

# A Proteomic Analysis of Human Hemodialysis Fluid\*<sup>§</sup>

Henrik Molina<sup>‡§</sup>, Jakob Bunkenborg<sup>‡§¶</sup>, G. Hanumanthu Reddy<sup>||</sup>,  
Babylakshmi Muthusamy<sup>||</sup>, Paul J. Scheel<sup>\*\*</sup>, and Akhilesh Pandey<sup>‡ ¶</sup>

The vascular compartment is an easily accessible compartment that provides an opportunity to measure analytes for diagnostic, prognostic, or therapeutic indications. Both serum and plasma have been analyzed extensively by proteomic approaches in an effort to catalog all proteins and polypeptides. Limitations of such approaches in obtaining a comprehensive catalog of proteins include the fact that a handful of proteins constitute over 90% of plasma protein content and that the renal glomeruli filter out proteins and polypeptides that are smaller than 66 kDa from blood. We chose to study hemodialysis fluid because it contains a higher concentration of small proteins and polypeptides and is also simultaneously depleted of the most abundant proteins present in blood. Using gel electrophoresis in combination with LC-MS/MS, we identified 292 proteins of which greater than 70% had not been previously identified from serum or plasma. More than half of the proteins identified from the hemodialysis fluid were smaller than 40 kDa. We also found 50 N-terminally acetylated peptides that allowed us to unambiguously map the N termini of mature forms of the corresponding proteins. Several identified proteins, including cytokines, were only present as predicted transcripts in data bases and thus represent novel proteins. The proteins identified in this study could serve as biomarkers in serum using more sensitive methods such as ELISA-specific antibodies. *Molecular & Cellular Proteomics* 4:637–650, 2005.

A comprehensive analysis of human serum and plasma has proven to be difficult, especially for low molecular weight and low abundance proteins, because of the wide range of concentrations with the 10 most abundant proteins constituting almost 90% of the serum proteome by mass (1). The dynamic range issue is exacerbated for low molecular weight species because the kidneys filter away molecules with molecular

mass of less than 66 kDa (2). The abundant proteins such as albumin, immunoglobulins, and transferrin hamper identification of less abundant proteins. Although several methods including ultracentrifugation (3), immunodepletion, solvent extraction/precipitation (4), and size exclusion chromatography (5) have been tried for removal of abundant components, it is still difficult to completely eliminate the interference from residual amounts of these abundant proteins.

Hemodialysis fluid has previously been used as a source of polypeptides (6) due to its higher concentration of low molecular weight components, but no comprehensive list of constituents present in the hemofiltrate has yet been published. We chose to analyze hemodialysis fluid because it is greatly reduced in the protein content, from 70 g/liter in the plasma to ~70 mg/liter. This is because the filtration cutoff used during hemodialysis results in selective depletion of proteins greater than ~60 kDa. It has been shown that the concentration of albumin in the hemodialysis fluid is 5,000-fold lower compared with its normal concentration in serum (7).

Our strategy involved one-dimensional gel electrophoresis separation and in-gel digestion of proteins followed by LC-MS/MS for identifying proteins in the hemodialysis fluid. Using this approach, we identified 292 proteins of which 205 had never been previously reported in the serum or plasma. Western blot analysis of a subset of these proteins revealed that they were also present in normal serum indicating that the sensitivity of detection might be the major reason why the majority of these proteins have never been identified previously in serum or plasma. We were also able to identify the N termini of a number of proteins based on peptides sequences that were acetylated at their N termini. A number of semitryptic peptides that were identified in this study were most likely derived from *in vivo* proteolysis. We were also able to identify a number of novel proteins including several cytokines in this analysis. The lack of a major overlap between the list of proteins identified in this study and previously reported proteins in serum or plasma reflects the difficulty of identifying these components using current proteomic methods. As demonstrated by our study, it is likely that more sensitive methods such as Western blotting or ELISA will be able to detect these proteins in serum or plasma.

## EXPERIMENTAL PROCEDURES

*Sample Collection and Preparation and Reagents*—The hemodialysis fluid was obtained from a 68-year-old white male with acute renal

From the <sup>‡</sup>McKusick-Nathans Institute for Genetic Medicine and Department of Biological Chemistry and Oncology and the <sup>\*\*</sup>Department of Nephrology, The Johns Hopkins University, Baltimore, Maryland 21205, the <sup>§</sup>Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense 5230, Denmark, and the <sup>||</sup>Institute of Bioinformatics, International Technology Park Ltd., Bangalore 560 066, India

Received, February 12, 2005

Published, MCP Papers in Press, February 20, 2005, DOI 10.1074/mcp.M500042-MCP200

failure following a coronary artery bypass surgery. The patient was not known to have any other diseases. Vascular access was obtained utilizing a dual lumen catheter in the femoral position. Continuous venovenous hemodialysis was performed with a Prisma® dialysis machine (Cobe Renal Intensive Care, Lakewood, CO). The dialysis membrane was an acrylonitrile and sodium methallyl sulfonate copolymer, AN69® (Cobe Renal Intensive Care). Sterile dialysate was prepared in 5.0-liter aliquots (140 meq/liter sodium, 111 meq/liter chloride, 3.5 meq/liter calcium, 3 meq/liter lactate, 1 meq/liter magnesium, 32 meq/liter bicarbonate, 2.0 meq/liter potassium), and the dialysis was carried out with a blood flow of 180 ml/min and a dialysate flow of 1.0 liter/hr. The spent dialysate was collected in 5.0-liter aliquots and analyzed after concentrating using a 3-kDa-cutoff filter (Centricon YM3, Millipore). The protein content in the hemodialysis fluid was measured using a modified Lowry protein assay kit (Bio-Rad). The concentrated hemodialysis fluid was run on precast NuPage 4–12% bis-tris<sup>1</sup> and 10–20% Tricine minigels (Invitrogen). The gels were silver-stained (8), and between 20 and 25 visible protein bands were excised. Protein gel bands were reduced with dithiothreitol (Fluka, Buch, Switzerland) and alkylated with iodoacetamide (Sigma) before digestion with trypsin (Promega, Madison, WI) in 100 mM NH<sub>4</sub>HCO<sub>3</sub> (Fluka). Solvents for liquid chromatography included heptafluorobutyric acid (Sigma), glacial acetic acid (Fisher Scientific), and HPLC-grade acetonitrile (J. T. Baker Inc.).

**Trypsin Digestion and LC-MS/MS Analysis**—The excised gel slices were digested with trypsin as follows. The gel bands were washed twice in 0.1 M NH<sub>4</sub>HCO<sub>3</sub>, washed twice in 50% acetonitrile, and subsequently cut into 2 × 2-mm pieces. The gel pieces were shrunk using 100% acetonitrile, and proteins were reduced by addition of 0.1 M dithiothreitol followed by an incubation step at 56 °C for 45 min. The washing procedure described above was repeated, and proteins were alkylated by adding 55 mM iodoacetamide and incubating for 30 min at room temperature in the dark. After an additional wash and shrinkage, 10 ng/μl trypsin in 0.1 M NH<sub>4</sub>HCO<sub>3</sub> sufficient to cover the gel pieces was added followed by an incubation on ice for 20 min. When the gel pieces were completely rehydrated, any excess trypsin solution was removed and replaced by 0.1 M NH<sub>4</sub>HCO<sub>3</sub>, and samples were incubated overnight at 37 °C. The digestion was stopped by adding 10 μl of glacial acetic acid, and the supernatant containing the tryptic peptides was harvested. An extraction step was carried out to recover the peptides from the gel slices by adding 50% acetonitrile and incubating at room temperature for 30 min. The supernatant was harvested again and pooled. The pooled peptide extracts were dried down to ~10 μl and subjected to LC-MS/MS analysis as follows. Samples were injected onto a 5-cm C<sub>18</sub> trap column (inner diameter, 75 μm) packed with YMC ODS-A 5–15-μm beads (Kanematsu USA Inc., New York, NY) using an autosampler (1100 microwell plate autosampler, Agilent Technologies, Palo Alto, CA). The peptides were eluted from the trap column onto an analytical 10-cm C<sub>18</sub> column (inner diameter, 75 μm) packed with Vydac MS218 5-μm beads (Vydac, Columbia, MD) with a gradient increasing from 10% solvent B, 90% solvent A (solvent A: 0.4% acetic acid, 0.005% heptafluorobutyric acid; solvent B: 90% acetonitrile, 0.4% acetic acid, 0.005% heptafluorobutyric acid) to 45% solvent B, 55% solvent A in 30 min. A flow of 4 μl/min during loading and 300 nl/min during elution was delivered by a nanoflow pump (Agilent Technologies 1100 nanopump). The LC setup was connected to either a quadrupole-time-of-flight mass spectrometer (QTOF API-US, Micromass, Manchester, UK) or an ion trap mass spectrometer (LC/MSD Trap XCT, Agilent

Technologies) using nanoelectrospray sources from Proxeon (Odense, Denmark).

**Western Blotting**—For Western blotting experiments, serum from a healthy person was first depleted of albumin, IgG, IgA, transferrin, haptoglobin, and antitrypsin using an Agilent Technologies multiple affinity removal kit. 20 μl of serum were diluted to 100 μl in Buffer A (Buffer A and Buffer B as supplied in the multiple affinity removal kit by Agilent Technologies) prior to loading onto the column at 250 μl/min in Buffer A. The elution from the column was monitored at 280 nm, and the depleted serum was collected in the interval from 2 to 4 min. Cleaning of the column was carried out by eluting the bound proteins as follows. 10 min after the injection, Buffer A was exchanged with Buffer B, and the flow rate was simultaneously increased to 1,000 μl/min. Following this, Buffer B at 1,000 μl/min was continued for 18 min after which the conditions were returned to the initial loading conditions. After conditioning the column under loading conditions for 5 min, the system was ready for a new depletion process.

