Cell Surface Labeling and Mass Spectrometry Reveal Diversity of Cell Surface Markers and Signaling Molecules Expressed in Undifferentiated Mouse Embryonic Stem Cells*†

Kazuto Nunomura‡, Koji Nagano‡, Chiharu Itagaki‡, Masato Taoka§, Nobuko Okamura‡¶, Yoshio Yamauchi§, Sumio Sugano§/H20648, Nobuhiro Takahashi**, Tomonori Izumi‡, and Toshiaki Isobe‡§‡‡

Although interactions between cell surface proteins and extracellular ligands are key to initiating embryonic stem cell differentiation to specific cell lineages, the plasma membrane protein components of these cells are largely unknown. We describe here a group of proteins expressed on the surface of the undifferentiated mouse embryonic stem cell line D3. These proteins were identified using a combination of cell surface labeling with biotin, subcellular fractionation of plasma membranes, and mass spectrometry-based protein identification technology. From 965 unique peptides carrying biotin labels, we assigned 324 proteins including 235 proteins that have putative signal sequences and/or transmembrane segments. Receptors, transporters, and cell adhesion molecules were the major classes of proteins identified. Besides known cell surface markers of embryonic stem cells, such as alkaline phosphatase, the analysis identified 59 clusters of differentiation-related molecules and more than 80 components of multiple cell signaling pathways that are characteristic of a number of different cell lineages. We identified receptors for leukemia-inhibitory factor, interleukin 6, and bone morphogenetic protein, which play critical roles in the maintenance of undifferentiated mouse embryonic stem cells. We also identified receptors for growth factors/cytokines, such as fibroblast growth factor, platelet-derived growth factor, ephrin, Hedgehog, and Wnt, which transduce signals for cell differentiation and embryonic development. Finally we identified a variety of integrins, cell adhesion molecules, and matrix metalloproteases. These results suggest that D3 cells express diverse cell surface proteins that function to maintain pluripotency, enabling cells to respond to various external signals that initiate differentiation into a variety of cell types. Molecular & Cellular Proteomics 4: 1968–1976, 2005.

Embryonic stem (ES)1 cells are a unique type of cultured cells defined by two functional properties, self-renewal and pluripotency. In cultured mouse ES cells, the soluble cytokine leukemia-inhibitory factor (LIF) can support the undifferentiated state and promote self-renewal, whereas the formation of embryoid bodies followed by the addition of growth factors induces differentiation of the cells to specific fates (1–4). Interactions between cell surface proteins and soluble factors or insoluble ligands play important roles in regulating ES cell functions. However, the molecular mechanisms involved in these cellular processes remain unclear because we lack a thorough understanding of the properties and functions of ES cell surface proteins. The study of ES cell surface proteins is also attractive because some of these proteins can be used as non-destructive markers to characterize and/or isolate specific cell types. Thus, a large scale identification of ES cell surface proteins is key to understanding the regulation of ES cell function and to developing new research tools.

Recent advances in MS-based proteomics have enabled us to identify a large number of proteins from a variety of membrane preparations (5–7). However, it is difficult to isolate...
plasma membranes in a pure form because the membranes lose their specific structure upon cell lysis, and a typical plasma membrane-rich fraction prepared by ultracentrifugation is heavily contaminated with other membrane components (8). Therefore, several methods have been developed to obtain relatively homogeneous preparations of plasma membranes, including coating of intact cells with silica derivatives (9). In this study, we describe an alternative approach for large scale and selective identification of ES cell surface proteins. First cell surface proteins of intact cells were selectively labeled with the membrane-impermeable reagent biotin, and biotinylated plasma membrane proteins were then enriched via affinity capture using immobilized avidin (10–14). The biotinylated proteins can be separated by gel electrophoresis and identified by MS (10–13); alternatively then can be proteolytically digested and identified by a MS-based “shotgun” approach (14).

We describe here a strategy for large scale and selective identification of cell surface proteins using cell surface labeling coupled with high resolution two-dimensional (2D) LC-MS/MS (Fig. 1). The method allowed us to identify more than 200 cell surface proteins expressed in mouse ES cells with simultaneous identification of the sites of biotinylation on each protein molecule. Analysis of the identified proteins indicated that ES cells express a wide variety of cell surface markers and signaling molecules (such as receptors, transporters, cell adhesion molecules, etc.), which are characteristic not only of ES cells but also of differentiated cell types, such as hematopoietic or neural cells. Our previous study identified ~1,800 proteins expressed in undifferentiated ES cells, thereby revealing the diversity of the ES cell proteome (15).

**EXPERIMENTAL PROCEDURES**

**Cell Surface Labeling—** A mouse embryonic stem cell line, D3 (American Type Culture Collection, Manassas, VA), was maintained on 0.1% gelatin-coated tissue culture dishes in Dulbecco’s modified Eagle’s medium (Invitrogen) supplemented with 15% heat-inactivated FCS (JRH Biosciences, Lenexa, KS), 0.1 mM β-mercaptoethanol, 100 units/ml penicillin, 100 μg/ml streptomycin, and 1,000 units/ml recombinant mouse LIF (ESGRO-Chemicon International, Temecula, CA). Under our culture conditions, ~95% of the cells were found to be undifferentiated as monitored by staining with alkaline phosphatase and stage-specific embryonic antigen-1 (SSEA-1), which are cell surface markers for undifferentiated ES cells. D3 cells grown to ~80% confluency on 150-mm tissue culture dishes were first incubated in serum-free Dulbecco’s modified Eagle’s medium for 1 h, rinsed twice with ice-cold PBS (10 mM NaH2PO4/Na2HPO4, pH 7.4, 138 mM NaCl, 2.7 mM KCl) supplemented with 0.1 mM CaCl2, 1 mM MgCl2 (PBS+), and then incubated with 1 mg/ml EZ-Link™ sulfo-NHS-LC-biotin (Pierce) in PBS+ for 20 min at 4 °C with gentle agitation. After removal of the supernatant, residual sulfo-NHS-LC-biotin was quenched with 100 mM glycine in PBS+, and the cells were harvested using a plastic scraper.

