The Human Erythrocyte Proteome

ANALYSIS BY ION TRAP MASS SPECTROMETRY*

David G. Kakhniashvili‡§¶, Lee A. Bulla, Jr.‡¶‖, and Steven R. Goodman‡§¶‖**‡‡

This report describes an analysis of the red blood cell proteome by ion trap tandem mass spectrometry in line with liquid chromatography. Mature red blood cells lack all internal cell structures and consist of cytoplasm within a plasma membrane envelope. To maximize outcome, total red blood cell protein was divided into two fractions of membrane-associated proteins and cytoplasmic proteins. Both fractions were divided into subfractions, and proteins were identified in each fraction separately through tryptic digestion. Membrane protein digests were collected from externally exposed proteins, internally exposed proteins, "spectrin extract" mainly consisting of membrane skeleton proteins, and membrane proteins minus spectrin extract. Cytoplasmic proteins were divided into 21 fractions based on molecular mass by size exclusion chromatography. The tryptic peptides were separated by reverse-phase high-performance liquid chromatography and identified by ion trap tandem mass spectrometry. A total of 181 unique protein sequences were identified: 91 in the membrane fractions and 91 in the cytoplasmic fractions. Glyceraldehyde-3-phosphate dehydrogenase was identified with high sequence coverage in both membrane and cytoplasmic fractions. Identified proteins include membrane skeletal proteins, metabolic enzymes, transporters and channel proteins, adhesion proteins, hemoglobins, cellular defense proteins, proteins of the ubiquitin-proteasome system, G-proteins of the Ras family, kinases, chaperone proteins, proteases, translation initiation factors, and others. In addition to the known proteins, there were 43 proteins whose identification was not determined. ***Molecular & Cellular Proteomics 3:501–509, 2004.***

A human red blood cell (RBC)[1] is in residence in the human circulatory system for 120 days carrying oxygen from the lungs to all tissues within the body and carbon dioxide from the tissues back to the lungs. An RBC is an 8-μm biconcave disk bounded by a plasma membrane. The major cytoplasmic constituent is hemoglobin, which is responsible for binding and releasing oxygen and carbon dioxide. On the cytoplasmic surface of the plasma membrane is a two-dimensional meshwork of proteins referred to as spectrin membrane skeleton. The spectrin membrane skeleton renders elasticity and flexibility to an RBC, allowing it to pass through vessels and capillaries that narrow to 1 μm in diameter (1).

Because of the ease in obtaining RBCs and because they lack internal organelles, the plasma membrane of this cell type has been studied extensively. The functions of hemoglobin are also well documented. Based on four decades of study, the identity, function, and topology of many RBC membrane proteins have been determined (1–3). With the advent of modern mass spectrometry (MS) and associated proteomic techniques, determination of the RBC proteome is now plausible. This kind of approach is a necessary first step in understanding how the RBC proteome becomes altered in various hematologic disorders. With this goal in mind, we utilized ion trap tandem MS to analyze the entire human erythrocyte proteome (plasma membrane and cytoplasmic proteins). We identified 181 unique RBC proteins, half of which reside in the plasma membrane and half in the cytoplasm. Moreover, we were able to not only identify the proteins but to also categorize most of them according to function.

EXPERIMENTAL PROCEDURES

Sample Preparation—Human peripheral whole blood was collected in vacutainer tubes containing lithium heparin, sufficient for 10 ml of blood, and used within 24 h. The RBCs were sedimented at 1000 × *g* for 10 min at 4 °C and resuspended in phosphate-buffered saline (PBS; 10 mM $NaPO_4$, pH 7.6, 150 mM NaCl) to the original volume (~10 ml) four times. Each time, the upper 1–2 mm layer of packed cells was aspirated along with liquid phase to remove white blood cells. RBCs were transferred to 50-ml centrifugal tubes (2–3 ml of packed cells per tube) and washed with PBS at 4 °C three times: the cells were resuspended in 10 volumes of PBS and sedimented at 2000 × *g* for 10 min.

To maximize the number of identified RBC proteins, membrane-associated proteins, as well as cytoplasmic proteins, were separated and digested into several fractions; each fraction was analyzed as a separate sample. The RBC membranes, cytoplasmic proteins, inside out vesicles (IOV), and membrane skeleton proteins ("spectrin extract") were prepared as described (4) with the following modifications. Five milliliters of PBS-washed and packed RBCs were resuspended in 10 ml of PBS and incubated (gently shaken) with 1.5 mg of tosylphenylalanyl chloromethyl ketone-treated trypsin (Worthington, Lakewood, NJ) for 2 h at room temperature to digest exposed do-

From the ‡Institute of Biomedical Sciences and Technology, §Sickle Cell Disease Research Center, ¶Department of Molecular and Cell Biology, and ‖Center for Biotechnology and Bioinformatics, University of Texas at Dallas, Richardson, TX 75083-0688; and **Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390

Received, December 8, 2003, and in revised form, February 12, 2004

Published, MCP Papers in Press, February 12, 2004, DOI 10.1074/mcp.M300132-MCP200

[1] The abbreviations used are: RBC, red blood cell; HPLC, high-performance liquid chromatography; IOV, inside out vesicles; MS, mass spectrometry; LC/MS/MS, liquid chromatography in line with tandem mass spectrometry; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; PBS, phosphate-buffered saline.

