Published in final edited form as:
J Proteome Res. 2011 July 1; 10(7): 3076–3088. doi:10.1021/pr200040j.

Comprehensive Identification of Glycated Peptides and Their Glycation Motifs in Plasma and Erythrocytes of Control and Diabetic Subjects

Qibin Zhang, **Matthew E. Monroe**, **Athena A. Schepmoes**, **Therese R. W. Clauss**, **Marina A. Gritsenko**, **Da Meng**, **Vladislav A. Petyuk**, **Richard D. Smith**, and **Thomas O. Metz***
Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA

Abstract

Non-enzymatic glycation of proteins sets the stage for formation of advanced glycation end-products and development of chronic complications of diabetes. In this report, we extended our previous methods on proteomics analysis of glycated proteins to comprehensively identify glycated proteins in control and diabetic human plasma and erythrocytes. Using immunodepletion, enrichment, and fractionation strategies, we identified 7749 unique glycated peptides, corresponding to 3742 unique glycated proteins. Semi-quantitative comparisons showed that glycation levels of a number of proteins were significantly increased in diabetes and that erythrocyte proteins were more extensively glycated than plasma proteins. A glycation motif analysis revealed that some amino acids were favored more than others in the protein primary structures in the vicinity of the glycation sites in both sample types. The glycated peptides and corresponding proteins reported here provide a foundation for potential identification of novel markers for diabetes, hyperglycemia, and diabetic complications in future studies.

Keywords

nonenzymatic glycation; Amadori compound; boronate affinity chromatography; electron transfer dissociation; type 2 diabetes mellitus; plasma; erythrocyte; red blood cell; glycation motif

Non-enzymatic glycation of proteins, also known as the Maillard or browning reaction, was primarily associated with food chemistry and processing[1, 2] until the discovery and structural elucidation of glycated human hemoglobin (HbA_{1c}) in the 1960s.[3–5] Although a less frequently studied protein post-translational modification, non-enzymatic protein glycation has recently attracted increased attention in proteomics research due to its clinical relevance to diabetes mellitus and the related complications,[6–8] as well as its effects on efficacy and stability of therapeutic antibodies.[9–12]

Protein glycation begins with nucleophilic attack of a protein primary amine on the carbonyl group of a reducing sugar (predominantly glucose *in vivo*) to form a reversible Schiff base intermediate, which then rearranges to form a relatively stable Amadori adduct.[13] The Amadori adduct further reacts to form advanced glycation end products (AGEs), which

*Corresponding author: P.O. Box 999, MS K8-98, Richland, WA 99352, Phone: (509) 371-6581, Fax: (509) 371-6546, thomas.metz@pnl.gov.

Supporting Information
Detailed tables of glycated peptides from low abundance proteins, glycation sites from high abundance proteins, and of all glycated peptides identified in this study are available as supporting information. This information is available free of charge via the Internet at http://pubs.acs.org

accumulate in long-lived tissue proteins, often altering their structure, function, and turnover.[14, 15] Age-adjusted levels of AGEs were shown to correlate with diabetic complications,[16] and AGEs are implicated in age-associated neurodegenerative diseases, such as Alzheimer's disease[17, 18] and amyotrophic lateral sclerosis.[18–20] For the purposes of this paper, we define "glycation" as the non-enzymatic reaction of glucose with side-chain amino groups on lysine residues to form the Amadori adduct.

HbA_{1c}, the stable Amadori compound formed between β-chain valine residues of hemoglobin and glucose, is used clinically to monitor long term glycemic control in diabetic patients. Although the oral glucose tolerance test (OGTT) is the gold standard for diagnosis of diabetes mellitus, measurement of HbA_{1c} has also been proposed as a more convenient and reliable diagnostic test due to the reported poor reproducibility of the OGTT.[21] However, individual differences in levels of HbA_{1c} often complicate its use in diagnosing diabetes,[22, 23] and large clinical studies have shown that levels of HbA_{1c} and results of the OGTT do not correlate well.[24] Therefore, there is a need for more sensitive and specific markers for risk or progression of diabetic complications at the earliest stage, as well as for better understanding of the role of protein glycation in development of diabetic complications.

Similarly, there is a need for a better understanding of which proteins are glycated to the greatest extent in diabetes and whether the difference in glycation may provide additional insights into risk for complications, as the overall level of protein glycation in an individual may be less important than glycation induced loss of function in a few key proteins. Free amino groups on proteins vary widely in their rate and extent of glycation; for example, acidic and basic neighboring groups affect the specificity of protein glycation, either through effects on the *p*Ka of the amino group and therefore its nucleophilicity and the kinetics of the formation of the Schiff base, or through catalysis of the Amadori rearrangement, which is considered as the rate limiting step in glycation.[13, 25, 26] Binding of anionic ligands (phosphate and bicarbonate) also affects the kinetics and specificity of glycation of proteins.[13] More recently, Johansen et al. performed a statistical analysis of glycation of lysine residues, resulting in a sequence-based glycation predictor;[27] however, that study was based on only 89 glycated lysine residues, limiting the statistical power of the analysis.

In this work, we have comprehensively identified glycated proteins in human plasma and erythrocyte samples from control and diabetic subjects in order to create a database of identified glycated peptides and corresponding proteins as a resource for the diabetes research community to facilitate the discovery of potential novel markers of diabetes, hyperglycemia, and diabetic complications in future studies. In addition, we have identified glycation motifs in the vicinity of the primary structures of the glycation sites in both sample types that may be useful in predicting potential glycation sites in other proteins.

EXPERIMENTAL

Chemicals and materials

All chemicals and peptide desalting SPE cartridges (Supelco Discovery DSC-18) were purchased from Sigma-Aldrich (St. Louis, MO), and the Micro-BCA protein assay kit was obtained from Pierce (Rockford, IL). Sequencing-grade trypsin was purchased from Promega (Madison, WI). Glycogel II boronate affinity gel (Pierce, Rockford, IL) was a gift from Dr. Bart Haigh of the Institute for Bioanalytics (Branford, CT).

Human plasma and erythrocytes

Human blood plasma and erythrocyte samples were received frozen on dry ice from the Screening for Impaired Glucose Tolerance (SIGT) study[28] via the National Institute for

Diabetes and Digestive and Kidney Diseases Central Repository. Samples were divided into three groups by the SIGT based on the results of the OGTT: normal glucose tolerance (NGT), impaired glucose tolerance (IGT) and type 2 diabetes mellitus (T2DM). To create representative samples for each group and to generate sufficient material for subsequent sample processing steps, two pooled samples were created for each specimen type as follows: 238 samples were pooled from individuals with NGT to create a NGT group and 97 samples from individuals with IGT were combined with 25 samples from individuals with T2DM to create an abnormal glucose tolerance (AGT) group. Aliquots of the pooled plasma samples were subjected to tryptic digestion without further manipulation, while high abundance proteins were removed from the remaining portions using immunodepletion prior to tryptic digestion (see below).

In vitro glycation of human plasma

Human blood plasmas from 20 healthy individuals (Bioreclamation Inc., Hicksville, NY) were pooled and then incubated with 1 M glucose in PBS containing 1 mM DTPA at a final concentration of 25 mg/mL for 48 h at 37°C. To remove excess glucose, glycated plasma proteins were dialyzed overnight against PBS containing 100 mM DTPA using 3500 molecular weight cut-off (MWCO) tubing. The dialyzed proteins were then subjected to either tryptic digestion directly or immunodepletion as outlined below.

