

Analysis of Cell Surface Proteome Changes via Label-free, Quantitative Mass Spectrometry*

Ralph Schiess^{‡§¶}, Lukas N. Mueller^{‡¶}, Alexander Schmidt^{‡¶}, Markus Mueller^{||}, Bernd Wollscheid^{‡**‡‡}, and Ruedi Aebersold^{‡¶§§¶¶|||}

We present a mass spectrometry-based strategy for the specific detection and quantification of cell surface proteome changes. The method is based on the label-free quantification of peptide patterns acquired by high mass accuracy mass spectrometry using new software tools and the cell surface capturing technology that selectively enriches glycopeptides exposed to the cell exterior. The method was applied to monitor dynamic protein changes in the cell surface glycoproteome of *Drosophila melanogaster* cells. The results led to the construction of a cell surface glycoprotein atlas consisting of 202 cell surface glycoproteins of *D. melanogaster* Kc167 cells and indicated relative quantitative changes of cell surface glycoproteins in four different cellular states. Furthermore we specifically investigated cell surface proteome changes upon prolonged insulin stimulation. The data revealed insulin-dependent cell surface glycoprotein dynamics, including insulin receptor internalization, and linked these changes to intracellular signaling networks. *Molecular & Cellular Proteomics* 8:624–638, 2009.

The multitude of cells and cell types that constitute multicellular organisms are organized in intricate higher order structures and organs. These cells also communicate with each other either via direct cell-to-cell contact or over longer distances via soluble mediators. In either form of communication the proteins at the surface of the cell, including adhesion molecules, channel transporter proteins, cell surface receptors, and enzymes, are of critical importance for sensing, inducing, and catalyzing responses to the changing environment of the cell. The ensemble of cell surface proteins, the cell surface proteome, therefore provides a unique molecular fingerprint to classify cells and cellular states. For these reasons, there has been considerable interest in a robust, sensitive,

specific, and quantitative technology to study the cell surface proteome.

MS is the method of choice for the identification and accurate quantification of the proteins contained in complex sample mixtures (1). Recent advances in MS-based proteomics, specifically improved instrumentation, software tools for the analysis of proteomics data sets (2), and emerging, more efficient data collection strategies (3), now routinely lead to the identification of hundreds to thousands of proteins in a single experiment. However, they still fall short of complete proteome analysis. As an alternative to the analysis of total cell or tissue extracts that leads to the identification and, if suitable quantification strategies are applied (4), to the quantification of a fraction of the proteins present in the sample, analysis of specific subproteomes that are enriched for proteins of particular types has been suggested (5). Implementations of this concept so far include the selective isolation and subsequent analysis of cysteine-containing peptides (6), phosphorylated peptides (7), *N*-glycosylated peptides (8), the set of N-terminal peptides (9), and specific subcellular fractions and organelles (10, 11). These enrichment technologies have in common that a particular subset of proteins or peptides is enriched and can be analyzed more comprehensively so that even low abundance proteins can potentially be detected.

Traditionally cells have often been classified by exploiting antibodies available for a limited number of molecules termed clusters of differentiation (CD)¹ (12). These antibodies also provide a powerful tool to investigate expression changes of CD molecules and are widely used in the fields of hematology, immunology, and pathology for research, diagnosis, and therapy. However, the generation and validation of such monoclonal antibodies is time-consuming, labor-intensive, and as-

From the [‡]Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule (ETH) Zurich, 8093 Zurich, Switzerland, [§]Ph.D. Program in Molecular Life Science and [¶]Faculty of Science, University of Zurich, 8057 Zurich, Switzerland, ^{||}Competence Center for Systems Physiology and Metabolic Disease, 8093 Zurich, Switzerland, ^{|||}Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland, ^{**}National Center for Competence in Research (NCCR) Neuro Center for Proteomics, ETH and University of Zurich, 8093 Zurich, Switzerland, and ^{§§}Institute for Systems Biology, Seattle, Washington 98103

Received, April 18, 2008, and in revised form, November 21, 2008
Published, MCP Papers in Press, November 25, 2008, DOI 10.1074/mcp.M800172-MCP200

¹ The abbreviations used are: CD, clusters of differentiation; CSC, cell surface capturing; NX(S/T) motif, amino acid sequence of asparagine followed by any amino acid (except proline) and either serine or threonine; TM, transmembrane; LTQ, linear ion trap; InR, insulin receptor (CG18402); TCEP, tris(2-carboxyethyl)phosphine; ICPL, isotope-coded protein label; *N*-glycosite, a peptide that is *N*-glycosylated in the intact protein in its deglycosylated form; MS1 feature, MS1 peak representing a potential peptide ion; rcf, relative centrifugal force; PNGaseF, peptide-*N*-glycosidase F; TPP, Trans-Proteomic Pipeline 3.5; RT, retention time; GO, gene ontology; CV, coefficient of variance; LPS, lipopolysaccharide; PTP, protein-tyrosine phosphatase.

sociated with high costs. Moreover many cell types and disease states cannot be unambiguously identified with the currently available set of CD molecules. Thus, clinical and basic biology research would greatly benefit from a broader selection of such differentiation markers. In the absence of a broader set of CD-specific reagents or a different approach to measure cell surface proteins more comprehensively, the characterization of this important class of proteins will remain incomplete and biased.

In contrast to the human and murine species where at least a subset of the cell surface proteins can be detected and quantified using the anti-CD antibodies, these critical reagents are almost completely lacking for organisms such as *Drosophila melanogaster* or *Caenorhabditis elegans*. However, extensive high quality data sets on proteomes or subproteomes are particularly useful in those species because they can be related to the readily accessible rich genomic and genetic resources. In prior work we have extensively mapped out the proteome and phosphoproteome of *D. melanogaster* (13, 14), and efforts by others have elucidated the metabolic protein network in the fly (15). In combination, these resources help to position this species for integrative studies in the emerging systems biology paradigm.

Despite the obvious interest in the cell surface proteome, technical limitations have so far precluded its comprehensive analysis. These include difficulties in efficiently separating membrane-associated from other cellular proteins, their frequently low abundance, and poor solubility (16). To facilitate the deep and specific analysis of cell surface proteins we recently developed a method for the selective identification of cell surface glycoproteins, the cell surface capturing (CSC) method.² CSC is based on the fact that the majority of proteins on the surface of cells are glycosylated (18). It comprises a highly selective procedure to enrich for the *N*-glycosylated peptides from glycoproteins via chemical reactivity of their carbohydrate moiety (8). By analyzing only the *N*-glycosites (peptides that are *N*-glycosylated in the intact protein in their deglycosylated form), the sample complexity is drastically reduced, and a relatively large abundance range of the cell surface proteome can be analyzed by mass spectrometry without the need for multidimensional separation. Therefore CSC represents a valuable tool to generate a comprehensive cell surface map in a single LC-MS analysis.

In this study we used the CSC method to characterize the cell surface proteome of the *D. melanogaster* Kc167 cell line and, in combination with label-free quantitative MS, to determine perturbation-induced changes in the surface proteome of the cell. Label-free quantification was achieved by comparing LC-MS feature maps using the software tool SuperHirn (19) and a new interactive software tool called JRatio with a graphical user interface for the relative protein quantification

of MS1 features detected in the different patterns. These experiments resulted in the identification of 202 glycoproteins, 183 (91%) of which contained at least one transmembrane (TM) domain. We determined that the variation of biological replicates was below 25%, which allowed distinguishing between different cellular states based on the cell surface protein patterns consisting of *N*-glycosites representing more than a hundred glycoproteins. Upon perturbation of specific pathways within the cell, quantitative analysis revealed abundance changes in the surface glycoproteome. Furthermore we monitored the internalization of the insulin receptor (InR) upon prolonged insulin stimulation (20) and thereby confirmed the down-regulation of the InR from the cell surface. Additionally we demonstrated that this change was due to InR internalization into endosomes and not degradation.

In conclusion, the work presented here describes a robust, sensitive, specific, and quantitative method to profile cell surface proteomes. Its application is well suited to monitoring quantitative protein changes in multiple samples and thus has the potential to facilitate initial biomarker discovery.

EXPERIMENTAL PROCEDURES

Chemicals—Porcine trypsin (modified, sequencing grade) was purchased from Promega (Madison, WI). Tris(2-carboxyethyl)phosphine (TCEP) and iodoacetamide were purchased from Fluka (Buchs, Switzerland). HPLC grade water and acetonitrile were purchased from Riedel-de Haën (Seelze, Germany). Sodium periodate was from Pierce, biocytin hydrazide was from Biotium, UltraLink immobilized streptavidin PLUS was from Pierce, the 15-ml Dounce Tissue Grinder was from Wheaton, RapiGest was from Waters, PNGaseF was from New England Biolabs, and Affi-Prep Hz hydrazide was from Bio-Rad.

Cell Culture—*D. melanogaster* embryonic Kc167 cells were maintained as described elsewhere (21). Briefly *D. melanogaster* Kc167 cells were maintained at 25 °C in Schneider's *D. melanogaster* medium (Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum, 100 units/ml penicillin, and 100 µg/ml streptomycin. Cells were first seeded in flasks at 5×10^6 cells/ml in a volume of 40 ml and were subcultured every 3rd or 4th day. Then 80 ml of Kc167 was diluted 1:3 in an Erlenmeyer flask. Cells were harvested at A_{600} 0.5–0.6.

Stimulation of Kc167 Cells—Cells were incubated in Schneider's *D. melanogaster* medium with either 1 µg/ml lipopolysaccharide (Sigma), 50 nM rapamycin (LC Laboratories), or 1 mM sodium vanadate (Sigma) (all final concentrations) for 1 h. Persistent insulin stimulation was achieved by first starving the cells in serum-free Schneider's medium overnight and then incubating the cells with 100 nM bovine insulin (Sigma) for 2 h.