mains of membrane proteins. The digested material and the trypsinized RBCs were separated by centrifugation at 2000 × *g* for 10 min. The digested material was collected (sample 1) and cleared by centrifugation at 32,000 × *g* for 15 min at 4 °C. The separated trypsin-treated RBCs were washed with 10 volume of PBS three times and resuspended (final volume 32 ml) in ice-cold lysis buffer (5 mM $NaPO_4$, pH 7.6, 1 mM EDTA). The lysed cells and the soluble cytoplasmic proteins were separated by centrifugation at 32,000 × *g* for 10 min at 2 °C. The cytoplasmic proteins (~35 mg protein/ml) were collected and cleared by recentrifugation. Lysed RBCs were further washed with 10 volume of ice-cold lysis buffer (six times) to prepare cell membranes (ghosts, ~4 mg protein/ml). Membrane skeleton proteins and IOVs were prepared as follows. The membranes (~5 ml) were washed in 10 volumes of the ice-cold spectrin extraction buffer (0.1 mM EDTA, pH 8.0), then resuspended in 20 ml of the same buffer and incubated at 37 °C for 30 min to dissociate membrane skeleton. The extracted membrane skeleton proteins (spectrin extract) and IOV were separated by centrifugation at 250,000 × *g* for 30 min at 4 °C. Spectrin extract (~0.2 mg/ml) was collected. The tight IOV pellet was gently resuspended in bicarbonate buffer (100 mM ammonium bicarbonate, pH 8.2) to the original membrane volume of 5 ml (~3 mg protein/ml).

The IOV were further diluted twice (~1.5 mg protein/ml) in the bicarbonate buffer and incubated with 0.1 mg/ml trypsin for 15 h at 37 °C to digest exposed domains of IOV proteins. The digested IOV were separated from corresponding supernatant containing tryptic peptides (sample 2) by sedimentation at 100,000 × *g* for 30 min at 4 °C. In the second set of experiments, the total IOV proteins were solubilized and digested with 0.2 mg/ml trypsin under similar conditions at 16 °C in the presence of 1% precondensed Triton X-114 (5). After digestion, the mixtures were incubated at 30 °C and centrifuged at 1000 × *g* for 10 min at room temperature to separate aqueous and detergent phases. The aqueous phase containing digested peptides (sample 3) released from IOV were collected and cleared by centrifugation at 100,000 × *g* for 30 min.

Spectrin extract, containing membrane skeleton proteins (0.2 mg/ml), was digested in bicarbonate buffer with 0.05 mg/ml trypsin at 37 °C for 15 h (sample 4).

The cytoplasmic proteins were diluted with 10 mM Tris-HCl, pH 7.8, to 3.7 mg/ml, and 4.6 ml was applied to a Sephacryl S100 HR (bed volume 92 ml) column (1 × 120 cm). The column was equilibrated and eluted with the Tris buffer at 6.0 ml/h flow rate. Three-milliliter fractions were collected, and protein concentrations were determined. Protein was detected in fractions 11–31 (samples 5–25). The samples (5–7, 10–25) were concentrated by vacuum centrifugation. All cytoplasmic samples were incubated with 8 M urea for 1 h at 37 °C, then diluted four times in the bicarbonate buffer (at this point, the protein concentration in different cytoplasmic samples varied from 0.08 to 1 mg/ml) and digested with 0.06 mg/ml trypsin for 15 h at 37 °C.

Digested sample aliquots were reduced with 2 mM dithiothreitol for 1 h and "alkylated" in the dark with 20 mM iodoacetamide for 1 h at 37 °C in the bicarbonate buffer. The reaction was stopped by addition of 200 mM 2-mercaptoethanol.

Mass Spectrometry—Digested samples were analyzed by microcapillary liquid chromatography in line with tandem MS (μLC/MS/MS) using a Surveyor high-performance liquid chromatography (HPLC) system connected to a LCQ DECA XP ion trap mass spectrometer with an electrospray ionization source (ThermoFinnigan, San Jose, CA). Proteins in the tryptic digest (20 μl) were separated by reverse-phase chromatography on a C18 column (2.1 × 150 mm; Thermo Hypersil-Keystone, Bellefonte, PA) at 200 μl/min flow rate. Water and acetonitrile with 0.1% formic acid each were used as solvents A and B, respectively. The gradient was started and kept for 5 min at 5% B, then ramped to 60% B in 110 or 165 min, and finally ramped to 90% B for another 15 min. The eluted peptides, singly, doubly, or triply ionized (charge state 1+, 2+, or 3+, respectively) at the electrospray source, were analyzed in data-dependent MS experiments ("big three" or "triple play") with dynamic exclusion. In the "big three" experiments, the data acquisition parameters were set such that each analytical event consisted of four consecutive scans: the first, full MS (*m/z* 300–2000) scan was followed by three MS/MS scans on the three most intense peptide ions from the full MS spectrum. In the "triple play" experiments, the first, full MS scan was followed by a zoom (high resolution) scan and an MS/MS scan on the most intense ion from the full scan. A peptide ion, analyzed twice within 30 s, was excluded from re-analysis for 2 min. The spray voltage was set at 4.5 kV; the ion transfer capillary temperature was set at 200 °C; and the normalized collision energy for MS/MS decomposition of peptides was set at 35%.