Immunodepletion of plasma

A ProteomeLab IgY-12 LC-10 affinity LC column (Beckman Coulter, Fullerton, CA)was used with an Agilent 1100 series HPLC system (Agilent, Palo Alto, CA) to separate twelve high abundance plasma proteins (albumin, IgG, α_1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α_1-acid glycoprotein, α_2-macroglobulin, apolipoprotein A-I, apolipoprotein A-II, and fibrinogen) from other low abundance species. All IgY-12 separations were performed according to the manufacturer's instructions in regards to column usage and loading capacity. The following buffers and gradient conditions were used in a separation scheme that consisted of sample loading, washing, stripping, and neutralization followed byre-equilibration for a total cycle time of 65 min: 1) washing: 0 – 25 min at a flow rate of 0.5 mL/min, followed by 25.01 – 30 min at flow rate of 2.0 mL/min, both with 10 mM Tris-HCl and 150 mM NaCl, pH 7.4; 2) stripping: 30.01 – 48 min at a flow rate of 2.0 mL/min with 100 mM glycine, pH 2.5; 3) neutralization: 48.01 – 56 min at a flow rate of 2.0 mL/min with 100 mM Tris-HCl, pH 8.0; and 4) re-equilibration: 56.01 – 65 min at flow rate of 2.0 mL/min with 10 mM Tris-HCl and 150 mM NaCl, pH 7.4. The eluent was monitored with a diode array detector at 280 nm and the corresponding flow-through fractions (low abundance proteins) were collected, pooled, and individually concentrated in Amicon™ Ultra-15 concentrators (Millipore, Billerica, MA) with MWCO of 5 KDa, followed by a buffer exchange to 50 mM NH_4HCO_3 in the same unit according to the manufacturer's instructions. Protein concentrations were then measured using the BCA proteinassay.

Erythrocyte membrane and cytosolic protein preparations

Pooled erythrocyte cell suspensions were diluted in 9 volumes of cold 5 mM sodium phosphate buffer, pH 8, and vortexed for 30 s followed by incubation at 4°C for 15 min. Cell lysates were then centrifuged at 600×g for 5 min to remove the remaining material, which was a tangle of fibrin, lymphocytes, platelets and unlysed erythrocytes. Supernatants were carefully transferred into round bottom ultracentrifuge tubes and centrifuged at 25,000×g for 15 min to separate membrane (pellets) from cytosolic proteins (supernatants). The cytosolic proteins were retained for protein concentration measurements. Membranes were washed three times with sodium phosphate buffer or until the supernatant became colorless and the membranes were slightly yellow in color. The membranes were then reconstituted in 50 mM

NH_4HCO_3, and protein concentrations were measured by the BCA assay prior to trypsin digestion.

Protein digestion

All soluble proteins (10 mg/mL) from whole plasma, immunodepleted plasma and erythrocyte cytosol samples were dissolved in 100 mM NH_4HCO_3 (pH 8.1) containing 8 M urea and reduced with 5 mM dithiothreitol for 1 h at 37°C; free sulfhydryl groups were then alkylated with 20 mM iodoacetamide at room temperature for 1 h in the dark. Samples were subsequently diluted with 50 mM NH_4HCO_3 (pH 8.1) to reduce the urea concentration to below 1 M, and $CaCl_2$ was added to a final concentration of 1.5 mM prior to the addition of sequencing-grade modified trypsin at a ratio of 1:50 (w/w, enzyme:protein). Samples were then digested at 37°C for 12 h. The final digestion mixture was passed through C_{18} SPE cartridges for desalting, and eluted peptide solutions were concentrated by a speed-vac (ThermoFisher Scientific, Milford, MA) before being processed further (see below). The insoluble erythrocyte membranes were processed similarly except that 100 mM NH_4HCO_3 (pH 8.1) containing 7 M urea and 2 M thiourea was used for solubilizing and denaturing the membrane proteins.

Enrichment of glycated peptides

Boronate affinity chromatography was used to enrich glycated peptides.[29] Briefly, peptides obtained from protein digestion and SPE clean-up were reconstituted in LC buffer A (*vide infra*) at 1μg/μL, and up to 500 μL was injected onto the Glycogel II boronate affinity column that was prepared in an empty Tricorn (GE Healthcare, Piscataway, NJ) liquid chromatography column (5 mm × 100 mm) under gravity flow. Chromatography was conducted using an Agilent 1100 series LC system equipped with a diode array detector and autosampler and fraction collector set at 4°C. The LC buffers consisted of (A) Loading buffer: 50 mM $MgCl_2$ and 250 mM NH_4OAc, pH 8.1, (B) Washing buffer: 50 mM NH_4HCO_3, pH 8.1, and (C) Eluting buffer: 0.1 M HOAc, all in water. A step-wise gradient totaling 40 min (100% A for 10 min, switch to 100% B for 10 min, switch to 100% C for 10 min, switch back to 100% A for 10 min, all switch times are 0.1 min) at 1.0 mL/min was used to both enrich and desalt glycated peptides. The LC effluent was monitored at 280 nm with a UV detector. Fractions corresponding to glycated peptides in the eluting step were collected, dried *in vacuo*, and reconstituted for subsequent SCX fractionation.

SCX fractionation of glycated peptides

Fractionation of peptides was performed using an Agilent 1100 HPLC System equipped as described above. Glycated peptides were separated with a PolySulfoethyl A (PolyLC Inc., Columbia, MD) column (200 mm × 2.1 mm; 5 μm particles with 300-Å pores) with a 10 mm × 2.1 mm guard column packed with the same material at a flow rate of 0.2 mL/min. Mobile phases consisted of (A) 10 mM NH_4COOH, pH 3.0, and 25% CH_3CN and (B) 500 mM NH_4COOH, pH 6.8, and 25% CH_3CN. The following linear gradient was used: 100% A for 10 min, ramp to 50% B over 50 min and then to 100% B over the next 10 min with a 15 min hold at 100% B. Fifteen fractions were collected for peptides from the native whole plasma and erythrocyte samples, and 10 fractions were collected for the native low abundance plasma protein sample. For *in vitro* glycated human plasma, 30 fractions were collected for peptides originating from both whole and depleted plasma. Fractions were dried *in vacuo* and then stored at −80°C prior to capillary LC-MS/MS analysis.0

LC-MS/MS analysis of glycated peptides

After reconstitution in 25 mM NH_4HCO_3, glycated peptide fractions were analyzed with an automated 4-column capillary LC system coupled on-line with an LTQ-XL ion trap mass

spectrometer (Thermo Scientific, San Jose, CA) via an electrospray ionization interface operated in positive ionization mode. For each analysis, approximately 1.5 μg of glycated peptides were loaded onto the reversed-phase capillary column (75 μm i.d. × 65 cm); the columns were slurry packed in-house with 3 μm Jupiter C_{18} particles (Phenomenex, Torrance, CA). The mobile phase solvents consisted of (A) 0.1% formic acid (FA) in water and (B) 0.1% FA in 90/10 CH_3CN/H_2O. Exponential gradient elution was initiated 50 min after sample loading with a column flow rate of 300 nL/min, and the mobile phase ramped from 0% to 60% mobile phase B over 100 min. The mass spectrometer was operated in data-dependent MS/MS mode with a precursor-ion scan range of 350 – 1500 *m/z*. Each sample was analyzed twice with data dependent ETD and data dependent neutral loss triggered ETD methods, respectively. In both methods, the 6 most intense ions detected from the MS survey scan were selected as precursor ions for fragmentation and tandem MS, and a dynamic exclusion of 90 s was applied to minimize repeated analyses of the same precursor-ions. The cation and anion populations were controlled by automatic gain control, and the target populations were set as follows: fluoranthene anions set to 2×10^5 and precursor-cations from full MS scan events set to 3×10^4. A 100 ms reaction time was selected for the ion-ion reaction in ETD. For neutral loss triggered ETD, the top 6 most intense ions were first fragmented with collision induced dissociation (CID), and precursor ions having neutral losses of 3 H_2O and 3 H_2O + HCHO (these are characteristic neutral losses for Amadori-modified peptides during CID[30, 31]) were further fragmented using ETD. Other parameters used during data acquisition were supplemental activation and rejection of singly charged precursor-ions during MS^2 scan events.[29, 32]

Data analysis

To identify peptides, raw data files were first converted to .dta files by Extract_MSn (version 3.0 in Bioworks Cluster 3.2; Thermo), and the SEQUEST algorithm (Version 27, revision 12) was then used to search all MS/MS spectra against a database composed of the human international protein index (IPI) database (Version 3.39 consisting of 69,731 protein entries). The following search parameters were used in the database search: 3 and 1 Da tolerances for precursor ion masses and fragment ion masses, respectively; partial tryptic enzyme restraint; maximum of three missed tryptic cleavages; static carbamidomethylation of cysteine and dynamic oxidation of methionine; and dynamic Amadori compound modification at lysine (162.05 Da).

