Published in final edited form as:

J Proteomics. 2012 December 5; 76(0 0): . doi:10.1016/j.jprot.2012.08.010.

A label-free proteome analysis strategy for identifying quantitative changes in erythrocyte membranes induced by red cell disorders

Esther N. Pesciotta^a, Sira Sriswasdi^{b,c}, Hsin-Yao Tang^b, Philip J. Mason^a, Monica Bessler^{a,d}, and David W. Speicher^{b,c,*}

^aDivision of Hematology, Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA, USA

^bCenter for Systems and Computational Biology and Molecular and Cellular Oncogenesis Program, The Wistar Institute, Philadelphia, PA, USA

^cGenomics and Computational Biology Graduate Group, The University of Pennsylvania, Philadelphia, PA, USA

^dDivision of Hematology/Oncology, Department of Internal Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Red blood cells have been extensively studied but many questions regarding membrane properties and pathophysiology remain unanswered. Proteome analysis of red cell membranes is complicated by a very wide dynamic range of protein concentrations as well as the presence of proteins that are very large, very hydrophobic, or heterogeneously glycosylated. This study investigated removal of other blood cell types, red cell membrane extraction, differing degrees of fractionation using 1-D SDS gels, and label-free quantitative methods to determine optimized conditions for proteomic comparisons of clinical blood samples. The results showed that fractionation of red cell membranes on 1-D SDS gels was more efficient than low-ionic-strength extractions followed by 1-D gel fractionation. When gel lanes were sliced into 30 uniform slices, a good depth of analysis was obtained that included identification of most well-characterized, low-abundance red cell membrane proteins including those present at 500 to 10,000 copies per cell. Furthermore, the size separation enabled detection of changes due to proteolysis or *in vivo* protein crosslinking. A combination of Rosetta Elucidator quantitation and subsequent statistical analysis enabled the robust detection of protein differences that could be used to address unresolved questions in red cell disorders.

1. Introduction

The red blood cell (RBC), unlike other eukaryotic cells, lacks a nucleus, other organelles, and the capacity to synthesize proteins. The RBC membrane and associated membrane skeleton proteins provide the flexibility that is required for red cells to pass through small

© 2012 Elsevier B.V. All rights reserved.

*Corresponding Author: David W. Speicher, Ph.D., The Wistar Institute, 3601 Spruce St., Room 272a, Philadelphia, PA 19104, USA. Phone: 215-898-3972. speicher@wistar.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

capillaries for gas transport and exchange. The ease of obtaining RBC, the readily obtained high purity of plasma membrane preparations, and the relatively simple protein composition make the RBC membrane one of the best-characterized membrane systems. Yet, despite years of research, there are many unanswered questions regarding RBC membrane properties and pathophysiology. In this context, unbiased proteomic studies provide a valuable platform for understanding how the red cell proteome is altered in erythrocyte disorders.

Previous red blood cell proteomic studies have used fractionation techniques to lower the complexity of the sample and enable more protein identifications. These techniques include white ghost (WG) analysis on 1-[1] or 2-DE gels,[2] in-solution digestion of four RBC fractions (white ghosts, cytoplasmic proteins, inside out vesicles (IOV), and membrane skeletal proteins),[3] and membrane protein extraction with detergents followed by in-solution digestion for multidimensional protein identification technology analysis (MudPIT). [4] Several studies used 2-DE as the preferred fractionation method for LC-MS/MS analysis of the RBC proteome.[5–8] However, 2-DE analysis of hydrophobic membrane proteins is not ideal and detection can vary significantly depending upon individual protein structures and characteristics.[9] Conversely, 1-D SDS gels provide the advantage of separation by protein size without the risk of losing highly hydrophobic proteins. Additionally, 1-D SDS gels ensure consistent protein loads for LC-MS/MS analysis and provide insight into size variations of individual proteins prior to digestion.

Our current approach was adapted from an in-depth analysis of the RBC proteome in which a stringent protocol was used to eliminate contaminating non-red cells.[10] In this robust study by Pasini et al, the peripheral blood was stored to allow maturation of reticulocytes, and then RBC were passed through a leukocyte depletion filter, density filter, and nylon nets prior to RBC washing and lysis. The white ghosts were then subjected to several processes such as a high pH extraction with sodium carbonate, ethanol solubilization, and membrane skeleton extraction to maximize protein identifications while obtaining information on interactions with the lipid bilayer. While such rigorous processing was successful in verifying the RBC proteome, a more streamlined approach is desirable when comparing red cell proteomes in multiple clinical specimens. Additionally, when studying rare hematological diseases, particularly those that affect young children, it is critical to generate methods that are compatible with small sample volumes and are highly consistent, so that label-free quantitative methods can be used. This study investigates RBC extraction, differing levels of fractionation, and label-free quantitative methods to determine optimized conditions for proteomic analysis for the comparison of clinical blood samples.