Database Search—A quality MS/MS spectrum, resulting from a peptide fragmentation, features a unique set of b and y fragment ions characteristic to the peptide. The peptide sequence can be identified through interpretation of its MS/MS spectrum and similarity to the MS/MS spectrum of a known peptide sequence. Each acquired MS/MS spectrum was searched against the nonredundant protein sequence database nr.fasta using the SEQUEST software tool (6, 7). The software creates theoretical peptides for all, or a limited group of, database proteins; calculates corresponding MS/MS spectra; and compares them to an experimental spectrum (submitted for the database search) to find the match. The database search was restricted to 700–3500 molecular mass tryptic peptides of human (*Homo sapiens*) origin. Up to two missed trypsin cleavage sites were allowed, and cysteines, where modified, were considered carbamidomethylated. The acceptable molecular mass difference (mass tolerance) between an experimental and database peptides was set to 1.5 mass unit. Mass tolerance for experimental and calculated MS/MS fragment ions was set to zero. Based on the similarity to the experimental MS/MS spectrum, the software assigns each database peptide the primary score (Sp), then the cross-correlation score (Xcorr) to filter candidate peptides and select a defined number of top hits: database peptides with highest Xcorr scores. Finally, a delta cross-correlation score (dCn, a difference between the top 2 Xcorr values normalized to 1) is calculated. The candidate database peptide with the highest Xcorr score was considered the match if the following identification criteria were met: 1) Xcorr of at least 2.0, 2.2, and 3.5 for singly, doubly, and triply charged peptides, respectively, and 2) dCn of at least 0.1 irregardless of charge state.

All identified peptides were grouped under the proteins of their origin. For each identified protein, the number of identified peptides was counted and the percentage of the covered sequence was calculated, and molecular mass was recorded.

RESULTS

Mature RBC consists of the plasma membrane, resting on the membrane skeleton and surrounding the cytoplasm. Extreme complexity of protein mixture, as well as highly abundant proteins (*e.g.* hemoglobin), limit the detection of low-abundance proteins by LC/MS/MS analysis of tryptic digest. A peptide coeluted from the HPLC column with more abundant peptides or a high number of other peptides may not be detected even if its absolute concentration in an eluate is within the sensitivity of the mass spectrometer.

To maximize the number of identified proteins, total RBC protein was divided into two major fractions: membrane-associated proteins, including the membrane skeleton, and cytoplasmic proteins. The two fractions were further divided into subfractions, and the proteins were identified in each subfrac-

TABLE I
Proteins identified in RBC membrane fractions

No.	Protein description	Molecular mass (Da)	Gi Number	Sequence coverage (%)	No. of identified peptides
1	Spectrin α chain, erythrocyte	279,916.5	1174412	48.0	77*
2	Spectrin β chain, erythrocyte	246,468.1	17476989	48.0	76*
3	Ankyrin 1, splice form 2	206,067.9	105337	45.0	55
4	Ankyrin 1, isoform 4, erythrocytic	203,416.6	10947036	45.0	50
5	Ankyrin 1, isoform 2, erythrocytic	189,011.2	10947042	46.0	48
6	Similar to ankyrin 1	206,264.8	13645508	51.0	46
7	Protein band 4.2, erythrocytic	79,946.5	107446	33.0	21
8	Protein band 4.1 (elliptocytosis 1, RH-linked)	66,398.5	4758274	45.0	17
9	Protein band 3, erythrocytic	101,792.3	4507021	28.0	17
10	Protein band 4.1, erythrocytic	97,016.9	14916944	32.0	16
11	Actin β chain	41,812.8	481515	47.0	12*
12	Flotillin 1, erythrocytic	47,355.3	5031699	47.0	12
13	Membrane protein p55, erythrocytic, (palmitoylated)	52,296.5	4505237	35.0	11
14	Flotillin 2	47,142.3	18587629	29.0	11
15	Protein band 4.9 (dematin), erythrocytic	45,514.4	13623437	40.0	10
16	Protein band 7.2b, stomatin	32,598.5	1103842	47.0	10
17	Glyceraldehyde-3-phosphate dehydrogenase	36,054.2	31645	51.0	10
18	Tropomyosin 3, cytoskeletal	29,032.7	136096	55.0	10
19	Solute carrier family 2 (facilitated glucose transporter), member 1	54,117.8	5730051	13.0	6
20	Similar to flotillin 2	42,565.9	13277550	15.0	6
21	Tropomyosin isoform	28,420.1	1082876	36.0	6
22	Glucose transporter glycoprotein	37,879.6	3387905	17.0	5
23	Tropomyosin α chain (smooth muscle)	26,576.7	136100	37.0	5
24	Actin α 2, aortic smooth muscle	42,108.1	1070613	20.0	5
25	Adducin α subunit, erythrocyte	80,955.1	12644231	10.0	5*
26	Rabphilin-3 A-integrating protein	80,858.2	1082757	8.0	5*
27	C-1-tetrahydrofolate synthase, cytoplasmic	101,559.2	115206	6.0	4
28	Translation initiation factor 2C, 2	66,252.2	18570004	10.0	4
29	Aldolase A	39,288.8	229674	17.0	4
30	Tropomodulin	40,569.2	4507553	16.0	3*
31	RAP2B, member of RAS oncogene family	20,504.4	11433346	43.0	3
32	Arginase type 1 erythroid variant	35,664.1	18535612	12.0	3
33	Arginase type 1	34,734.9	10947139	12.0	3
34	Creatine kinase, muscle	43,101.1	14763181	21.0	3
35	B-CAM protein	63,566.7	2134798	8.0	3
36	ATP-binding cassette half-transporter	99,712.3	11245444	5.0	2
37	RAP1A, member of RAS oncogene family or RAP1B	20,987.1 20,824.7	4506413 7661678	14.1 14.1	2 2
38	Calcium transporting ATPase 4	137,920.2	14286105	2.5	2
39	Rh blood D group antigen polypeptide	45,136.5	10800054	4.0	2
40	Channel-like integral membrane protein	16,239.5	1314306	15.0	2
41	Glycophorin A precursor	16,429.6	1070639	21.0	2
42	Solute carrier family 29 (nucleoside transporter), member 1	50,219.4	4826716	3.5	2
43	Glycophorin A	14,784.8	106140	23.0	2
44	Glutathione transferase	27,053.4	809436	19.0	2
45	Glycophorin C, isoform 1	13,810.6	4504229	20.0	1
46	Aquaporin 1	28,526.0	4502177	7.0	1
47	Erythroblast membrane-associated protein	52,604.8	17489129	3.0	1
48	Similar to glycophorin A	16,371.6	13529077	20.0	1
49	Cell surface glycoprotein CD44	39,433.8	7512338	4.0	1
50	Vesicle-associated membrane protein 2 (synaptobrevin 2)	12,648.7	7657675	15.0	1
51	Similar to adhesive plaque matrix protein precursor	106,879.1	17481669	1.9	1
52	Poly (A)-specific ribonuclease	73,451.0	4505611	3.0	1
53	Similar to RAS-related protein RAL-A	23,566.8	14740792	7.0	1
54	Presenilin-associated protein	39,862.4	6409316	6.0	1
55	Duodenal cytochrome *b*	31,611.2	13376257	3.5	1
56	bA421H8.2 (novel protein)	16,743.7	17402228	9.0	1
57	Similar to RAS-related protein RAB-15	23,517.9	18596861	5.0	1