To estimate the number of false positive peptide identifications, representative datasets from each sample type were searched against a decoy database composed of the same human IPI database and its sequence-reversed version. Peptide false discovery rate (FDR) was calculated as $FDR = N_R/N_F$, where N_R and N_F are the numbers of peptides identified from the reverse and forward database searches respectively.[33] Based on the results of the decoy database search, the following criteria were applied to the first hit search results in order to achieve an FDR ≤2%: minimum peptide length of 6 amino acids and delta correlation (ΔC_n2) ≥ 0.1; for fully tryptic peptides, cross-correlation score (X_{Corr}) ≥ 2.1, 2.7, 3.5, and 4.3 for 2^+, 3^+, 4^+, and 5^+ ions, respectively; and for partially tryptic peptides or non-tryptic protein terminal peptides, X_{Corr} ≥ 3.0, 4.0, 4.7, and 5.1 for 2^+, 3^+, 4^+, and 5^+ ions, respectively. The Venn Diagram Plotter (freely available at http://ncrr.pnl.gov/software/) was used to plot the overlap between glycated peptides identified from different samples. In the case of multiple proteins identified from a single peptide, only the first IPI accession number was reported.

Glycation sites were annotated from the original protein sequences in the IPI database without removing the signaling peptides or propeptides, unless specifically mentioned otherwise. To investigate the glycation motif, unique glycated peptides were extracted from the SEQUEST results, and peptide sequences surrounding the glycated lysine residue (± 30

amino acids with respect to the glycation site) were reconstructed from the original protein sequences. The occurring frequencies of neighboring amino acids surrounding the glycated lysines were analyzed by the WebLogo 3.0 program (http://weblogo.threeplusone.com),[34] which was installed on a local desktop computer to facilitate analyses of large sequence data.

RESULTS

Overview of glycated peptide analysis

The flow chart for sample processing and analysis is shown in Scheme 1. To increase coverage of low abundance proteins, human blood plasma samples, both native and *in vitro* glycated, were subjected to immunodepletion to remove the twelve most abundant plasma proteins. For the same reason, erythrocyte samples were divided into membrane and cytosolic protein fractions. Proteins from each sample type were then digested with trypsin, and, since glycation is a low abundance, relatively non-specific modification of proteins *in vivo*, the resulting glycated peptides were enriched using a previously developed boronate affinity chromatography method.[29, 35, 36] To overcome the common under-sampling issue related to sample complexity in LC-MS/MS-based bottom-up proteomics, enriched peptides were further fractionated off-line using SCX chromatography. Glycated peptides from each fraction were analyzed using ETD MS/MS and neutral loss triggered ETD-MS/MS, because Amadori-modified peptide molecular ions are labile and readily lose up to 4 water molecules by conventional CID;[31, 37] however, other CID methods (neutral loss triggered MS^3, multi-stage activation, higher energy collision dissociation in an Orbitrap mass spectrometer, and precursor ion scanning methods) have also been successful in sequencing glycated peptides.[8, 30, 37, 38]

The results from each sample type are summarized in detail in Table 1. In total, 7749 unique glycated peptides were identified, corresponding to 3742 unique glycated proteins (based on IPI number). Compared with native human plasma, *in vitro* glycated human plasma showed a ~74% increase in the number of unique glycated peptides identified, which is expected as a result of the artificial modification with high concentration of glucose. It should be noted that there were more glycated peptides identified from erythrocytes than from native plasma, and that significantly more glycated peptides were identified from the erythrocyte membrane compared to the cytosolic fraction.

To further explore the origins of the glycated peptides, the degrees of overlap among sample types were plotted as shown in the Venn Diagrams of Figure 1. In general, there is little overlap between different sample types, i.e. plasma versus erythrocyte (Fig. 1A) or erythrocyte membrane versus cytosol fractions (Fig. 1B). Although in Figure 1A–B, only a small fraction of the glycated peptides were identified as overlapping between the two sample types, a careful review revealed that these peptides typically exist primarily only in one type of sample, as demonstrated by peptide count information, which can be used semi-quantitatively to reflect peptide/protein abundance. For example, among the 226 unique glycated peptides that were identified overlapping between the erythrocytes and native human plasma (Fig. 1A), albumin peptide SLHTLFGDK#LCTVATLR (# denotes Amadori modification) had 556 spectrum counts in the native plasma pool but only 29 in erythrocytes; similarly, hemoglobin peptide K#VLGAFSDGLAHLDNLK had 529 spectrum counts in erythrocytes but only 6 in the native human plasma pool. This low level of cross contamination is hardly avoidable, particularly considering the nature of the samples, but demonstrates the relative quality of sample preparation.

There were 191 glycated peptides shared between non-depleted and depleted fractions of native human plasma (Fig. 1C). This overlap is not high, and a close look at the shared

peptides showed that some were preferentially found in one fraction; for example, hemopexin peptide RLEK#EVGTPHGIILDSVDAAFICPGSSR had 130 spectrum counts in depleted plasma but only 17 in whole plasma. In contrast, some glycated peptides from albumin were primarily found in whole plasma and scarcely in the depleted sample. Meanwhile, some glycated peptides from medium abundance proteins were approximately equally present in both samples. The percentage of overlap was higher between whole and depleted *in vitro* glycated plasma compared with the overlap between the analogous fractions in native plasma (Fig. 1D vs 1C) - the nature of the shared peptides was the same as outlined previously. Although there were many more glycated peptides found in *in vitro* glycated versus native plasma, and also relatively large overlap between these two samples types (Fig. 1E), there was a distinct number of glycated peptides present only in one sample type. It should be noted that there were 14% more glycated peptides found in the depleted fraction than in whole plasma in the *in vitro* glycated plasma samples (Fig. 1D). Compared with native plasma, this high number of glycated peptides found in the depleted fraction is likely due to the amount of glycated peptides enriched from this fraction, which is not as limited to the downstream SCX fractionation as compared with the peptides obtained from native plasma. In addition, low abundance proteins in native plasma may have shorter half-lives compared with high abundance plasma proteins and thus may not accumulate significant levels of the Amadori compound before *in vivo* proteolytic digestion and being removed from the blood stream.

Immunodepletion almost doubled the number of glycated peptides/proteins identified from the blood plasma samples. In total, 837 unique glycated peptides were found exclusively in depleted native plasma (Fig 1C), corresponding to 296 unique proteins, 60 of which were identified either with more than one glycated peptide or with ≥2 observations of the same glycated peptide. A selected list of low abundance glycated proteins and their associated glycated peptides are shown in Table 2 (a complete list is shown in the Supporting Information). Many of these proteins are in the concentration range of low μM to nM,[39] such as various complement component precursors (sub μM range), as well as pigment epithelium-derived factor (0.1 μM), sex-hormone-binding protein (0.03–0.2 μM), and hyaluronan-binding protein (0.07–0.2 μM).

Because of their high concentrations in plasma, erythrocyte cytosol and membrane fractions, glycated peptides derived from albumin, hemoglobin and spectrin were detected with highest frequency by LC-MS/MS analysis. Details of the glycation sites identified within these three proteins are summarized in Table 3. A more detailed list of all major glycated proteins, their respective glycated peptides, and the annotation of their glycation sites is in the Supporting Information. For comparison, the glycation sites identified from *in vitro* glycated blood plasma are listed in the same table as the native plasma (Supporting Information). Notably, many of the glycation sites have not been previously reported in the literature. While the majority of glycated peptides appeared in both native and *in vitro* glycated plasma, significantly more peptide counts were observed for the *in vitro* glycated sample. However, there were also some sites unique to the *in vitro* glycated sample. For example, peptide

RPCFSALEVDETYVPK#EFNAETFTFHADICTLSEK (corresponding to K^{500} in albumin) had 168 spectrum counts in plasma glycated *in vitro*, but was not detected in native plasma (Supporting Information). This might be explained by protein unfolding during *in vitro* glycation, which enables unique access of glucose to that particular lysine residue.