Depleted serum was resolved by SDS-PAGE, and the proteins were transferred onto a nitrocellulose membrane. The membrane was blocked at 4 °C overnight with 5% BSA in phosphate-buffered saline containing 0.1% Tween 20. The membrane was incubated with the relevant antibodies for 2 h, washed, and incubated with horseradish peroxidase-conjugated secondary antibody for 1 h. The proteins were visualized using enhanced chemiluminescence detection according to the manufacturer's instructions (Amersham Biosciences). The sources of primary antibodies were as follows: Cathepsin D, connective tissue growth factor, Galectin 3, and Lipocalin 2 (R&D Systems, Inc. Minneapolis, MN); Nucleophosmin 1 (Zymed Laboratories Inc.); α-defensins 1–3 (BD Biosciences); Cathepsin H (Serotec, Oxford, UK); and Cofilin 2 (Upstate Biotechnology, Lake Placid, NY). Conjugated secondary anti-mouse and anti-rabbit were from Amersham Biosciences, and anti-goat antibody was from Santa Cruz Biotechnology (Santa Cruz, CA).

**Data Base Search and Analysis**—The mass spectrometry data files from individual LC-MS/MS experiments were merged and then searched against the RefSeq data base (human: 27,975 entries; July 28, 2004) using Mascot (Matrix Sciences Ltd., London, UK). The search parameters were as follows: mass accuracy of the monoisotopic precursor and peptide fragments was set to 1.5 and 0.5 Da, respectively, for the data acquired on the ion trap mass spectrometer and 0.2 and 0.2 Da, respectively, for the data acquired on the quadrupole-time-of-flight mass spectrometer. The following variable modifications were permitted: oxidation of methionine, histidine, and tryptophan residues; N-terminal acetylation of proteins; and cyclization of N-terminal glutamine. Two missed tryptic cleavages were allowed. Additional searches using data acquired on the quadrupole time-of-flight mass spectrometer were performed with the following criteria: 1) semitryptic constraints, 2) N-acetylhexosamine modification of asparagine residues, and 3) hydroxylation of proline residues. For validation purposes, the retrieved peptide sequences were divided into three groups as follows: 1) peptides with a low score (quadrupole-time-of-flight data, <25; ion trap data, <30), 2) peptides with intermediate to high scores (quadrupole-time-of-flight data, >25; ion trap data, >30), and 3) proteins identified with only one intermediate or high scoring peptide. All of the low scoring peptides were discarded. All intermediate and high scoring peptides as well as single peptide hits used to identify proteins were manually validated if the following criteria were met: 1) several consecutive y-ions although absence of y-ions after proline and glycine, 2) the existence of lower a- and b-ions, 3) none or few unassigned fragments ions, and 4) a charge state of the precursor ion and fragment ions that are in accordance with basic amino acids in the assigned peptide sequence. Supplemental Fig. 6 pro-

<sup>1</sup> The abbreviations used are: bis-tris, 2-[bis(2-hydroxyethyl)-amino]-2-(hydroxymethyl)propane-1,3-diol; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; FABP, fatty acid-binding protein; HIV-1, human immunodeficiency virus, type 1.

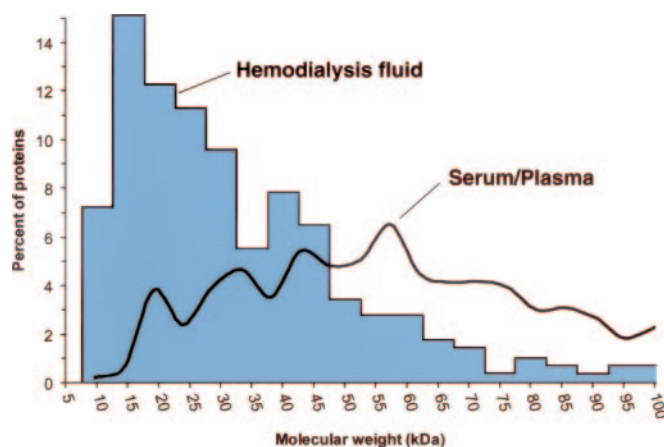


FIG. 1. Distribution of molecular masses of proteins identified from hemodialysis fluid. Superimposed are the similar distribution for serum and plasma proteins described in the literature (9, 10). Proteins were first sorted into bins of 5 kDa each, and the numbers in each bin were plotted as a percentage of the total number of entries.

vides the MS/MS spectra of all proteins that were identified on the basis of a single peptide.

#### RESULTS AND DISCUSSION

Our strategy to characterize the hemodialysis fluid proteome by mass spectrometry using a gel LC approach was as follows. The hemodialysis fluid was first desalted and concentrated using a 3-kDa-cutoff filter. For fractionating the proteins and polypeptides in the hemodialysis fluid, the desalted sample was resolved by SDS-PAGE and silver-stained, and the bands were excised, digested with trypsin, and analyzed by LC-MS/MS. Supplemental Table 1 lists all the proteins that were identified in this study along with the number of peptides that matched each protein, molecular weight, and assignments of the approved Human Genome Organization (HUGO) gene symbols, wherever available. Supplemental Table 1 provides a count of how many gene products were actually identified from this sample as the count on the basis of proteins can be misleading because of the presence of a large number of protein isoforms.

#### *Hemodialysis Fluid Proteome Versus the Plasma Proteome*

Of the proteins identified in this study, 85% were smaller than 60 kDa, and more than half were <30 kDa. This is expected because of the hemodialysis process in which blood is filtered through a membrane that has a molecular mass cutoff of ~60 kDa, which should only allow proteins less than 60 kDa into the hemodialysis fluid. Fig. 1 shows a histogram of the number of proteins identified plotted against the corresponding molecular masses. Most of the proteins are found between 15 and 30 kDa. Superimposed on the histogram is a similar plot from the combined studies by Anderson *et al.* (9) and Chan *et al.* (10). Those two studies resulted in more than 2,100 proteins of which ~50% were below 65 kDa,

and only 16% were below 30 kDa. Another distinguishing feature of the hemodialysis fluid data set is the absence of the 40–80-kDa hump that is conspicuous in the serum/plasma proteome data set. A subset of the serum/plasma proteins were identified by Tirumalai *et al.* (3), who used serum that had been filtered through a 30-kDa filter prior to analysis. In terms of the number of proteins identified, this data set (340 proteins) is comparable to our hemodialysis data set (292 proteins), although significant differences are observed. Comparing the molecular weight distribution of the proteins identified in these two studies, it is clear that a considerably greater fraction of the proteins found in the hemodialysis fluid (56% of proteins with molecular mass < 30 kDa) had a lower molecular weight than in the analysis by Tirumalai *et al.* (18% of proteins with molecular mass < 30 kDa).

We compared the known serum/plasma proteins with those identified in this study to determine the degree of overlap based on the genes that encoded the expressed proteins and their isoforms. We first mapped all the protein entries to HUGO-approved gene symbols wherever possible ([www.gene.ucl.ac.uk/nomenclature/](http://www.gene.ucl.ac.uk/nomenclature/)), which allowed us to eliminate redundancy and facilitated this comparison. Although the serum/plasma catalog contained more than 7 times as many protein entries, it was remarkable that 205 of 292 proteins found in the hemodialysis fluid were not reported in the serum/plasma proteome. Supplemental Fig. 1 shows a distribution of proteins identified in this study according to their molecular function. Table I lists the proteins that were identified from hemodialysis fluid but not previously reported in serum/plasma. Not surprisingly, several of these proteins are of low molecular mass, validating our use of hemodialysis fluid to identify smaller proteins and polypeptides. Fig. 2 shows a Venn diagram indicating the relative numbers of overlapping and non-overlapping gene products.

It can be argued that this difference observed between serum and hemodialysis fluid is attributable to the underlying condition in the patient undergoing dialysis. In that case, the proteins specific to our list of hemodialysis fluid would not be observed in serum from normal individuals. It should be noted that the components that are normally filtered away by the kidneys are excreted in the urine. Therefore, we also compared our results with two proteomic analyses of the human urinary proteome (11, 12) with a total of 358 proteins. Of the proteins specifically found in hemodialysis fluid, 29 of those have been identified from urine of normal individuals. This indicates that a significant fraction of the proteins are indeed found in normal serum/plasma but have not been identified earlier because of the technical limitations mentioned earlier.

#### *Western Blot Analysis to Confirm Expression in Normal Serum*

To test whether the expression of the proteins found in hemodialysis fluid was also detectable in normal serum, we

TABLE I  
Proteins found in hemodialysis fluid that were not previously described in serum/plasma

