiple functions depending on its environment. For example, according to the DAVID functional annotation tool, histones are classified as proteins involved in cell organization. However, as discussed earlier, they also have antimicrobial properties when secreted outside the cell. Thus, the role of many of the other proteins found in CVF during pregnancy is still unclear and warrants further investigation.

Prior to this study, the relative abundance of proteins that are native to CVF during pregnancy was largely unknown. The proteins in Table 1 are arranged by their decreasing order of normalized spectral counts. The generic ratio of IgG/IgA protein abundance in our analysis matches well with previous studies.[52] Interestingly, the protein abundance profile of CVF and serum differ significantly. Among the top 15 abundant CVF proteins, six proteins are known to be either non-native and/or low-abundance in serum (squamous cell carcinoma antigens, calgranulins A and B, small proline rich protein 3, fatty acid-binding protein epidermal, and mucin 5B).[20,21,53,54] Furthermore, proteins that are known to be in medium abundance in serum (complement factor C4, complement factor H, and apolipoprotein A-1) were found to be in low abundance in CVF.[20] Inspection of Table 1 suggests that 40% of the top 10 most-abundant proteins in the CVF are inflammatory response molecules. This further supports the assertion that CVF has an effective and aggressive cytokine response system in order to deal with pathogenesis.

A quantitative analysis of proteome overlap between the AF, serum, and CVF was carried out, and the last column in Table 1 and Supporting Information Table 1 denotes the CVF proteins that were also observed in AF (A) and serum (S). Active serum transport and local synthesis are known to be the sources of serum proteins in the cervix.[55] Confirming this, we found the sIgA complex, which is locally synthesized in the cervix.[56] In addition, we detected several abundant serum proteins[20,30,55] like serum albumin, alpha-1-antitrypsin precursor, apolipoprotein A1 precursor, serotransferrin, lactotransferrin, apolipoprotein A1 precursor, α-2-HS glycoprotein, Ig γ 1, 2, and 4 chain C regions, and β-2-glycopotein 1 precursor in CVF. It is interesting to note that we also detected several proteins in CVF, like small proline-rich protein 3, CD59 glycoprotein precursor, cystatin A, cystatin B, cornifin A, involucrin, and thioredoxin, which are found in AF but not serum. Parallel secretions of the chorionic−decidual membrane could be a source of these proteins in CVF. Among the proteins that were present in all three biological fluids, A1AT and ceruloplasmin (copper transporter) are known to have diagnostic importance. The abundance ceruloplasmin in maternal vaginal secretions and serum has been inversely correlated with incidence of premature rupture of membranes (PROM),[17,57] and increased expression of A1AT in serum has been correlated with cervical cancer.[51] This suggests that serial assessment of easily accessible body fluids like CVF or serum could be used in maternal and fetal health diagnostics. However, a caveat of the above conclusion is that the HUPO plasma proteome used in the comparison consists of both nonmaternal and maternal proteins. A separate comparison of the CVF proteome with second-trimester maternal serum proteome consisting of 392 proteins (data not shown) yielded 53 common proteins.

In summary, we have employed a combinatorial proteomics approach using multiple biological replicates, multiple experimental techniques for protein and peptide fractionation, and multiple search engines for data mining to characterize the CVF proteome. This multiplexed approach identified a large set of proteins that were not previously known to be present in CVF. The functional classification of the CVF proteome suggested the presence of a wide variety of cytokine response proteins that play a vital role in fighting pathogens and protecting the fetus. A quantitative analysis of proteome overlap between serum, AF, and CVF identified several serum and AF proteins as present in the CVF during pregnancy. Differential abundance of some of those proteins has already been linked to PROM and cervical cancer. This initial characterization of the CVF proteome involved a small sample size covering limited GA. Further investigation with serial first-, second-, and third-trimester sampling could enhance the proteome coverage and provide potential GA-dependent profiles. Large-scale, high-throughput proteomics technologies are vital to further our understanding of the CVF proteome during pregnancy and its development as a potential diagnostic tool for monitoring maternal and fetal health.