CSC—Cell surface capturing was performed as described by B. Wollscheid *et al.*² *D. melanogaster* Kc167 cells were harvested by spinning them down at 450 rcf for 5 min in a centrifuge. Then the cells were reconstituted in labeling buffer (1× PBS, 0.1% fetal bovine serum, pH 6.5) and oxidized with 1.25 mM sodium periodate for 15 min at room temperature. Cells were then washed twice with labeling buffer to remove dead cells and sodium periodate. Afterward the cells were incubated with 25 mg/ml biocytin hydrazide (Biotium) to label the oxidized carbohydrates of the cell surface molecules with biotin. Biotin-labeled cells were lysed in detergent-free lysis buffer (10 mM Tris-HCl, 0.5 mM MgCl₂, pH 7.5) at 4 °C for 10 min. Then the cells were additionally homogenized with 30 strokes using a Dounce Tis-

² B. Wollscheid, D. Bausch-Fluck, C. Henderson, R. O'Brien, R. Schiess, R. Aebersold, and J. D. Watts, submitted manuscript.

sue Grinder (Wheaton). Cell debris and nuclei were removed by centrifugation at 2800 rcf for 10 min, and membranes were pelleted from the supernatant by ultracentrifugation at 210,000 rcf for 1 h. The membrane fraction was solubilized with 0.1% RapiGest (Waters) in 100 mM ammonium bicarbonate. After reduction of disulfide bonds with 5 mM TCEP for 30 min at 37 °C and alkylation of cysteines with 10 mM iodoacetamide for 30 min at room temperature, the proteins were digested with trypsin at an enzyme to protein ratio of 1:100 at 37 °C overnight. 400 μ l of a 50% slurry of UltraLink immobilized streptavidin PLUS beads in 100 mM ammonium bicarbonate was added to the protein digest and incubated for 1 h in a head over head shaker at room temperature. Unbound peptides and lipids were then washed away with various buffers (5 M NaCl; 0.5% Triton in PBS, pH 8.0; and 100 mM sodium carbonate, pH 11.0) followed by the specific release of *N*-glycosites by PNGaseF (New England Biolabs) in 50 mM ammonium bicarbonate, pH 7.8, at 37 °C overnight. Released *N*-glycosites were dried in a SpeedVac concentrator and resolubilized in 0.1% formic acid for mass spectrometric analysis.

Whole Membrane Glycocapturing—Glycoproteins were enriched from cell culture using a modified version of the protocol published by Zhang *et al.* (8). *D. melanogaster* Kc167 cells were harvested by spinning them down at 450 rcf for 5 min in a centrifuge. The cells were lysed in detergent-free lysis buffer (10 mM Tris-HCl, 0.5 mM MgCl₂, pH 7.5) at 4 °C for 10 min. Then the cells were additionally homogenized with 30 strokes using a Dounce Tissue Grinder (Wheaton). Cell debris and nuclei were removed by centrifugation at 2800 rcf for 10 min, and membranes were pelleted from the supernatant by ultracentrifugation at 210,000 rcf for 1 h. The membrane fraction was solubilized with 0.1% RapiGest in 100 mM ammonium bicarbonate. After reduction of disulfide bonds with 5 mM TCEP for 30 min at 37 °C and alkylation of cysteines with 10 mM iodoacetamide for 30 min at room temperature, the proteins were digested with trypsin at an enzyme to protein ratio of 1:100 at 37 °C overnight. The peptides were cleaned by reversed phase chromatography, oxidized with 20 mM sodium periodate, and coupled to Affi-Prep Hz hydrazide support (Bio-Rad). Unbound peptides and lipids were then washed away with various buffers (1.5 M NaCl, pure methanol, 80% acetonitrile, and 50 mM ammonium bicarbonate) followed by the specific release of *N*-glycosites by PNGaseF in 50 mM ammonium bicarbonate, pH 7.8, at 37 °C overnight. Released *N*-glycosites were dried in a SpeedVac concentrator and resolubilized in 0.1% formic acid for mass spectrometric analysis.

Isotopic Labeling—Isotopic labeling was carried out as described by Schmidt *et al.* (22). The membrane protein fractions were prepared and biotin-labeled as described above and proteolyzed. The resulting peptides were labeled with either the heavy or the light version of the isotope-coded protein label (ICPL) reagent, respectively, and the biotinylated glycopeptides were then specifically enriched as described above.

Mass Spectrometry—Samples were analyzed on a hybrid linear ion trap (LTQ)-FT mass spectrometer (Thermo Electron, San Jose, CA) equipped with a nanoelectrospray ion source. Chromatographic separation of peptides was performed on an Agilent 1100 micro HPLC system (Waldbronn, Germany) equipped with a 15-cm fused silica emitter (150- μ m inner diameter, packed with a Magic C₁₈ AQ 5- μ m resin; Michrom Bioresources, Auburn, CA). Peptides were loaded on the column from a cooled (4 °C) Agilent autosampler and separated with a linear gradient of acetonitrile/water containing 0.1% formic acid at a flow rate of 1.2 μ l/min. A linear gradient from 2 to 28% acetonitrile in 60 min that was optimized for the number of peptide features detected was used. The MS instrument was operated to maximize the quality of LC-MS feature maps as opposed to maximizing the number of identifications. Therefore, for each peptide sample a standard data-dependent acquisition on the three most intense ions per MS scan was performed. Three MS/MS spectra were acquired in

the linear ion trap per FT-MS scan; the latter was acquired at 100,000 full-width at half-maximum (at 350 *m/z*) nominal resolution, resulting in an overall cycle time of \sim 1 s. Charge state screening was used, allowing fragmentation of doubly and higher charged ions and rejecting ions of single and unknown charge state. A threshold of 200 ion counts was set to trigger an MS/MS attempt.

Data Analysis—The raw data acquired by the LTQ-FT instrument (software, Xcalibur 2.0 SR1) were converted to mzXML using ReAdW 3.5.1 (23) applying default parameters. MS/MS scans were then exported as dta files without further processing using the program mzXML2Other (23). MS/MS spectra were searched against the Berkeley *D. melanogaster* Genome Project BDGP5.2 (20,981 entries) database using SEQUEST version 27 (24). The SEQUEST database search criteria included variable modifications of 57.02146 Da for cysteines (for the alkylation with iodoacetamide to combine search results of both reduced/alkylated and non-reduced/non-alkylated samples), 15.99491 Da for methionines (for oxidation), and 0.98406 Da for potential formerly *N*-glycosylated asparagines (which are converted to aspartic acid by PNGaseF release), respectively. The following additional search constraints were applied: monoisotopic parent and fragment masses; precursor ion mass tolerance, 0.05 Da; fragment ion mass tolerance, 0.5 Da; at least one tryptic terminus; and one missed cleavage. The identified peptides were processed and analyzed through the mass spectrometry Trans-Proteomic Pipeline 3.5 (TPP) (25). In the TPP, the database search results were validated using the PeptideProphet software (26), which uses various SEQUEST scores (XCORR, Δ Cn, and Sp) to calculate a probability score for each identified peptide by linear discriminant analysis. *N*-Glycosylation motif information and accurate mass binning were used in PeptideProphet. The peptides were then assigned for protein identification using the ProteinProphet software (27). ProteinProphet allowed filtering of large scale data sets with assessment of predictable sensitivity and false positive identification error rates. In this study, we used a PeptideProphet probability score \geq 0.9 and a ProteinProphet probability score \geq 0.9. This resulted in an overall false positive error rate below 1% as determined by ProteinProphet (27).

For the quantitation of the ICPL differential labeling experiment, the XPRESS software integrated in the TPP was used (28). Protein ratios were calculated by accurately quantifying the relative abundance of ICPL-labeled peptides from their chromatographic co-elution profiles. Starting with the peptide identification, XPRESS isolates *d*₀ and *d*₄ peptide elution profiles, determines the area of each peptide peak, and calculates the abundance ratio based on these areas.

Transmembrane Prediction—Transmembrane domain predictions were obtained from the SOSUI Web server for classification and secondary structure prediction of membrane proteins (29) and TM-HMM, a hidden Markov model-based predictor for transmembrane helices in protein sequences (30).

Label-free Quantification of Peptide and Protein Ratios—Data from LC-MS runs were converted from raw to the mzXML data format (23) and processed by the software tool SuperHirn (19). SuperHirn performs feature detection on acquired LC-MS feature maps whereby isotopic patterns of peptides are extracted and tracked along their chromatographic elution profile. It centroids raw peak data and reduces *m/z* signals of an MS1 scan to the corresponding monoisotopic masses along with the charge state (*z*). The integrated peak area of the detected MS1 feature is calculated from the intensity values of the detected monoisotopic peak areas over the chromatographic elution period. Acquired peptide identifications in pepXML format (26) are then mapped to the extracted MS1 features via their accurate precursor mass and retention time coordinate. MS1 features are then mapped across the different LC-MS maps and integrated into a general repository of aligned MS1 features, designated MasterMap. The MasterMap represents a framework for further data analysis such

as intensity normalization and the extraction of MS1 feature, peptide, and protein ratios.

To increase the number of proteins identified, we used the inclusion list annotation feature of SuperHirn where peptide identifications from targeted MS/MS experiments can be mapped back to detected MS1 features (3). Thereby the MasterMap was further updated with MS/MS information to assign MS1 features that had not been annotated in a particular experiment with the corresponding peptide sequence. The acquired peptide identifications were mapped back to corresponding MS1 features using accurate mass, normalized retention time ($\Delta RT = 1.0$ min), and the peptide charge information (15).