**Preparation of Biotinylated Membrane Fractions—** Biotinylated D3 cells (~4.8 × 10⁷ cells) were washed twice with PBS+; suspended in 10 mM Hapes-NaOH, pH 7.5, 0.25 mM sucrose (8.5% (w/v)), and protease inhibitor mixture (Roche Diagnostics); and then lysed by nitrogen cavitation (at 800 p.s.i. on ice for 20 min) (16). The cell lysates were then centrifuged at 3,000 × g for 10 min to remove large cell debris and nuclei. The supernatant was layered on a discontinuous sucrose density gradient, containing layers of 15, 30, 45, and 60% sucrose (w/v) in 10 mM Hepes-NaOH, pH 7.5, and centrifuged at 100,000 × g for 17 h (17). Resultant fractions were analyzed by SDS-PAGE followed by Western blotting using alkaline phosphatase-conjugated avidin (Pierce) or organelle-specific antibodies (Organelle Sampler kit, BD Biosciences). Fractions containing biotinylated proteins (fractions 3 and 4) were then combined, diluted 4-fold with distilled water, and centrifuged at 120,000 × g for 2 h to obtain plasma membrane-rich pellets.

**Tryptic Digestion and Avidin Affinity Enrichment of Biotinylated Peptides—** The biotinylated membrane fractions were delipidated twice with cold acetone, dried, and solubilized with 8 mM urea in 400 mM NH4HCO3, pH 8.5. Protein samples were reduced with 2 mM dithiothreitol at room temperature for 30 min, alkylated with 2 mM iodoacetamide for 30 min, and diluted 4-fold with distilled water. Protein samples were then digested with Na-L-phenylalanine chloromethyl ketone-treated porcine trypsin (Promega, Madison, WI) at an enzyme:substrate ratio of 1:100 (w/w) at 37 °C for 16 h. Digestion was monitored by SDS-PAGE/Western blotting using alkaline phosphatase-conjugated avidin. Biotin-labeled peptides were enriched from the tryptic digest by avidin affinity chromatography. Approximately 1 mg of the digest was applied to a column packed with 1 ml of Immunopure immobilized monomeric avidin (Pierce) pre-treated with 30% CH3CN in 0.4% trifluoroacetic acid and equilibrated with 2 mM urea in 100 mM NH4HCO3, pH 8.5. After sequential washing with (i) 2 ml urea in 100 mM NH4HCO3, (ii) 2 ml urea in 100 mM NH4HCO3 containing 0.5 mM NaCl, (iii) 2 ml urea in 100 mM NH4HCO3 containing 30% CH3CN, and (iv) 100 mM NH4HCO3, the bound peptides were eluted with 30% CH3CN in 0.4% trifluoroacetic acid and concentrated using a vacuum concentrator. The amounts of peptides and biotin labels were determined using the BCA protein assay kit (Pierce) and 2-(4'-hydroxyazobenzene)-benzoic acid (Pierce), respectively, according to the manufacturer’s instructions.

**Automated Multidimensional LC-MS/MS Analysis—** The peptide mixtures were analyzed on an automated 2D LC-MS/MS system (15, 18–20) using a combination of first dimensional cation exchange and second dimensional reverse-phase chromatography. The peptide mixture (50 μg) was separated on a SP-SPW column (1-mm internal diameter × 40 mm long, 20-μm particles; Tosoh, Tokyo, Japan) by 15-min stepwise elution (20 μm acetate buffer, pH 4.0, containing 0, 25, 50, 100, or 400 mM NaCl) at a flow rate of 10 μl/min. Eluted peptides in each step were captured on a trap column (Mightysil C18-0.5 mm internal diameter × 1 mm long, 3-μm particles; Kanto Chemicals, Tokyo, Japan) for desalting and then separated on a Mightysil C18 column (0.15-mm internal diameter × 40 mm long, 3-μm particles; Kanto Chemicals) using a three-step linear gradient (0–30% CH3CN in 0.1% formic acid for 120 min, 30–70% CH3CN in 0.1% formic acid for 40 min, and 70% CH3CN in 0.1% formic acid for an additional 10 min) at a flow rate of 50 μl/min. The eluted peptides were directly sprayed into a high resolution quadrupole time-of-flight hybrid mass spectrometer (Q-ToF-2; Micromass, Manchester, UK). The total analysis time for a single 2D LC-MS/MS operation was 22.5 h.