2. Materials and Methods

2.1 RBC sample preparation

Typically, 5–10 mL of whole blood was obtained with informed consent and collected in K₂EDTA. Samples and buffers were kept at 0–4°C throughout the procedure to minimize proteolysis. The procedure was modified from previously published methods.[10] RBC were isolated from the plasma by centrifugation for 10 min at 150 x g before being resuspended and passed through a leukocyte depletion filter (Plasmodipur®, Accurate Chemical & Scientific Corp., Westbury, NY). Red cells were washed four times with PBS at 1700 x g for 20 min and stored overnight on ice at 4°C in PBS with 10 mM glucose and 0.15 mM PMSF. Cells were lysed with 10 volumes of hypotonic lysis buffer (5 mM sodium phosphate, 1mM EDTA, 0.1 mM DFP [diisopropylfluorophosphate], pH 8.0). White ghosts were isolated by washing with lysis buffer and membranes were collected by centrifugation at 31,000 x g for 35 min using a total of four to five washes. WG were fractionated further into IOV and membrane skeletal proteins, an extract referred to as dilute crude spectrin (DCS), by

incubating WG with five volumes of 100 μ M EDTA, 0.5 mM β -mercaptoethanol, and 0.5 mM DFP, pH 9.6 at 37°C for 20 min. IOV were pelleted at 45,000 $\times$ g for 40 min at 4°C, and DCS was collected as the supernatant. Protein concentrations of WG, IOV, and DCS were determined using a Modified Lowry Protein Assay Kit (Thermo Scientific, Rockford, IL). Samples were analyzed with SDS-PAGE on 15-well, 3–8% Tris-acetate gels (Invitrogen, Carlsbad, CA) with Tris-Acetate SDS buffer, using maximized optimal loads of 8 μ g for WG, 4 μ g for IOV, and 1.5 μ g for DCS, followed by staining with colloidal Coomassie (Invitrogen) (Supplemental Information Figure S1). Aliquots of each fraction were stored at –80°C for potential future use.

2.2 In-gel digestion and LC-MS/MS

Preparative gels were run to 3 cm on 10-well, 10% Bis-Tris mini gels with MOPS SDS buffer (Invitrogen). Protein loads for WG, IOV, and DCS samples were increased 50% relative to the analytical gels described above due to the wider lane width on 10-lane gels (Supplemental Information Figure S2). Gel lanes from each sample were cut into 30 $\times$ 1-mm slices. For some samples, three adjacent slices were combined to yield a 10-fraction digest. Alternatively, corresponding slices from three gel lanes were combined to produce a 30-fraction digest. Gel slices were reduced with 20 mM Tris(2-carboxyethyl)phosphine (TCEP) for 15 min at 37°C prior to alkylation with 40 mM iodoacetamide for 30 min at 37°C. Gel slices were then dried and digested with 0.8 μ g sequence grade modified trypsin (Promega Corp., Madison, WI) overnight at 37°C. The digestion was quenched the following day with 40 mM ammonium bicarbonate/3% formic acid. To evaluate the effects of different extents of fractionation, aliquots of a 30-fraction WG digestion were pooled to give 10 fractions (three adjacent digests were pooled) and 15 fractions (two adjacent fractions were pooled). Digests were stored at –20°C prior to LC-MS/MS analysis, which was carried out using a nanoACQUITY HPLC (Waters, Milford, MA) and LTQ-Orbitrap XL (Thermo Scientific) mass spectrometer. For each analysis, 8 μ L of tryptic digest was loaded onto a 180 μ m $\times$ 20 mm trap column packed with 5 μ m Symmetry C18 resin (Waters) using solvent A (0.1% formic acid in Milli-Q H₂O [Millipore, Billerica, MA]) for 5 min followed by separation in a 75 μ m $\times$ 250 mm analytical 1.7 μ m BEH130 C18 column (Waters). The peptides were eluted with a gradient of solvent B (0.1% formic acid in acetonitrile): 5–28% B over 42 min, 28–50% B over 25.5 min, 50–80% B over 5 min, and constant 80% B for 5 min. A 25-min blank gradient was run in between each sample injection to minimize peptide carryover. Full scans were carried out from 400–2000 m/z with 60,000 resolution. MS2 data were acquired through data-dependent analysis of the top six most intense ions with dynamic exclusion enabled for 60 sec, monoisotopic precursor selection enabled, and single charged ions were rejected.