Semi-quantitative comparison of protein glycation between NGT and AGT states

Despite having identified a large number of glycated peptides, further analysis of the data showed that a small portion of these peptides have significant increases in spectral counts in

AGT (samples pooled from individuals with IGT and T2DM) compared with NGT. Eight proteins were found as increasingly glycated in AGT versus NGT in native plasma (Table 4). Notably, hemopexin was heavily glycated in AGT samples compared with NGT. This result suggests that either hemopexin is up-regulated or its glycation level is higher in diabetic patients. In this respect, it was reported that hemopexin concentration in blood plasma was higher in type 2 diabetes,[40] and the level of Amadori compound modification of hemopexin was increased in type 2 diabetes as well.[41]

Compared with native human plasma, there were more glycated peptides up-regulated in AGT versus NGT erythrocytes. These proteins were primarily membrane or membrane associated proteins and hemoglobin. A selected list of glycated peptides found in spectrin is highlighted in Table 5 (the full list of up regulated glycated peptides and proteins is shown in Supplemental Information). Due to missed cleavages by trypsin (trypsin does not cleave at glycated lysine residues) and oxidation of methionine, the same glycation site can be identified by more than one glycated peptide; however, all peptides corresponding to the same glycation site follow the same trend. For example, five different peptides corresponding to glycation site K^{1695} within spectrin beta chain were identified, and their spectral counts in AGT samples were consistently higher than in NGT samples. Increased glycation of erythrocyte membrane proteins and hemoglobins is well-established during hyperglycemia,[42–44] and the higher number of erythrocyte proteins glycated relative to plasma proteins is likely due to the much longer life span of erythrocytes.

Glycation motif analyses

To study whether there is a characteristic amino acid motif in the vicinity of glycated lysine residues, we conducted motif analyses on the primary structures of the peptide sequences identified in our study. To reduce background noise from infrequently identified peptides, only unique glycated peptides having peptide spectrum counts of ≥5 were included in the motif analysis. These criteria resulted in 361, 903, and 443 unique glycated peptide sequences identified from native human plasma, *in vitro* glycated human plasma, and erythrocytes, respectively. Glycation motifs were identified using the corresponding protein sequences and the WebLogo program, specifying 30 amino acids prior to and after the glycation site.

In sequence logo, the frequency of occurrence of each amino acid at a specific position is graphically represented by its size/height. The amino acids are then stacked on top of each other in increasing order of their frequencies.[34] In general, hydrophobic short chain or uncharged side chain amino acids, such as alanine (A), valine (V), leucine (L), and serine (S) were identified as occurring the most frequently in peptides from proteins glycated *in vivo*. For example, in addition to threonine (T), A, V, L, and S occurred most frequently in glycated peptides identified in native human plasma (Figure 2A). Similarly, A, V, L, and S also occurred more frequently in glycated peptides identified in erythrocytes (Figure 2B), although glutamic acid (E) also was highly represented at certain amino acid positions relative to the glycation site. Alternatively, S, L and E were more frequently observed for peptides from *in vitro* glycated human plasma (Figure 2C).

In general, histidine (H), tyrosine (Y), methionine (M), tryptophan (W), cysteine (C) and phenylalanine (F) did not appear to be among the major amino acids that are near the vicinity of the glycation sites for all three glycation motif plots. Although other amino acids did occur, they were present far less frequently than the hydrophobic short chain or uncharged side chain amino acids mentioned previously. It should be noted that, at position +1, alanine (A) and valine (V) were always in the top 4 occurring amino acids in every sample type.

It has been suggested that the properties of nearby amino acids in either the primary or the three dimensional structure play a role in determining whether a given lysine has the propensity to be glycated. For example, the positively charged amino acids lysine and histidine have been proposed to catalyze the formation of the Amadori compound.[45] In our study, lysines were indeed observed around the N- or C- termini of the glycation site; however, histidine was not significantly present in the proximity of glycated lysine residues on a primary structure level. Nevertheless, due to the frequency of threonine, particularly in the native human plasma sample, our data is consistent with the notion that histidine catalyzes the formation of Amadori compound only in 3D space and is mediated by threonine in the vicinity of the glycation site.[26] Acidic amino acids have also been proposed to catalyze glycation;[46] in this respect, glutamic acid was observed more frequently in the glycated peptides from all three sample types, particularly in erythrocytes and *in vitro* glycated plasma. This modest increase of neighboring glutamic acid in these two sample types seems to correlate with the extent of glycation, since either long term glycation (the erythrocyte life span is 120 days), or incubation of plasma with a high concentration of glucose occurred. The notion that basic residues are favored N-terminal to the site of glycation and acidic residues are favored C-terminal is weakly supported by our data,[27] as lysine and glutamic acid did not show a strong preference to either terminus with respect to the glycation site (for example, as shown in Figure 2B).

DISCUSSION

The goal of our study was to develop a comprehensive database of glycated peptides in human blood, so it can be utilized as a resource by the diabetes research community in future quantitative analyses of protein glycation in large cohorts of clinical blood samples. To this end, we used a combination of immunodepletion (to remove the 12 most abundant proteins from human plasma samples) and SCX fractionation to reduce sample complexity and enrich subsets of peptides, so as to further increase the identification of glycated peptides.

Immunodepletion has been used previously to enhance the proteomic coverage of plasma samples,[47, 48] and we found that immunodepletion nearly doubled the number of unique glycated peptides identified – because of this, many low abundance proteins (μM to nM range) were identified as being glycated. Because glycation occurs to a relatively low extent *in vivo*,[49] some glycation sites will be difficult to identify, particularly when sample amounts are limited. For this reason, we included plasma glycated *in vitro* in our study. Although *in vitro* and *in vivo* glycation are distinct, *in vitro* glycation emulates uncontrolled hyperglycemia. A significant number of glycated peptides were shared between native and *in vitro* glycated plasma; however, there were peptides identified as unique to either sample, and more glycated peptides were identified from the *in vitro* glycated plasma. In total, we identified 7749 unique glycated peptides, which, due to the use of SCX fractionation, greatly surpassed our previous report.[35]

The much larger number of glycated peptides identified in our study provides better statistics for identification of glycation motifs. This data should enhance the power of future efforts to predict ‘glycation hot spots’ and further modeling of which biological processes or molecular functions could be affected by glycation. However, it is important to note that sequence-based protein primary structure information does not correlate with the spatial proximity of the potential reactive sites and glycation motif analyses preferably should be carried out on the 3D structure of proteins, as it truly reflects the spatial proximity of amino acids to the reactive lysines. This is particularly important when considering the formation of cross-linked AGEs; however, this is not yet technically feasible for a large set of proteins. In addition, confident identification of a protein based on a single peptide is still a topic of

debate,[50] but we have included these in the list of identifications due to the low level of glycation *in vivo* and our rigorous filtering criteria. However, we did not include them in the semi-quantitative comparison of glycation levels between NGT and AGT samples or in the analyses of the glycation motifs, because of the possibility that they would contribute noise to the glycation patterns.

It is not unexpected that we observed higher levels of glycated proteins in pooled samples representing AGT versus pooled samples representing NGT. In individuals with hyperglycemia, the protein glycation level increases to a new steady state roughly proportional to the increase in blood glucose concentration.[51] However, because the samples were pooled from a large number of individual samples without matching the donors in terms of age, gender, body weight, etc., it is somewhat irresponsible to draw solid biological implications of the identified glycated proteins, as genetically determined biological variations in blood glucose metabolism and HbA_{1c} levels[22] may result in the pooled samples being biased by a few individuals. Further complicating any speculation on the biological significance of our findings is the fact that, in contrast to AGEs, little evidence is available that glycation (i.e. the formation of the Amadori compound on side chain lysine residues) itself is pathologically significant. As mentioned in the Introduction, while glycation is generally considered to be the first step of the Maillard reaction *in vivo*, it is the AGEs generated in later stages of the reaction that are implicated in the pathogenesis of various chronic diseases, including diabetes mellitus. For example, AGEs interact with various cell receptors to induce oxidative stress and to activate inflammatory cascades.[52–57] In addition, the accumulation of AGEs on collagen and elastin affects the elasticity of the extracellular matrix (ECM),[58–60] the binding of cells,[61–63] and the turnover of ECM proteins.[63] Finally, the accumulation of AGEs is implicated in the development of vascular, renal, retinal, and neural complications of diabetes.[64–66] However, we emphasize that the formation of the Amadori modification often leads to the formation of AGEs at the same site within proteins. Therefore, our data represents a list of glycated proteins which may eventually be modified by AGEs and whose structures, turnovers, and functions may eventually be impacted, leading to the development of diabetic complications. Further, the glycated proteins identified in our study may represent potential new and better markers of long term glycemic control compared to HbA_{1c}. Both of these possibilities can now be evaluated by the diabetes research community based on our results.