TABLE I—*continued*

No.	Protein description	Molecular mass (Da)	Gi Number	Sequence coverage (%)	No. of identified peptides
58	CD59antigenp18–20	17,067.4	17473237	5.0	1
59	Rhesus D category VI type III protein	45,247.7	2765839	1.9	1
60	RAB 35, RAS oncogene family	23,025.2	5803135	6.0	1
61	Ral A binding protein	76,063.4	5803145	2.1	1
62	Hypothetical protein XP_100510	8,049.3	18577723	16.0	1
63	ATP-binding cassette, subfamily C, member 6	164,904.4	6715561	0.9	1
64	Phosphoribosyl pyrophosphate synthetase	34,834.2	4506127	7.0	1
65	Unknown protein	46,884.2	18089137	3.6	1
66	Similar to Lutheran blood group	59,287.7	18589892	3.1	1
67	Phosphatidylinositol-4-phosphate 5 kinase, type III	46,078.6	1730569	4.4	1
68	Hypothetical protein XP_100665	35,877.6	18604339	7.0	1
69	Hypothetical protein XP_100619	18,567.6	18604359	15.0	1
70	Block of proliferation 1	83,629.5	23830903	1.5	1
71	Similar to tropomyosin	10,804.3	18590249	13.0	1
72	Hypothetical protein XP_061743 or	48,719.0	17472555	2.5	1
	XP_089854	31,487.9	18577194	4.0	1
73	Hypothetical protein XP_106269	12,703.8	18558481	22.0	1
74	Hypothetical protein XP_100925	22,863.4	18601384	8.0	1
75	Zona pellucida binding protein	40,169.8	5902116	3.4	1
76	2′,3′-cyclic-nucleotide 3′-phosphodiesterase	4,242.9	7435185	60.0	1
77	Lyn B protein	56,033.3	2117805	4.3	1
78	KIAA0340	117,819.0	2224621	2.1	1
79	Hypothetical protein XP_091724	144,900.8	18588504	1.1	1
80	Hypothetical protein XP_091430	27,641.2	18586430	6.0	1*
81	Similar to tropomyosin 4	18,426.8	14729747	6.0	1*
82	HGTD-P	17,342.4	9295192	10.0	1*
83	Hypothetical protein XP_095819	291,206.1	18572484	0.6	1*
84	Far upstream element binding protein	67,534.4	1082624	2.5	1*
85	Hypothetical protein XP_103707	13,374.6	18551195	12.0	1*
86	Hypothetical protein XP_092517	41,409.4	18552304	2.6	1*
87	Enhancer protein	41,289.8	1345400	4.6	1*
88	Hypothetical protein	15,770.3	18551736	12.0	1*
89	KIAA1741 protein	123,305.7	12698027	1.7	1*
90	Ig heavy chain V-V region	10,995.4	87863	16.0	1*
91	DC 38	31,691.4	12005984	4.7	1*

*, The proteins found primarily in the low-ionic-strength spectrin extract from RBC membranes.

tion separately as tryptic fragments. For membrane subfractions, trypsin digests were collected from externally exposed proteins, internally exposed proteins, spectrin extract consisting mainly of membrane skeleton proteins, and soluble membrane proteins (missing the spectrin extract). Cytoplasmic proteins were divided into 21 subfractions based on molecular mass using size exclusion chromatography. Thus, the complexity of each sample analyzed was minimized and the abundant hemoglobins were separated from the majority of cytoplasmic proteins.

We were able to identify 181 protein sequences, which are organized in Tables I and II. Proteins found in the membrane fractions are presented in Table I with spectrin extract proteins indicated by an asterisk. Proteins found in the cytoplasmic fractions are listed in Table II. Data for each identified protein include protein description assigned by SEQUEST, protein identification number (Gi), percent of the covered amino acid sequence, number of identified peptides, and molecular mass. Ninety one unique sequences are listed in Table I for the plasma membrane fractions and 91 unique sequences are itemized in Table II for the cytoplasmic fractions. As expected, some proteins were found in more than one fraction. For example, spectrin subunits, protein 4.1, and tropomyosin 3 were found in the spectrin extract as well as in other membrane subfractions. The sequence coverage (number of identified peptides) for protein 4.1 or tropomyosin 3 was of comparable level in the spectrin extract and IOV fractions. As for spectrin subunits, the sequence coverage was much higher in the spectrin extract than in the other fractions. Hemoglobin subunits were found in both cytoplasmic and membrane fractions. However, the sequence coverage was much higher in the cytoplasmic fractions than in the membrane fractions. For this reason, hemoglobin subunits are listed only in Table II. Glyceraldehyde-3-phosphate dehydrogenase, on the other hand, is listed in both tables because coverage of its sequence was of comparable level in both the membrane and cytoplasmic fractions.