**Protein Identification by Database Search—** The MS/MS signals were acquired by MassLynx (Micromass) and converted to text files by ProteinLynx software (Micromass). The database search was performed in triplicate with MASCOT (Matrix Science Ltd.,. London, UK) against the Refseq mouse, human, and rat sequence databases with the following parameters: fixed modification, carbamoylmethylation (Cys); variable modifications, oxidation (Met) and sulfo-NHS-LC-biotin (Lys); maximum missed cleavages, 3; peptide mass tolerance, 150 ppm; and MS/MS tolerance, 0.5 Da. For peptide and protein identification, the search results were processed as follows. (i) The candi-
date peptide sequences were screened with the probability-based MOWSE scores that exceeded their thresholds ($p < 0.05$) and with MS/MS signals for y- or b-ions $>3$. (ii) Redundant peptide sequences were removed. (iii) Each peptide sequence was assigned to a protein that gave the maximal number of peptide assignments among the candidates. (iv) The mouse, human, and rat datasets were combined. (v) Interspecies redundancy of proteins was removed. Details of the methods of protein identification and data processing are described elsewhere (21). In this study, the following additional criteria were applied during visual inspection of individual MS/MS spectra for reliable identification of labeled peptide sequences: (vi) the presence of an MS/MS signal corresponding to the labeled lysine residue and/or (vii) the presence of one of fragment ions ($M^+ = 227.1$ and $340.2$) derived from the labeling reagent (14). Peptide sequences identified without biotin modifications ($<10\%$) were excluded regardless of their scores.

**Cell Staining**—D3 cells grown on gelatin-coated coverglass were labeled as described above in the presence or absence of sulfo-NHS-LC-biotin. The labeled cells were washed with PBS; fixed with 4% paraformaldehyde for 20 min, treated with 0.5% (v/v) Triton X-100 for 10 min, and then treated with 3% BSA for 1 h. After being washed, cells were incubated with 20 $\mu$g/ml of FITC-conjugated avidin (Pierce) for 30 min. The cells were co-stained with 2 $\mu$g/ml of Hoechst 33342 (Invitrogen) and observed under a laser scanning confocal microscope (Radiance 2100, Bio-Rad).

**RNA Extraction, cDNA Synthesis, and Plasmid Construction**—Total RNA was extracted from D3 cells using the RNeasy minikit (Qiagen, GmbH, Germany). cDNA was synthesized from the total RNA using the Advantage RT-for-PCR kit (BD Biosciences) and random hexamers as primers. PCR amplification was performed using Advantage 2 Polymerase Mix (BD Biosciences). Primer pairs used for the amplification of individual fragments were as follows: RIKEN cDNA 430119L13, 5’-cccCCCGGAtgcgtgagctctcgatgc-3’ and 5’-cccAAAGTCTTcctacagttactttc-3’; trophoblast glycoprotein, 5’-cgGGATCCatggtctc-3’; and hypothetical protein D7Ertd458e, 5’-ggGAATTCttggaaaggctggtctgga-3’; glycoprotein A33, 5’-CGGAATTCgtgctgtttggctccc-3’; and hypothetical protein D7Ertd458e, 5’-ggGAATTCgtgctgtttggctccc-3’.

PCR-amplified DNA fragments were treated with appropriate restriction enzymes (RIKEN cDNA 430119L13, Smal-HindIII; trophoblast glycoprotein, BamHi-EcoRI; glycoprotein A33, BamHi-EcoRI; hypothetical protein D7Ertd458e, SacII-EcoRI), subcloned into the FLAG tag expression vector (pCMV-tag4A, Invitrogen), and verified by nucleotide sequencing.

**Transfection of D3 Cells and Immunofluorescence Staining**—D3 cells were transiently transfected by seeding cells on collagen-coated glass slides at a density of $6 \times 10^5$ cells/ml 1 day before transfection. Constructs to express cell surface proteins (0.5 $\mu$g of each) were transfected into cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer’s recommendations and then incubated at 37 °C for 24 h. After washing with PBS, cells were incubated in PBS containing 3.7% formaldehyde for 30 min at room temperature. Fixed cells were washed with PBS and then permeabilized by incubating in 0.2% (v/v) Triton X-100 at room temperature for 6 min. Cells were washed with PBS and blocked with PBS containing 1% BSA and 5% FCS for 30 min. A primary monoclonal antibody against FLAG (Sigma) was applied directly to coverslides at 1:1,000 dilution in PBS; the goat anti-mouse secondary antibody labeled with FITC (Invitrogen) was diluted 1:1,000. Hoechst 33342 was used at 2 $\mu$g/ml to detect nuclei. Cells were visualized with a Radiance 2100 confocal microscope (Bio-Rad).

**Characterization of Identified Proteins**—Protein annotations, such as subcellular localization and molecular function, were obtained using Expression Analysis Systematic Explorer (EASE) version 2.0.