Acknowledgment. This work has been supported in part by ProteoGenix, Inc. Oregon Health & Science University and Drs. Gravett, Roberts, and Nagalla all have a significant

financial interest in ProteoGenix, Inc., a company that may have a commercial interest in the results of this research and technology. This potential conflict of interest has been reviewed and a management plan approved by the OHSU Conflict of Interest in Research Committee.

Supporting Information Available: Figure showing the protein and peptide identification workflow; table listing the single-peptide protein identifications of the human CVF proteome. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Blandau, R. J.; Moghissi, K. S. *The Biology of the Cervix*; University of Chicago Press: Chicago, IL, 1973; p xi, p 450.

(2) Bigelow, J. L.; Dunson, D. B.; Stanford, J. B.; Ecochard, R.; Gnoth, C.; Colombo, B. Mucus observations in the fertile window: a better predictor of conception than timing of intercourse. *Hum. Reprod.* **2004**, *19* (4), 889−92.

(3) Mattsby-Baltzer, I.; Platz-Christensen, J. J.; Hosseini, N.; Rosen, P. IL-1beta, IL-6, TNFalpha, fetal fibronectin, and endotoxin in the lower genital tract of pregnant women with bacterial vaginosis. *Acta Obstet. Gynecol. Scand.* **1998**, *77* (7), 701−6.

(4) Eschenbach, D. A.; Davick, P. R.; Williams, B. L.; Klebanoff, S. J.; Young-Smith, K.; Critchlow, C. M.; Holmes, K. K. Prevalence of hydrogen peroxide-producing Lactobacillus species in normal women and women with bacterial vaginosis. *J. Clin. Microbiol.* **1989**, *27* (2), 251−6.

(5) Mikamo, H.; Sato, Y.; Hayasaki, Y.; Kawazoe, K.; Izumi, K.; Ito, K.; Yamada, Y.; Tamaya, T. Intravaginal bacterial flora in patients with uterine cervical cancer. High incidence of detection of Gardnerella vaginalis. *J. Infect. Chemother.* **1999**, *5* (2), 82−5.

(6) Ness, R. B.; Kip, K. E.; Hillier, S. L.; Soper, D. E.; Stamm, C. A.; Sweet, R. L.; Rice, P.; Richter, H. E. A cluster analysis of bacterial vaginosis-associated microflora and pelvic inflammatory disease. *Am. J. Epidemiol.* **2005**, *162* (6), 585−90.

(7) Haggerty, C. L.; Hillier, S. L.; Bass, D. C.; Ness, R. B. Bacterial vaginosis and anaerobic bacteria are associated with endometritis. *Clin. Infect. Dis.* **2004**, *39* (7), 990−5.

(8) Morris, M.; Nicoll, A.; Simms, I.; Wilson, J.; Catchpole, M. Bacterial vaginosis: a public health review. *BJOG* **2001**, *108* (5), 439−50.

(9) Gravett, M. G.; Nelson, H. P.; DeRouen, T.; Critchlow, C.; Eschenbach, D. A.; Holmes, K. K. Independent associations of bacterial vaginosis and Chlamydia trachomatis infection with adverse pregnancy outcome. *JAMA, J. Am. Med. Assoc.* **1986**, *256* (14), 1899−903.

(10) Poli, G.; Fauci, A. S. The effect of cytokines and pharmacologic agents on chronic HIV infection. *AIDS Res. Hum. Retroviruses* **1992**, *8* (2), 191−7.

(11) Zara, F.; Nappi, R. E.; Brerra, R.; Migliavacca, R.; Maserati, R.; Spinillo, A. Markers of local immunity in cervico-vaginal secretions of HIV infected women: implications for HIV shedding. *Sex Transm. Infect.* **2004**, *80* (2), 108−12.

(12) John, M.; Keller, M. J.; Fam, E. H.; Cheshenko, N.; Hogarty, K.; Kasowitz, A.; Wallenstein, S.; Carlucci, M. J.; Tuyama, A. C.; Lu, W.; Klotman, M. E.; Lehrer, R. I.; Herold, B. C. Cervicovaginal secretions contribute to innate resistance to herpes simplex virus infection. *J. Infect. Dis.* **2005**, *192* (10), 1731−40.