The in-house developed Java software JRatio was used for the calculation and visual assessment of peptide and protein ratios. Specifically JRatio imports aligned MS1 features from the MasterMap and quantifies -fold changes of MS1 features across LC-MS measurements. In a first step, -fold changes of aligned MS1 features were quantified by computing ratios between sample states from the average MS1 feature intensities $\bar{V}_{A/B}$ from replicate measurements of treatment group A and B, respectively (Equation 1). Accordingly standard deviations of the computed ratios were derived as described in Equation 2.

$$R_{\text{Feature}_B^A} = \frac{\bar{V}_A}{\bar{V}_B} \quad (\text{Eq. 1})$$

$$\sigma_{\text{Feature}_B^A} = R_{\text{Feature}_B^A} \cdot \sqrt{\left(\frac{\sigma_{V_A}}{\bar{V}_A}\right)^2 + \left(\frac{\sigma_{V_B}}{\bar{V}_B}\right)^2} \quad (\text{Eq. 2})$$

JRatio provides visual support for the assessment of the computed results where MS1 feature ratios can be explored to verify the extracted chromatographic elution profiles and intensity reproducibility of aligned MS1 features between replicate runs etc. Proteins were then assembled from MS1 feature ratios characterized by a high quality MS/MS peptide identification (PeptideProphet probability >0.9 (26)). Importantly only fully tryptic *N*-glycosites were taken into account for protein quantification. Robust protein ratios and standard deviations were derived for each protein from its associated MS1 feature ratios (Equations 3 and 4). A normal distribution was used to describe the calculated protein ratios, and Student's *t* test statistics was applied to assess the significance of a protein -fold change. Proteins with a *p* value smaller than 0.1 were considered to be significantly regulated. JRatio is freely available.

$$R_{\text{Protein}_B^A} = \frac{\sum_{i=1}^{N_{\text{Feature}}} \frac{R_{\text{Feature},i}}{\sqrt{\sigma_{\text{Feature},i}}}}{\sum_{i=1}^{N_{\text{Feature}}} \frac{1}{\sqrt{\sigma_{\text{Feature},i}}}} \quad (\text{Eq. 3})$$

$$\sigma_{\text{Protein}_B^A} = \frac{\sqrt{\sum_{i=1}^{N_{\text{Feature}}} (\sigma_{\text{Feature},i})^2}}{N_{\text{Feature}}} \quad (\text{Eq. 4})$$

RESULTS

The goal of this study was to test the hypothesis that different cellular states can be distinguished by the comprehensive, fast, and quantitative analysis of the cell surface proteome. We approached this by combining the selective enrichment of cell surface glycoproteins with a label-free, quantitative proteomics method that provides the sample throughput required to analyze multiple samples. To test the specificity and the sensitivity of the method, cell surface pro-

teins in *D. melanogaster* Kc167 cells were selectively isolated, and a comprehensive catalogue of 202 cell surface glycosylated proteins was generated. We then assessed the reproducibility of the label-free quantification approach. Subsequently changes in the cell surface glycoproteome induced by specific perturbations were quantitatively monitored, indicating that the cell surface glycoproteome indeed changed as a function of the state of the cell. Finally we validated the technique by applying it to a well studied biological system, the regulation of InR action.

The D. melanogaster Kc167 Cell Surface Glycoproteome Atlas—We first identified the *N*-glycosites from the *D. melanogaster* Kc167 cell surface proteome to generate a reference map for further comparative analyses by label-free quantitative MS. The *N*-glycosites were identified by LC-MS/MS after their isolation via the CSC method that is based on selective affinity labeling and solid phase capturing of glycosylated cell surface peptides (Fig. 1).² The data were stored in a database (Kc167 glycoproteome atlas) and are represented in an LC-MS (retention time (RT) versus *m/z*) feature map (Fig. 2 and supplemental Table S1) in which the identified *N*-glycosites are annotated with their amino acid sequence.

To generate the data set we combined the results of 12 experiments (for more information see supplemental Table S2) in which cell surface glycoproteins from *D. melanogaster* Kc167 cells were isolated and subjected to LC-MS/MS. The fragment ion spectra acquired from a total of 90 LC-MS/MS runs were searched against the database BDGP5.2 (for details see “Experimental Procedures”). A total of 20,608 MS/MS spectra were assigned at a PeptideProphet probability (26) threshold of ≥ 0.9 (false discovery rate, 1%) to peptide sequences of which 84% (17,397) matched to peptides containing the NX(S/T) motif, indicating the presence of a glycan at that site in the intact protein. These results indicate the high selectivity of the method for *N*-glycosites.

Overall the assigned spectra represented 1002 unique *N*-glycosites (PeptideProphet probability score ≥ 0.9 ; false discovery rate, 1%) matching to 202 unique glycoproteins (supplemental Tables S3 and S4). 578 different sites of *N*-glycosylation could be unambiguously assigned due to the mass shift caused by enzymatic deamidation at the site of glycan attachment (mass difference, 0.98604 Da). 183 (91%) of the identified cell surface glycoproteins contained at least one TM domain as predicted by SOSUI, a classification and secondary structure prediction algorithm for membrane proteins (29). 126 of the 202 identified proteins could be gene ontology (GO)-annotated using Babelomics (31) software. 108 (86%) of the GO-annotated proteins belonged to the group “membrane,” seven (5%) belonged to the group “extracellular matrix,” and only 11 (9%) were annotated as “intracellular” (Fig. 3). Interestingly five of the 11 proteins annotated as intracellular contain at least one

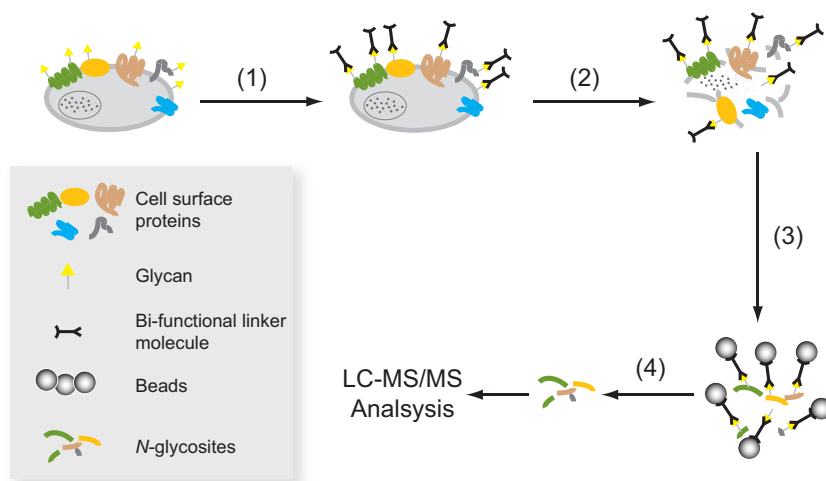


FIG. 1. Scheme of selective enrichment of cell surface glycoproteins. 1, the carbohydrate moieties of glycoproteins on the surface of intact, living cells were oxidized with sodium periodate. The thus chemically activated glycoproteins were then coupled to a biotin-containing linker molecule that does not damage or penetrate the cells. 2, the tagged cells were then lysed, cell debris and nuclei were removed by centrifugation, and a fraction enriched in membrane proteins was isolated from the lysate via ultracentrifugation. 3, upon digestion of the isolated membrane proteins, the tagged glycopeptides were affinity-purified using a streptavidin-covered solid support. 4, *N*-glycosites were then enzymatically released and analyzed by an LTQ-FT mass spectrometer. The obtained spectra were searched against the database BDGP5.2 using SEQUEST.

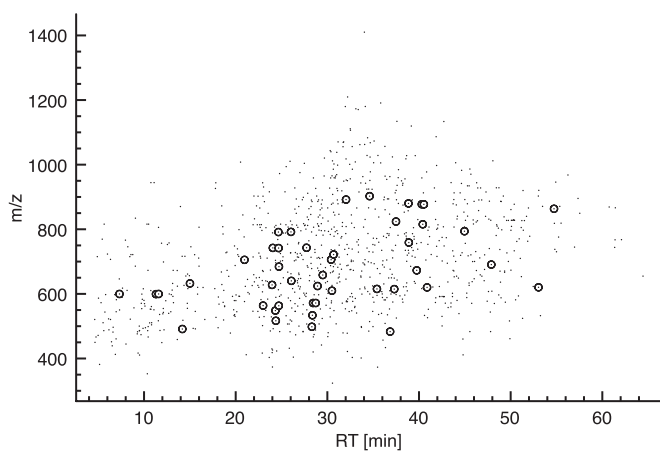


FIG. 2. Cell surface LC-MS map of *N*-glycosites. A graphical representation of the LC-MS (RT versus *m/z*) feature map is shown. Only the identified *N*-glycosites are shown. Circled dots represent *N*-glycosites originating from the InR. The InR has 25 potential *N*-glycosylation sites of which 19 could be verified by the 44 *N*-glycosites described here. The data represented here can be found in supplemental Table S1.

predicted transmembrane domain and therefore are likely to also be constituents of the plasma membrane. Additionally three proteins, oligosaccharyltransferase 3 (CG7748), a protein with oligosaccharyltransferase activity (CG1518), and prolyl-4-hydroxylase- α EFB (CG31022), which are according to their GO-annotated molecular function constituents of the endoplasmic reticulum, were identified. The repetitive identification of these proteins via CSC technology led us to the conclusion that they were *N*-glycosylated and very likely accessible at the plasma membrane during the short time period of labeling. However, we cannot exclude that these

proteins as well as the ones predicted to be intracellular are contaminants.

The proteins identified cover all major classes of cell surface receptor proteins. For example, the proteins methuselah (CG6936) and methuselah-like 3 (CG6530), members of the class of the G-protein-coupled receptors, which are known to be generally of low abundance (32), were found. We also identified several enzyme-linked receptors, including the receptor tyrosine kinases InR (CG18402), Eph receptor tyrosine kinase (CG1511), and the platelet-derived growth factor/vascular endothelial growth factor receptor (CG8222) as well as the membrane-spanning phosphotyrosine phosphatase receptors PTP69D (CG10975), PTP10D (CG1817), and PTP4E (CG6899). Further the cytokine receptors activin receptor (CG7904) and domeless (CG14226), the integrin adhesion molecules (CG1560, CG1771, CG8095, and CG9623) that are involved in cell-cell interaction, and membrane transporters like the β subunits 1 and 2 of the sodium/potassium-transporting ATPase (CG9258 and CG9261) as well as the organic cation (CG13610) and anion transporter proteins (CG3380 and CG7571) were identified. These results demonstrate that in terms of function and abundance a wide range of cell surface molecules was covered implying that our approach is limited neither to a certain subclass of cell surface proteins nor by the dynamic range.