The number of peptides identified for the listed proteins

TABLE II
Proteins identified in RBC cytoplasmic fractions

No.	Protein description	Molecular mass (Da)	Gi Number	Sequence coverage (%)	No. of identified peptides
1	Catalase	59,756.2	4557014	58.0	24
2	Carbonic anhydrase II	29,246.1	4557395	74.0	14
3	Hemoglobin, β chain	15,867.2	122615	85.0	12
4	Carbonic anhydrase I	28,870.2	4502517	71.0	12
5	Hemoglobin δ	15,907.2	229172	70.0	11
6	Lactate dehydrogenase H chain	36,507.3	13786847	40.0	10
7	Similar to valosin-containing protein	89,337.8	15082580	19.0	10
8	Ubiquitin-activating enzyme	117,849.0	14758460	18.0	10
9	Similar to oxidized protein hydrolase	81,224.6	13640732	20.0	10
10	Hemoglobin α chain	15,126.4	122412	72.0	8
11	Glyceraldehyde-3-phosphate dehydrogenase	36,054.2	31645	45.0	8
12	Peroxiredoxin 2	21,891.9	13631440	44.0	8
13	Biliverdin reductase B	22,119.3	4502419	49.0	8
14	ATP citrate lyase	120,839.2	14774921	9.0	6
15	Aminolevulinate, δ-, dehydratase	37,229.1	12654313	27.0	6
16	Phosphoglucose isomerase chain A	63,015.9	14488680	18.0	6
17	Prostatic binding protein (neuropolypeptide h3	21,056.8	4505621	64.0	6
18	Aldehyde dehydrogenase 1A1	54,861.8	14745882	20.0	5
19	Purine nucleoside phosphorylase	32,117.9	14751808	29.0	5
20	Similar to suppression of tumorigenicity 13	41,264.6	17434275	18.0	5
21	Epsilon polypeptide, mono-oxygenase activation protein	29,173.9	5803225	27.0	5
22	Carbonic anhydrase III	29,571.4	115462	30.0	5
23	Cypa complexed with Hagpia, chain A	17,881.3	2981734	38.0	5
24	Heat shock protein HSP 90-α	84,673.7	123678	10.0	4
25	Heat shock 70 kDa protein 8	70,898.1	5729877	10.0	4
26	Nm23 protein	20,411.5	35068	32.0	4
27	Site specific mutant of carbonic anhydrase Ii, chain A	29,180.0	11514573	21.0	4
28	Hemoglobin γ-G chain	16,969.5	106221	21.0	3
29	Hydroxyacyl glutathione hydrolase	28,860.0	4885389	22.0	3
30	Polyubiquitin	68,497.5	2627129	42.0	3
31	Calpain inhibitor (Calpastatin)	76,484.7	1352419	12.0	3
32	Ubiquitin isopeptidase T	93,307.7	14784166	9.0	3
33	Transglutaminase 2	77,328.8	14770762	9.0	3
34	Proteasome subunit, β type, 4	29,204.2	16165126	21.0	2
35	DNA-damage inducible protein 2	46,571.0	33695074	8.0	2
36	OCP2	17,271.4	1017813	27.0	2
37	Similar to proteasome subunit, α type, 3	27,647.3	13528948	10.0	2
38	Embrionic Gower Ii carbonmonoxy hemoglobin F chain	16,071.6	3318951	16.0	2
39	2, 3-bisphosphoglycerate mutase	30,005.2	4502445	14.0	2
40	S100 calcium-binding protein A3	10,204.9	16159792	21.0	2
41	Horf6, a novel human peroxydase enzyme, chain A	25,034.7	3318841	15.0	2
42	Thioredoxin	11,737.5	14740408	23.0	2
43	Ubiquitin conjugating enzyme E2 variant 1	16,495.0	11323320	17.0	2
44	SH3 domain binding glutamic acid-rich like protein	12,801.4	4506925	20.0	2
45	Transgelin 2	22,406.4	12803567	16.0	2
46	Cofilin 2	18,736.6	14719392	17.0	2
47	RAD23 homolog A	39,609.2	4826964	8.0	1
48	Valosin-containing protein	89,321.8	6005942	3.3	1
49	Cationic trypsinogen	9,148.5	1616766	12.0	1
50	Proteasome subunit, β type, 3	22,948.9	11424309	7.0	1
51	Proteasome subunit, α type, 2	25,898.6	4506181	8.0	1
52	Similar to proteasome subunit α type 7	26,928.0	19263962	6.0	1
53	Hypothetical protein XP_093639	46,800.7	18555427	4.0	1
54	Placental ribonuclease inhibitor	1,530.6	251989	100.0	1
55	RuvB-like 2	51,156.6	5730023	3.5	1
56	Proteasome 26S subunit, ATPase, 6	44,160.9	4506215	3.6	1