(13) Venkataraman, N.; Cole, A. L.; Svoboda, P.; Pohl, J.; Cole, A. M. Cationic polypeptides are required for anti-HIV-1 activity of human vaginal fluid. *J. Immunol.* **2005**, *175* (11), 7560−7.

(14) Maher, D.; Wu, X.; Schacker, T.; Horbul, J.; Southern, P. HIV binding, penetration, and primary infection in human cervicovaginal tissue. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (32), 11504−9.

(15) Quinones-Mateu, M. E.; Lederman, M. M.; Feng, Z.; Chakraborty, B.; Weber, J.; Rangel, H. R.; Marotta, M. L.; Mirza, M.; Jiang, B.; Kiser, P.; Medvik, K.; Sieg, S. F.; Weinberg, A. Human epithelial beta-defensins 2 and 3 inhibit HIV-1 replication. *AIDS* **2003**, *17* (16), F39−48.

(16) Helmig, B. R.; Romero, R.; Espinoza, J.; Chaiworapongsa, T.; Bujold, E.; Gomez, R.; Ohlsson, K.; Uldbjerg, N. Neutrophil elastase and secretory leukocyte protease inhibitor in prelabor rupture of membranes, parturition and intra-amniotic infection. *J. Matern.−Fetal Neonat. Med.* **2002**, *12* (4), 237−46.

(17) Ogino, M.; Hiyamuta, S.; Takatsuji-Okawa, M.; Tomooka, Y.; Minoura, S. Establishment of a prediction method for premature rupture of membranes in term pregnancy using active ceruloplasmin in cervicovaginal secretion as a clinical marker. *J. Obstet. Gynaecol. Res.* **2005**, *31* (5), 421−6.

(18) Swamy, G. K.; Simhan, H. N.; Gammill, H. S.; Heine, R. P. Clinical utility of fetal fibronectin for predicting preterm birth. *J. Reprod. Med.* **2005**, *50* (11), 851−6.

(19) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242−7.

(20) Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **2004**, *3* (4), 311−26.

(21) Wilmarth, P. A.; Riviere, M. A.; Rustvold, D. L.; Lauten, J. D.; Madden, T. E.; David, L. L. Two-dimensional liquid chromatography study of the human whole saliva proteome. *J. Proteome Res.* **2004**, *3* (5), 1017−23.

(22) Resing, K. A.; Meyer-Arendt, K.; Mendoza, A. M.; Aveline-Wolf, L. D.; Jonscher, K. R.; Pierce, K. G.; Old, W. M.; Cheung, H. T.; Russell, S.; Wattawa, J. L.; Goehle, G. R.; Knight, R. D.; Ahn, N. G. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal. Chem.* **2004**, *76* (13), 3556−68.

(23) Setsuko Komatsu, X. Z. Naoki Tanaka. Comparision of two proteomics techniques used to identify proteins regulated by gibberellin in rice. *J. Proteome Res.* **2006**, *5* (2), 270−6.

(24) Fenyo, D.; Beavis, R. C. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* **2003**, *75* (4), 768−74.

(25) Searle, B. C.; Dasari, S.; Turner, M.; Reddy, A. P.; Choi, D.; Wilmarth, P. A.; McCormack, A. L.; David, L. L.; Nagalla, S. R. High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS de novo sequencing results. *Anal. Chem.* **2004**, *76* (8), 2220−30.

(26) Searle, B. C.; Dasari, S.; Wilmarth, P. A.; Turner, M.; Reddy, A. P.; David, L. L.; Nagalla, S. R. Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm. *J. Proteome Res.* **2005**, *4* (2), 546−54.

(27) Gu, C.; Tsaprailis, G.; Breci, L.; Wysocki, V. H. Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal. Chem.* **2000**, *72* (23), 5804−13.

(28) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics.* **2005**, 4 (10), 1487−502.

(29) Dennis, G., Jr.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A. DAVID: database for annotation, visualization, and integrated discovery. *GenomeBiology* **2003**, *4* (5), P3.

(30) States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M. Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat. Biotechnol.* **2006**, *24* (3), 333−8.