	Protein name	Accession number	Gene symbol	Molecular (Daltons) mass	
1	Thymosin, $\beta$ 10	NP_066926.1	TMSB10	5,023	
2	Thymosin-like 2	NP_877594.1	TMSL2	5,036	
3	Thymosin-like 6	NP_852093.1	TMSL6	5,039	
4	Thymosin, $\beta$ 4	NP_066932.1	TMSB4X	5,050	
5	Thymosin-like 4	NP_877595.1	TMSL4	5,109	
6	Metallothionein 1H	NP_005942.1	MT1H	7,175	
7	Metallothionein 1G	NP_005941.1	MT1G	7,205	
8	Metallothionein 1F	NP_005940.1	MT1F	7,221	
9	Metallothionein 1A	NP_005937.1	MT1A	7,268	
10	Metallothionein 1K	NP_789846.1	MT1K	7,272	
11	Metallothionein 1B	NP_005938.1	MT1B	7,307	
12	Antioxidant protein 1	NP_004036.1	ATOX1	7,397	
13	Adipose-specific 2	NP_006820.1	C10orf116	7,850	
14	Ribosomal protein S28	NP_001022.1	RPS28	7,893	
15	Trefoil factor 3 (intestinal)	NP_003217.1	TFF3	8,476	
16	Neural precursor cell expressed, developmentally down-regulated 8	NP_006147.1	NEDD8	9,066	
17	Hypothetical protein BM-002	NP_057701.1	Ufm1	9,169	
18	Secretoglobin, family 1A, member 1 (uteroglobin)	NP_003348.1	SCGB1A1	10,215	Previously identified in urine
19	S100 calcium-binding protein A6	NP_055439.1	S100A6	10,230	Previously identified in urine
20	Cytochrome c oxidase subunit Vib	NP_001854.1	COX6B1	10,413	
21	SH3 domain binding glutamic acid-rich protein-like 2	NP_113657.1	SH3BGRL2	10,431	
22	Dynein light chain 2	NP_542408.1	Dlc2	10,457	
23	Defensin, $\alpha$ 1, preproprotein	NP_004075.1	DEFA1	10,536	Previously identified in urine
24	Ly-6 neurotoxin-like protein 1 isoform b	NP_076435	SLURP2	10,723	
25	Similar to SMT3 suppressor of mif two 3 homolog 2	XP_376612	LOC401340	10,864	
26	Heat shock 10-kDa protein 1 (chaperonin 10)	NP_002148.1	HSPE1	10,925	
27	Cystatin B	NP_000091.1	CSTB	11,133	
28	Death-associated protein	NP_004385.1	DAP	11,158	
29	Muscle-type acylphosphatase 2	NP_612457.1	ACYP2	11,190	
30	Diazepam binding inhibitor	NP_065438.1	DBI	11,786	
31	ARS component B precursor	NP_065160.1	ARS	11,805	
32	Erythroid-associated factor	NP_057717.1	ERAF	11,833	
33	S100 calcium binding protein A11 (calgizzarin)	NP_005611.1	S100A11	11,847	Previously identified in urine
34	FK506-binding protein 1A	NP_000792.1	FKBP1A	12,000	
35	Glutaredoxin (thioltransferase)	NP_002055.1	GLRX	12,053	
36	Resistin	NP_065148.1	RETN	12,096	
37	SH3 domain binding glutamic acid-rich protein-like 2	NP_113657.1	SH3BGRL2	12,375	
38	Colipase preproprotein	NP_001823.1	CLPS	12,516	
39	Macrophage migration inhibitory factor (glycosylation-inhibiting factor)	NP_002406.1	MIF	12,639	Previously identified in urine
40	SH3 domain binding glutamic acid-rich protein-like	NP_003013.1	SH3BGRL	12,766	
41	D-Dopachrome tautomerase	NP_001346.1	DDT	12,818	Previously identified in urine
42	$\beta$ -tubulin cofactor A	NP_004598.1	TBCA	12,904	
43	Hypothetical protein MGC14353	NP_116120.1	TXNL5	13,932	
44	CD59 antigen p18-20	NP_000602.1	CD59	14,168	Previously identified in urine
45	Programmed cell death 5	NP_004699.1	PDCD5	14,276	
46	Gastrotropin	NP_001436.1	FABP6	14,362	
47	Translational inhibitor protein p14.5	NP_005827.1	HRSP12	14,542	Previously identified in urine
48	Allograft inflammatory factor 1 isoform 2	NP_004838.1	AIF1	14,608	
49	Metallothionein 1E	NP_783316.1	MT1E	14,651	
50	Fatty acid-binding protein 4, adipocyte	NP_001433.1	FABP4	14,824	
51	Profilin 1	NP_005013.1	PFN1	15,045	Previously identified in urine
52	Fatty acid-binding protein 5 (psoriasis-associated)	NP_001435.1	FABP5	15,155	
53	$\alpha$ 2 globin	NP_000508.1	HBA2	15,248	
54	Retinol-binding protein 2, cellular	NP_004155.2	RBP2	15,868	
55	Coactosin-like 1	NP_066972.1	COTL1	15,935	
56	$\delta$ globin	NP_000510.1	HBD	16,045	

TABLE I—continued

	Protein name	Accession number	Gene symbol	Molecular (Daltons) mass	
57	Niemann-Pick disease, type C2 precursor	NP_006423.1	NPC2	16,559	
58	Similar to Ig $\kappa$ variable region	XP_376073	LOC400967	16,693	
59	Ubiquitin-conjugating enzyme E2D 2 isoform 1	NP_003330.1	UBE2D2	16,724	
60	Cystatin M precursor	NP_001314.1	CST6	16,785	
61	Glia maturation factor, $\beta$	NP_004115.1	GMFB	16,874	
62	Ubiquitin-conjugating enzyme E2N	NP_003339.1	UBE2N	17,127	Previously identified in urine
63	Pancreatic ribonuclease precursor	NP_002924.1	RNASE1	17,633	Previously identified in urine
64	Ubiquitin-conjugating enzyme E2L 3 isoform 1	NP_003338.1	UBE2L3	17,850	
65	Similar to peptidyl-Pro cis-trans isomerase	XP_292085.1	LOC341457	17,901	
66	Ubiquitin and ribosomal protein S27a precursor	NP_002945.1	RPS27A	17,953	
67	Peptidylprolyl isomerase A (cyclophilin A)	NP_066953.1	PPIA	18,001	Previously identified in urine
68	Regenerating islet-derived 1 $\alpha$ precursor	NP_002900.2	REG1A	18,719	Previously identified in urine
69	Cofilin 2	NP_068733.1	CFL2	18,725	
70	Chromosome 19 open reading frame 10	NP_061980.1	C19orf10	18,783	
71	Regenerating islet-derived 1 $\beta$ precursor	NP_006498.1	REG1B	19,052	
72	Breast cancer membrane protein 11	NP_789783.1	BCMP11	19,273	
73	Peptidylprolyl isomerase H	NP_006338.1	PPIH	19,481	
74	Ubiquitin-conjugating enzyme E2G 1 isoform 1	NP_003333.1	UBE2G1	19,497	
75	Similar to peptidylprolyl isomerase A (cyclophilin A)	XP_372328	LOC390006	19,603	
76	Hypothetical protein HSPC155	NP_057490.1	Ufc1	19,626	
77	Ferredoxin 1 precursor, adrenodoxin	NP_004100.1	FDX1	19,666	
78	Tumor protein, translationally controlled 1	NP_003286.1	TPT1	19,697	
79	18-kDa antrum mucosa protein	NP_062563.2	GKN1	20,546	
80	ADP-ribosylation factor 3	NP_001650.1	ARF3	20,645	Previously identified in urine
81	ADP-ribosylation factor 1	NP_001649.1	ARF1	20,684	Previously identified in urine
82	GM2 ganglioside activator protein precursor	NP_000396.2	GM2A	20,825	Previously identified in urine
83	Prostatic binding protein	NP_002558.1	PBP	21,044	Previously identified in urine
84	Chromosome 7 open reading frame 24	NP_076956.1	C7orf24	21,222	
85	Cysteine- and glycine-rich protein 1	NP_004069.1	CSRP1	21,409	
86	Apoptosis-associated speck-like protein containing a CARD isoform a	NP_037390.2	PYCARD	21,613	
87	Peptidylprolyl isomerase F (cyclophilin F)	NP_005720.1	PPIF	22,026	
88	Peroxiredoxin 1	NP_002565.1	PRDX1	22,096	
89	H1 histone family, member 3	NP_005311.1	HIST1H1D	22,336	
90	Hypothetical protein MGC15429	NP_116139.1	MGC15429	22,446	Previously identified in urine
91	Transgelin	NP_003177.1	TAGLN	22,461	
92	Huntingtin-interacting protein 2	NP_005330.1	HIP2	22,507	
93	H1 histone family, member 5	NP_005313.1	HIST1H1B	22,566	
94	Lipocalin 2 (oncogene 24p3)	NP_005555.2	LCN2	22,574	
95	Human Ig rearranged $\gamma$ chain mRNA, V-J-C region and complete cds	CAA41998	IGKV2D-40	22,986	
96	Rho GDP dissociation inhibitor (GDI) $\beta$	NP_001166.1	ARHGDI B	23,031	
97	Similar to midline 1	NP_997227	LOC283377	23,114	
98	Rho GDP dissociation inhibitor (GDI) $\alpha$	NP_004300.1	ARHGDI A	23,250	
99	Glutathione transferase	NP_000843.1	GSTP1	23,569	Previously identified in urine
100	BCL2-associated X protein isoform $\beta$	NP_004315.1	BAX	24,375	
101	Protein-L-isoaspartate (D-aspartate) O-methyltransferase	NP_005380.1	PCMT1	24,792	
102	Predicted osteoblast protein	NP_055703.1	FAM3C	24,950	
103	Lysophospholipase I	NP_006321.1	LYPLA1	24,996	
104	Lysophospholipase II	NP_009191.1	LYPLA2	25,063	
105	Peroxiredoxin 6	NP_004896.1	PRDX6	25,133	Previously identified in urine
106	Glutathione transferase A5	NP_714543.1	GSTA5	25,763	
107	Ubiquitin-conjugating enzyme E2 Kue-UEV isoform 2	NP_003340.1	Kua-UEV	25,780	
108	UMP-CMP kinase	NP_057392.1	UMP-CMPK	26,180	
109	Galectin-3	NP_002297.1	LGALS3	26,229	
110	Methionine sulfoxide reductase A	NP_036463.1	MSRA	26,401	
111	Adenylate kinase 2 isoform a	NP_001616.1	AK2	26,461	
112	Chloride intracellular channel 1	NP_001279.2	CLIC1	26,906	Previously identified in urine
113	Protease, serine, 2 preproprotein	NP_002761.1	PRSS2	26,927	