2.3 Data processing

MS/MS raw files were extracted for MS2 and searched using the SEQUEST algorithm (Ver. 28, rev. 13, University of Washington, Seattle, WA) in BioWorks (Ver. 3.3.1 SP1, Thermo Fisher Scientific). The FASTA database (human UniRef 100, Ver. June, 2011) was downloaded from Protein Information Resource (PIR), Georgetown University, Washington, DC. A decoy database was generated by using the reverse sequences from all of the proteins listed in UniRef 100 which was then appended to the forward database. This combined dataset was then indexed to include either full or partial tryptic peptide specificity. SEQUEST search parameters were set to include two missed cleavages, a 100 ppm mass tolerance, variable methionine oxidation (+15.9949) and static cysteine modification (+57.0215). Protein lists were generated using DTASelect (Ver. 2.0, licensed from Scripps Research Institute, La Jolla, CA) after searching in the partial tryptic database. Data filters for full tryptic peptides, 10 ppm mass tolerance, a minimum of two peptides per locus, and a minimum Δ Cn score of 0.05 were applied in DTASelect to minimize the false discovery rate

(FDR). The FDR was estimated as the ratio of decoy database peptide or protein hits divided by the total of forward and reverse matches. Rosetta Elucidator software (Ver. 3.3, Rosetta Biosoftware, Seattle, WA) was utilized to align full mass scan signals and identify features across all raw data files.[11–13] The raw files were trimmed to a 20–85-min window and only signals within a m/z range of 400–2000 were considered. DTAs were created for high-quality features with a peak time score > 0.7 , a peak m/z score > 0.8 , and $1 > z > 5$. These DTAs were searched using the database indexed for full tryptic cleavage using SEQUEST, as previously described.[13] Protein and Peptide Tellers were used to compute the probability that a protein is present in a sample based on the combined probability of the associated peptides. The data was then filtered to include ProteinTeller scores > 0.95 and PeptideTeller scores > 0.8 , thereby minimizing the FDR.

2.4 Label-free quantitation

Label-free quantitation of protein levels was assessed using protein intensity levels provided by the Rosetta Elucidator software, as previously described.[13] The experiment was setup in Elucidator to include three treatment groups (WG, IOV, DCS) with alignment based on the WG sample B data. Technical replicates were performed on a IOV digest (IOV-1 and IOV-2) and the 10-fraction WG analyses of samples A and B from the same donor were biological replicates. To evaluate inter-subject variability we compared two pools of WG from three age, sex, and race-matched donors, denoted Group-1 and Group-2 (biological variants). Protein quantitative comparisons were only performed for the pairwise IOV technical replicates, WG biological replicates, and the pooled WG samples. This is because label-free software, such as Elucidator, is designed to compare samples with substantial overlap in protein composition that have been through identical processing steps. In contrast, IOV and DCS were extracted from WG and contain quite different protein compositions, making label-free quantitative comparisons between these samples sub-optimal. Conversely, any samples of the same type that has been processed in the same manner, e.g. analysis of two or more WG samples prepared from RBC of different donors, contain similar protein compositions and thereby enable proper alignment for quantitative comparisons. The capacity of this label-free analysis approach to detect clinically meaningful differences in patients has been previously demonstrated[13] Data was normalized based on the median $\log_{10}$ intensity ratio of either IOV-1/IOV-2, WG-A/WG-B, or Group-1/Group-2 (see Supplementary Information). An intensity cut-off of 1×10^5 was used to minimize noise. This threshold was determined based on a maximum coefficient of variation of 5%. Differences in protein levels were quantified using a 95% confidence interval and, based on this analysis, fold-changes greater than 3 were considered significant.

3. Results

3.1 Proteome coverage of RBC membrane extracts

Efficient, reproducible RBC processing and sample handling is critical for obtaining a consistent and comprehensive proteome with minimal contamination from other cell types. When working with small volumes of peripheral blood it is important to find a balance that maximizes protein yield while minimizing proteolysis and contaminants. In the current workflow, proteolysis was reduced by keeping all reagents and samples at 0–4°C during processing and using non-protein protease inhibitors, i.e., EDTA and DFP. A practical and straightforward approach to remove contaminant cells involves an initial removal of blood plasma and the bulk of the non-RBC cells, followed by passing the red cells through a leukocyte depletion filter. The flow diagram in Figure 1 depicts the strategy used to isolate RBC membranes and fractionate the white ghosts into inside-out-vesicles and membrane skeletal extracts. Extracting IOV and DCS from WG was expected to reduce the complexity

of the samples, potentially allowing for a greater depth of analysis with an increase in protein identifications.