(31) Park, S. J.; Yoon, W. G.; Song, J. S.; Jung, H. S.; Kim, C. J.; Oh, S. Y.; Yoon, B. H.; Jung, G.; Kim, H. J.; Nirasawa, T. Proteome analysis of human amnion and amniotic fluid by two-dimensional electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proteomics* **2006**, *6* (1), 349−63.

(32) Michel, P. E.; Crettaz, D.; Morier, P.; Heller, M.; Gallot, D.; Tissot, J. D.; Reymond, F.; Rossier, J. S. Proteome analysis of human plasma and amniotic fluid by Off-Gel isoelectric focusing followed by nano-LC-MS/MS. *Electrophoresis* **2006**, *27* (5−6), 1169−81.

(33) Michaels, J.-E. A.; Dasari, S.; Pereira, L.; Reddy, A. P.; Lapidus, J. A.; Lu, X.; Jacob, T.; Thomas, A.; Rodland, M.; Roberts, C. T., Jr.; Gravett, M. G.; Nagalla, S. R. Comprehensive proteomic analysis of the human amniotic fluid proteome: gestational age-dependent changes. *J. Proteome Res.* **2007**, *6*, 1277−1285.

(34) Kerkhoff, C.; Klempt, M.; Sorg, C. Novel insights into structure and function of MRP8 (S100A8) and MRP14 (S100A9). *Biochim. Biophys. Acta* **1998**, *1448* (2), 200−11.

(35) Gravett, M. G.; Novy, M. J.; Rosenfeld, R. G.; Reddy, A. P.; Jacob, T.; Turner, M.; McCormack, A.; Lapidus, J. A.; Hitti, J.; Eschenbach, D. A.; Roberts, C. T. Jr.; Nagalla, S. R. Diagnosis of intra-amniotic infection by proteomic profiling and identification of novel biomarkers. *JAMA, J. Am. Med. Assoc.* **2004**, *292* (4), 462–9.

(36) Ruetschi, U.; Rosen, A.; Karlsson, G.; Zetterberg, H.; Rymo, L.; Hagberg, H.; Jacobsson, B. Proteomic analysis using protein chips to detect biomarkers in cervical and amniotic fluid in women with intra-amniotic inflammation. *J. Proteome Res.* **2005**, *4* (6), 2236–42.

(37) Lachapelle, M. H.; Miron, P.; Hemmings, R.; Roy, D. C. Endometrial, T, B, and NK cells in patients with recurrent spontaneous abortion. Altered profile and pregnancy outcome. *J. Immunol.* **1996**, *156* (10), 4027–34.

(38) Borzychowski, A. M.; Croy, B. A.; Chan, W. L.; Redman, C. W.; Sargent, I. L. Changes in systemic type 1 and type 2 immunity in normal pregnancy and pre-eclampsia may be mediated by natural killer cells. *Eur. J. Immunol.* **2005**, *35* (10), 3054–63.

(39) Genc, M. R.; Karasahin, E.; Onderdonk, A. B.; Bongiovanni, A. M.; Delaney, M. L.; Witkin, S. S. Association between vaginal 70-kd heat shock protein, interleukin-1 receptor antagonist, and microbial flora in mid trimester pregnant women. *Am. J. Obstet. Gynecol.* **2005**, *192* (3), 916–21.

(40) MasCasullo, V.; Fam, E.; Keller, M. J.; Herold, B. C. Role of mucosal immunity in preventing genital herpes infection. *Viral Immunol.* **2005**, *18* (4), 595–606.

(41) Brinkmann, V.; Reichard, U.; Goosmann, C.; Fauler, B.; Uhlemann, Y.; Weiss, D. S.; Weinrauch, Y.; Zychlinsky, A. Neutrophil extracellular traps kill bacteria. *Science* **2004**, *303* (5663), 1532–5.

(42) Buchanan, J. T.; Simpson, A. J.; Aziz, R. K.; Liu, G. Y.; Kristian, S. A.; Kotb, M.; Feramisco, J.; Nizet, V. DNase expression allows the pathogen group A Streptococcus to escape killing in neutrophil extracellular traps. *Curr. Biol.* **2006**, *16* (4), 396–400.