**RESULTS**

**Cell Surface Labeling and Preparation of Plasma Membrane**

The method presented here consisted of (i) in situ biotinylation of surface proteins on intact cells using the membrane-impermeable reagent sulfo-NHS-LC-biotin, (ii) cell lysis and subcellular fractionation of the biotinylated membranes on sucrose gradients, (iii) tryptic digestion of the protein mixture, (iv) affinity capture of the biotinylated peptides with avidin, and (v) 2D LC-MS/MS analysis of the peptide mixture (Fig. 1). First D3 cells grown on culture dishes were labeled in situ with sulfo-NHS-LC-biotin. The majority (>95%) of D3 cells displayed clear outlines as visualized by FITC-conjugated avidin, suggesting that the cell surface was selectively biotinylated with sulfo-NHS-LC-biotin (Fig. 2A). Diffuse staining, however, was also observed in cells inside the colony, suggesting that cytoplasmic labeling occurred in non-viable cells (Fig. 2A). Although scraping cells prior to labeling increased the labeling efficiency, it caused undesirable labeling of cytoplasmic proteins (data not shown). After lysis of the labeled cells by nitrogen cavitation and centrifugation to remove large debris, the subcellular components in the cell lysate were fractionated on a discontinuous sucrose density gradient. Western blotting analyses of the fractions revealed that biotinylated proteins were recovered in fractions 3 ($d = 1.14$ g/ml) and 4 ($d = 1.11$ g/ml) exclusively (Fig. 2B). To characterize the contents of these subcellular fractions, all fractions were analyzed by Western blotting with antibodies against known molecular markers for several organelles: annexin II for the plasma membrane, GM130 for the Golgi apparatus, Bip/GRP78 for the endoplasmic reticulum, and nucleoporin p62 for the nucleus. The distribution of these molecular markers indicated that fractions 3 and 4 were enriched in plasma membrane as they were labeled with the annexin II antibody but showed little staining with the Bip/GRP78 and nucleoporin.
antibodies (Fig. 2B). These fractions were combined and digested to generate a tryptic peptide mixture for MS identification of cell surface proteins.

**Protein Identification by 2D LC-MS/MS**

The automated 2D LC-MS/MS analysis of the avidin-purified biotin-labeled peptide mixture (50 μg) generated 5,871 MS/MS spectra in a single 22.5-h analysis. These spectra were assigned to 608 unique peptides by the sequential MASCOT search of Refseq mouse, human, and rat sequence databases. Of those, 551 peptides were biotinylated. A typical MS/MS spectrum assigned to a peptide from the transferrin receptor is shown in Fig. 3. Subjecting biotinylated peptides to MS generated fragment ions at m/z = 227.1 and 340.2 due to collision-induced dissociation of the labeling reagent as indicated by arrows in Fig. 3. Thus, all MS/MS spectra were carefully inspected, and peptide assignments without biotin labels were excluded. The 551 peptides were attributed to 240 unique proteins (2.3 biotin-labeled peptides/protein on average). Finally the 2D LC-MS/MS analysis was performed twice with different peptide preparations from two different cell labeling experiments, yielding a composite subproteome consisting of 324 proteins derived from 965 unique biotinylated peptide sequences (Supplemental Table 1). The number of peptides used to identify a single protein ranged from one to 37 with an average of 3.0 peptides/protein, and 151 proteins (47%) were identified by multiple peptide assignments.

**Enrichment of Cell Surface Proteins by the Protein Labeling Reaction**

The analysis of identified proteins by the SOSUI program predicted transmembrane (TM) segments for 84 proteins and 68 cell surface proteins. The 2D LC-MS/MS analysis of the avidin-purified biotin-labeled peptide mixture (50 μg) generated 5,871 MS/MS spectra in a single 22.5-h analysis. These spectra were assigned to 608 unique peptides by the sequential MASCOT search of Refseq mouse, human, and rat sequence databases. Of those, 551 peptides were biotinylated. A typical MS/MS spectrum assigned to a peptide from the transferrin receptor is shown in Fig. 3. Subjecting biotinylated peptides to MS generated fragment ions at m/z = 227.1 and 340.2 due to collision-induced dissociation of the labeling reagent as indicated by arrows in Fig. 3. Thus, all MS/MS spectra were carefully inspected, and peptide assignments without biotin labels were excluded. The 551 peptides were attributed to 240 unique proteins (2.3 biotin-labeled peptides/protein on average). Finally the 2D LC-MS/MS analysis was performed twice with different peptide preparations from two different cell labeling experiments, yielding a composite subproteome consisting of 324 proteins derived from 965 unique biotinylated peptide sequences (Supplemental Table 1). The number of peptides used to identify a single protein ranged from one to 37 with an average of 3.0 peptides/protein, and 151 proteins (47%) were identified by multiple peptide assignments.

**Enrichment of Cell Surface Proteins by the Protein Labeling Reaction**

The analysis of identified proteins by the SOSUI program predicted transmembrane (TM) segments for 84 proteins and 68 cell surface proteins. The 2D LC-MS/MS analysis of the avidin-purified biotin-labeled peptide mixture (50 μg) generated 5,871 MS/MS spectra in a single 22.5-h analysis. These spectra were assigned to 608 unique peptides by the sequential MASCOT search of Refseq mouse, human, and rat sequence databases. Of those, 551 peptides were biotinylated. A typical MS/MS spectrum assigned to a peptide from the transferrin receptor is shown in Fig. 3. Subjecting biotinylated peptides to MS generated fragment ions at m/z = 227.1 and 340.2 due to collision-induced dissociation of the labeling reagent as indicated by arrows in Fig. 3. Thus, all MS/MS spectra were carefully inspected, and peptide assignments without biotin labels were excluded. The 551 peptides were attributed to 240 unique proteins (2.3 biotin-labeled peptides/protein on average). Finally the 2D LC-MS/MS analysis was performed twice with different peptide preparations from two different cell labeling experiments, yielding a composite subproteome consisting of 324 proteins derived from 965 unique biotinylated peptide sequences (Supplemental Table 1). The number of peptides used to identify a single protein ranged from one to 37 with an average of 3.0 peptides/protein, and 151 proteins (47%) were identified by multiple peptide assignments.
both signal sequences and TM segments for 116 proteins (Fig. 4A). Thus, a total of 200 proteins among 324 identified proteins (62%) had molecular characteristics typical of integral membrane proteins. The number of TM domains in the molecules ranged from one (122 proteins) to 13 (four proteins) (Supplemental Table 1). Because 20–30% of all open reading frames encoded by the genome have been predicted to be integral membrane proteins (23, 24), the method presented here concentrated potential TM proteins efficiently. By including the 35 proteins with signal sequences but no detectable TM domains suggestive of a secreted protein, potential cell surface membrane and secreted proteins accounted for ~73% of total identified proteins.