Two small volume (less than 10 ml) blood samples (labeled A and B) were collected from the same healthy donor approximately one month apart, to analyze the variability between biological replicates. Separation of WG-B into IOV and DCS was used to analyze the impact of low-ionic-strength extraction on depth of analysis. IOV technical replicates (IOV-1 and IOV-2) were also used to monitor analytical reproducibility. Interestingly, the spectral counts were similar for all 10-fraction runs (Supplemental Information Table S1). Figure 2 shows the unique peptides and proteins identified in each dataset. WG-B provides the most unique peptide and protein identifications with 37, 38, and 22% more peptides and 38, 39, and 37% more proteins (two or more peptides) than in IOV-1, IOV-2, and DCS, respectively. The IOV technical replicates show good reproducibility with similar peptide and protein counts. Comparison of the protein lists from IOV-1 and IOV-2 show approximately 45 proteins unique to each dataset. While this number may seem large, a recent study from our group showed that when comparing protein lists, the majority of apparently unique proteins could actually be attributed to either variations in assignments of peptides to homologous proteins or to slight variations in the number of high-confidence peptides identified in replicate datasets, i.e., two peptides identified one dataset (protein identification retained) vs. one peptide in the alternate dataset (protein filtered out).[14]

Comparison of protein lists for IOV-1 and DCS showed far less overlap, with only 104 common proteins of the approximately 300 proteins identified in each dataset. Given that WG was divided into IOV and DCS during the extraction procedure, and this extraction is known to be incomplete, it is not surprising that there is substantial but incomplete overlap between these sub-proteomes. To directly compare potential benefits of fractionating WG into DCS and IOVs, the DCS and IOV-1 datasets were combined using DTASelect. Comparison of this combined dataset with WG-B shows 519 more peptide identifications in the IOV-1 and DCS dataset (Figure 2A). However, the overall protein numbers were similar with a total of 492 proteins in the WG-B dataset and 507 proteins in the IOV-1 and DCS combined dataset. Thus, carrying out 10 LC-MS/MS runs of WG alone provides roughly the same results as 10 LC-MS/MS runs each of IOV and DCS. Taken together, it is clear that WG extraction into IOV and DCS fractions does not significantly improve proteome coverage and does not justify the additional mass spectrometer analysis time.

3.2 Optimizing throughput by fractionation

When designing proteomics discovery workflows, there is often a balance between maximizing the number of proteins and minimizing LC-MS/MS runs in order to efficiently utilize instrument time. Typically, more extensive fractionation and even repeat injections result in more protein identifications of complex proteomes. Since WG proteomes are less complex than biological fluids or cell lysates, the effects of analyzing differing numbers of fractions were assessed using the WG-A sample. For this experiment, 3-cm gels were sliced into 30×1 -mm fragments and corresponding slices from three gel lanes were pooled for digestion with trypsin. Aliquots of these 30 digests were then pooled to create 15 and 10 fractions.

As expected, the total spectra increased linearly with the number of LC-MS/MS runs with a 1.5-fold increase from 10 to 15 runs and a two-fold increase from 15 to 30 runs (Supplemental Information Table S1). The total number of unique proteins also increased as sample fractionation increased (Figure 3), with an additional 71 proteins identified in the 15-run dataset and an additional 294 proteins in the 30-run dataset when compared to the 10-run data (with a two peptide per protein minimum requirement). These data indicate that increasing gel slices and the number of MS runs substantially improves depth of analysis.

Maximizing depth of analysis is usually desirable because low-abundance proteins are often of greater interest than easily detectable higher abundance proteins. The distribution of protein abundance can be inferred by normalizing the spectral counts of each protein to its molecular weight because using spectral counts alone is biased towards larger proteins. Examination of normalized spectral counts shows that WG-A (30) provides the greatest number of spectra for lower abundance proteins therefore providing a larger dynamic range and more extensive protein sequence coverage (Supplementary Information Figure S3).

In addition to a greater dynamic range, the 30-slice proteome also provides greater resolution of different molecular forms for abundant proteins. Figure 4 illustrates the increased resolution of variations in spectral counts across gel slices for eight abundant proteins in the 10-, 15-, and 30-slice proteomes. For example, the spectral count distribution of Band 3 shows the expected broad peak in gel slices 11–16 of the 30-slice proteome due to heterogeneous glycosylation. In addition, there are three distinct lower molecular weight peaks indicative of Band 3 proteolysis in the 30-slice proteome, which becomes less resolved in the 10- and 15-slice proteomes. This increased size resolution can have a profound impact on the detection of changes in mRNA alternative splicing, post-translational modifications, and proteolysis. Additionally, enhanced size resolution can provide insights into changes in *in vivo* covalent crosslinking, presumably due to the activity of transglutaminases.[15] Examples include the detection of Protein 4.1 in higher molecular weight regions (gel slices 4–5 and 10–13 of the 30-slice proteome) and the large spectral counts on the leading ends of α and β -spectrin (gel slices 1–4 and 1–6, respectively). Thus, separation of intact proteins by 1-D SDS gels for LC-MS/MS analysis can provide critical information about the state of a protein.[16]