(43) Silphaduang, U.; Hincke, M. T.; Nys, Y.; Mine, Y. Antimicrobial proteins in chicken reproductive system. *Biochem. Biophys. Res. Commun.* **2006**, *340* (2), 648–55.

(44) Jacobsen, F.; Baraniskin, A.; Mertens, J.; Mittler, D.; Mohammadi-Tabrisi, A.; Schubert, S.; Soltau, M.; Lehnhardt, M.; Behnke, B.; Gatermann, S.; Steinau, H. U.; Steinstraesser, L. Activity of histone H1.2 in infected burn wounds. *J. Antimicrob. Chemother.* **2005**, *55* (5), 735–41.

(45) Kim, H. S.; Cho, J. H.; Park, H. W.; Yoon, H.; Kim, M. S.; Kim, S. C. Endotoxin-neutralizing antimicrobial proteins of the human placenta. *J. Immunol.* **2002**, *168* (5), 2356–64.

(46) Rose, F. R.; Bailey, K.; Keyte, J. W.; Chan, W. C.; Greenwood, D.; Mahida, Y. R. Potential role of epithelial cell-derived histone H1 proteins in innate antimicrobial defense in the human gastrointestinal tract. *Infect. Immun.* **1998**, *66* (7), 3255–63.

(47) Joseph, K.; Tholanikunnel, B. G.; Kaplan, A. P. Heat shock protein 90 catalyzes activation of the prekallikrein-kininogen complex in the absence of factor XII. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (2), 896–900.

(48) Pudney, J.; Quayle, A. J.; Anderson, D. J. Immunological microenvironments in the human vagina and cervix: mediators of cellular immunity are concentrated in the cervical transformation zone. *Biol. Reprod.* **2005**, *73* (6), 1253–63.

(49) Bhuvarahamurthy, V.; Govindasamy, S. Extracellular matrix components and proteolytic enzymes in uterine cervical carcinoma. *Mol. Cell. Biochem.* **1995**, *144* (1), 35–43.

(50) Makarewicz, R.; Drewa, G.; Szymanski, W.; Skonieczna-Makarewicz, I. Cathepsin B in predicting the extent of the cervix carcinoma. *Neoplasma* **1995**, *42* (1), 21–4.

(51) Benitez-Bribiesca, L.; Freyre-Horta, R.; Gallegos-Vargas, G. Protease and antiprotease concentrations in serum and vaginal fluid of patients with carcinoma of the cervix. *Arch. Invest. Med.* **1980**, *11* (4), 523–45.

(52) Mestecky, J.; Moldoveanu, Z.; Russell, M. W. Immunologic uniqueness of the genital tract: challenge for vaccine development. *Am. J. Reprod. Immunol.* **2005**, *53* (5), 208–14.

(53) Qin, S.; Ferdinand, A. S.; Richie, J. P.; O'Leary, M. P.; Mok, S. C.; Liu, B. C. Chromatofocusing fractionation and two-dimensional difference gel electrophoresis for low abundance serum proteins. *Proteomics* **2005**, *5* (12), 3183–92.

(54) Katz, A. B.; Taichman, L. B. A partial catalog of proteins secreted by epidermal keratinocytes in culture. *J. Invest. Dermatol.* **1999**, *112* (5), 818–21.

(55) Bard, E.; Riethmuller, D.; Biichle, S.; Meillet, D.; Pretet, J. L.; Mougin, C.; Seilles, E. Validation of a high sensitive immunoenzymatic assay to establish the origin of immunoglobulins in female genital secretions. *J. Immunoassay Immunochem.* **2002**, *23* (2), 145–62.

(56) Hocini, H.; Barra, A.; Belec, L.; Iscaki, S.; Preud'homme, J. L.; Pillot, J.; Bouvet, J. P. Systemic and secretory humoral immunity in the normal human vaginal tract. *Scand. J. Immunol.* **1995**, *42* (2), 269–74.

(57) Kiilholma, P.; Gronroos, M.; Erkkola, R.; Pakarinen, P.; Nanto, V. The role of calcium, copper, iron and zinc in preterm delivery and premature rupture of fetal membranes. *Gynecol. Obstet. Invest.* **1984**, *17* (4), 194–201.

(58) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31* (13), 3784–8.

PR0605419