TABLE I—continued

	Protein name	Accession number	Gene symbol	Molecular (Daltons) mass	
114	Triosephosphate isomerase 1	NP_000356.1	TPI1	26,938	Previously identified in urine
115	Arginine-rich, mutated in early stage tumors	NP_006001.2	ARMET	27,402	
116	6-Phosphogluconolactonase	NP_036220.1	PGLS	27,815	Previously identified in urine
117	Electron-transfer-flavoprotein, $\beta$ polypeptide	NP_001976.1	ETFB	27,826	
118	Insulin-like growth factor-binding protein 4	NP_001543.1	IGFBP4	27,923	
119	Hypothetical protein BC001573	NP_620164.1	LOC134147	28,030	
120	3'(2'), 5'-bisphosphate nucleotidase 1	NP_006076.3	BPNT1	28,395	
121	WNT1 inducible signaling pathway protein 2 precursor	NP_003872.1	WISP2	28,460	
122	Thiopurine S-methyltransferase	NP_000358.1	TPMT	28,618	
123	Nucleophosmin 1	NP_954654.1	NPM1	29,617	Previously identified in urine
124	Dimethylarginine dimethylaminohydrolase 2	NP_039268.1	DDAH2	29,911	
125	PTD012 protein	NP_054758.2	PTD012	29,929	
126	Carbonyl reductase 1	NP_001748.1	CBR1	30,641	
127	Dimethylarginine dimethylaminohydrolase 1	NP_036269.1	DDAH1	31,444	Previously identified in urine
128	Mitochondrial short-chain enoyl-coenzyme A hydratase 1 precursor	NP_004083.2	ECHS1	31,807	
129	Vesicle-associated membrane protein-associated protein A isoform 1	NP_003565.3	VAPA	32,245	
130	3-hydroxyanthranilate 3,4-dioxygenase	NP_036337.1	HAAO	32,522	
131	Isopentenyl-diphosphate $\delta$ isomerase	NP_004499.2	IDI1	32,978	
132	Lactamase, $\beta$ 2	NP_057111.1	LACTB2	33,070	
133	3-Mercaptopyruvate sulfurtransferase	NP_066949.1	MRST	33,158	
134	CGI-150 protein	NP_057164.2	C17orf25	33,228	
135	Thiosulfate sulfurtransferase	NP_003303.2	TST	33,408	
136	Biliverdin reductase A	NP_000703.2	BLVRA	33,692	
137	Regucalcin	NP_004674.1	RGN	33,802	
138	v-crK sarcoma virus CT10 oncogene homolog isoform a	NP_058431.2	CRK	33,908	
139	Secreted phosphoprotein 1	NP_000573.1	SPP1	33,994	Previously identified in urine
140	Mitochondrial malate dehydrogenase precursor	NP_005909.2	MDH2	35,481	
141	Aldo-keto reductase family 1, member B1	NP_001619.1	AKR1B1	35,827	Previously identified in urine
142	Hypothetical protein FLJ11151	NP_060810.1	FLJ11151	35,942	
143	Heterogeneous nuclear ribonucleoprotein A2/B1 isoform A2	NP_002128.1	HNRPA2B1	36,041	
144	Glyceraldehyde-3-phosphate dehydrogenase	NP_002037.2	GAPD	36,201	Previously identified in urine
145	Poly(rC) binding protein 3	NP_065389.1	PCBP3	36,201	
146	3-Hydroxymethyl-3-methylglutaryl-CoA lyase-like 1	NP_061909	HMGCLL1	36,760	
147	Aldo-keto reductase family 1, member A1	NP_006057.1	AKR1A1	36,892	
148	Hypothetical protein BM-009	NP_057707.3	BM-009	37,024	
149	Cathepsin H isoform b precursor	NP_683880.1	CTSH	37,369	
150	Aldose 1-epimerase	NP_620156.1	GALM	37,742	
151	Capping protein (actin filament), gelsolin-like	NP_001738.1	CAPG	38,779	
152	Hypothetical protein FLJ20446	NP_060295.1	ADPRHL2	39,264	
153	Protein tyrosine kinase 9-like	NP_009215.1	PTK9L	39,523	
154	H2A histone family, member Y isoform 2	NP_004884.1	H2AFY	39,635	
155	Tropomodulin 3 (ubiquitous)	NP_055362.1	TMOD3	39,741	
156	Class I alcohol dehydrogenase, $\alpha$ subunit	NP_000658.1	ADH1A	39,832	
157	Aldo-keto reductase family 1, member C1	NP_001344.2	AKR1C1	39,833	
158	$\alpha$ -Actinin-2-associated LIM protein	NP_055291.1	PDLIM3	39,862	
159	Connective tissue growth factor	NP_001892.1	CTGF	40,267	
160	$\alpha$ 1 actin precursor	NP_001091.1	ACTA1	42,366	
161	Pepsinogen 5, group I (pepsinogen A)	NP_055039.1	PGA5	42,366	
162	Class II alcohol dehydrogenase 4 $\pi$ subunit	NP_000661.1	ADH4	42,584	
163	Similar to peroxisomal acyl-coenzyme A thioester hydrolase 2a	XP_090885.5	C14orf42	42,898	
164	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 9	NP_004146.1	SERPINB9	43,004	
165	UV excision repair protein RAD23 homolog B	NP_002865.1	RAD23B	43,202	
166	Chitinase 3-like 2	NP_003991.1	CHI3L2	43,240	
167	Alanine-glyoxylate aminotransferase	NP_000021.1	AGXT	43,325	

TABLE I—continued

	Protein name	Accession number	Gene symbol	Molecular (Daltons) mass	
168	3-Hydroxyisobutyryl-coenzyme A hydrolase isoform 1	NP_055177	HIBCH	43,797	
169	Ig superfamily protein	NP_009199.1	Z39IG	44,529	
170	Butyrate-induced transcript 1	NP_057479.1	HSPC121	44,622	
171	Hypothetical protein MGC10940	NP_115679.2	C9orf99	45,651	
172	Protease inhibitor 14	NP_006208.1	SERPINI2	46,287	
173	Pancreatic carboxypeptidase A1 precursor	NP_001859.1	CPA1	47,225	
174	Enolase 1	NP_001419.1	ENO1	47,481	Previously identified in urine
175	Hypothetical protein FLJ20291	NP_060218.1	FLJ20291	49,674	
176	GDP dissociation inhibitor 2	NP_001485.2	GDI2	51,087	
177	Keratin 14	NP_000517.2	KRT14	51,875	
178	Adenylyl cyclase-associated protein	NP_006358.1	CAP1	51,926	
179	Selenium-binding protein 1	NP_003935.2	SELENBP1	52,358	
180	Peroxisomal long-chain acyl-CoA thioesterase	NP_006812.2	ZAP128	53,631	
181	Similar to bA13B9.3 (novel protein similar to KRT8)	XP_294590	LOC347265	55,459	
182	Keratin protein K6irs	NP_542785.1	K6IRS2	56,470	
183	Glucose-regulated protein, 58 kDa	NP_005304.3	GRP58	57,146	
184	ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, $\alpha$ subunit, isoform 1	NP_004037.1	ATP5A1	59,714	
185	Immunoglobulin heavy chain	BAC87432.1	IGH@	60,769	
186	LIM protein (similar to rat protein kinase C-binding enigma)	NP_006448.1	LIM	65,185	
187	LIM domain binding 3	NP_009009	LDB3	77,135	
188	Hypothetical protein FLJ13946	NP_689488.2	FLJ13946	76,900	
189	Ubiquitin C	NP_066289.1	UBC	76,991	
190	Solute carrier family 15 (H <sup>+</sup> /peptide transporter), member 2	NP_066568.2	SLC15A2	82,385	
191	Desmocollin 1 isoform Dsc1b preproprotein	NP_004939.1	DSC1	94,916	
192	Eukaryotic translation elongation factor 2	NP_001952.1	EEF2	96,246	
193	Rap1 guanine nucleotide exchange factor directly activated by cA	NP_006096.2	RAPGEF3	100,317	
194	Desmocollin 3 isoform Dsc3a preproprotein	NP_001932.1	DSC3	101,262	
195	Adaptor-related protein complex 2, $\beta$ 1 subunit	NP_001273.1	AP2B1	105,398	
196	Tolloid-like 1	NP_036596.3	TLL1	114,635	
197	Similar to pote protein	XP_292982	LOC344227	116,896	
198	Serine protease inhibitor	NP_006837.1	SPINK5	124,399	
199	BAI1-associated protein 3	NP_003924.2	BAIAP3	133,129	
200	Catenin (cadherin-associated protein), $\delta$ 2	NP_001323.1	CTNND2	133,658	
201	$\alpha$ 1 type XV collagen precursor	NP_001846.2	COL15A1	142,356	
202	Serine/threonine kinase 36 (fused homolog, <i>Drosophila</i> )	NP_056505.1	STK36	145,729	
203	Agrin	NP_940978	AGRN	214,706	
204	Myosin X	NP_036466.1	MYO10	239,234	
205	Nuclear receptor binding SET domain protein 1 isoform b	NP_071900.2	NSD1	302,109	

obtained antibodies that work in Western blotting experiments against a subset of proteins. As shown in Fig. 3, we were able to observe the expression of Galectin 3, Cofilin 2, Cathepsin H,  $\alpha$ -defensins 1 and 3, Cathepsin D, Nucleophosmin 1, connective tissue growth factor, and Lipocalin 2 in normal serum. Thus, although we cannot formally rule out the possibility that some of the proteins are found in the serum because of the underlying disease of the patient undergoing hemodialysis, we think that the majority of the proteins identified in this study are normal components of serum/plasma and have not been previously identified due to their lower concentrations in the blood because of clearance by the kidneys.

#### Secreted Proteins Identified from the Hemodialysis Fluid

We identified several proteins that were either known or predicted to be secreted proteins. Below we will discuss a subset of these proteins that are likely to be especially interesting from a biological perspective. Many of them have not been previously identified in serum/plasma. All accession numbers are from the RefSeq data base.

*A Novel Predicted Osteoblast Protein (FAM3C; NP\_055703)*—This protein was initially identified from genomic data bases using structure-based methods in search of novel “four-helix bundle” cytokines. Four genes were iden-

tified in this family, although functional prediction has only been done for a related protein encoded by a related gene, *FAM3B*, which inhibits basal insulin secretion. This novel protein encoded by the *FAM3C* gene is expressed in all the tissues examined but was named as a predicted osteoblast protein as it was initially observed in osteoblasts. It contains an N-terminal signal peptide and no transmembrane domain, which suggests that it is a secreted protein (13).