Cell Surface Selectivity of Protein Labeling Reaction

Because sulfo-NHS-LC-biotin is impermeable to the plasma membrane and labels primary amines in proteins mainly at the ε-amino group of lysine residues exposed to the extracellular space, the labeled peptides should reside in the extracellular domain of a transmembrane protein. However, if non-selective labeling reactions took place, labeled peptides might be found on both sides of the TM domain. Thus, we examined the biotinylation sites of the 122 single TM domain proteins found in this study. On the assumption that the SOSUI program predicts TM segments correctly, none of these proteins were biotinylated on both sides of the TM segment. Namely among the 122 proteins assigned from 522 unique biotinylated peptides, 98 proteins (80%) had biotinylated lysine(s) only on the amino-terminal side of the predicted TM segment and were thereby defined as type I transmembrane proteins (Fig. 4B). Conversely the remaining 24 proteins (20%) contained biotinylated lysine(s) on the carboxyl-terminal side of the predicted TM domain and were defined as type II transmembrane proteins (Fig. 4B). Furthermore most of the amino-terminally labeled TM proteins (87 proteins) had potential signal sequences at the amino terminus of the polypeptide as expected for a type I transmembrane protein (Supplemental Table 1). Thus, the results suggest that the labeling reaction took place rather specifically at the surface of D3 cells and that the labeled peptides represented the extracellular domains of cell surface transmembrane proteins.

To further validate the present procedure, the subcellular localization of several proteins with no annotated localization in the Gene Ontology (GO) Database was studied by transiently expressing individual FLAG-tagged proteins in D3 cells. We selected RIKEN cDNA B430119L13, a trophoblast plasma membrane glycoprotein, glycoprotein A33, and the hypothetical protein D7Erd458e, all of which have a single predicted transmembrane domain and are present in pre- and/or postimplantation embryos or in neoplastic tissues. Immunofluorescence staining of transfected D3 cells revealed that all four proteins were localized extensively on the plasma membrane as judged by co-localization with CD9, a cell surface marker of ES cells (Fig. 5).

Characterization of ES Cell Surface Proteins

Functional Classification—The molecular functions of the 324 proteins identified in this study were classified according to the GO database and literature surveys (Fig. 6). Receptors were the largest subgroup, consisting of 63 proteins (19% of identified proteins) with cellular abundance ranging from several hundred copies to hundreds of thousands of copies. The
identified proteins (33 proteins) had no annotated functions.

Approximately 10% of these proteins had various functions such as nucleic acid metabolism, protein chaperones (eight proteins) as well as "other" 56 proteins with structural molecules (15 proteins), histones (10 proteins), and ribosomal constituents (21 proteins), hypothetical protein D7Erd458e, and empty vector. Cells were stained with anti-FLAG, anti-CD9 (an ES cell plasma membrane marker), and Hoechst 33342.

The classification is according to the GO database (Expression Analysis Systematic Explorer version 2.0) and literature surveys.

CD Antigens—CD molecules are protein or carbohydrate antigens recognized by specific antibodies, and they are utilized to distinguish particular cell types. Of the ~300 proteinaceous CD molecules that have been identified in a variety of cell types and are listed in Protein Reviews On The Web (mpr.nci.nih.gov/prow/), we found 59 molecules on the surface of D3 cells (Fig. 7A). We found CD9, which is down-regulated upon differentiation of ES cells (25); CD130 (gp130) and CD222, which have potential roles in embryo development (26, 27); and CD87 and CD146 (melanoma cell adhesion molecule), which are associated with cancer development (28, 29) (Fig. 7A). In addition to ubiquitous cell surface molecules such as CD55 and CD71, we found a number of CD molecules that serve as markers for neural cells (CD56; neural cell adhesion molecule), hematopoietic stem cells (CD117; c-Kit), epithelial cells (CD142; coagulation factor III), and endothelial cells (CD31; platelet endothelial cell adhesion molecule) (Fig. 7A). Interestingly immunohistochemistry demonstrated that D3 cells expressing CD9 and SSEA-1, which have been used as cell surface markers for undifferentiated ES cells, co-expressed some of these CD markers of differentiated cells (Fig. 7B).