Reproducible, Label-free Quantification of the Kc167 Cell Surface Glycoproteome—To assess the accuracy and reproducibility of quantifying cell surface proteome changes detected by the analysis of LC-MS maps of isolated *N*-glycosites, we performed biological and technical replicates of the Kc167 cell surface proteins. The CSC method was

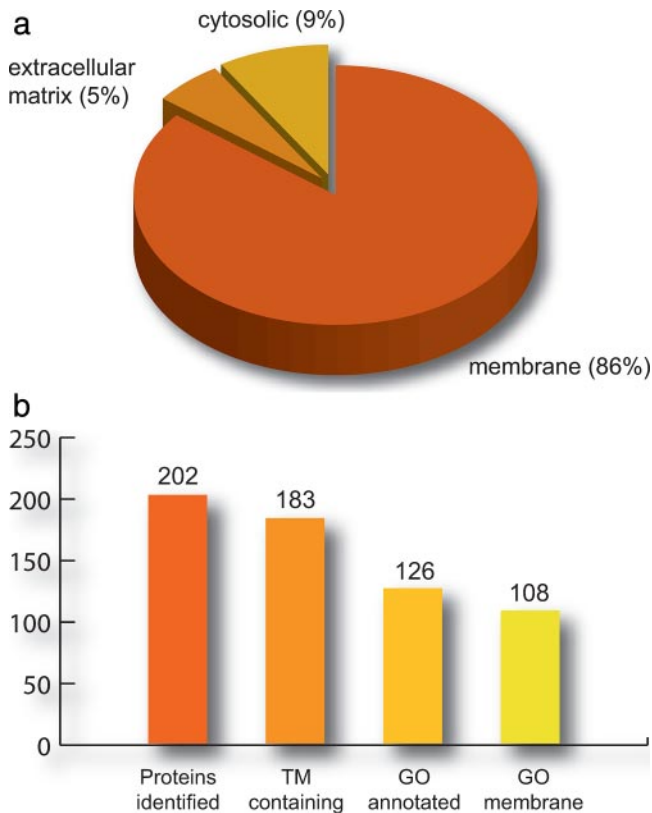


FIG. 3. Analysis of identified cell surface glycoproteins. *a*, GO cellular component analysis of the identified proteins. GO annotation for 126 proteins of the 202 glycoproteins identified was available. 86% belonged to the membrane, 5% belonged to the extracellular matrix, and 9% belonged to the cytoplasm. *b*, 183 of the 202 glycoproteins identified contain one or more TM domains as predicted by SOSUI (29). Furthermore 108 of the 126 GO-annotated proteins are membrane constituents.

applied to cells from three parallel Kc167 cell cultures. Each of the three biological isolates was analyzed in triplicate on a high mass resolution LTQ-FT instrument. The nine LC-MS feature maps thus generated were processed by the software SuperHirn (19). SuperHirn identified the peptide features in each MS1 feature map and aligned them across the different maps to generate an intensity-normalized Master-Map. In total, 5166 MS1 features were detected of which 1210 could be aligned over all nine LC-MS runs (Table I). To quantify the reproducibility of the method, the coefficient of variance (CV) of feature intensity values across all nine aligned LC-MS patterns was computed. 76% of the quantified features showed a CV equal to or below 30%, and the average coefficient of variance of all 1210 aligned features was 24%. For technical replicates CVs between 10 and 13% were achieved (supplemental Table S5).

Fig. 4 illustrates the reproducibility between the biological and the technical replicates using scatter plots. Peak intensities obtained from aligned MS1 features of three technical replicates acquired from the same biological sample were plotted against each other (Fig. 4a). The high squared Pear-

TABLE I

Alignment of MS1 features across replicate LC-MS measurements

The total of 5166 MS1 features extracted from nine LC-MS runs originating from three biological and three technical replicates each were aligned with SuperHirn software. The table illustrates the common MS1 features shared by different LC-MS runs.

Replicates	Common MS1 features
9	1210
8	260
7	235
6	367
5	276
4	341
3	633
2	585
1	1259
Total	5166

son correlation R^2 (0.986–0.989) and the near straight lines indicated the nearly optimal linear relationship between the replicates. Similarly the mean peak intensities of MS1 features obtained from the three biological replicates showed very high correlation ($R^2 = 0.925$ – 0.962) as illustrated in Fig. 4b. Furthermore the plots also illustrate that the intensities of the detected MS1 features span a dynamic range of more than 3 orders of magnitude. As can be noted from the scatter plots, the variability of signal intensities among replicates increases with decreasing signal intensity.

The variation among biological experiments and their technical LC-MS replicates was further investigated by comparing the respective peak ratios. These are expected to be 1.0 because the same amount of protein was anticipated to be present in each sample. For each biological experiment, the average intensity of each aligned MS1 feature was calculated across the three technical replicates and then divided by the average feature intensity measured over all nine consecutive runs belonging to the three experimental replicates. Only MS1 features detected in all nine runs were considered for this statistical analysis. More than 95% of all MS1 feature ratios were below a 2-fold variance, and 85% of all MS1 features varied by less than 40%. The same analysis was done for the three technical replicates of each experiment. 90% of the aligned MS1 features showed a ratio variation of less than 25%, which clearly demonstrates the high reproducibility of MS1 feature intensities acquired by the LTQ-FT mass spectrometer and calculated by our label-free quantification approach.

Finally the quantification error between replicated runs of this experiment showed a variation similar to the one reported by Wang *et al.* (33) in two duplicate LC-MS runs. In conclusion, the combination of the method for the selective isolation of cell surface *N*-glycosites with label-free quantification is feasible. We demonstrate the capacity of the technique to highly enrich for the cell surface glycoproteome and to quantify peptide features over a dynamic

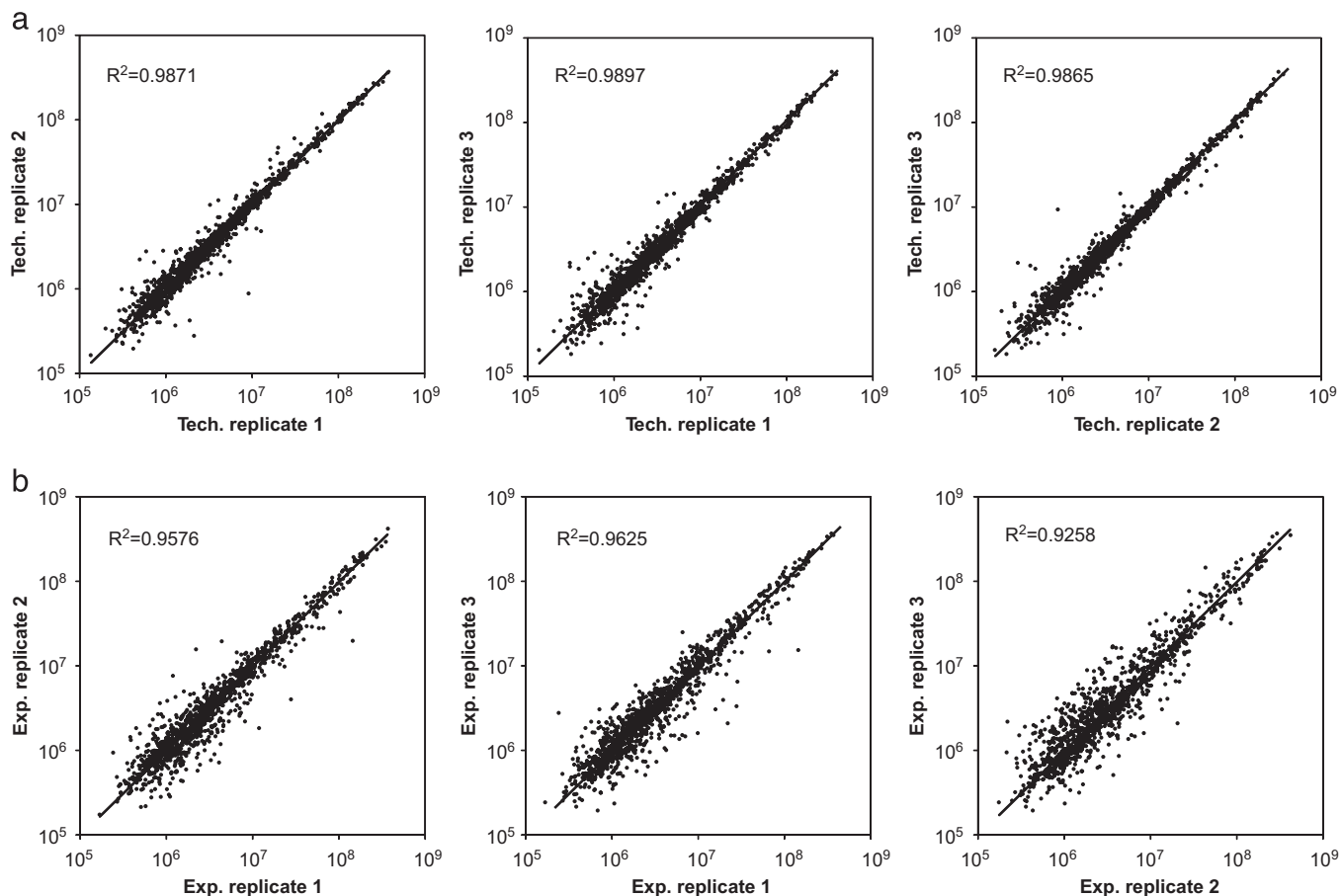


FIG. 4. **Reproducibility of label-free quantification of the Kc167 cell surface glycoproteome.** A scatterplot of LC-MS features of isolated *N*-glycosites using CSC and the respective squared Pearson correlation R^2 are shown. *a*, the peptide peak areas of aligned MS1 features in the three replicate runs of the first experiment were plotted against each other. *b*, the mean peptide peak areas of aligned MS1 features in the three experiments were plotted against each other. *Tech.*, technical; *Exp.*, experimental.

range of more than 3 orders of magnitude and for multiple samples, thus making it possible to detect and quantify even low abundance cell surface glycoproteins in serial comparisons.

The Cell Surface Proteome Changes as a Function of Cellular State—We hypothesize that the composition of the cell surface proteome reflects the state of the intracellular signaling systems and thus perturbations in intracellular signaling systems can be detected by changes in the cell surface proteome.

To test this hypothesis Kc167 cell cultures were subjected to an array of different perturbations to induce changes in cellular state and to monitor the resulting changes in the cell surface glycoproteome. The cells were treated either with lipopolysaccharide (LPS), which elicits strong activation of c-Jun N-terminal kinase (JNK), a stress-activated protein kinase downstream of mitogen-activated protein kinases (MAPKs) (34); rapamycin, an inhibitor of the target of rapamycin (TOR), which is an important component of the insulin signaling pathway and affects cell growth by modulating the

activity of S6K kinase (21); or vanadate, which generally inhibits protein phosphatases and therefore triggers a whole cascade of stress-activated protein phosphorylation changes (35). These three selected stimuli are known to have a strong effect on intracellular signaling events and thus were believed to also affect the cell surface proteome. To quantify the cell surface abundance changes introduced by the respective perturbations, each cell sample was subjected to *N*-glycosite enrichment, the resulting samples were analyzed on an LTQ-FT instrument, and glycoprotein abundance ratios were calculated against the intensity values of an untreated control cell culture. LC-MS feature maps were analyzed as described above and combined into a MasterMap. Additionally MS/MS-based peptide identifications of cell surface glycoproteins from the cell surface atlas were used to annotate MS1 features not sequenced in these particular LC-MS/MS runs. JRatio was then used to obtain protein ratios from MS1 features with available high quality peptide information (PeptideProphet $p > 0.9$) (for details see “Experimental Procedures”).