*Fatty Acid-binding Protein 3 (FABP3) (NP\_004093)*—This

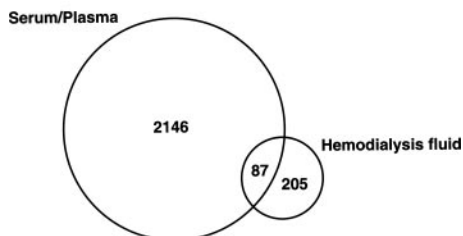


FIG. 2. A Venn diagram showing the degree of overlap based on the non-redundant set of genes that encoded the proteins identified in this study and those previously reported in serum/plasma (9, 10).

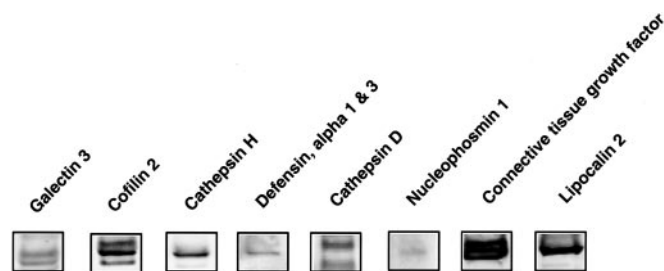


FIG. 3. Western blot analysis to confirm expression in normal serum. The figure shows a Western blot analysis of serum from a healthy volunteer that was depleted of albumin, IgG, IgA, transferrin, haptoglobin, and antitrypsin. The antibodies used are directed against the proteins whose names are indicated above each panel.

was originally identified by screening a human adult muscle  $\lambda$  gt11 expression library with an antibody to muscle FABP (14). The same molecule was also identified as mammary-derived growth inhibitor based on its activity as a growth inhibitor in lactating bovine mammary gland. This protein plays a role in the intracellular transport of long-chain fatty acids and their acyl-CoA esters. *FABP3* is a candidate tumor suppressor gene involved in breast cancer (15).

*Antrum Mucosa Protein 18 kDa (NP\_062563)*—This molecule, also designated as CA11, was initially isolated using differential display in human gastric cancer tissue. The expression of CA11 was observed to be down-regulated in gastric cancer tissue as compared with the normal gastric mucosa. Northern blot and RACE indicated that it is predominantly expressed in stomach and at low levels in the uterus and placenta (16). It has been suggested that the loss of its expression in gastric tissues may play an important role in gastric carcinogenesis (17). This protein is localized to the secretory granules of mucosal epithelium lining the stomach lumen (18).

*Resistin (NP\_065148)*—Resistin (resistance to insulin) belongs to a family of proteins that is involved in inflammatory processes and in regulating metabolism. Human resistin (also referred to as FIZZ3) was identified by searching sequence data bases with a related mouse protein called FIZZ1 (19). The expression of resistin is induced during adipogenesis, and it is normally secreted by adipocytes. Elevated resistin levels are seen in serum in both genetic and diet-induced obesity suggesting that it could potentially link obesity to diabetes (20).

*Dermcidin (NP\_444513)*—Dermcidin is a novel human antimicrobial peptide secreted by the sweat glands. Screening of a subtracted cDNA library of primary melanoma and benign melanocytic nevus tissues with cDNA arrays led to isolation of dermcidin. It has been shown to possess antimicrobial activity against *Escherichia coli*, *Enterococcus faecalis*, *Staphylococcus aureus*, and *Candida albicans* (21). It may also play a role

TABLE II

A list of semitryptic peptides identified from the hemodialysis fluid

Peptide sequences identified by mass spectrometry are bold and underlined.

Protein name	Accession number	Peptide sequence
$\beta$ globin	NP_000509	KVLGAFSDGLAHL <del>DLNLK</del> DNLKGT <b>FATLS</b> <b>ELHCDK</b> AYQKVV <b>AGVAN</b> <b>ALAHK</b>
Gelsolin	NP_000168	DGLGLSY <b>LSSHIANVER</b>
Prostatic binding protein	NP_002558	VKNRPT <b>SISWDGLDSGK</b> SGKLYTL <b>VLTDPDAPSR</b> SGKLYTL <b>VLTDPDAPSR</b>
Albumin precursor	NP_000468	<b>SHCIAEVENDEMPADLPSLA</b> <b>SHCIAEVENDEMPADLPSLA</b>
Aldo-keto reductase family 1	NP_006057	<b>QIDDILSVASVR</b> PAVLQVEC
$\delta$ globin	NP_000510	EALG RLL <b>VVYPWTQR</b> KVLGAFSDGLAHL <del>DLNLK</del> QAAYQKVV <b>AGVANALAHK</b>
Peptidylprolyl isomerase A	NP_066953	ILSMAN <b>AGPNTNGSQFFICTAK</b>
$\alpha$ 2 globin	NP_000508	DDMPNAL <b>SALSDLHAHK</b>
Ubiquitin and ribosomal protein S27a precursor	NP_002945	AKI QDKEGI <b>PPDQQR</b>

in tumorigenesis by enhancing cell growth and survival in a subset of breast carcinomas (22).

*$\alpha$ -Defensin 1 (NP\_004075)*—Defensins are a family of microbicidal and cytotoxic polypeptides involved in host defense.  $\alpha$ -Defensin 1 was identified by screening a cDNA library constructed from HL-60, a human promyelocytic leukemia cell line, with an oligonucleotide probe based on the C-terminal sequence of human neutrophil peptides (23). It was found in different tissues including bone marrow, blood, neutrophils, and plasma (24).  $\alpha$ -Defensins have been shown to inhibit the replication of HIV-1 (25).

*$\alpha$ -Defensin 3 (NP\_005208)*—Several proteins secreted by activated CD8<sup>+</sup> T cells from long term non-progressors with HIV-1 were identified. One of them was  $\alpha$ -defensin 3 encoded by the *DEFA3* gene (25).  $\alpha$ -Defensins 1, 2, and 3 collectively account for much of the anti-HIV-1 activity of CD8 antiviral factor that is not attributable to  $\beta$ -chemokines (24). It is known to be expressed in bone marrow, leukocytes, and neutrophils (24, 26). All the family members are known to be present in plasma (27).

*Thymosin  $\beta$  10 (NP\_066926)*—Thymosin  $\beta$  10 was isolated from a kidney cDNA library and is an actin-sequestering protein (28). Thymosin  $\beta$  10 has been shown to be a putative progression marker for human cutaneous melanoma (28). It is likely that this protein is released into the plasma after lysis of cells.

*Chromosome 19 Open Reading Frame 10 (NP\_061980)*—A novel secreted protein was identified in a murine system using an expression cloning strategy (29). We have identified the human ortholog of this murine secreted protein. The function of this protein is not known, although its sequence suggests that it is likely to be a cytokine.

#### *Semitryptic Versus Full Tryptic Data Base Searching*

All of the 292 proteins in this study were identified by searching data bases using tryptic constraints for peptides. However, one would also expect to observe fragments derived after proteolysis *in vivo*. Thus, we tested this by searching a subset of our data having the highest mass accuracy (data from quadrupole-time-of-flight mass spectrometer) with semitryptic constraints. Searching our data with semitryptic instead of fully tryptic constraints resulted in a higher total score for nearly all of the proteins identified. However, for most of the entries, the higher score was a result of the contribution of low scoring peptides that did not pass our validation criteria. Semitryptic high scoring peptides that passed our manual validation are listed in Table II. Although some of the peptides presented in Table II could result from in-source fragmentation of labile proline-containing peptides (30), we expect that the majority arise from proteolytic cleavage events that occurred *in vivo*. Fig. 4A shows an example of an MS/MS spectrum of a semitryptic peptide derived from Gelsolin.

#### *Proteolytic Activity in Hemodialysis Fluid*

Theoretically the hemodialysis fluid should not contain proteins greater in size than 60 kDa. Nevertheless it is possible for larger proteins to be detected in the hemodialysis fluid if they undergo proteolytic cleavage *in vivo*. Thus, we decided to examine the distribution of peptides in greater detail. Of the proteins greater than 90 kDa, more than half exhibited a grouping of several peptides in either the N or C terminus of the protein, suggesting the presence of a fragment or isoform. Smaller fragments of proteins can be the result of proteolytic activity in plasma. Fig. 5 shows five examples of such groupings within a larger protein. For most of the proteins shown in Fig. 5, all peptides are clustered in regions that constitute only 4–8% of the whole protein sequence. For the 260-kDa protein fibronectin 1 and the 189-kDa protein complement component 3, a clustering of peptides was found in the middle of the proteins. This observation could be explained if these proteins were cleaved twice resulting in the observed fragment. For plasminogen (93 kDa) and desmocollin 1 (100 kDa), a clustering of peptides was observed in the N-terminal part of the protein, whereas a C-terminal clustering was observed in the case of Perlecan (480 kDa). Furthermore all of the above mentioned proteins were identified from bands that were at least a third of their expected molecular masses. For instance, fibronectin 1 was identified from the 15–19-kDa gel band, and desmocollin 1 was identified from the 6–15-kDa gel band. It is also worth noting that although several proteolytic cleavage sites are located in the C-terminal region of plasminogen (31), our data suggest that a cleavage occurred in the N terminus of the protein to generate Angiostatin (32), an important player in the regulation of angiogenesis (33). The probability of clustering of five peptides within a region that is 5% of the total protein length is  $10^{-6}$  strongly suggesting that we are observing true *in vivo* proteolytic events. Of the above described proteins, all were retrieved from protein bands migrating in the ~6–15-kDa region, except for complement component 3 that was retrieved from the ~60–80-kDa region. For the latter this observation is in support of a single cleavage rather than a multicleavage event and is in agreement with literature where C3 convertase has been reported to cleave complement component 3 at residue 748, generating C3a anaphylatoxin (34).