Cell Signaling Molecules—This study identified 82 proteins that have potential roles in a wide variety of cell signaling pathways, including 59 cell surface receptors and adhesion molecules, eight ligands, and 15 signal modulators in D3 cells (Table I). Among these components, we found Lifr and its co-receptor, interleukin 6 receptor (Il6st), which, together with ciliary neurotrophic factor receptor (Cnfr), transduce LIF signals required for the maintenance of undifferentiated mouse ES cells (26). We also identified BMP receptor type 1A (Bmpr1a), a component that transduces BMP-4 signals to sustain LIF-mediated self-renewal of mouse ES cells by inducing the expression of Id protein in serum-free cultures (30), and integrin α6β1, the major laminin receptor on mouse ES cells that promotes self-renewal in both mouse and human ES cells (31, 32). Besides these components, the D3 cell line expressed receptors for PDGF (Pdgfra), VEGF (Flt4; Nrp2), and FGFR (Fgfr2), which direct the differentiation of ES cells to neural, hematopoietic, endothelial, epidermal, or neural cell lineages (33). We also found many cell signaling molecules that have roles in embryonic development, such as Notch (Notch3) and receptors for transforming growth factor β (Eng), Wnt (Fzd2 and Fzd10), Hedgehog (Pch), and ephrin (Epha1, Epha2, Epha4, and Ephb4) and their proteinaceous ligand (Efnb1). The other receptors for growth factors or cytokines we identified were Erbb2, interferon α and β receptor 1 (Ifnar1), IGFR 1 and 2 receptors (IGF1R and Igf2r), and IL1 receptor (Il1rap). Erbb2 is a co-receptor of epidermal growth factor that is up-regulated in certain types of cancer cells (34), whereas IGF1R plays a role in mesenchymal stem cell differentiation into adipocytes (35). This study also identified many...
cell adhesion molecules that participate in cell-cell and cell-
matrix interactions and integrins that transduce extracellular
matrix signals to the cell together with their potential ligands
(COL4A1 and FN1, etc.) and modulators, including various
members of the ADAM family and other metalloproteases
(Adam 9, 10, 17, and 23; Mme; and Mmp14).

**TABLE I**

| Signaling pathway | Receptor or adhesion molecules identified | Ligand identified | Modulator identified |
|-------------------|-----------------------------------------|------------------|----------------------|
| EGF family | Erbb2 | Tdgf1 |
| Ephrin | Epha1, Epha2, Epha4, Ephb4 | Efnb1 |
| FGF | Fgfr2 | Il17rd |
| Hedgehog | Ptcn | Disp1 |
| Interferon (IFN) | Iftn1 | |
| IGF | IGF1R, Igl2r | |
| IL1 | Il1rap | |
| IL6 | Il6st, Lilr, Cntfr | |
| Insulin | Insr | |
| Integrin | Itga3, Itgax, Itga5, Itgaa, Itgav, ITGB1, Itgb3, Itgb5 | Col18a1, COL4A1, Lama1, FN1, FLRT3 | Cd9, Cd47, Cyr61, Nid1, Nid2, Adam9, Adam23, Mme, Mmp14 |
| Mdkine | Notch3 | |
| Notch | Pdgfra | |
| PDGF | Kit | |
| Stem cell factor | Bmpr1a, Eng | |
| TGF-β/BMP superfamily | Tlr2 | |
| Toll-like receptor (TLR) | Tnf11 | |
| Tumor necrosis factor (TNF) | Flt4, NRP2 | |
| VEGF | Fzd2, Fzd10 | |
| Wnt | Celsr1, Celsr3, Cdx2, Gpr56, LPHN1, LPHN2, Pthr1 | |
| Cell-cell adhesion | Ptk7, Ror2 | |
| G-protein-coupled receptor | Ptpn1, Ptprg, Ptpr, Ptpk | |
| Other pathways | Podxl, Unc5h2, Amfr | |

**DISCUSSION**

The strategy described here illustrates that the combination
of cell surface labeling with sulfo-NHS-LC-biotin, membrane
fractionation, isolation of peptides with immobilized avidin,
and large scale 2D LC-MS/MS analysis of the resulting bioti-
nylated peptides is a powerful means of characterizing the cell
surface subproteome. We were able to identify 200 known or putative cell surface proteins using this approach. The study also identified minor cellular components such as cytokine/growth factor receptors, ErbB2, Fgfr2, Fli4, and the LIF receptor, Lifr. Lif has been reported to be present at only 250–300 copies/cell (36), suggesting that this labeling and isolation method enabled us to identify relatively low abundance plasma membrane proteins. In addition, the MS-based identification of labeled peptides allowed the determination of the sites of biotinylation in each protein. When biotinylation can be shown to specifically label only the amino-terminal or only carboxyl-terminal side of the potential TM domain, the site of biotinylation can be used to assign the putative TM protein as a type I or type II TM protein. We identified 122 proteins for which we could make such assignments (Fig. 4). Furthermore four proteins of unknown function and localization identified in this study localized to the plasma membrane when the FLAG-tagged versions of these proteins were expressed in D3 cells (Fig. 5).

In early attempts, we prepared biotinylated peptides by direct tryptic digestion of total cell lysates of biotin-labeled D3 cells without prior subcellular fractionation. Although the LC-MS/MS analysis of this peptide preparation identified a comparable number of biotinylated peptides and ~50% of identified proteins had potential signal sequences and/or TM domains, many peptides were derived from a number of highly abundant intracellular proteins, probably originating from non-viable cells (data not shown). To minimize contamination that interfered with the selective and sensitive identification of low abundance cell surface proteins, we incorporated subcellular fractionation by sucrose density centrifugation prior to the tryptic digestion of the protein samples.