CONCLUSIONS

The samples used in our study included blood plasma and erythrocytes from individuals with NGT, IGT and T2DM, as well as *in vitro* glycated human plasma. Through immunodepletion, boronate affinity enrichment, and SCX fractionation, we have identified 7749 unique glycated peptides from these proteomes. This represents by far the most comprehensive profiling of glycated peptides in these sample types to date, and it will facilitate subsequent high throughput, quantitative proteomics studies of individual samples in large cohorts of clinical studies using the LC-MS-based accurate mass and time (AMT) tag approach developed in our laboratory.[67]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Bart Haigh of the Institute for Bioanalytics for kindly providing the Glycogel™ II boronate affinity gel and Dr. John W. Baynes of the University of South Carolina for critically reading the manuscript. This research was supported by NIH grant DK071283; portions of this research were supported through the National

Center for Research Resources (RR018522), and work was performed at the Environmental Molecular Sciences Laboratory, a national scientific user facility located at Pacific Northwest National Laboratory (PNNL) and sponsored by the U. S. Department of Energy (DOE) Office of Biological and Environmental Research. PNNL is operated by Battelle for the DOE under Contract No. DE-AC06-76RLO-1830.

Abbreviations

LC-MS	liquid chromatography-mass spectrometry
MS/MS	tandem mass spectrometry
SCX	strong cation exchange
CID	collision-induced dissociation
ETD	electron transfer dissociation
SPE	solid phase extraction
PBS	phosphate buffered saline
DTPA	diethylene triamine pentaacetic acid

References

1. Finot PA. Historical perspective of the Maillard reaction in food science. Ann N Y Acad Sci. 2005; 1043:1–8. [PubMed: 16037216]
2. Chuyen NV. Maillard reaction and food processing. Application aspects Adv Exp Med Biol. 1998; 434:213–235.
3. Bookchin RM, Gallop PM. Structure of hemoglobin AIc: nature of the N-terminal beta chain blocking group. Biochem Biophys Res Commun. 1968; 32 (1):86–93. [PubMed: 4874776]
4. Bunn HF, Haney DN, Gabbay KH, Gallop PM. Further identification of the nature and linkage of the carbohydrate in hemoglobin A1c. Biochem Biophys Res Commun. 1975; 67 (1):103–109. [PubMed: 1201013]
5. Rahbar S. The discovery of glycated hemoglobin: a major event in the study of nonenzymatic chemistry in biological systems. Ann N Y Acad Sci. 2005; 1043:9–19. [PubMed: 16037217]
6. Lapolla A, Fedele D, Seraglia R, Traldi P. The role of mass spectrometry in the study of non-enzymatic protein glycation in diabetes: an update. Mass Spectrom Rev. 2006; 25 (5):775–797. [PubMed: 16625652]
7. Zhang Q, Ames JM, Smith RD, Baynes JW, Metz TO. A perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: probing the pathogenesis of chronic disease. J Proteome Res. 2009; 8 (2):754–769. [PubMed: 19093874]
8. Priego-Capote F, Scherl A, Muller M, Waridel P, Lisacek F, Sanchez JC. Glycation isotopic labeling with 13C-reducing sugars for quantitative analysis of glycated proteins in human plasma. Mol Cell Proteomics. 2010; 9 (3):579–592. [PubMed: 19955080]
9. Quan C, Alcala E, Petkovska I, Matthews D, Canova-Davis E, Taticek R, Ma S. A study in glycation of a therapeutic recombinant humanized monoclonal antibody: where it is, how it got there, and how it affects charge-based behavior. Anal Biochem. 2008; 373 (2):179–191. [PubMed: 18158144]
10. Zheng X, Wu SL, Hancock WS. Glycation of interferon-beta-1b and human serum albumin in a lyophilized glucose formulation. Part III: application of proteomic analysis to the manufacture of biological drugs Int J Pharm. 2006; 322 (1–2):136–145.
11. Zhang B, Yang Y, Yuk I, Pai R, McKay P, Eigenbrot C, Dennis M, Katta V, Francissen KC. Unveiling a glycation hot spot in a recombinant humanized monoclonal antibody. Anal Chem. 2008; 80 (7):2379–2390. [PubMed: 18307322]
12. Fischer S, Hoernschemeyer J, Mahler HC. Glycation during storage and administration of monoclonal antibody formulations. Eur J Pharm Biopharm. 2008

13. Baynes JW, Watkins NG, Fisher CI, Hull CJ, Patrick JS, Ahmed MU, Dunn JA, Thorpe SR. The Amadori product on protein: structure and reactions. Prog Clin Biol Res. 1989; 304:43–67. [PubMed: 2675036]

14. Baynes JW. The role of AGEs in aging: causation or correlation. Exp Gerontol. 2001; 36 (9):1527–1537. [PubMed: 11525875]

15. Brownlee M. Advanced protein glycosylation in diabetes and aging. Annu Rev Med. 1995; 46:223–234. [PubMed: 7598459]

16. Monnier VM, Vishwanath V, Frank KE, Elmets CA, Dauchot P, Kohn RR. Relation between complications of type I diabetes mellitus and collagen-linked fluorescence. N Engl J Med. 1986; 314 (7):403–408. [PubMed: 3945267]

17. Reddy VP, Obrenovich ME, Atwood CS, Perry G, Smith MA. Involvement of Maillard reactions in Alzheimer disease. Neurotox Res. 2002; 4 (3):191–209. [PubMed: 12829400]

18. Sasaki N, Fukatsu R, Tsuzuki K, Hayashi Y, Yoshida T, Fujii N, Koike T, Wakayama I, Yanagihara R, Garruto R, Amano N, Makita Z. Advanced glycation end products in Alzheimer's disease and other neurodegenerative diseases. Am J Pathol. 1998; 153 (4):1149–1155. [PubMed: 9777946]

19. Vicente Miranda H, Outeiro TF. The sour side of neurodegenerative disorders: the effects of protein glycation. J Pathol. 2010; 221 (1):13–25. [PubMed: 20186922]

20. Kaufmann E, Boehm BO, Sussmuth SD, Kientsch-Engel R, Sperfeld A, Ludolph AC, Tumani H. The advanced glycation end-product N epsilon-(carboxymethyl)lysine level is elevated in cerebrospinal fluid of patients with amyotrophic lateral sclerosis. Neurosci Lett. 2004; 371 (2–3): 226–229. [PubMed: 15519762]

21. Peters AL, Davidson MB, Schriger DL, Hasselblad V. A clinical approach for the diagnosis of diabetes mellitus: an analysis using glycosylated hemoglobin levels. Meta-analysis Research Group on the Diagnosis of Diabetes Using Glycated Hemoglobin Levels JAMA. 1996; 276 (15): 1246–1252.

22. Leslie RD, Cohen RM. Biologic variability in plasma glucose, hemoglobin A1c, and advanced glycation end products associated with diabetes complications. J Diabetes Sci Technol. 2009; 3 (4):635–643. [PubMed: 20144305]

23. Cohen RM. A1C: does one size fit all? Diabetes Care. 2007; 30 (10):2756–2758. [PubMed: 17901536]

24. Dong XL, Liu Y, Sun Y, Sun C, Fu FM, Wang SL, Chen L. Comparison of HbA1c and OGTT Criteria to Diagnose Diabetes Among Chinese. Exp Clin Endocrinol Diabetes. 2010

25. Bunn HF, Shapiro R, McManus M, Garrick L, McDonald MJ, Gallop PM, Gabbay KH. Structural heterogeneity of human hemoglobin A due to nonenzymatic glycosylation. J Biol Chem. 1979; 254 (10):3892–3898. [PubMed: 438166]

26. Shilton BH, Campbell RL, Walton DJ. Site specificity of glycation of horse liver alcohol dehydrogenase in vitro. Eur J Biochem. 1993; 215 (3):567–572. [PubMed: 8354263]

27. Johansen MB, Kiemer L, Brunak S. Analysis and prediction of mammalian protein glycation. Glycobiology. 2006; 16 (9):844–853. [PubMed: 16762979]

28. Phillips LS, Weintraub WS, Ziemer DC, Kolm P, Foster JK, Vaccarino V, Rhee MK, Budhwani RK, Caudle JM. All pre-diabetes is not the same: metabolic and vascular risks of impaired fasting glucose at 100 versus 110 mg/dl: the Screening for Impaired Glucose Tolerance study 1 (SIGT 1). Diabetes Care. 2006; 29 (6):1405–1407. [PubMed: 16732034]