3.3 Label-free quantitation for comparative analysis

Previous label-free methods have been employed on the red blood cell proteome for comparative quantitation of mouse models using CRAWDAD[17] and MaxQuant[18] software. In this study, label-free quantitation of protein differences using the LC-MS signal is facilitated by the use of Rosetta Elucidator software. This software aligns MS features across multiple raw files and calculates the protein intensity based on peak heights of associated peptides. To evaluate the reproducibility of our analytical pipeline, we quantitatively compared the IOV technical replicates and WG biological replicates described above, as well as WG pools from two sets of three donors to test the inter-variability between different groups of normal donors. Figure 5A illustrates the fold change distribution between IOV-1 and IOV-2, showing most proteins (with two or more peptides) fall within a three-fold change, indicating good technical reproducibility. In addition, a 95 % confidence interval derived from the standard deviation of the mean protein intensities (Supplemental Figure S4–6) was applied to determine significant changes in protein levels (Figure 5B). A total of 42 proteins fall outside of these 95% boundaries for the IOV replicates. Figure 5B also shows an increase in noise as the protein intensity decreases. Consequently, the coefficient of variance increases rapidly for protein intensities below 1×10^5 (Figure 5C). Therefore, an intensity cut-off of 1×10^5 was applied to all datasets analyzed by this method to reduce the noise and variation in low-abundance proteins. After applying this threshold, the number of proteins identified outside of the 95% confidence interval decreased to 14 for the IOV replicates (Figure 5D).

While the IOV technical replicates indicate good analytical performance, the WG biological replicates include the variability that is intrinsically present when comparing proteomes from different blood draws. First, the differences in the biological replicates can be assessed by applying the 95% confidence intervals and three-fold change ratios from the IOV replicate analysis for all proteins with intensities above 1×10^5 (Figure 6A). This results in 18

proteins that are significantly different between biological WG replicates. However, when carrying out the analogous statistical characterization on the WG samples alone, the 95% confidence interval changes drastically to accommodate the moderately increased variability between the two datasets in the higher intensity region (Figure 6B). This results in only nine proteins that fall outside of the boundary.

To illustrate the versatility of this technique and its ability to account for biological variation between different donor samples, we compared two control pools. In each pool, equal WG protein amounts were combined from three age, sex, and race-matched donors and subjected to the same proteomics pipeline as described above. Label-free comparison of different donor pools, designated Group-1 and Group-2, showed more variation in protein intensities. Table 1 provides the percentage of proteins that are maintained within different fold change ratios for the IOV-1 and -2 technical replicates, WG-A and -B biological replicates, and Group-1 and -2 biological variants. The technical and biological replicates are highly consistent, with over 97% of proteins exhibiting less than a 3-fold change. However, when comparing different groups of controls, less than 84% of the proteins were within a 3-fold change. Therefore, this label-free analysis pipeline and corresponding statistical analysis provides a straightforward approach to quantify differences in protein levels between any two blood samples, without the requirement of isotopic labels.

4. Discussion

One important consideration for RBC proteomics is the presence of contaminating non-red cells and the detection of proteins that are not native to the RBC. Hence, even if contaminating cells are present at very low levels, such as one macrophage or platelet per million red cells, abundant macrophage or platelet proteins may be detected in the RBC proteome analysis as low-abundance proteins. A recent review listed 50 blood-cell-type-specific proteins that were only found in purified constituents.[19] Using the data filters described above, only five of 41 non-erythrocyte proteins were found in our most extensive WG-A (30) dataset (plasma membrane calcium transporting ATPase and tyrosine-protein kinase SYK from monocytes, and junctional adhesion molecule A precursor, Ras-related protein Rab-6B, and Syntaxin-11 from platelets) (Supplemental Information Table S2). Three of these five proteins (plasma membrane calcium-transporting ATPase 4, junctional adhesion molecule A precursor, and Ras-related protein Rab-6B) were also present in the comprehensive RBC proteome reported by Pasini et al in which exhaustive measures were used to eliminate contaminant cells.[10] Moreover, we did not identify any other well-known leukocyte or platelet proteins such as lactoferrin A, elastase 2, myeloperoxidase, eosinophil peroxidase, or any histone proteins.