In total, 112 *N*-glycosites were quantified, collectively representing 61 different glycoproteins (supplemental Tables S6 and S7). For each stimulus about 80% of the glycoproteins quantified showed less than 2-fold regulation. This is expected because each stimulus triggers only certain specific signaling cascades. The proteins fascilin 1 and 2 (CG6588 and CG3665), tetraspanin 86D (CG4591), and multiple integrins (CG8095, CG1771, and CG1560), which are responsible for maintaining cell structure and cell adhesion and thought to have constant abundance levels, did not change at any of the conditions tested. In contrast, the glycosylated cell surface proteins guanylate cyclase (CG8742) and 26–29-kDa proteinase (CG8947) showed lower and higher abundance, respectively, after stimulation independent of the stimulus. Other glycoproteins exhibited specific abundance changes following one or two particular stimuli. These include the macroglobulin complement-related protein (CG7586) that was up-regulated after rapamycin and vanadate treatment, a protein with similarity to the ATP-binding cassette transporter family (CG5789) up-regulated after vanadate treatment only, and the Niemann-Pick Type C-1 protein (CG5722) that was down-regulated after LPS stimulation. The abundance ratios detected in these perturbation experiments along with the corresponding standard deviation and *p* value are shown in Table II, and changes in protein abundance are illustrated with a heat map in Fig. 5. In summary, the results illustrate that cell samples representing differentially perturbed states can be distinguished by their specific cell surface proteome patterns.

Insulin-induced Internalization of InR—To test the ability of the described method to measure the quantitative behavior of specific cell surface proteins, we used a well studied biological system, the regulation of the InR. Its primary function is to maintain glucose homeostasis through signaling induced by the interaction of the receptor with insulin. Signaling is modulated by reversible phosphorylation of cellular substrates catalyzed in part by the InR tyrosine kinase and protein-tyrosine phosphatases (PTPs), respectively. The InR activity is dependent on a dynamic equilibrium between surface (active) and internal (inactive) receptor pools (36). To measure this redistribution in response to InR stimulation, Kc167 cells were starved in serum-free Schneider's medium overnight. Half of the cells were then incubated with 100 nM bovine insulin for 2 h to simulate persistent insulin stimulation. Insulin stimulation was verified by pS6K Western blotting (data not shown) (37). In parallel, the two populations of cells were subjected to the CSC method followed by the quantitative analysis of the isolated *N*-glycosites using the LTQ-FT instrument and subsequent comparative analysis of the LC-MS maps. The experiment was repeated to account for biological variation between experiments, and three technical replicates were carried out and used to generate the respective MasterMaps.

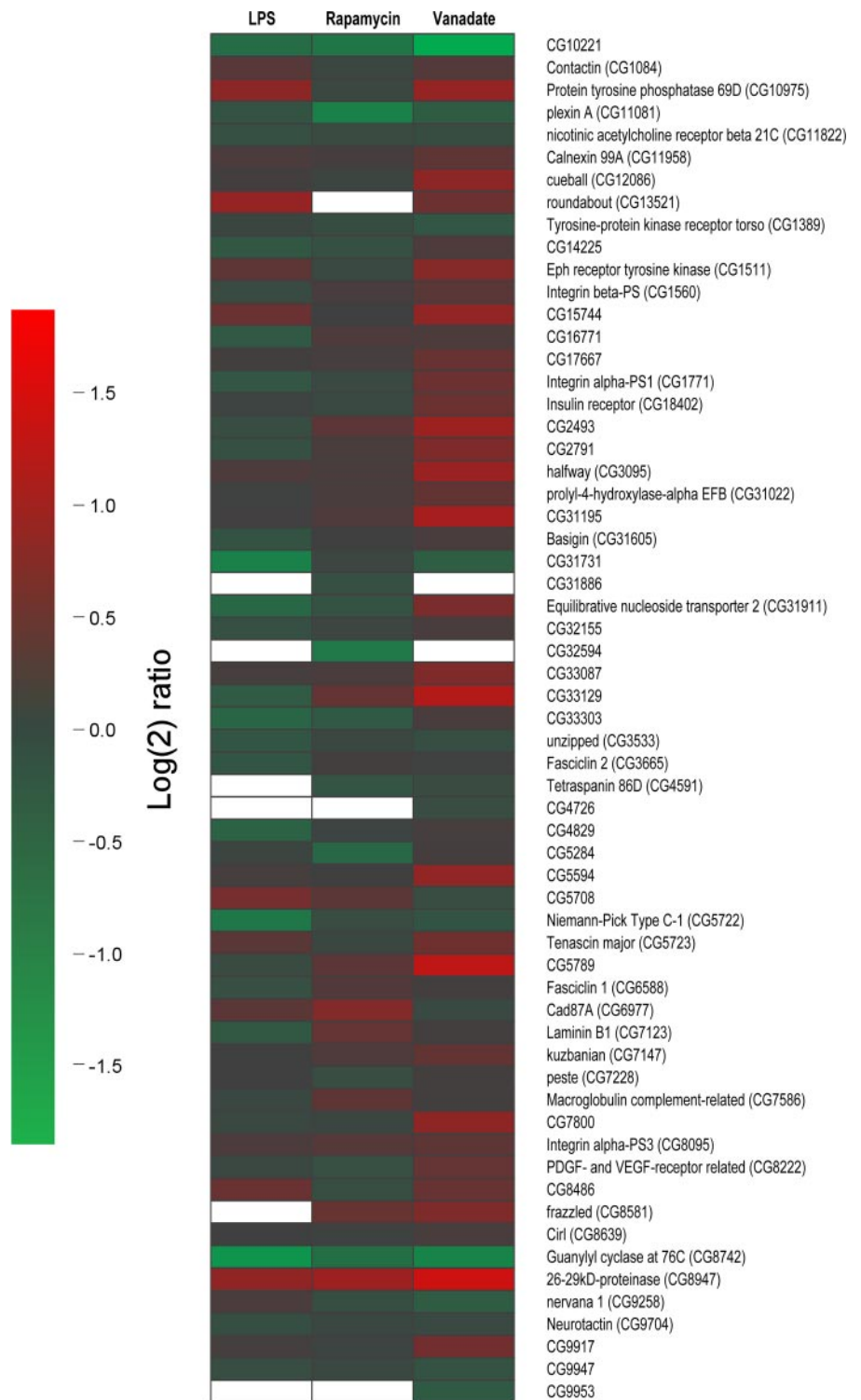
Computation of peptide and protein ratios was carried out as described above. The results showed an unambiguous 2-fold down-regulation (ratio, 0.41 ± 0.07) of the InR

TABLE II
Quantitative pattern of selected glycoproteins following different perturbations

The protein ratios (perturbed/non-perturbed) and corresponding S.D. including the *p* values of identified and quantified glycoproteins following different perturbations are given. It is demonstrated that each stimuli changes abundance levels of a specific subset of cell surface glycoproteins. NA, not applicable.

FlyBase ID	Annotation symbol	Protein name	LPS-stimulated cell culture		Rapamycin-stimulated cell culture		Vanadate-stimulated cell culture		Biological process
			Protein ratio, S.D.	<i>p</i> value	Protein ratio, S.D.	<i>p</i> value	Protein ratio, S.D.	<i>p</i> value	
FBgn0000634	CG6588	Fascilin 1	0.90 ± 0.18	0.65	1.26 ± 0.25	0.21	1.13 ± 0.17	0.52	Cell adhesion
FBgn0000635	CG3665	Fascilin 2	0.83 ± 0.14	0.39	1.13 ± 0.22	0.60	1.08 ± 0.18	0.44	Cell adhesion
FBgn0003328	CG8095	Integrin α PS3	1.20 ± 0.24	0.36	1.28 ± 0.20	0.17	1.37 ± 0.23	0.95	Cell adhesion
FBgn0004456	CG1771	Integrin α PS1	0.82 ± 0.10	0.37	0.98 ± 0.13	0.62	1.50 ± 0.01	0.84	Cell adhesion
FBgn0004657	CG1560	Integrin β PS	0.95 ± 0.04	0.85	1.17 ± 0.32	0.45	1.32 ± 0.04	0.86	Cell adhesion
FBgn0037848	CG4591	Tetraspanin 86D	NA	NA	0.84 ± 0.13	0.12	0.96 ± 0.28	0.27	Cell adhesion
FBgn0013974	CG8742	Guanylyl cyclase at 76C	0.36 ± 0.17	<0.01	0.60 ± 0.17	<0.01	0.45 ± 0.14	<0.01	Metabolic process
FBgn0028967	CG8947	26–29-kDa proteinase	1.90 ± 0.52	<0.01	2.09 ± 0.46	<0.01	2.88 ± 0.42	0.03	Metabolic process
FBgn0039207	CG5789	CG5789	0.96 ± 0.24	0.89	1.35 ± 0.14	0.08	2.6 ± 0.4	0.07	Unknown
FBgn0020240	CG7586	Macroglobulin complement-related	0.99 ± 0.09	1.00	1.36 ± 0.14	0.07	1.13 ± 0.14	0.52	Unknown
FBgn0024320	CG5722	Niemann-Pick Type C-1	0.53 ± 0.04	<0.01	0.94 ± 0.09	0.43	0.85 ± 0.04	0.14	Multicellular organismal process

FIG. 5. Protein abundance changes on the cell surface upon differential perturbation. Glycoproteins from different perturbation experiments are shown in color according to their log ratio from *green* (4-fold down-regulated) to *red* (4-fold up-regulated). Glycoprotein ratios were built comparing each stimulated sample to a control. *White fields* indicate features that were not detected or quantified in the respective sample. Individual glycoprotein changes are listed in supplemental Table S6. *PDGF*, platelet-derived growth factor; *VEGF*, vascular endothelial growth factor.