In a previous study, we identified 1,790 proteins expressed in mouse ES cells by applying the MS-based protein identification technology to tryptic digests of whole-cell lysates (15). The identified proteins included many housekeeping proteins found in common with other cell types as well as a group of proteins unique to ES cell function, such as the transcription factors Oct-3/4 and Sox-2. The set of previously identified proteins also contained 260 potential membrane proteins, including cell surface antigens such as CD9 and CD81. This represented 15% of the total proteins identified. Eighty-two of these were again identified in this study. Thus, 118 of the 200 potential cell surface proteins identified in the present study of utilizing specific cell surface labeling coupled with MS had not been previously identified by our previous study (Supplemental Table 1).

However, our labeling and isolation strategy does not yield plasma membrane proteins completely free of contamination. First, intracellular proteins from non-viable cells (usually ~5% of the cell culture) were labeled with the biotinylating reagent. In this study, about one-fourth of the identified proteins had no potential signal sequences or TM segments (Fig. 4A) and contained abundant housekeeping proteins, such as ribosomal constituents, structural molecules, histones, and chaperones (Fig. 6). Although some of these proteins might potentially be cell surface components (37), it was still difficult to distinguish them from the intracellular components of non-viable cells that might be labeled and retained in the membrane fractions. Second, the biotinylating reagent might fail to label proteins that have few reactive lysine residues, small extracellular regions, or many post-translational modifications, such as glycosylation.

To our knowledge, this study identified the largest subset of CD molecules on the surface of a single cell line to date. These CD molecules, as well as many cell surface proteins identified in this study, suggest that the D3 cell line expresses a wide variety of proteins on the plasma membrane. In our previous analysis of the ES cell proteome, we also found that these cells express a number of ES-specific proteins, such as Oct-3/4 and UTF-1, as well as many cell signaling molecules that are characteristic of differentiated cell lineages, such as hematopoietic and neural cells (15). Although we could not exclude the possibility that a small portion of cells were differentiated to a variety of cell lineages during culture, our present and previous results imply that ES cells express multiple proteins considered unique to a number of differentiated cell lineages, enabling cells to respond to a variety of stimuli leading to differentiation to different specific cell lineages. Indeed our present study shows that a number of cell surface markers for differentiated cells are co-expressed with known markers of undifferentiated ES cells in D3 cells (Fig. 7). Future applications of this method in combination with quantitative proteomic approaches, such as stable isotope labeling, should identify stage- and lineage-specific expression of ES cell surface proteins and provide a catalogue of candidates for molecular markers of ES cells and of potential targets for controlled differentiation of ES cells in tissue engineering.

Acknowledgments—We thank Drs. N. Yoshida and M. Satoh for advice in ES cell culture and S. Hamada for data analyses.

* This work was supported in part by a grant from Pioneer Research on Genome the Frontier, Ministry of Education, Culture, and Sports of Japan and by a grant-in-aid for Research for the Future Program from the Japan Society for the Promotion of Science. 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: Division of Proteomics Research, Inst. of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. E-mail: isobe-toshiaki@cc.metro-u.ac.jp.

REFERENCES

1. Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Rogers, D. (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 336, 688–690
2. Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L.,
ES Cell Surface Proteins

Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A., and Gough, N. M. (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336, 684–687

3. Rathjen, J., and Rathjen, P. D. (2001) Mouse ES cells: experimental exploitation of pluripotent differentiation potential. Curr. Opin. Genet. Dev. 11, 587–594

4. Donovan, P. J., and Gearhart, J. (2001) The end of the beginning for pluripotent stem cells. Nature 414, 92–97

5. Wu, C. C., and Yates, J. R., III (2003) The application of mass spectrometry to membrane proteomes. Nat. Biotechnol. 21, 262–267

6. Foster, L. J., De Hoog, C. L., and Mann, M. (2003) Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. Proc. Natl. Acad. Sci. U. S. A. 100, 5813–5818

7. Schirmer, E. C., Florens, L., Quan, T., Yates, J. R., III, and Gerace, L. (2003) Nuclear membrane proteins with potential disease links found by subtractive proteomics. Science 303, 1380–1382

8. Pasquall, C., Fialka, I., and Huber, L. A. (1999) Subcellular fractionation, electromigration analysis and mapping of organelles. J. Chromatogr. B Biomed. Sci. Appl. 722, 89–102

9. Durr, E., Yu, J., Krasinska, K. M., Carver, L. A., Yates, J. E., Oh, P., and Schnitzer, J. E. (2004) Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. Nat. Biotechnol. 22, 985–992

10. Zhao, Y., Zhang, W., White, M. A., and Zhao, Y. (2003) Capillary high-performance liquid chromatography/mass spectrometric analysis of proteins from affinity-purified plasma membrane. Anal. Chem. 75, 3751–3757

11. Zhang, W., Zhou, G., Zhao, Y., White, M. A., and Zhao, Y. (2003) Affinity enrichment of plasma membrane for proteomics analysis. Electrophoresis 24, 2855–2863

12. Sabath, N., Lamer, S., Zimny-Arndt, U., Jungblut, P. R., Meyer, T. F., and Umann, D. (2002) Identification of surface proteins of Helicobacter pylori by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis. J. Biol. Chem. 277, 27986–27992