The drawback to eliminating low-abundance proteins from the dataset by requiring more peptides or spectral counts or by using less fractionation to minimize contaminating proteins is that lower abundance RBC proteins may be lost. We evaluated our capacity to detect low-abundance RBC proteins by comparing the WG-A 30-fraction proteome to a review by Burton et al in which they searched the literature and compiled a list of 49 well-characterized RBC membrane proteins ranging in copy number from 500 to 1.2 million per cell.[20] We were able to identify 11 of the 13 known lower abundance proteins with less than 10,000 copies per cell, including CD99, Kell glycoprotein, complement receptor 1, CD58, CD44, Lutheran glycoprotein, LW glycoprotein, and NaK-ATPase. Importantly, these proteins were all detected in the WG-A (30) dataset by three or more peptides, further justifying the data filters used here. The exceptions were the Duffy antigen, which was detected by a single peptide and was filtered from the final dataset on this basis, as well as the XG glycoprotein, which was not found. In addition, all of the more abundant proteins in

the Burton et al list were identified in our WG-A 30-fraction proteome except for several minor glycophorin isoforms (glycophorin B, D, and E) and carbonic anhydrase 4.

The depth of analysis of the current study compares favorably with prior RBC proteomics studies. Our most extensive dataset from 30 fractions of WG-A provided an excellent depth of analysis with 842 proteins identified by two peptides or more. Early proteomics studies of red cell membranes were limited by instrumental capabilities and minimal fractionation prior to LC-MS/MS, as seen in the pilot study by Kakhniashvili et al[3] in which IOV and spectrin extracts were analyzed to identify 91 erythroid membrane proteins. An extensive study done by Pasini and coworkers[10] utilized WG extraction and 1-D SDS-PAGE fractionation prior to LC-MS/MS to identify 340 membrane proteins. Using an alternative MudPit approach, DePalma et al[4] identified 275 membrane proteins from WG extracts. A recent publication by van Gestel et al.[21] utilized 2-D blue-native/SDS-PAGE to fractionate WG and identify 524 membrane proteins. Comparison of our WG-A 30 fraction protein list to these previous studies showed variable overlap ranging from 29–53% based upon protein descriptions. While these percentages may seem low, the comparison of datasets from different search engines and particularly using databases is not straightforward when the only basis for comparison is protein descriptions. The large redundancy in databases and highly variable names for the same protein or closely related isoforms limit the capacity to match hits across studies, thereby leading to apparent anomalously low overlaps.

One surprising result that emerged from this study was the observation that further fractionation of WG into a low-ionic-strength extract and the residual IOVs was not superior to direct proteome analysis of WG alone. In general, further fractionation of complex proteomes results in greater depth of analysis with identification of more proteins and greater sequence coverage. However, typically, further fractionation enables the downstream analysis of a larger percentage of the original starting sample. This is not the case for red cell membranes when using the 1-D gel approach, as is reflected by the images in Figure 1. The maximum load that can be applied to a 1-D gel without extensive band broadening is limited in white ghosts by the very abundant spectrin and Band 3 proteins. Since the majority of spectrin is segregated into the DCS fraction, proportionately larger protein loads are not practical. Similarly, Band 3 is primarily separated into the IOV fraction, thus limiting larger protein loads. While the results might be somewhat different for alternative downstream fractionation methods, the advantage of being able to evaluate different molecular forms of higher abundance proteins outweighs any minor advantage that might be achieved by utilizing further fractionation. In addition, further fractionation introduces an additional variable, as the relatively crude extraction and centrifugation may lead to varying degrees of cross-contamination between these two fractions.

5. Conclusion

Low ionic strength extraction of WG with separate analysis of the resulting pellet and supernatant was not superior to direct proteome analysis of WG. The high abundance Band 3 and spectrin proteins, present in the IOV and DCS fractions, limit the gel protein load and therefore do not provide an advantage over WG analysis. Separation of WG into 30 fractions using 1-D SDS gels yielded an excellent depth of analysis and provided insights into protein size heterogeneity. The combination of Rosetta Elucidator quantitation and subsequent statistical analysis enabled the determination of significant changes in protein abundance that can be used to address unresolved questions in red cell disorders. As the field of RBC proteomics and subsequent data analysis progresses, more information can be extracted from the data such as insights into the RBC interactome[22] and models for erythrocyte

metabolism.[23] This will allow us to generate hypotheses to identify key pathways and mechanisms involved when performing quantitative red cell comparisons.

The method described in this study showed robust reproducibility with minimal biological variation between multiple blood draws from the same donor. Hence, this analytical platform should provide a sensitive method for detecting changes in RBC membrane proteomes related to red cell disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the administrative assistance of Mea Fuller and the assistance of the Wistar Institute Proteomics Core Facility. This work was supported in part by NIH grant HL38794 (to D. W. S.) and DK084188 to (M. B.), and an Institutional Cancer Center Support Grant CA010815.