29. Zhang Q, Schepmoes AA, Brock JW, Wu S, Moore RJ, Purvine SO, Baynes JW, Smith RD, Metz TO. Improved methods for the Enrichment and Analysis of Glycated Peptides. Anal Chem. 2008; 80 (24):9822–9829. [PubMed: 18989935]

30. Zhang Q, Petyuk VA, Schepmoes AA, Orton DJ, Monroe ME, Yang F, Smith RD, Metz TO. Analysis of non-enzymatically glycated peptides: neutral-loss-triggered MS(3) versus multi-stage activation tandem mass spectrometry. Rapid Commun Mass Spectrom. 2008; 22 (19):3027–3034. [PubMed: 18763275]

31. Zhang Q, Frolov A, Tang N, Hoffmann R, van de Goor T, Metz TO, Smith RD. Application of electron transfer dissociation mass spectrometry in analyses of non-enzymatically glycated peptides. Rapid Commun Mass Spectrom. 2007; 21 (5):661–666. [PubMed: 17279487]

32. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JE, Coon JJ. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. Anal Chem. 2007; 79 (2):477–485. [PubMed: 17222010]
33. Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, Petritis K, Camp DG 2nd, Smith RD. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res. 2005; 4 (1):53–62. [PubMed: 15707357]
34. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990; 18 (20):6097–6100. [PubMed: 2172928]
35. Zhang Q, Tang N, Schepmoes AA, Phillips LS, Smith RD, Metz TO. Proteomic profiling of nonenzymatically glycated proteins in human plasma and erythrocyte membranes. J Proteome Res. 2008; 7 (5):2025–2032. [PubMed: 18396901]
36. Zhang Q, Tang N, Brock JW, Mottaz HM, Ames JM, Baynes JW, Smith RD, Metz TO. Enrichment and analysis of nonenzymatically glycated peptides: boronate affinity chromatography coupled with electron-transfer dissociation mass spectrometry. J Proteome Res. 2007; 6 (6):2323–2330. [PubMed: 17488106]
37. Frolov A, Hoffmann P, Hoffmann R. Fragmentation behavior of glycated peptides derived from D-glucose, D-fructose and D-ribose in tandem mass spectrometry. J Mass Spectrom. 2006; 41 (11): 1459–1469. [PubMed: 17063450]
38. Gadgil HS, Bondarenko PV, Treuheit MJ, Ren D. Screening and sequencing of glycated proteins by neutral loss scan LC/MS/MS method. Anal Chem. 2007; 79 (15):5991–5999. [PubMed: 17571855]
39. Hortin GL, Sviridov D, Anderson NL. High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. Clin Chem. 2008; 54 (10):1608–1616. [PubMed: 18687737]
40. Van Campenhout A, Van Campenhout C, Lagrou AR, Abrams P, Moorkens G, Van Gaal L, Manuel-y-Keenoy B. Impact of diabetes mellitus on the relationships between iron-, inflammatory- and oxidative stress status. Diabetes Metab Res Rev. 2006; 22 (6):444–454. [PubMed: 16506275]
41. Jaleel A, Halvatsiotis P, Williamson B, Juhasz P, Martin S, Nair KS. Identification of Amadori-modified plasma proteins in type 2 diabetes and the effect of short-term intensive insulin treatment. Diabetes Care. 2005; 28 (3):645–652. [PubMed: 15735202]
42. Rahbar S. Glycosylated hemoglobins. Tex Rep Biol Med. 1980; 40:373–385. [PubMed: 6798703]
43. Schwartz RS, Madsen JW, Rybicki AC, Nagel RL. Oxidation of spectrin and deformability defects in diabetic erythrocytes. Diabetes. 1991; 40 (6):701–708. [PubMed: 2040386]
44. Jiang M, Jia L, Jiang W, Hu X, Zhou H, Gao X, Lu Z, Zhang Z. Protein disregulation in red blood cell membranes of type 2 diabetic patients. Biochem Biophys Res Commun. 2003; 309(1):196–200. [PubMed: 12943682]
45. Iberg N, Fluckiger R. Nonenzymatic glycosylation of albumin in vivo. Identification of multiple glycosylated sites J Biol Chem. 1986; 261 (29):13542–13545.
46. Watkins NG, Thorpe SR, Baynes JW. Glycation of amino groups in protein. Studies on the specificity of modification of RNase by glucose J Biol Chem. 1985; 260 (19):10629–10636.
47. Liu T, Qian WJ, Mottaz HM, Gritsenko MA, Norbeck AD, Moore RJ, Purvine SO, Camp DG 2nd, Smith RD. Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. Mol Cell Proteomics. 2006; 5 (11):2167–2174. [PubMed: 16854842]
48. Qian WJ, Kaleta DT, Petritis BO, Jiang H, Liu T, Zhang X, Mottaz HM, Varnum SM, Camp DG 2nd, Huang L, Fang X, Zhang WW, Smith RD. Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. Mol Cell Proteomics. 2008; 7 (10):1963–1973. [PubMed: 18632595]
49. Danze PM, Tarjoman A, Rousseaux J, Fossati P, Dautrevaux M. Evidence for an increased glycation of IgG in diabetic patients. Clin Chim Acta. 1987; 166 (2–3):143–153. [PubMed: 3621595]

50. Veenstra TD, Conrads TP, Issaq HJ. What to do with “one-hit wonders”? Electrophoresis. 2004; 25 (9):1278–1279. [PubMed: 15174049]
51. Thornalley PJ, Battah S, Ahmed N, Karachalias N, Agalou S, Babaei-Jadidi R, Dawnay A. Quantitative screening of advanced glycation endproducts in cellular and extracellular proteins by tandem mass spectrometry. Biochem J. 2003; 375 (Pt 3):581–592. [PubMed: 12885296]
52. Schmidt AM, Vianna M, Gerlach M, Brett J, Ryan J, Kao J, Esposito C, Hegarty H, Hurley W, Clauss M, et al. Isolation and characterization of two binding proteins for advanced glycosylation end products from bovine lung which are present on the endothelial cell surface. J Biol Chem. 1992; 267 (21):14987–14997. [PubMed: 1321822]
53. Yan SF, Ramasamy R, Naka Y, Schmidt AM. Glycation, inflammation, and RAGE: a scaffold for the macrovascular complications of diabetes and beyond. Circ Res. 2003; 93 (12):1159–1169. [PubMed: 14670831]
54. Schmidt AM, Yan SD, Yan SF, Stern DM. The biology of the receptor for advanced glycation end products and its ligands. Biochim Biophys Acta. 2000; 1498 (2–3):99–111. [PubMed: 11108954]
55. Skolnik EY, Yang Z, Makita Z, Radoff S, Kirstein M, Vlassara H. Human and rat mesangial cell receptors for glucose-modified proteins: potential role in kidney tissue remodelling and diabetic nephropathy. J Exp Med. 1991; 174 (4):931–939. [PubMed: 1655949]
56. Zill H, Gunther R, Erbersdobler HF, Folsch UR, Faist V. RAGE expression and AGE-induced MAP kinase activation in Caco-2 cells. Biochem Biophys Res Commun. 2001; 288 (5):1108–1111. [PubMed: 11700025]
57. Basta G, Schmidt AM, De Caterina R. Advanced glycation end products and vascular inflammation: implications for accelerated atherosclerosis in diabetes. Cardiovasc Res. 2004; 63 (4):582–592. [PubMed: 15306213]
58. Bruel A, Oxlund H. Changes in biomechanical properties, composition of collagen and elastin, and advanced glycation endproducts of the rat aorta in relation to age. Atherosclerosis. 1996; 127 (2): 155–165. [PubMed: 9125305]
59. Corman B, Duriez M, Poitevin P, Heudes D, Bruneval P, Tedgui A, Levy BI. Aminoguanidine prevents age-related arterial stiffening and cardiac hypertrophy. Proc Natl Acad Sci U S A. 1998; 95 (3):1301–1306. [PubMed: 9448326]
60. Konova E, Baydanoff S, Atanasova M, Velkova A. Age-related changes in the glycation of human aortic elastin. Exp Gerontol. 2004; 39 (2):249–254. [PubMed: 15036419]
61. Pedchenko VK, Chetyrkin SV, Chuang P, Ham AJ, Saleem MA, Mathieson PW, Hudson BG, Voziyan PA. Mechanism of perturbation of integrin-mediated cell-matrix interactions by reactive carbonyl compounds and its implication for pathogenesis of diabetic nephropathy. Diabetes. 2005; 54 (10):2952–2960. [PubMed: 16186398]
62. Dobler D, Ahmed N, Song L, Eboigbodin KE, Thornalley PJ. Increased dicarbonyl metabolism in endothelial cells in hyperglycemia induces anoikis and impairs angiogenesis by RGD and GFOGER motif modification. Diabetes. 2006; 55 (7):1961–1969. [PubMed: 16804064]
63. Chong SA, Lee W, Arora PD, Laschinger C, Young EW, Simmons CA, Manolson M, Sodek J, McCulloch CA. Methylglyoxal inhibits the binding step of collagen phagocytosis. J Biol Chem. 2007; 282 (11):8510–8520. [PubMed: 17229729]
64. Genuth S, Sun W, Cleary P, Sell DR, Dahms W, Malone J, Sivitz W, Monnier VM. Glycation and carboxymethyllysine levels in skin collagen predict the risk of future 10-year progression of diabetic retinopathy and nephropathy in the diabetes control and complications trial and epidemiology of diabetes interventions and complications participants with type 1 diabetes. Diabetes. 2005; 54 (11):3103–3111. [PubMed: 16249432]
65. Ahmed N, Thornalley PJ. Advanced glycation endproducts: what is their relevance to diabetic complications? Diabetes Obes Metab. 2007; 9 (3):233–245. [PubMed: 17391149]
66. Goh SY, Cooper ME. Clinical review: The role of advanced glycation end products in progression and complications of diabetes. J Clin Endocrinol Metab. 2008; 93 (4):1143–1152. [PubMed: 18182449]
67. Zimmer JS, Monroe ME, Qian WJ, Smith RD. Advances in proteomics data analysis and display using an accurate mass and time tag approach. Mass Spectrom Rev. 2006; 25 (3):450–482. [PubMed: 16429408]