TABLE II—*continued*

No.	Protein description	Molecular mass (Da)	Gi Number	Sequence coverage (%)	No. of identified peptides
57	Chaperonin containing TCP1, subunit 7	59,366.6	5453607	2.0	1
58	Lactate dehydrogenase A chain	36,721.6	5031857	3.0	1
59	Proteasome subunit, β type, 2	22,836.3	4506195	6.0	1
60	Proteasome subunit β type 5 precursor	22,896.9	1172607	4.8	1
61	TIP120 protein	136,361.7	8924260	1.1	1
62	KIHUA adenylate kinase	21,704.9	66932	6.2	1
63	Bleomycin hydrolase, chain A	52,275.9	7245509	2.4	1
64	Complexed carbonic anhydrase I	28,739.0	515109	4.6	1
65	Hypothetical protein XP_089225	59,627.6	18548268	3.5	1
66	D-dopachrome tautomerase	12,711.8	4503291	9.0	1
67	ADP-ribosylation factor 3	20,600.8	4502203	4.4	1
68	UEV1Bs	17,083.6	2689608	11.0	1
69	Peptidylprolyl isomerase A	18,012.5	10863927	7.0	1
70	SH3BGRL3-like protein	10,437.7	13775198	16.0	1
71	Adenylate kinase isoenzyme 1	21,676.9	125150	7.0	1
72	θ_1 globin	15,376.6	225891	8.0	1
73	Translation initiation factor 5A	16,832.2	4503545	15.0	1
74	Acid phosphatase isoenzyme Af	17,939.3	251370	8.0	1
75	Thioredoxin mutant with Cys-73 replaced by Ser	11,721.4	1827675	9.0	1
76	Similar to prostatic binding protein	21,266.0	18543899	8.0	1
77	Similar to ubiquitin conjugating enzyme E2L 3	14,134.2	18582039	18.0	1
78	Cofilin 1 (nonmuscle)	18,502.5	5031635	10.0	1
79	Similar to discs, large (*Drosophila*) homolog 3 (neuroendocrine-dlg)	90,314.0	18595882	1.7	1
80	Glutaredoxin	11,775.7	4504025	10.0	1
81	Similar to olfactory receptor-like protein F6	103,640.8	18583537	1.4	1
82	FK506-binding protein 1A	8,847.2	11182150	18.0	1
83	17 kDa cyclophilin A	3,199.4	1041969	6.2	1
84	Small nuclear ribonucleoprotein D3 polypeptide	13,916.2	4759160	12.0	1
85	Hypothetical protein XP_095686	71,925.3	18571465	1.7	1
86	Similar to glyceraldehyde-3-phosphate dehydrogenase	19,577.4	17475471	9.0	1
87	Carbonic anhydrase Ii	29,069.9	2098448	6.2	1
88	Early endosome antigen 1	145,099.7	13650577	0.6	1
89	Zinc finger protein 180	79,040.1	7019579	2.6	1
90	Diacylglycerol kinase, γ	89,123.9	11434723	1.3	1
91	FLJ00257 protein	93,314.7	18676716	1.8	1

varies from 1 to 77. Although abundant average-size proteins are identified by several peptides, very low-level proteins usually are detected by a single peptide (8). Thus, the numbers of identified peptides provides a semiquantitative estimate of relative amounts of the different proteins. In this regard, the following should be noted. Proteins are identified through a unique set of detected peptides, characteristic exclusively to that given protein. However, similar proteins (*e.g.* isoforms) contain regions of identical sequence and may produce a number of identical peptides. The analytical approach used can not distinguish identical peptides originated from different proteins. A peptide detected in an analyzed mixture is assigned to all potential parent proteins found in the same mixture. As a result, the number of peptides identified for the similar proteins, especially for ones of low abundance, may be overestimated. If a set of identified peptides could originate from more than one protein, we list all potential parent proteins of human origin (Table I, positions 37 and 72). Furthermore, 93 of the proteins identified in this article are based on single peptide assignment, which should be regarded as tentative.

We grouped the 181 identified proteins into different categories as summarized in Fig. 1. Number and percent of proteins included in each category as well as protein positions in Tables I and II are presented in Table III. Proteins that are described as similar to X (example similar to tropomyosin 4) are included under unknown proteins in Table III and Fig. 1.

The two largest groups of identified RBC proteins are membrane skeletal proteins and metabolic enzymes (Fig. 1). Proteins listed in category 5 in Table III (band 7.2b and flotillins) most likely act as separate scaffolding components at the cytoplasmic face of erythrocyte lipid rafts (9). The globins represent the most abundant group of proteins specific to erythrocytes. Identified globins include α, β, γ-G, and δ chains, embryonic Gower Ii carbonmonoxy hemoglobin F chain, and θ_1 globin. It should be noted that θ_1 globin had not