13. Zhao, Y., Zhang, W., Kyo, H., and Zhao, Y. (2004) Proteomic analysis of integral plasma membrane proteins. Anal. Chem. 76, 1817–1823

14. Chen, W. N., Yu, L. R., Strittmatter, E. F., Thrall, B. D., Camp, D. G., II, and Smith, R. D. (2003) Detection of in situ labeled cell surface proteins by mass spectrometry: application to the membrane subproteome of human mammary epithelial cells. Proteomics 3, 1647–1651

15. Naganu, K., Taoka, M., Yamauchi, Y., Itagaki, C., Shinkawa, T., Nunomura, K., Okamura, N., Takahashi, N., Izumi, T., and Isobe, T. (2005) Large-scale identification of proteins expressed in mouse embryonic stem cells. Proteomics 5, 1346–1361

16. Watarai, H., Inagaki, Y., Kubota, N., Fuji, K., Nagafu, J., Yamaguchi, Y., and Kadoya, T. (2000) Proteomic approach to the identification of cell membrane proteins. Electrophoresis 21, 460–464

17. Adam, P. J., Boyd, R., Tyson, K. L., Fletcher, G. C., Stamps, A., Hudson, L., Poyser, H. R., Redpath, N., Griffiths, M., Steers, G., Harris, A. L., Patel, S., Berry, J., Loader, J. A., Townsend, R. R., Daviet, L., Legrain, P., Parek, R., and Terrett, J. A. (2003) Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. J. Biol. Chem. 278, 6482–6489

18. Maxuvenyega, K. G., Kaji, H., Yamauchi, Y., Shinkawa, T., Saito, H., Taoka, M., Takahashi, N., and Isobe, T. (2003) Large-scale identification of Caenorhabditis elegans proteins by multidimensional liquid chromatography-tandem mass spectrometry. J. Proteome Res. 2, 23–35

19. Taoka, M., Yamauchi, Y., Shinkawa, T., Kaji, H., Motohashi, W., Nakayama, H., Takahashi, N., and Isobe, T. (2004) Only a small subset of the horizontally transferred chromosomal genes in Escherichia coli are translated into proteins. Mol. Cell. Proteomics 3, 780–787

20. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., and Isobe, T. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. Nat. Biotechnol. 21, 667–672

21. Shinkawa, T., Taoka, M., Yamauchi, Y., Ichimura, T., Kaji, H., Takahashi, N., and Isobe, T. (2005) STEM: a software tool for large-scale proteomic data analyses. J. Proteome Res. 4, 1826–1831

22. Hossack, D. A., Dennis, G. J., Sherman, B. T., Lane, H. C., and Lempicki, R. A. (2003) Identifying biological themes within lists of genes with EASE. Genome Biol. 4, R70

23. Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562

24. The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420, 563–573

25. Oka, M., Tagoku, K., Russell, T. L., Nakano, Y., Hamazaki, T., Meyer, E. M., Yokota, T., and Terada, N. (2002) CD9 is associated with leukemia inhibitory factor-mediated maintenance of embryonic stem cells. Mol. Biol. Cell 13, 1274–1281

26. Taga, T., and Kishimoto, T. (1997) Gp130 and the interleukin-6 family of cytokines. Annu. Rev. Immunol. 15, 797–819

27. Wang, Z. O., Fung, M. R., Barlow, D. P., and Wagner, E. F. (1994) Regulation of embryonic growth and lysosomal targeting by the imprinted Igf2/Mpr gene. Nature 372, 464–467

28. Luther, T., Magdolen, V., Albrecht, S., Kasper, M., Riemer, C., Kessler, H., Graef, H., Muller, M., and Schmitt, M. (1997) Epitope-mapped monoclonal antibodies as tools for functional and morphological analyses of the human urokinase receptor in tumor tissue. Am. J. Pathol. 150, 1231–1244

29. Lehmann, J. M., Riethmüller, G., and Johnson, J. P. (1989) MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily. Proc. Natl. Acad. Sci. U. S. A. 86, 9891–9895

30. Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. J. Biol. Chem. 278, 27896–27902

31. Cooper, H. M., Tamura, R. N., and Quaranta, V. (1991) The major laminin receptor of mouse embryonic stem cells is a novel isoform of the cell self-renewal in collaboration with STAT3. J. Cell Biol. 115, 1647–1651

32. The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420, 563–573

33. Loebel, D. A., Watson, C. M., De Young, R. A., and Tam, P. P. (2003) Lineage choice and differentiation in mouse embryos and embryonic stem cells. Dev. Biol. 264, 1–14

34. Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., Levin, W. J., Stuart, S. G., Udoive, J., Ullrich, A., and Press, M. F. (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 244, 707–712

35. Scavo, L. M., Karas, M., Murray, M., and Leroith, D. (2004) Insulin-like growth factor-I stimulates both cell growth and lipogenesis during differentiation of human mesenchymal stem cells into adipocytes. Clin. Endocrinol. Metab. 89, 3543–3553

36. Hilton, D. J., and Nicola, N. A. (1992) Kinetic analyses of the binding of leukemia inhibitory factor to receptor on cells and membranes and in detergent solution. J. Biol. Chem. 267, 10238–10247

37. Watson, K., Gooderham, N. J., Davies, D. S., and Edwards, R. J. (1999) Nucleosomes bind to cell surface proteoglycans. J. Biol. Chem. 274, 21707–21713