(CG18402) that was quantified by six *N*-glycopeptides. In addition, we were able to quantify an additional 100 cell surface glycoproteins (supplemental Tables S8 and S9); this was about half the glycoproteins described previously in the cell surface glycoprotein atlas. We determined that the abundance of fasciclin 1 (CG6588), integrins (CG8095, CG1771, and

CG1560), and tetraspanins 86D and 3A (CG4591 and CG10742), proteins responsible for maintaining cell structure and cell adhesion, did not change upon insulin stimulation. In contrast, some glycoproteins known to be involved in developmental processes like InR exhibited clear abundance changes. As an example, the protein frazzled (CG8581) that

TABLE III
Quantitative pattern of selected glycoproteins following InR stimulation

Shown are the abundance ratios (insulin-stimulated/non-stimulated) including S.D. and *p* value of selected glycoproteins discussed here specifically that were identified and quantified upon insulin stimulation. Glycoproteins involved in cell adhesion exhibit clearly less regulation than ones carrying out developmental processes.

FlyBase ID	Annotation symbol	Protein name	Protein ratio	<i>p</i> value	Peptides quantified	Biological process
FBgn0000634	CG6588	Fasciclin 1	0.87 ± 0.34	0.91	7	Cell adhesion
FBgn0003328	CG8095	Integrin α PS3	1.20 ± 0.58	0.38	1	Cell adhesion
FBgn0004456	CG1771	Integrin α PS1	1.03 ± 0.66	0.63	1	Cell adhesion
FBgn0004657	CG1560	Integrin β PS	0.91 ± 0.37	0.83	4	Cell adhesion
FBgn0037848	CG4591	Tetraspanin 86D	0.88 ± 0.24	0.91	2	Cell adhesion
FBgn0040334	CG10742	Tetraspanin 3A	1.05 ± 0.39	0.59	1	Cell adhesion
FBgn0013984	CG18402	InR	0.41 ± 0.07	0.09	6	Developmental process
FBgn0033134	CG12840	Tetraspanin 42EI	0.21 ± 0.15	<0.01	1	Developmental process
FBgn0005631	CG13521	Roundabout	2.73 ± 0.68	<0.01	1	Developmental process
FBgn0011592	CG8581	Frazzled	0.29 ± 0.34	<0.01	2	Developmental process

has netrin receptor activity and thus is involved in axon guidance (38) was 3-fold down-regulated after insulin stimulation, whereas roundabout (CG13521), a membrane receptor with positive regulation of cell-cell adhesion, was almost 3-fold up-regulated in insulin-stimulated cells. Tetraspanin 42EI (CG12840) was 5-fold down-regulated upon insulin stimulation. In contrast to the tetraspanins mentioned above, tetraspanin 42EI is not involved in cell-cell adhesion but in developmental processes through receptor signaling. The glycoprotein changes mentioned here are summarized in Table III.

To verify the results obtained via label-free quantification for the InR abundance changes caused by insulin treatment, we related these data obtained on parallel samples using a differential isotope labeling approach. Therefore we used half of the insulin-stimulated and the control membrane fractions from the label-free quantification experiment just described above. Half of each fraction was isotopically labeled using either heavy (control) or light (insulin-stimulated) ICPL (22). *N*-Glycosites were specifically enriched by CSC from the combined peptide samples as described before and analyzed twice on the LTQ-FT instrument.

The data obtained by differential stable isotope labeling showed that the InR was 2-fold down-regulated (ratio, 0.42 ± 0.08), confirming the previous findings from the label-free approach (ratio, 0.41 ± 0.07). Fig. 6 shows an MS spectrum (Fig. 6a) as well as an elution profile (Fig. 6b) of the glycopeptide VDLEHAN***NTESPVR**³ originating from the InR confirming the 2-fold difference. The peptide is representative for the other four *N*-glycosites identified from the InR (supplemental Table S10). The data indicate that the label-free quantification method reaches accuracy similar to that achieved by stable isotope labeling.

These data are in agreement with the model proposed for the regulation of InR action. Under starving conditions (in the

absence of serum) and therefore in the absence of ligand, the InR accumulates at the plasma membrane. Upon insulin binding, the ligand-receptor complex is sequestered from the plasma membrane and internalized into endosomes (39). The acidic pH of endosomes induces the dissociation of insulin from InR. Although insulin is degraded by endosomal acidic insulinase (40), the InR is recycled back to the cell surface. However, under conditions of prolonged stimulation with saturating levels of insulin, a subset of the InRs are transported to the late endosome and lysosome for degradation (41).

The characteristic of the CSC method that selects for glycoproteins present on the cell surface at the time of labeling, if related to a whole cell membrane analysis of *N*-glycosites, should allow for the differentiation of cell surface *versus* internalized proteins and thus for the fraction of InR that is active on the cell surface and the fraction that is inactivated by internalization into endosomes. To demonstrate that the InR is indeed internalized but not yet degraded, we performed a whole membrane glyco-capture experiment in which the cells were lysed and their membrane fraction was prepared by ultracentrifugation. From this sample we then specifically enriched for glycopeptides using a modified version of the original solid phase extraction protocol for serum glycoproteins (8). In contrast to the CSC method where glycoproteins from the cell surface are isolated, this approach enriches for all glycoproteins present in the membrane fraction, *i.e.* also proteins present in membranes of internal organelles. Therefore, if the InR was only internalized into vesicles but not degraded, then the abundance ratio of InR measured in insulin-stimulated and non-stimulated cells should not change.

Four biological replicates of Kc167 cells were starved in serum-free Schneider's medium overnight followed by insulin stimulation with 100 nM bovine insulin for 2 h for half of the cells. The two cell samples were then subjected to the whole membrane glyco-capture method followed by quantitative analysis using the LTQ-FT instrument. Two technical repli-

³ Nomenclature for glycosylation sites: an *N*-glycosylated residue is indicated in bold font and followed by an asterisk (**N***).

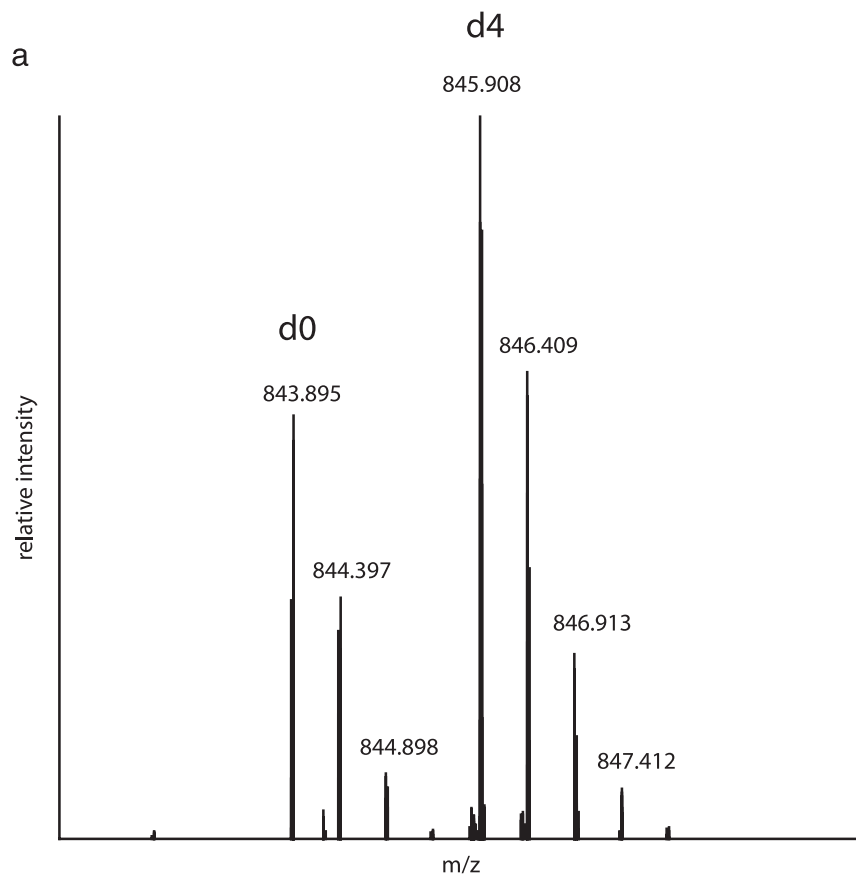
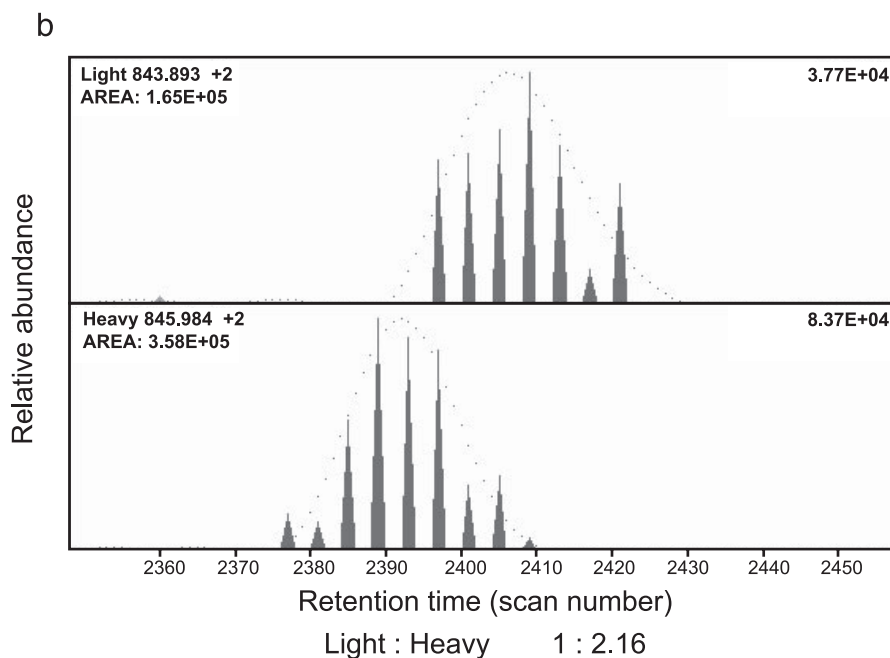


FIG. 6. Quantitation of InR using isotope-coded protein labeling. *a*, MS spectra obtained over the whole elution time of VDLEHAN*NTESPVR, a formerly glycosylated peptide originating from the InR. The doubly charged species shows a lower signal for the light labeled peptide (d_0 ; insulin-treated) than for the heavy labeled peptide (d_4 ; control). *b*, relative abundance and the calculated $d_0:d_4$ ratio obtained using XPRESS software.