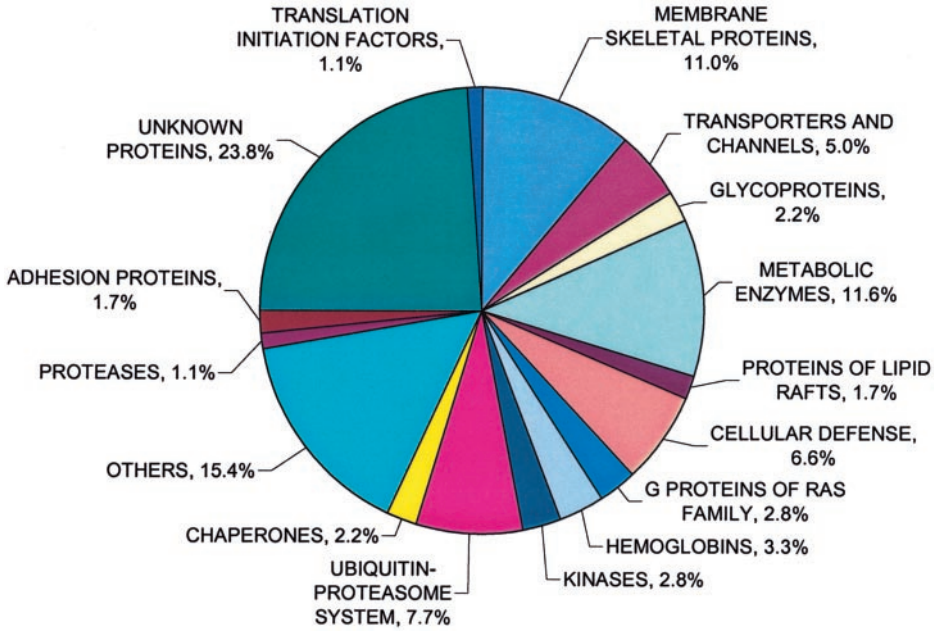

FIG. 1. **Functional categories of 181 identified RBC proteins.** The percent of proteins included in each category is indicated.

TABLE III

Functional categories of the identified RBC proteins listed in Tables I and II

No.	Protein groups	*n*[a]	%[b]	Protein positions in Table I	Protein positions in Table II
1	Membrane skeletal proteins	20	11.0	1–5, 7, 8, 10, 11, 15, 18, 21, 23–26, 30	45, 46, 78
2	Transporters and channels	9	5.0	9, 19, 22, 36, 38, 40, 42, 46, 63	
3	Glycoproteins	4	2.2	41, 43, 45, 58	
4	Adhesion proteins	3	1.7	35, 49, 51	
5	Proteins of lipid rafts	3	1.7	12, 14, 16	
6	G proteins of Ras family and related proteins	5	2.8	31, 37, 60, 61	67
7	Metabolic enzymes	21	11.6	17, 27, 29, 32–34, 64	2, 4, 6, 11, 14–16, 18, 19, 22, 27, 39, 58, 64, 87
8	Cellular defense proteins	12	6.6	44	1, 12, 13, 24, 25, 29, 41, 42, 70, 75, 80
9	Globins	6	3.3		3, 5, 10, 28, 38, 72
10	Ubiquitin-proteasome system	14	7.7		8, 30, 32, 34, 36, 43, 47, 50, 51, 56, 59–61, 68
11	Kinases	5	2.8	67	26, 62, 71, 90
12	Chaperones	4	2.2		23, 57, 69, 83
13	Proteases	2	1.1		49, 63
14	Translation initiation factors	2	1.1	28	73
15	Others	28	15.4	13, 39, 47, 50, 52, 54, 55, 59, 75–77, 84, 87, 90	17, 21, 31, 33, 40, 54, 55, 66, 74, 79, 82, 84, 86, 89
16	Unknown proteins	43	23.8	6, 20, 48, 53, 56, 57, 62, 65, 66, 68–74, 78–83, 85, 86, 88, 89, 91	7, 9, 20, 35, 37, 44, 48, 52, 53, 65, 76, 77, 81, 85, 86, 91

[a] Number of proteins in the group.

[b] Percent of total number of identified proteins.

previously been found in adult human erythroid or nonerythroid tissues. It should be also noted that though mutants of α hemoglobin (Gi 3212437) and β hemoglobin (Gi 1431650, 18418633) are not included in the tables, mutated peptides specific to these proteins were identified in the analyzed samples. As expected, transporters and channel proteins were found only in the membrane fractions (Table I). Eleven of the 12 identified cellular defense proteins were found in the cytoplasmic fractions. Approximately half of the proteins from these groups were identified with three or more peptides.

All 14 identified proteins of the ubiquitin-proteasome system, including six proteasome subunits, were found in the cytoplasmic fractions (Table II). Ubiquitin-activating enzyme was detected with highest sequence coverage (10 identified

peptides), followed by polyubiquitin and ubiquitin isopeptidase T with three identified peptides each. All but two of eight identified proteasome subunits were detected with a single peptide.

The category of unknown proteins includes hypothetical proteins whose existence was predicted from the genome sequence; proteins whose existence was shown only at the transcriptome level of different human cell types where the sequences were deduced from cDNAs; and unknown proteins showing similarity to the sequences of known proteins. The majority of the unknown proteins listed in both tables were determined by a single peptide. The primary reasons why some proteins are difficult to identify are low level of expression or extensive post-translational modifications. Nevertheless, we were able to specify 43 unknown proteins, which represents a significant portion (~24%) of all the erythrocyte proteins described in Tables I and II.

DISCUSSION

The present study represents the first attempt at determining the complete proteome of the human erythrocyte. The human RBC provides an ideal model for proteomic analysis because it combines simplicity with physiologic significance. Although the RBC is simple (*e.g.* it is devoid of a transcriptome), the proteome of the human erythrocyte provides further insight into its physiological make-up, which can be applied to more metabolically complex cells. A significant insight into the metabolism of the RBC, provided by the present study, is the realization that the cell contains a large number of proteasomal proteins. While six of these proteasomal proteins were identified based on a single-peptide assignment, two (Table II, positions 34 and 37) were based on two-peptide assignment with 21 and 10% sequence coverage, respectively. Previous studies have shown that the human RBC contains many ubiquitinated proteins including spectrin (10), ankyrin (11), and band 3 (12). Spectrin has an E2/E3 ubiquitin-conjugating/ligating activity that targets itself (10) as well as ankyrin (11) and band 3 (12). Ubiquitination of spectrin down-regulates the spectrin-protein 4.1-actin interaction (13) and the spectrin-adducin-actin interaction (14). An earlier report (15) suggested that mature erythrocytes have no ubiquitin- and ATP-dependent protein degradation capacity due to the lack of proteasomes, although they do maintain significant levels of ubiquitin conjugates. The results of our proteomic analysis reveal the likelihood that RBCs do contain at least a remnant of proteasomes, which could maintain a low level of ubiquitin-proteasomal activity.