Synopsis

Using immunodepletion, enrichment, and fractionation strategies, we identified 7749 unique glycated peptides, corresponding to 3742 unique glycated proteins. Semi-quantitative comparisons showed that glycation levels of a number of proteins were significantly increased in diabetes and that erythrocyte proteins were more extensively glycated than plasma proteins. A glycation motif analysis revealed that some amino acids were favored in the protein primary structures in the vicinity of the glycation sites in both sample types.

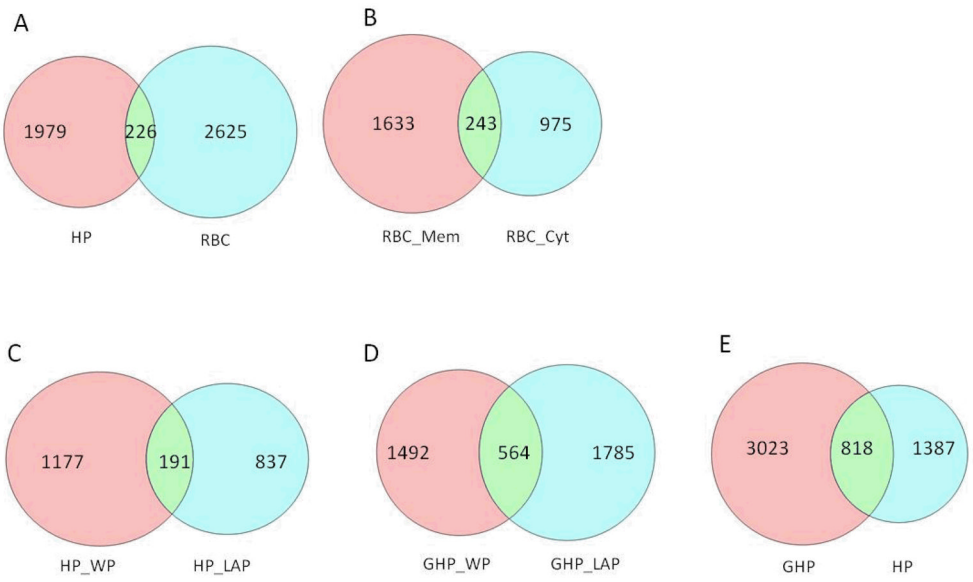
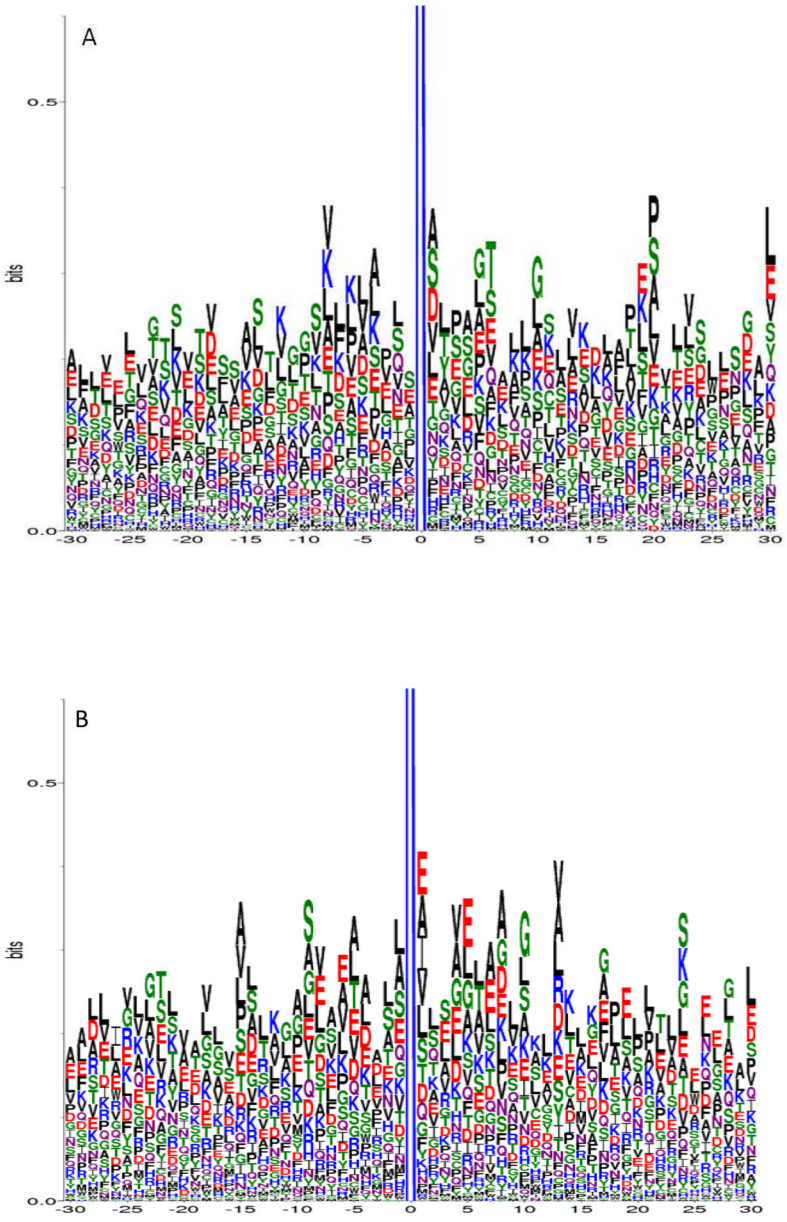


Figure 1.

Overlap of unique glycated peptides identified from various samples. GHP, *in vitro* glycated human plasma; HP, human plasma; RBC, erythrocytes; Mem, membrane fraction; Cyt, cytosolic fraction; WP, whole plasma; LAP, low abundance proteins collected from immunodepletion of whole plasma with IgY-12 column.