References

1. Bosman GJ, Lasonder E, Luten M, Roerdinkholder-Stoelwinder B, Novotny VM, Bos H, De Grip WJ. The proteome of red cell membranes and vesicles during storage in blood bank conditions. *Transfusion*. 2008; 48:827–35. [PubMed: 18346024]
2. Low TY, Seow TK, Chung MCM. Separation of human erythrocyte membrane associated proteins with one-dimensional and two-dimensional gel electrophoresis followed by identification with matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Proteomics*. 2002; 2:1229–39. [PubMed: 12362340]
3. Kakhniashvili DG, Bulla LA, Goodman SR. The human erythrocyte proteome - Analysis by ion trap mass spectrometry. *Molecular & Cellular Proteomics*. 2004; 3:501–9. [PubMed: 14963112]
4. De Palma A, Roveri A, Zaccarin M, Benazzi L, Daminelli S, Pantano G, Buttarello M, Ursini F, Gion M, Mauri PL. Extraction methods of red blood cell membrane proteins for Multidimensional Protein Identification Technology (MudPIT) analysis. *Journal of Chromatography A*. 2010; 1217:5328–36. [PubMed: 20621298]
5. Peker S, Akar N, Demiralp DO. Proteomic identification of erythrocyte membrane protein deficiency in hereditary spherocytosis. *Molecular Biology Reports*. 2012; 39:3161–7. [PubMed: 21706353]
6. Kakhniashvili DG, Griko NB, Bulla LA Jr, Goodman SR. The proteomics of sickle cell disease: profiling of erythrocyte membrane proteins by 2D-DIGE and tandem mass spectrometry. *Experimental Biology and Medicine* (Maywood). 2005; 230:787–92.
7. D'Amici GM, Rinalducci S, Zolla L. Proteomic analysis of RBC membrane protein degradation during blood storage. *Journal of Proteome Research*. 2007; 6:3242–55. [PubMed: 17585793]
8. Bruschi M, Seppi C, Arena S, Musante L, Santucci L, Balduini C, Scaloni A, Lanciotti M, Righetti PG, Candiano G. Proteomic analysis of erythrocyte membranes by soft Immobiline gels combined with differential protein extraction. *Journal of Proteome Research*. 2005; 4:1304–9. [PubMed: 16083280]
9. Rabilloud T. Membrane proteins and proteomics: love is possible, but so difficult. *Electrophoresis*. 2009; 30 (Suppl 1):S174–80. [PubMed: 19517508]
10. Pasini EM, Kirkegaard M, Mortensen P, Lutz HU, Thomas AW, Mann M. In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood*. 2006; 108:791–801. [PubMed: 16861337]
11. Pawletz CP, Wiener MC, Bondarenko AY, Yates NA, Song QH, Liaw A, Lee AYH, Hunt BT, Henle ES, Meng FY, Sleph HF, Holahan M, Sankaranarayanan S, Simon AJ, Settlege RE, Sachs JR, Shearman M, Sachs AB, Cook JJ, Hendrickson RC. Application of an End-to-End Biomarker Discovery Platform to Identify Target Engagement Markers in Cerebrospinal Fluid by High

- Resolution Differential Mass Spectrometry. *Journal of Proteome Research*. 2010; 9:1392–401. [PubMed: 20095649]
12. Neubert H, Bonnert TP, Rumpel K, Hunt BT, Henle ES, James IT. Label-free detection of differential protein expression by LC/MALDI mass spectrometry. *Journal of Proteome Research*. 2008; 7:2270–9. [PubMed: 18412385]
 13. Beer LA, Tang HY, Sriswasdi S, Barnhart KT, Speicher DW. Systematic Discovery of Ectopic Pregnancy Serum Biomarkers Using 3-D Protein Profiling Coupled with Label-free Quantitation. *Journal of Proteome Research*. 2011; 10:1126–38. [PubMed: 21142075]
 14. Wang H, Chang-Wong T, Tang HY, Speicher DW. Comparison of Extensive Protein Fractionation and Repetitive LC-MS/MS Analyses on Depth of Analysis for Complex Proteomes. *Journal of Proteome Research*. 2010; 9:1032–40. [PubMed: 20014860]
 15. Bennett V, Baines AJ. Spectrin and ankyrin-based pathways: Metazoan inventions for integrating cells into tissues. *Physiological Reviews*. 2001; 81:1353–92. [PubMed: 11427698]
 16. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews Molecular Cell Biology*. 2004; 5:699–711.
 17. Wooden JM, Finney GL, Rynes E, MacCoss MJ, Lambert AJ, Robledo RF, Peters LL, Gilligan DM. Comparative proteomics reveals deficiency of SLC9A1 (sodium/hydrogen exchanger NHE1) in beta-adducin null red cells. *British Journal of Haematology*. 2011; 154:492–501. [PubMed: 21689084]
 18. Zelenak C, Foller M, Velic A, Krug K, Qadri SM, Viollet B, Lang F, Macek B. Proteome Analysis of Erythrocytes Lacking AMP-Activated Protein Kinase Reveals a Role of PAK2 Kinase in Eryptosis. *Journal of Proteome Research*. 2011; 10:1690–7. [PubMed: 21214270]
 19. Haudek VJ, Slany A, Gundacker NC, Wimmer H, Drach J, Gerner C. Proteome Maps of the Main Human Peripheral Blood Constituents. *Journal of Proteome Research*. 2009; 8:3834–43. [PubMed: 19580323]
 20. Burton NM, Bruce LJ. Modelling the structure of the red cell membrane. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire*. 2011; 89:200–15. [PubMed: 21455271]
 21. van Gestel RA, van Solinge WW, van der Toorn HW, Rijkssen G, Heck AJ, van Wijk R, Slijper M. Quantitative erythrocyte membrane proteome analysis with Blue-native/SDS PAGE. *Journal of Proteomics*. 2010; 73:456–65. [PubMed: 19778645]
 22. D'Alessandro A, Righetti PG, Zolla L. The red blood cell proteome and interactome: an update. *Journal of Proteome Research*. 2010; 9:144–63. [PubMed: 19886704]
 23. Bordbar A, Jamshidi N, Palsson BO. iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. *BMC Systems Biology*. 2011;5. [PubMed: 21219666]