#### *Post-translational Modifications*

Many proteins and peptides found in plasma and hemodialysis fluid have undergone a number of changes since the formation of nascent polypeptide chains. These changes have to be taken into account when interpreting the mass spectrometry data partly to get a better description of the biological sample and partly to avoid false-positive identifications. Basically the modifications can be grouped into enzyme-catalyzed changes and spontaneously occurring modifications. A large number of peptides were oxidized at methionine and tryptophan residues or exhibited pyroglutamine formation

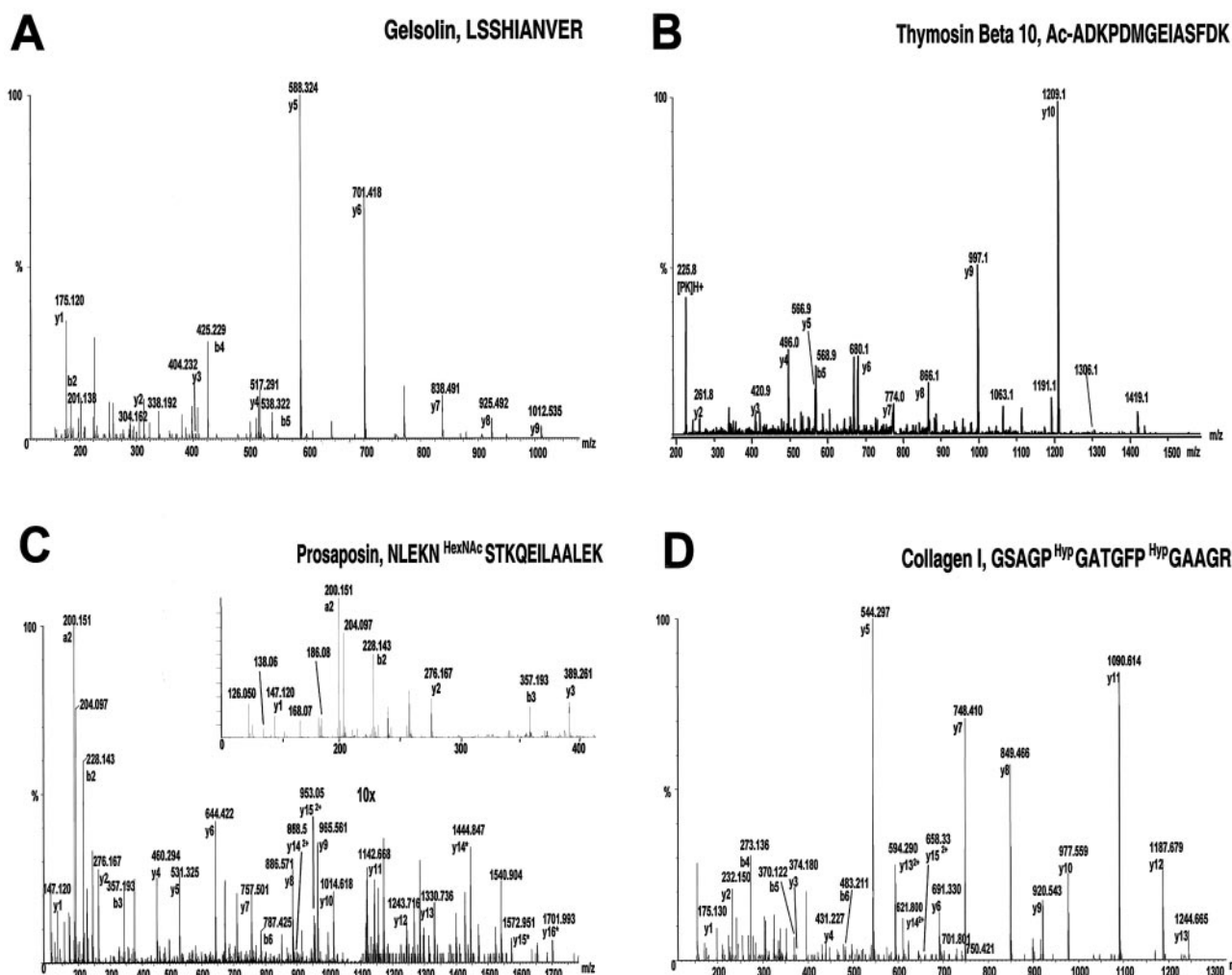


FIG. 4. Tandem mass spectra of a semitryptic peptide, LSSHIANVER, from Gelsolin (A), an N-terminally acetylated peptide, Ac-ADKPDMEIASFDK, from thymosin  $\beta$  10 (B), an N-glycosylated peptide, NLEKN<sup>HexNAc</sup>STKQEILAALEK, from prosaposin (C), a peptide containing hydroxyproline, GSAGP<sup>Hyp</sup>GATGFP<sup>Hyp</sup>GAAGR, derived from Collagen I (D). The corresponding peptide sequence is indicated in each case. Ac refers to acetylation, Hyp refers to hydroxyproline, and HexNAc refers to N-acetylhexosamine.

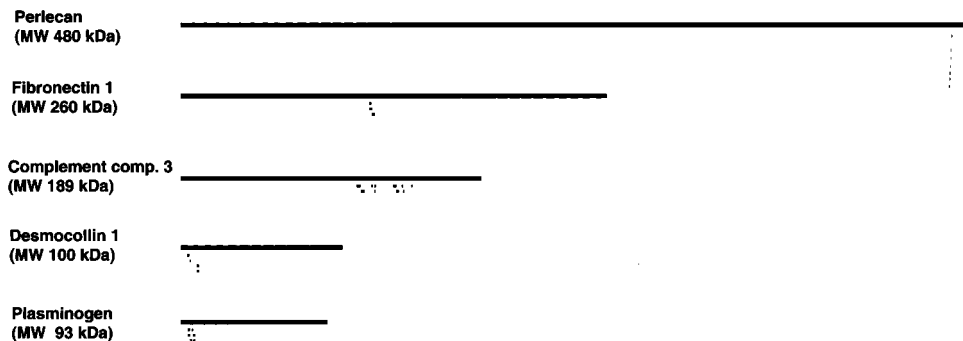


FIG. 5. A distribution of peptides for five large proteins. The proteins shown in the figure were identified from bands corresponding to molecular masses that were significantly lower than the calculated molecular masses. The peptides corresponding to the proteins are indicated by short thick lines. comp., component.

by cyclization by an N-terminal glutamine residue. It is difficult to infer any biological significance from these spontaneously occurring modifications, and they will not be discussed in

detail. We searched the data from the quadrupole time-of-flight mass spectrometer to identify some of the commonly occurring post-translational modifications.

TABLE III  
A list of N-terminally acetylated peptides and the corresponding proteins

Protein name	Accession number	Gene symbol	Peptide sequence	Instrument
1 Thymosin, $\beta$ 10	NP_066926.1	TMSB10	Ac-ADKPDMEIASFDK	Ion trap
2 Thymosin-like 2	NP_877594.1	TMSL2	Ac-SDKPDMAEIEK Ac-SDKPDMAEIEKFDK	Q-TOF Q-TOF
3 Thymosin, $\beta$ 4	NP_066932.1	TMSB4X	Ac-SDKPDMoxAEIEKFDK Ac-SDKPDMAEIEK	Q-TOF Q-TOF
4 Thymosin-like 4	NP_877595.1	TMSL4	Ac-SDKPDMAEIEK Ac-SDKPDMAEIEKFDK	Q-TOF Q-TOF
5 Metallothionein 2A	NP_005944.1	MT2A	Ac-MDPNCSCAAGDSCTCAGSCK	Q-TOF
6 Metallothionein 1X	NP_005943.1	MT1X	Ac-MoxDPNCSCSPVGSACAGSCK	Q-TOF
7 Metallothionein 1G	NP_005941.1	MT1G	Ac-MDPNCSCAAGVSCTCASSCK	Q-TOF
8 Metallothionein 1F	NP_005940.1	MT1F	Ac-MoxDPNCSCAAGVSCTCAGSCK	Q-TOF
9 Metallothionein 1B	NP_005938.1	MT1B	Ac-MDPNCSCCTGGSCACAGSCK	Q-TOF
10 Heat shock 10-kDa protein 1 (chaperonin 10)	NP_002148.1	HSPE1	Ac-AGQAFRK	Q-TOF
11 Cystatin B	NP_000091.1	CSTB	Ac-MMCGAPSATQPATAETQHIADQVR	Q-TOF
12 Death-associated protein	NP_004385.1	DAP	Ac-SSPPEGKLETK	Ion trap
13 Glutaredoxin (thioltransferase)	NP_002055.1	GLRX	Ac-AQEFVNCK	Ion trap
14 Macrophage migration-inhibitory factor (glycosylation-inhibiting factor)	NP_002406.1	MIF	Ac-PMoxFIVNTNVPR	Ion trap
15 Gastrotropin	NP_001436.1	FABP6	Ac-AFTGKFESESEKYNDEFMK	Ion trap
16 Metallothionein 1E	NP_783316.1	MT1E	Ac-MDPNCSCATGGSCCTCAGSCK	Q-TOF
17 Fatty acid-binding protein 4, adipocyte	NP_001433.1	FABP4	Ac-CDAFVGTWK	Ion trap
18 Fatty acid-binding protein 3	NP_004093.1	FABP3	Ac-VDAFLGTWK	Ion trap
19 Profilin 1	NP_005013.1	PFN1	Ac-AGWNAYIDNLMADGTCQDAAIVGYK	Q-TOF
20 $\beta$ -Galactosidase binding lectin precursor	NP_002296.1	LGALS1	Ac-ACGLVASNLNLKPGECLR	Ion trap
21 Fatty acid-binding protein 5 (psoriasis-associated)	NP_001435.1	FABP5	Ac-ATVQQLEGR	Ion trap
22 Coactosin-like 1	NP_066972.1	COTL1	Ac-ATKIDKEACR	Ion trap
23 Glia maturation factor, $\beta$	NP_004115.1	GMFB	Ac-SESLVVCDAEDLVEKLR Ac-SESLVVCDAEDLVEK	Ion trap Ion trap
24 Peptidylprolyl isomerase A (cyclophilin A)	NP_066953.1	PPIA	Ac-MVNPTVFFDIAVDGPELGR Ac-VNPTVFFDIAVDGPELGR*	Ion trap Q-TOF
25 Cofilin 1 (non-muscle)	NP_005498.1	CFL1	Ac-ASGVAVSDGVIK	Q-TOF
26 Cofilin 2	NP_068733.1	CFL2	Ac-ASGVTVNDEVIK	Ion trap
27 Ubiquitin-conjugating enzyme E2G 1 isoform 1	NP_003333.1	UBE2G1	Ac-TELQSALLLR	Q-TOF
28 ADP-ribosylation factor 1	NP_001649.1	ARF1	Ac-GNIFANLFK	Q-TOF
29 Transgelin 2	NP_003555.1	TAGLN2	Ac-ANRGPAYGLSR	Q-TOF
30 Transgelin	NP_003177.1	TAGLN	Ac-ANKGPSYGMRSR	Ion trap
31 Chloride intracellular channel 1	NP_001279.2	CLIC1	Ac-AEEQPQVELFVK	Ion trap
32 Carbonic anhydrase I	NP_001729.1	CA1	Ac-ASPDWGYDDKNGPEQWSK Ac-ASPDWGYDDK	Q-TOF Q-TOF
33 Carbonic anhydrase II	NP_000058.1	CA2	Ac-SHHWGYGK	Ion trap
34 Dimethylarginine dimethylaminohydrolase 1	NP_036269.1	DDAH1	Ac-AGLGHPAAFGR	Q-TOF
35 Regucalcin	NP_004674.1	RGN	Ac-SSIKIECVLPENCR	Ion trap
36 Hypothetical protein FLJ11151	NP_060810.1	FLJ11151	Ac-SAAEAGGVFHR	Ion trap
37 Aldo-keto reductase family 1, member A1	NP_006057.1	AKR1A1	Ac-AASCVLLHTGQK	Q-TOF
38 $\beta$ actin	NP_001092.1	ACTB	Ac-DDDIAALVVDNGSGMCK	Ion trap
39 Phosphoglycerate kinase 1	NP_000282.1	PGK1	Ac-SLSNKLTLDKLDVK	Ion trap
40 LIM protein (similar to rat protein kinase C-binding enigma)	NP_006448.1	LIM	Ac-SNYSVSLVGPAPWGFR	Ion trap
41 LIM domain binding 3	NP_009009	LDB3	Ac-SYSVTLTGPGPWGFR	Q-TOF
42 Tollid-like 1	NP_036596.3	TLL1	Ac-GLGTLSPR	Q-TOF
43 Chromosome 6 open reading frame 115	XP_371848	C6orf115	Ac-MoxNVDHEVNLLVEEIIHR Ac-MNVDHEVNLLVEEIIHR	Q-TOF Q-TOF