Using gel filtration, we separated the abundant hemoglobins from the majority of the cytoplasmic proteins to facilitate the detection of lower-abundance proteins. Hemoglobin peak fractions, however, would in addition contain a significant number of other cytoplasmic proteins with molecular mass close to that of hemoglobin. In these fractions, where relative abundance of hemoglobin would remain high, many proteins were most likely not detected. An approach that specifically subtracts hemoglobin (for example immunoaffinity) could further expand the analysis of the RBC proteome.

A recent study combining one- and two-dimensional electrophoresis with matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) MS identified 84 unique RBC membrane proteins (3). They made no attempt to study RBC cytoplasmic proteins. Interestingly, two major RBC glycoproteins, glycophorin A (~600,000 copies/RBC) and glycophorin C (~50,000 copies/RBC) (1, 2) were not identified in the recent study (3). Glycosylated transmembrane proteins are known to be underrepresented on one- and two-dimensional gels (16), which makes the μLC/MS/MS technique of great value when trying to obtain a complete proteome analysis.

We believe that our study provides a strong basis for analysis and interpretation of the physiological competence of the RBC and sets the stage for further protein expression and function-based activity profiling not only of normal healthy erythrocytes but also for RBC pathology as well. Indeed, a more thorough proteomic examination involving μLC/MS/MS combined with isoelectrofocusing-SDS-PAGE/MALDI-TOF approaches should afford a means to elucidate the human erythrocyte proteome in its entirety. In turn, a more general understanding and appreciation of the metabolic capability of the RBC and other cells will be realized.

Acknowledgments—We thank Jonathan Bernauer for his assistance on these studies.

* This work was supported by a National Institutes of Health Sickle Cell Center Grant (HL070588) Project 1, which was awarded to S. R. G. 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.

‡‡ To whom correspondence should be addressed. Department of Molecular and Cell Biology, Room 3.610, University of Texas at Dallas, 2601 N. Floyd Road, Richardson, TX 75083-0688. Tel.: 972-883-4872; Fax: 972-883-4871; E-mail: sgoodmn@utdallas.edu.

REFERENCES

1. Goodman, S. R., Krebs, K. E., Whitfield, C. F., Riederer, B. M., and Zagon, I. S. (1988) Spectrin and related molecules. *Crit. Rev. Biochem.* **23,** 171–234
2. Steck, T. L. (1974) The organization of proteins in the human red blood cell membrane. *J. Cell Biol.* **62,** 1–19
3. Low, T. Y., Seow, T. K., and Chung, M. C. M. (2002) Separation of human erythrocyte membrane associated proteins with one dimensional and two-dimensional gel eletrophoresis followed by identification with matrix assisted laser desorption/ionization-time of flight mass spectrometry. *Proteomics* **2,** 1229–1239
4. Karinch, A. M., Zimmer, W. E., and Goodman, S. R. (1990) The identification and sequence of the actin-binding domain of human red cell β-spectrin. *J. Biol. Chem.* **65,** 11833–11840
5. Bordier, C. (1981) Phase separation of integral membrane proteins in Triton X-114 solution. *J. Biol. Chem.* **256,** 1604–1607
6. Eng, J. K., McCormack, A. L., and Yates, J. R., 3rd. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989
7. Yates, J. R., 3rd, Eng, J. K., McCormack, A. L., and Schieltz, D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67,** 1426–1436

8. Ducret, A., Van Oostveen, I., Eng, J. K., Yates, J. R., 3rd, and Aebersold, R. (1998) High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Prot. Sci.* **7,** 706–719
9. Salser, U., and Prohaska R. (2001) Stomatin, flotillin-1, and flotillin-2 are major integral proteins of erythrocyte lipid rafts. *Blood* **97,** 1141–1143
10. Kakhniashvili, D. G., Chaudhary, T., Zimmer, W. E., Bencsath, F. A., Jardine, I., and Goodman, S. R. (2001) Erythrocyte spectrin is an E2 ubiquitin conjugating enzyme. *Biochemistry* **40,** 11630–11642
11. Chang, T. L., Cubillos, F. F., Kakhniashvili, D. G., and Goodman, S. R. (2003) Ankyrin is a target of spectrin's E2/E3 ubiquitin-conjugating/ligating activity. *Cell Mol. Biol.* **50,** 59–66
12. Chang, T. L., Cubillos, F. F., Kakhniashvili, D. G., and Goodman, S. R. (2004) Band 3 is a target protein of spectrin's E2/E3 activity: Implication for sickle cell disease and normal red blood cell aging. *Cell Mol. Biol.* **50,** 171–177
13. Ghatpande, S., and Goodman, S. R. (2003) Ubiquitination of spectrin regulates the erythrocyte spectrin-protein 4.1-actin ternary complex dissociation: Implications for the sickle cell membrane skeletons. *Cell Mol. Biol.* **50,** 67–74
14. Mishra, R., and Goodman, S. R. (2003) Ubiquitination of spectrin regulates the dissociation of spectrin-adducin-F-actin ternary complex *in vitro. Cell Mol. Biol.* **50,** 75–80
15. Pickart, C. M., and Vella, A. T. (1988) Levels of active ubiquitin carrier proteins decline during erythroid maturation. *J. Biol. Chem.* **263,** 12028–12035
16. Santoni, V., Molloy, M., and Rabilloud, T. (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21,** 1054–1070