Highlights

- Low-ionic-strength extraction of red cell ghosts with separate analysis of the membrane skeleton protein extract and inside-out vesicles was not superior to direct proteome analysis of the red cell ghosts.
- Separation of red cell ghosts into 30 fractions using 1-D SDS gels yielded an excellent depth of analysis and provided insights into protein size heterogeneity.
- Rosetta Elucidator quantitation enables the determination of protein differences that can be used to address unresolved questions in red cell disorders.

Figure 1.
Schematic of processing pipeline for RBC membrane proteomics.

Figure 2.

Peptide and protein identifications for WG-B, IOV-1, IOV-2, DCS, and a combined dataset of IOV-1 and DCS. A) The WG-B dataset identified more unique peptides compared to either IOV or DCS alone. The combined IOV-1 and DCS dataset identified 591 more peptides than WG alone. B) At the protein level, WG-B provides a substantial increase in unique proteins compared to IOV and DCS, with approximately 190 more proteins identified in WG-B. WG-B and the combined dataset of IOV-1 and DCS identified a similar number of proteins.

Figure 3. Number of unique proteins identified with increasing LC-MS/MS runs from gels sliced into 10, 15 or 30 segments. Increasing fractionation greatly enhances protein coverage.

Figure 4. Distribution of spectral counts across gel slices for eight abundant proteins in A) 30-slice proteome, B) 15-slice proteome, and C) 10 slice proteome.

Figure 5.

Rosetta Elucidator analysis of IOV technical replicates. A) Normalized fold change ratio of protein intensities for IOV-1 compared to IOV-2. Red dashed lines define the three-fold boundary. B) Distribution of protein intensities for IOV-1 versus IOV-2, indicating the 95% confidence interval (red dotted lines) and 3-fold change (black dotted line). C) Coefficient of variation for IOV technical replicates. A maximum coefficient of variation of 5 indicates an intensity cut off of 1×10^5 . D) Distribution of protein intensities for IOV-1 versus IOV-2 for proteins with intensities greater than 1×10^5 . The red dotted lines indicate the 95% confidence interval and black dotted lines correspond to a 3-fold change.

Figure 6.

Rosetta Elucidator software analysis of WG biological replicates. A) Distribution of protein intensities for WG-A versus WG-B with an intensity cut off at 1×10^5 . The 95% confidence interval (red dotted lines) and 3-fold change (black dotted lines) were derived from the IOV technical replicates. B) Distribution of protein intensities for WG-A versus WG-B with an intensity cut off at 1×10^5 . The red dotted lines correspond to 95% confidence interval and black dotted lines indicate a 3-fold change derived from the WG biological replicates alone.

Table 1

Percentage of Proteins That Fall Within Specific Fold Change Ratios

Fold Change	IOV-1 vs IOV-2 Technical Replicate	WG-A vs WG-B Biological Replicate	Group-1 vs Group-2 Biological Variant
2	93.70%	85.60%	62.70%
2.5	96.80%	94.80%	74.50%
3	97.60%	97.50%	83.90%
4	98.40%	99.10%	91.80%
5	99.70%	99.50%	95.40%
6	100.00%	99.60%	96.90%
7	100.00%	99.80%	97.20%
8	100.00%	99.80%	97.90%
9	100.00%	99.80%	98.20%
10	100.00%	99.80%	98.80%