### Acetylation

The N terminus of most proteins *in vivo* is processed by aminopeptidases and N-acetyltransferases (35). Aminopeptidases generally cleave the initiator methionine if the penultimate residue has a small radius of gyration (e.g. Gly, Ala, Ser,

Cys, Thr, Pro, or Val). N-terminal acetylation is a very common co- and post-translational process where an acetyl group is transferred from acetyl-CoA to the N-terminal  $\alpha$ -amino acid of a protein (35). In our study, the acetylated N termini of 43 proteins were detected. The proteins and the corresponding

TABLE IV  
A list of peptides with other post-translational modifications

Protein name	Gene symbol	Accession number	Type of post-translational modification	Peptide sequence <sup>a</sup>	Instrument
Prosaposin	PSAP	NP_002769.1	HexNAc	<b>TN</b> *STFVQALVEHVKEECDR	Q-TOF
			HexNAc	NLEKN*STKQEILAALEK	Q-TOF
			HexNAc	<b>N</b> *STKQEILAALEK	Q-TOF
Fibrinogen, $\alpha$ chain	FGA	NP_000499.1	Hydroxylation	TFPGFFSP*MLGEFVSETESR	Q-TOF
$\alpha$ 2 type I collagen	COL1A2	NP_000080	Hydroxylation	GEP*GNIGFP*GPK	Q-TOF
$\alpha$ 1 type I collagen preproprotein	COL1A1	NP_000079.1	Hydroxylation	GSAGPP*GATGFP*GAAGR	Q-TOF
			Hydroxylation	GFP*GADGVAGPK	Q-TOF
			Hydroxylation	GVVGLP*GQR	Q-TOF
			Hydroxylation	GFP*GLP*GPSGEPGK	Q-TOF

<sup>a</sup> Bold X\* indicates the residue that is modified.

acetylated peptides are listed in Table III. We were able to find two different N termini for peptidylprolyl isomerase A; in one instance, the initiator methionine was acetylated, and in another, the initiator methionine was removed and the next amino acid, valine, was acetylated. Fig. 4B shows an MS/MS spectrum of an acetylated peptide derived from thymosin  $\beta$  10 protein. Our data on N-terminal acetylation is in good agreement with the preference of methionine aminopeptidases that the ultimate amino acid is small or unmodified. Of the 43 acetylated proteins, one N-acetylated peptide was found in the middle of a predicted protein (XP\_371848), suggesting that this could be a wrongly predicted protein. Indeed a Blast analysis of orthologous proteins confirms our new assignment of the translational initiation site (see Supplemental Fig. 2), which is in agreement with sequence conservation in five different species beginning at this methionine. Supplemental Fig. 3 shows the MS/MS spectra of all peptides containing an acetylated residue.

#### N-Glycosylation of Peptides

Three peptides were found to be covalently modified with an N-acetylhexosamine moiety, which corresponds to a mass increase of 203 Da (Table III and Supplemental Fig. 4). The CID-induced fragmentation spectrum of the triply charged ion with  $m/z$  711.396 is shown in Fig. 4C. The N-acetylhexosamine gives rise to an intense oxonium ion at  $m/z$  204, and further fragmentation of the oxonium ion gives rise to ions with  $m/z$  186, 168, 138, and 126 (see Fig. 4C, inset). Interpretation of the mass spectrum revealed that the second asparagine in the peptide was covalently linked to an N-acetylhexosamine moiety. This modified asparagine occurs in the glycosylation consensus sequence NX(S/T) where X can be any amino acid. This is an unusual modification because the common N-glycan is cotranslationally transferred to the asparagine residue en bloc as the oligosaccharide GlcNAc<sub>2</sub>Man<sub>9</sub>Glc<sub>3</sub> and seldom trimmed beyond the trimannosyl-chitobiose core. It is interesting that only three peptides are found with this modification and come from a protein called prosaposin. Because N-glycosylation takes place only on asparagines in the glycosylation sequon NX(S/T), there will always be an adjacent

serine of threonine. It is difficult to find peptides where O-GlcNAc is at least not a theoretically possible explanation (except in the rare instances where X is arginine or lysine, and trypsin cleavage is not hindered by the glycan moiety). However, there is a doubly charged ion in the spectrum that shows, even though it is small, that asparagine is the site of modification.

Prosaposin is a 524-amino acid glycoprotein that gives rise to four saposins that are predominantly localized in late endosomal/lysosomal compartment. A large number of endo- and exoglycosidases are also present in the lysosome, and it is possible that the extensive trimming of the glycan down to a single N-acetylglucosamine takes place there. Whether this trimming of the glycan structure has any biological significance is not known, but it should be noted that two cases of metachromatic leukodystrophy have been reported where the mutations in the glycosylation sequon, either N215H (36) or T217I (37) corresponding to glycosylation sequon in the identified peptide TNSTFVQALVEHVKEECDR, led to a dysfunctional saposin B protein emphasizing the importance of glycosylation.

#### Proline Hydroxylation

A number of structural molecules in the extracellular matrix are known to undergo extensive post-translational modifications. Hydroxylation of prolines by prolyl hydroxylase is a common modification of collagen that confers structural stability to the collagen triple helix (38). We identified one proline hydroxylation site in fibrinogen, two sites in collagen  $\alpha$  2 type I, and six sites in collagen  $\alpha$  1 type I (see Fig. 4D for the MS/MS spectrum of a peptide containing two hydroxylated proline residues). Table IV lists all the hydroxylation sites identified in this study, and the MS/MS spectra are shown in Supplemental Fig. 5.

#### Conclusions

Using one-dimensional gel electrophoresis and LC-MS/MS, we have identified 292 proteins from hemodialysis fluid of which more than half were proteins smaller than 30 kDa.

Analysis of the modified peptides led to identification of 43 N-terminally acetylated proteins and three proteins hydroxylated on prolines. We also found three peptides from prosaposin to be modified with a single HexNAc at two different glycosylation sequons. We were able to map the identified peptides onto larger proteins, which showed groupings of peptides within limited regions. A comparison of our results with previously published studies that examined serum and plasma proteomes showed that two-thirds of the proteins identified in this study had not been identified previously as components of serum or plasma. We feel that this is mainly due to two major contributing factors: the first is the greater dynamic range of protein concentrations in serum/plasma samples, and the second is enrichment of the lower molecular weight proteins in the hemodialysis fluid. The proteins identified in this study will allow further investigations into their detection in serum/plasma and possible use as biomarkers of disease states.

This study presents the first comprehensive list of hemofiltrate proteins and in-depth analysis of the post-translational modifications. This proteomic survey is by no means exhaustive, and there are probably many proteins in the low molecular weight region we have not identified. This is not an uncommon phenomenon. In the Anderson *et al.* (9) list where literature and three proteomic studies have been compared, 196 of 1,275 proteins were reported in more than one study, and only 46 were reported in all four studies. It is currently not possible to extrapolate from the hemofiltrate to plasma or urine constituents, but it represents a little step toward mapping the enormous unknown of the human body fluid proteomes in health and disease.

**Acknowledgments**—We thank John Kloss for help with data base searching and Raghunath Reddy for creating helpful scripts. We thank Sun Microsystems for providing us a computer cluster under the Academic Equipment Grant Program.

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this manuscript (available at <http://www.mcponline.org>) contains supplemental material.

¶ Supported by a grant from the Danish Natural Sciences Research Council.

‡ Supported by Beckman Young Investigator and Sidney Kimmel Scholar awards. Serves as Chief Scientific Advisor to the Institute of Bioinformatics. The terms of this arrangement are being managed by The Johns Hopkins University in accordance with its conflict of interest policies. To whom correspondence should be addressed. Tel.: 410-502-6662; Fax: 410-502-7544; E-mail: [pandey@jhmi.edu](mailto:pandey@jhmi.edu).