cates were generated for all four samples, and data analysis was carried out as described above. The 132 *N*-glycosites identified and quantified in this experiment corresponded to 78 glycoproteins (supplemental Tables S11 and S12). Two distinct *N*-glycosites originating from the InR were identified

and quantified. The calculated abundance ratio from insulin-stimulated *versus* non-stimulated cells was 0.94 ± 0.24 , indicating that the InR was not degraded after 2 h of insulin stimulation. Together with the quantitative data of InR obtained by cell surface glycoprotein analysis we conclude that

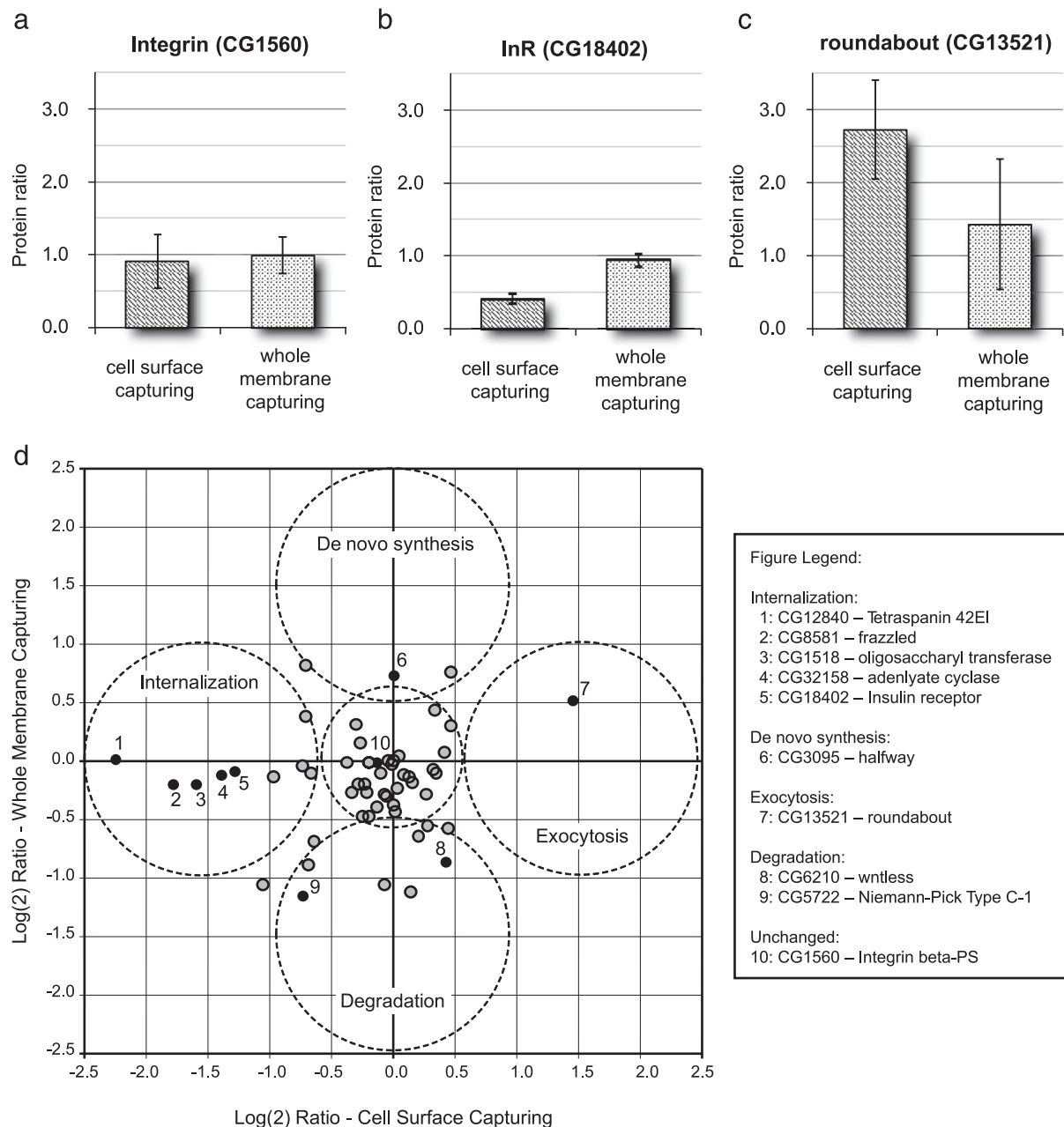


FIG. 7. Protein abundance changes on the cell surface versus the whole cell membrane upon insulin stimulation. Protein ratios from insulin-treated versus non-treated cells obtained by CSC and whole membrane capturing are shown for integrin (a), InR (b), and baboon (c). d, protein log ratios obtained from both CSC (x axis) and whole membrane capturing (y axis) are plotted. Only protein ratios obtained in both experiments are shown. Proteins mentioned in the text are depicted as *black dots*. The corresponding abundance ratios of proteins quantified by both the CSC and the whole membrane capturing method are shown in bold in supplemental Tables S8 and S11.

~50% of the total receptor was compartmentalized into endosomes.

Fig. 7 summarizes the different protein changes after insulin stimulation obtained by either the cell surface capturing or the whole membrane capturing approach. The abundance ratios of most glycoproteins observed in both analyses changed neither on the cell surface nor in the whole cell membrane. An example of this class is integrin β PS (CG1560) (Fig. 7a). There

were at least seven proteins, among them the protein frazzled (CG8581), tetraspanin 42 EI (CG12840), a membrane-spanning adenylate cyclase (CG32158), and a potential oligosaccharyltransferase (CG1518), that were down-regulated on the cell surface, whereas the whole membrane capturing revealed no significant abundance change, thus showing the same behavior as InR (Fig. 7b). Another interesting protein, roundabout (CG13521) (Fig. 7c), was up-regulated on the cell sur-

face, whereas the overall membrane content of this protein was found to be constant, indicating insulin-stimulated redistribution from an internal reservoir to the cell surface. The multipass membrane protein *wntless* (CG6210) that is specifically required for the secretion of *wingless* and thus participates in multiple development events was also slightly up-regulated on the cell surface, whereas the overall *wntless* protein content in the membrane decreased. The protein *halfway* (CG3095), involved in the antagonistic ecdysone signaling pathway, was found to be increased almost 2-fold in the whole cell membrane; however, the protein level of the same protein did not change on the cell surface, indicating insulin-stimulated synthesis and storage in the cell interior. Further NPC1 (Niemann-Pick Type C-1) protein (CG5722), which is involved in sterol metabolic processes, was decreased on the cell surface (0.60 ± 0.09) as well as in the whole membrane extract (0.45 ± 0.05) suggesting that the protein was degraded after insulin stimulation. In combination, these findings, summarized in Fig. 7d, show that the combination of CSC and whole membrane glyco-capturing enables us to study the dynamic distribution of membrane proteins after cell perturbation. In the case of insulin stimulation we observed multiple quantitative patterns for specific proteins that suggest cell internal protein redistribution, *de novo* synthesis, and degradation.

DISCUSSION

In this study we show that the highly selective CSC method allowed for the unambiguous identification of cell surface glycoproteins in *D. melanogaster* cells, including their sites of carbohydrate linkage, at a selectivity of almost 90% at the peptide level. Nonspecifically isolated peptides could be readily identified due to their lack of an NX(S/T) glycosylation motif. This led to the identification of 202 glycoproteins from *D. melanogaster* Kc167 cells of which 91% had at least one predicted transmembrane domain and 86% of the GO-annotated proteins belonged to the membrane. These results are consistent with results obtained using the same method on mammalian cells.² Insect cells, in contrast to mammalian cells, do not contain sialic acid, which is the outermost sugar residue in mammalian carbohydrate structures and, due to its *cis*-diols, readily accessible for hydrazide linkage via periodate oxidation (42). Furthermore we identified three peptides containing the NXC motif (43), which has not been described in insect cells to date and occurs very rarely compared with the NX(S/T) motif in mammalian cells. The data therefore indicate that the lack of terminal sialic acid does not limit our glycopeptide isolation strategy, a fact that extends the CSC method to other species lacking sialic acid.

To identify and quantify the isolated *N*-glycosites the samples were analyzed by ESI-MS/MS. Because one objective of the study was the quantitative comparison of the cell surface proteome of Kc167 cells in multiple perturbed states the sample throughput and accuracy of the quantitative proteom-

ics method were critical issues. Although quantification by spectral counting would be possible using the data obtained we decided to quantify based on the precursor intensities because it is more accurate, especially for proteins of lower abundance with few spectral counts, and is not limited to the quantification of the peptides identified by MS/MS. Moreover we chose to explore label-free quantification via feature pattern matching because this gave us the flexibility to compare any pattern with any other pattern acquired in the context of the study. This is unlike the situation in studies depending on stable isotope labeling where the samples to be compared have to be anticipated during the design of the study. Initially we compared the performance of ICPLs and label-free quantification whereby we evaluated the relative errors arising from the biological (difference between different biological isolates) and technical (difference caused by repeat analyses of identical samples) reproducibility. The data indicated that the observed protein abundance changes of the two methods were very similar and, consistent with this notion, that the major source of variance was rooted in the biological rather than technical reproducibility. The technical reproducibility of the label-free method was very high at CV 10–13%, whereas the biological reproducibility (including the whole *N*-glycosite isolation process) was somewhat lower at CV below 25%. Therefore, protein abundance changes of 50% and more could be reliably assigned. These data, achieved with samples isolated via a complex sample preparation method, are consistent with data obtained with much simpler sample preparation protocols (33), indicating that the CSC method has a level of reproducibility that is compatible with label-free quantification. Several aspects of the data also suggest that the method chosen is sensitive. First, we were able to quantify a higher number of *N*-glycosites and proteins than with the isotope labeling method. This was the result of the possibility of propagating MS1 features identified in one pattern or for that matter represented in the cell surface atlas over all the other patterns in the study (19). Second, we were able to identify and quantify G-protein-coupled receptor proteins that are known to be expressed in levels numbering in the upper hundreds to low thousands of copies per cell (32). Third, the method displayed a dynamic range exceeding 3 orders of magnitude. Furthermore a higher number of peptides could be quantified using the label-free method. At least in part this excellent sensitivity and dynamic range could be due to reduced ion suppression effects that are the result of the reduced sample complexity.