#### REFERENCES

- Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell Proteomics* **1**, 845–867
- Maack, T. (2000) in *The Kidney: Physiology and Pathophysiology* (Seldin, D. W., and Giebisch, G. G., eds) 3rd Ed., pp. 2235–2267, Lippincott Williams & Wilkins, Philadelphia, PA
- Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P., and Veenstra, T. D. (2003) Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2**, 1096–1103
- Alpert, A. J., and Shukla, A. K. (2003) in *Association of Biomolecular Resource Facilities Conference, Denver, Colorado, February 10–13, 2003*, Abstr. P111-W, Association of Biomolecular Resource Facilities, Santa Fe, NM
- Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S., Schatz, C. R., Miller, S. S., Su, Q., McGrath, A. M., Estock, M. A., Parmar, P. P., Zhao, M., Huang, S. T., Zhou, J., Wang, F., Esquer-Blasco, R., Anderson, N. L., Taylor, J., and Steiner, S. (2003) The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* **3**, 1345–1364
- Raida, M., Schulz-Knappe, P., Heine, G., and Forssmann, W. G. (1999) Liquid chromatography and electrospray mass spectrometric mapping of peptides from human plasma filtrate. *J. Am. Soc. Mass Spectrom.* **10**, 45–54
- Schepky, A. G., Bensch, K. W., Schulz-Knappe, P., and Forssmann, W. G. (1994) Human hemofiltrate as a source of circulating bioactive peptides: determination of amino acids, peptides and proteins. *Biomed. Chromatogr.* **8**, 90–94
- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858
- 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., and Lobley, A. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3**, 311–326
- Chan, K. C., Lucas, D. A., Hise, D., Schaefer, C. F., Xiao, Z., Janini, G. M., Buetow, K. H., Issaq, H. J., Veenstra, T. D., and Conrads, T. P. (2004) Analysis of the human serum proteome. *Clin. Proteomics* **1**, 101–226
- Davis, M. T., Spahr, C. S., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Yu, W., Luethy, R., and Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. II. Limitations of complex mixture analyses. *Proteomics* **1**, 108–117
- Pisitkun, T., Shen, R. F., and Knepper, M. A. (2004) Identification and proteomic profiling of exosomes in human urine. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13368–13373
- Zhu, Y., Xu, G., Patel, A., McLaughlin, M. M., Silverman, C., Knecht, K., Sweitzer, S., Li, X., McDonnell, P., Mirabile, R., Zimmerman, D., Boyce, R., Tierney, L. A., Hu, E., Livi, G. P., Wolf, B., Abdel-Meguid, S. S., Rose, G. D., Aurora, R., Hensley, P., Briggs, M., and Young, P. R. (2002) Cloning, expression, and initial characterization of a novel cytokine-like gene family. *Genomics* **80**, 144–150
- Peeters, R. A., Veerkamp, J. H., Geurts van Kessel, A., Kanda, T., and Ono, T. (1991) Cloning of the cDNA encoding human skeletal-muscle fatty-acid-binding protein, its peptide sequence and chromosomal localization. *Biochem. J.* **276**, 203–207
- Phelan, C. M., Larsson, C., Baird, S., Futreal, P. A., Ruttledge, M. H., Morgan, K., Tonin, P., Hung, H., Korneluk, R. G., Pollak, M. N., and Narod, S. A. (1996) The human mammary-derived growth inhibitor (MDGI) gene: genomic structure and mutation analysis in human breast tumors. *Genomics* **34**, 63–68
- Yoshikawa, Y., Mukai, H., Hino, F., Asada, K., and Kato, I. (2000) Isolation of two novel genes, down-regulated in gastric cancer. *Jpn. J. Cancer Res.* **91**, 459–463
- Shiozaki, K., Nakamori, S., Tsujie, M., Okami, J., Yamamoto, H., Nagano, H., Dono, K., Umeshita, K., Sakon, M., Furukawa, H., Hiratsuka, M., Kasugai, T., Ishiguro, S., and Monden, M. (2001) Human stomach-specific gene, CA11, is down-regulated in gastric cancer. *Int. J. Oncol.* **19**, 701–707
- Martin, T. E., Powell, C. T., Wang, Z., Bhattacharyya, S., Walsh-Reitz, M. M., Agarwal, K., and Toback, F. G. (2003) A novel mitogenic protein that is highly expressed in cells of the gastric antrum mucosa. *Am. J. Physiol.* **285**, G332–G343
- Holcomb, I. N., Kabakoff, R. C., Chan, B., Baker, T. W., Gurney, A., Henzel, W., Nelson, C., Lowman, H. B., Wright, B. D., Skelton, N. J., Frantz, G. D., Tumas, D. B., Peale, F. V., Jr., Shelton, D. L., and Hebert, C. C. (2000)

- FIZZ1, a novel cysteine-rich secreted protein associated with pulmonary inflammation, defines a new gene family. *EMBO J.* **19**, 4046–4055
20. Banerjee, R. R., Rangwala, S. M., Shapiro, J. S., Rich, A. S., Rhoades, B., Qi, Y., Wang, J., Rajala, M. W., Poci, A., Scherer, P. E., Stepan, C. M., Ahima, R. S., Obici, S., Rossetti, L., and Lazar, M. A. (2004) Regulation of fasted blood glucose by resistin. *Science* **303**, 1195–1198
  21. Cunningham, T. J., Hodge, L., Speicher, D., Reim, D., Tyler-Polsz, C., Levitt, P., Eagleson, K., Kennedy, S., and Wang, Y. (1998) Identification of a survival-promoting peptide in medium conditioned by oxidatively stressed cell lines of nervous system origin. *J. Neurosci.* **18**, 7047–7060
  22. Porter, D., Weremowicz, S., Chin, K., Seth, P., Keshaviah, A., Lahti-Domenici, J., Bae, Y. K., Monitto, C. L., Merlos-Suarez, A., Chan, J., Hulette, C. M., Richardson, A., Morton, C. C., Marks, J., Duyao, M., Hruban, R., Gabrielson, E., Gelman, R., and Polyak, K. (2003) A neural survival factor is a candidate oncogene in breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10931–10936
  23. Daher, K. A., Lehrer, R. I., Ganz, T., and Kronenberg, M. (1988) Isolation and characterization of human defensin cDNA clones. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 7327–7331
  24. Date, Y., Nakazato, M., Shiomi, K., Toshimori, H., Kangawa, K., Matsuo, H., and Matsukura, S. (1994) Localization of human neutrophil peptide (HNP) and its messenger RNA in neutrophil series. *Ann. Hematol.* **69**, 73–77
  25. Zhang, L., Yu, W., He, T., Yu, J., Caffrey, R. E., Dalmasso, E. A., Fu, S., Pham, T., Mei, J., Ho, J. J., Zhang, W., Lopez, P., and Ho, D. D. (2002) Contribution of human  $\alpha$ -defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. *Science* **298**, 995–1000
  26. Linzmeier, R., Michaelson, D., Liu, L., and Ganz, T. (1993) The structure of neutrophil defensin genes. *FEBS Lett.* **326**, 299–300
  27. Ashitani, J., Nakazato, M., Mukae, H., Taniguchi, H., Date, Y., and Matsukura, S. (2000) Recombinant granulocyte colony-stimulating factor induces production of human neutrophil peptides in lung cancer patients with neutropenia. *Regul. Pept.* **95**, 87–92
  28. McCreary, V., Kartha, S., Bell, G. I., and Toback, F. G. (1988) Sequence of a human kidney cDNA clone encoding thymosin  $\beta$  10. *Biochem. Biophys. Res. Commun.* **152**, 862–866
  29. Tulin, E. E., Onoda, N., Nakata, Y., Maeda, M., Hasegawa, M., Nomura, H., and Kitamura, T. (2001) SF20/IL-25, a novel bone marrow stroma-derived growth factor that binds to mouse thymic shared antigen-1 and supports lymphoid cell proliferation. *J. Immunol.* **167**, 6338–6347
  30. Olsen, J. V., Ong, S. E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **3**, 608–614
  31. Ohyama, S., Harada, T., Chikanishi, T., Miura, Y., and Hasumi, K. (2004) Nonlysine-analog plasminogen modulators promote autolytic generation of plasmin(ogen) fragments with angiostatin-like activity. *Eur. J. Biochem.* **271**, 809–820
  32. Patterson, B. C., and Sang, Q. A. (1997) Angiostatin-converting enzyme activities of human matrilysin (MMP-7) and gelatinase B/type IV collagenase (MMP-9). *J. Biol. Chem.* **272**, 28823–28825
  33. O'Reilly, M. S., Holmgren, L., Shing, Y., Chen, C., Rosenthal, R. A., Moses, M., Lane, W. S., Cao, Y., Sage, E. H., and Folkman, J. (1994) Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma. *Cell* **79**, 315
  34. Hugli, T. E. (1975) Human anaphylatoxin (C3a) from the third component of complement. Primary structure. *J. Biol. Chem.* **250**, 8293–8301
  35. Polevoda, B., and Sherman, F. (2003) N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. *J. Mol. Biol.* **325**, 595–622
  36. Wroble, D., Henseler, M., Huettler, S., Pascual Pascual, S. I., Chabas, A., and Sandhoff, K. (2000) A non-glycosylated and functionally deficient mutant (N215H) of the sphingolipid activator protein B (SAP-B) in a novel case of metachromatic leukodystrophy (MLD). *J. Inher. Metab. Dis.* **23**, 63–76
  37. Kretz, K. A., Carson, G. S., Morimoto, S., Kishimoto, Y., Fluharty, A. L., and O'Brien, J. S. (1990) Characterization of a mutation in a family with saposin B deficiency: a glycosylation site defect. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2541–2544
  38. Pihlajaniemi, T., Myllyla, R., and Kivirikko, K. I. (1991) Prolyl 4-hydroxylase and its role in collagen synthesis. *J. Hepatol.* **13**, Suppl. 3, S2–S7