We used the combination of label-free quantitative mass spectrometry with CSC to test the hypothesis that the perturbation of signaling systems in the cell could be detected by quantitative changes in the cell surface proteome. Importantly the method does not necessarily show differences in protein abundance but rather differences in the amount of the respective glycopeptides between samples that could arise from changes in the protein concentration or changes in glycosy-

lation site occupancy. Perturbation of cultured *D. melanogaster* cells with different well characterized chemicals known to change important intracellular signaling systems indeed affected the cell surface glycoproteome in a specific manner. With one of the perturbations, the stimulation of Kc167 cells with insulin, we further attempted to distinguish different fates of proteins that showed quantitative changes on the cell surface. This was accomplished by comparing the quantitative pattern of cell surface proteins via CSC following insulin stimulation with the pattern of the same glycoproteins in the whole cell membrane using *N*-glycosite capturing. For InR we observed internalization but not significant degradation, a pattern that is in agreement with the findings observed by Knutson *et al.* (17) who used a heavy isotope density shift technique. Furthermore we observed additional cell surface glycoproteins exhibiting behavior similar to that of InR and proteins showing different fates. Notable is the protein roundabout, which showed a pattern opposite to InR suggesting that it might undergo exocytosis. By combining whole cell membrane and cell surface *N*-glycosite measurement we could therefore monitor biologically relevant quantitative changes and at least in part determine the underlying rationale for different patterns such as internalization without degradation, protein degradation, or intracellular redistribution.

In conclusion, we could assign a specific fingerprint indicating the origin, function, and physical state of a cell using our proteomics and informatics pipeline. Cell surface proteins are carrying out various important functions and are therefore important targets in many pharmacological studies. The findings in this study indicate that the cell surface proteome mirrors changes occurring within the cell. Because at least some of the cell surface glycoproteins are also secreted, shed, or otherwise released by the cell the systematic analysis of this important subproteome therefore opens the possibility of detecting molecular signatures in body fluids indicating the state of a cell or tissue. These results presented here are therefore of interest for the wide field of biomarker discovery where large research efforts are currently being invested.

Acknowledgments—We thank Martin Jünger and Matthias Gstaiger for helpful discussions and James Eddes for bioinformatics support.

* This work was supported, in whole or in part, by National Institutes of Health Grant N01-HV-28179 from NHLBI. This work was also supported by Swiss National Science Foundation Grant 31000-10767.

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

‡ Supported with funds from the NCCR Neural Plasticity and Repair.

|| To whom correspondence should be addressed. E-mail: aebersold@imsb.biol.ethz.ch.

REFERENCES

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
2. de Godoy, L. M., Olsen, J. V., de Souza, G. A., Li, G., Mortensen, P., and Mann, M. (2006) Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **7**, R50
3. Schmidt, A., Gehlenborg, N., Bodenmiller, B., Mueller, L. N., Campbell, D., Mueller, M., Aebersold, R., and Domon, B. (2008) An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures. *Mol. Cell. Proteomics* **7**, 2138–2150
4. Mueller, L. N., Brusniak, M. Y., Mani, D. R., and Aebersold, R. (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J. Proteome Res.* **7**, 51–61
5. Zhang, H., Yi, E. C., Li, X. J., Mallick, P., Kelly-Spratt, K. S., Masselon, C. D., Camp, D. G., II, Smith, R. D., Kemp, C. J., and Aebersold, R. (2005) High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **4**, 144–155
6. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999
7. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., and Aebersold, R. (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat. Methods* **4**, 231–237
8. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21**, 660–666
9. McDonald, L., Robertson, D. H., Hurst, J. L., and Beynon, R. J. (2005) Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. *Nat. Methods* **2**, 955–957
10. Marelli, M., Nesvizhskii, A. I., and Aitchison, J. D. (2008) Identifying bona fide components of an organelle by isotope-coded labeling of subcellular fractions: an example in peroxisomes. *Methods Mol. Biol.* **432**, 357–371
11. Kislinger, T., Cox, B., Kannan, A., Chung, C., Hu, P., Ignatchenko, A., Scott, M. S., Gramolini, A. O., Morris, Q., Hallett, M. T., Rossant, J., Hughes, T. R., Frey, B., and Emili, A. (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **125**, 173–186
12. Zola, H., Swart, B., Nicholson, I., Aasted, B., Bensussan, A., Boumsell, L., Buckley, C., Clark, G., Drbal, K., Engel, P., Hart, D., Horejsi, V., Isacke, C., Macardle, P., Malavasi, F., Mason, D., Olive, D., Saalmueller, A., Schlossman, S. F., Schwartz-Albiez, R., Simmons, P., Tedder, T. F., Ugucioni, M., and Warren, H. (2005) CD molecules 2005: human cell differentiation molecules. *Blood* **106**, 3123–3126
13. Brunner, E., Ahrens, C. H., Mohanty, S., Baetschmann, H., Loevenich, S., Potthast, F., Deutsch, E. W., Panse, C., de Lichtenberg, U., Rinner, O., Lee, H., Pedrioli, P. G., Malmstrom, J., Koehler, K., Schrimpf, S., Krijgsvelde, J., Kregenow, F., Heck, A. J., Hafen, E., Schlapbach, R., and Aebersold, R. (2007) A high-quality catalog of the *Drosophila melanogaster* proteome. *Nat. Biotechnol.* **25**, 576–583
14. Bodenmiller, B., Malmstrom, J., Gerrits, B., Campbell, D., Lam, H., Schmidt, A., Rinner, O., Mueller, L. N., Shannon, P. T., Pedrioli, P. G., Panse, C., Lee, H. K., Schlapbach, R., and Aebersold, R. (2007) PhosphoPep—a phosphoproteome resource for systems biology research in *Drosophila Kc167* cells. *Mol. Syst. Biol.* **3**, 139
15. Rinner, O., Mueller, L. N., Hubalek, M., Muller, M., Gstaiger, M., and Aebersold, R. (2007) An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nat. Biotechnol.* **25**, 345–352
16. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9390–9395
17. Knutson, V. P., Ronnett, G. V., and Lane, M. D. (1982) Control of insulin receptor level in 3T3 cells: effect of insulin-induced down-regulation and dexamethasone-induced up-regulation on rate of receptor inactivation. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 2822–2826
18. Gahmberg, C. G., and Tolvanen, M. (1996) Why mammalian cell surface proteins are glycoproteins. *Trends Biochem. Sci.* **21**, 308–311
19. Mueller, L. N., Rinner, O., Schmidt, A., Letarte, S., Bodenmiller, B., Brusniak, M. Y., Vittek, O., Aebersold, R., and Muller, M. (2007) SuperHirn—a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* **7**, 3470–3480
20. Arsenis, G., Hayes, G. R., and Livingston, J. N. (1985) Insulin receptor

- cycling and insulin action in the rat adipocyte. *J. Biol. Chem.* **260**, 2202–2207
21. Radimerski, T., Mini, T., Schneider, U., Wettenhall, R. E., Thomas, G., and Jenö, P. (2000) Identification of insulin-induced sites of ribosomal protein S6 phosphorylation in *Drosophila melanogaster*. *Biochemistry* **39**, 5766–5774
 22. Schmidt, A., Kellermann, J., and Lottspeich, F. (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5**, 4–15
 23. Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Paton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W., and Aebersold, R. (2004) A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.* **22**, 1459–1466
 24. Eng, J., McCormack, A. L., and Yates, J. R., III (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
 25. Keller, A., Eng, J., Zhang, N., Li, X. J., and Aebersold, R. (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* **1**, 2005.0017
 26. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392
 27. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658
 28. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951
 29. Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379
 30. Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182
 31. Al-Shahrour, F., Minguez, P., Vaquerizas, J. M., Conde, L., and Dopazo, J. (2005) BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucleic Acids Res.* **33**, W460–W464
 32. Stanasila, L., Pattus, F., and Massotte, D. (1998) Heterologous expression of G-protein-coupled receptors: human opioid receptors under scrutiny. *Biochimie (Paris)* **80**, 563–571
 33. Wang, G., Wu, W. W., Zeng, W., Chou, C. L., and Shen, R. F. (2006) Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: reproducibility, linearity, and application with complex proteomes. *J. Proteome Res.* **5**, 1214–1223
 34. Chen, W., White, M. A., and Cobb, M. H. (2002) Stimulus-specific requirements for MAP3 kinases in activating the JNK pathway. *J. Biol. Chem.* **277**, 49105–49110
 35. Botella, J. A., Baines, I. A., Williams, D. D., Goberdhan, D. C., Proud, C. G., and Wilson, C. (2001) The *Drosophila* cell shape regulator c-Jun N-terminal kinase also functions as a stress-activated protein kinase. *Insect Biochem. Mol. Biol.* **31**, 839–847
 36. Standaert, M. L., and Pollet, R. J. (1984) Equilibrium model for insulin-induced receptor down-regulation. Regulation of insulin receptors in differentiated BC3H-1 myocytes. *J. Biol. Chem.* **259**, 2346–2354
 37. Radimerski, T., Montagne, J., Rintelen, F., Stocker, H., van der Kaay, J., Downes, C. P., Hafen, E., and Thomas, G. (2002) dS6K-regulated cell growth is dPKB/dPI(3)K-independent, but requires dPK1. *Nat. Cell Biol.* **4**, 251–255
 38. Kolodziej, P. A., Timpe, L. C., Mitchell, K. J., Fried, S. R., Goodman, C. S., Jan, L. Y., and Jan, Y. N. (1996) frazzled encodes a *Drosophila* member of the DCC immunoglobulin subfamily and is required for CNS and motor axon guidance. *Cell* **87**, 197–204
 39. Burgess, J. W., Wada, I., Ling, N., Khan, M. N., Bergeron, J. J., and Posner, B. I. (1992) Decrease in β -subunit phosphotyrosine correlates with internalization and activation of the endosomal insulin receptor kinase. *J. Biol. Chem.* **267**, 10077–10086
 40. Authier, F., Rachubinski, R. A., Posner, B. I., and Bergeron, J. J. (1994) Endosomal proteolysis of insulin by an acidic thiol metalloprotease unrelated to insulin degrading enzyme. *J. Biol. Chem.* **269**, 3010–3016
 41. Doherty, J. J., II, Kay, D. G., Lai, W. H., Posner, B. I., and Bergeron, J. J. (1990) Selective degradation of insulin within rat liver endosomes. *J. Cell Biol.* **110**, 35–42
 42. Tomiya, N., Narang, S., Lee, Y. C., and Betenbaugh, M. J. (2004) Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. *Glycoconj. J.* **21**, 343–360
 43. Medzhradzky, K. F. (2008) Characterization of site-specific N-glycosylation. *Methods Mol. Biol.* **446**, 293–316