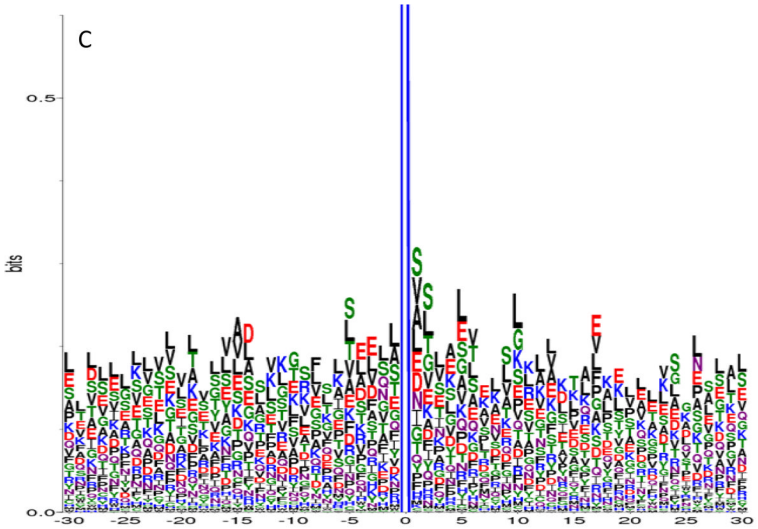
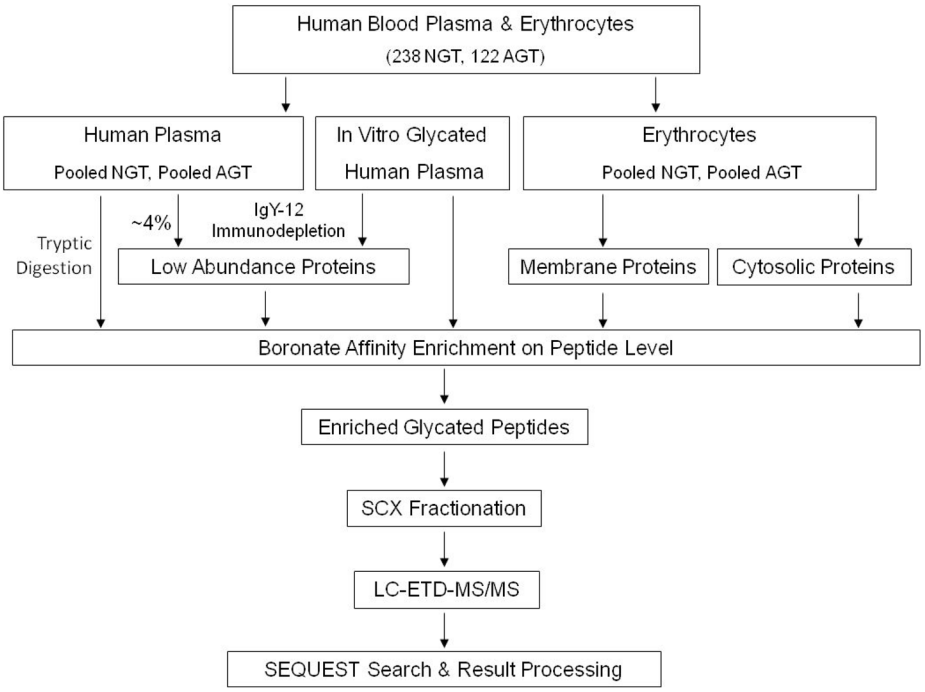


Figure 2.

Glycation motif logos for glycated peptides identified from A) native human plasma, B) *erythrocytes*, and C) *in vitro* glycated human plasma. Each glycated peptide was required to have a peptide observation count of ≥5 in order to be included in the analysis. The X-axis represents the neighboring amino acids (−30 to +30) around the glycated lysine (at position 0) in the protein primary sequence; the Y-axis represents the frequency of each amino acid observed at a particular position. Note that the full length of each logo is out of scale on the Y-axis at 4.3 bits; therefore, it is truncated to show the details of the neighboring amino acids.



Scheme 1.

Flow chart depicting the sample preparation and analyses processes for identification of glycated peptides. NGT samples (blood plasma or erythrocytes) were pooled from individuals with NGT, and AGT samples were pooled from individuals with both IGT and T2DM, respectively. See experimental section for details. NGT, normal glucose tolerance; AGT, abnormal glucose tolerance; IGT, impaired glucose tolerance; T2DM, type 2 diabetes mellitus.

Table 1

Unique glycated peptides identified from each sample type. HP, native human plasma; WP, whole plasma; LAP, low abundance proteins collected from immunodepletion of whole plasma with IgY-12 column; GHP, *in vitro* glycated human plasma; RBC, erythrocytes; Mem, membrane proteins; Cyt, cytosolic proteins

Sample Type	Number of Glycated Peptides	Number of Glycated Proteins
HP_WP	1368	799
HP_LAP	1028	380
HP	2205	1095
GHP_WP	2056	823
GHP_LAP	2349	940
GHP	3841	1592
RBC_Mem	1876	952
RBC_Cyt	1218	853
RBC	2851	1664
All Sample Types	7749	3742

Table 2

A selected list of low abundance glycated proteins and their associated glycated peptides specifically observed in the immunodepleted fraction of native human blood plasma

Gene Symbol	IPI No	Protein Name	Peptide Sequence	Peptide Observation Count	Protein concentration (μM)[¥]
AZGP1	IPI:IPI00166729.4	alpha-2-glycoprotein 1, zinc	R.QVEGM* EDWK#QDSQLQK.A	1	0.8–1.6
AZGP1	IPI:IPI00166729.4	alpha-2-glycoprotein 1, zinc	R.AK#AYLEEECPATLR.K	1	0.8–1.6
AZGP1	IPI:IPI00166729.4	alpha-2-glycoprotein 1, zinc	K.LK#CLAYDFYPGK.I	1	0.8–1.6
C2	IPI:IPI00303963.1	Complement C2 (Fragment)	R.LCK#SSGQWQTPGATR.S	1	0.1–0.3
C2	IPI:IPI00303963.1	Complement C2 (Fragment)	K.SNMGGSPK#TAVDHIR.E	1	0.1–0.3
C4B	IPI:IPI00418163.3	C4B1	R.VEASISK#ASSFLGEK.A	5	0.3–1
C4B	IPI:IPI00418163.3	C4B1	K.VLSLAQEQVGGSPEK#LQETSNWLLSQQQADGS FQDLSPVIHR.S	1	0.3–1
C4BPB	IPI:IPI00025862.2	Isoform 1 of C4b-binding protein beta chain	K.YSLELK#KAELK.A	2	0.5–1
C5	IPI:IPI00032291.2	Complement C5	R.LHM* K#TLLPVSKPEIR.S	4	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	R.GYGNSDYK#R.I	1	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.KVTCTNAELVK#GR.Q	4	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.KIEEIAAK#YK.H	3	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.TGEAVAEK#DSEITFIKK.V	5	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.K#LKEGMLSIM* SYR.N	1	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.VFK#DVFLEM* NIPYSVVR.G	1	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	R.FWKDNLQHK#DSSVPNTGTAR.M	2	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.K#IEEIAAK.Y	1	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	R.K#QTACKPEIAYAYK.V	4	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	R.LHMK#TLLPVSKPEIR.S	2	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	R.RK#EFPYRIPLDLVPK.T	1	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.GALHNYK#MTDK.N	2	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.GALHNYK#M* TDK.N	5	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.DNLQHK#DSSVPNTGTAR.M	3	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.K#LKEGM* LSIM* SYR.N	2	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.YK#ATLLDIYK.T	3	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.TGEAVAEK#DSEITFIK.K	7	0.4–0.6 for each chain

Gene Symbol	IPI No	Protein Name	Peptide Sequence	Peptide Observation Count	Protein concentration (μM)[¥]
C5	IPI:IPI00032291.2	Complement C5	K.VTCTNAELVK#GR.Q	2	0.4–0.6 for each chain
C5	IPI:IPI00032291.2	Complement C5	K.SPYIDK#ITHYNYLILSK.G	2	0.4–0.6 for each chain
C6	IPI:IPI00009920.2	Complement C6	K.GGNQLYCVK#M[*] GSSTSEK.T	2	0.5–0.9
C6	IPI:IPI00009920.2	Complement C6	K.GGNQLYCVK#MGSSTSEK.T	2	0.5–0.9
C6	IPI:IPI00009920.2	Complement C6	K.LSEK#HEGSFIQGAEK.S	8	0.5–0.9
C6	IPI:IPI00009920.2	Complement C6	K.LSEKHEGSFIQGAEK#SISLIR.G	4	0.5–0.9
C6	IPI:IPI00009920.2	Complement C6	K.AK#DLHLSDVFLK.A	3	0.5–0.9
C8B	IPI:IPI00294395.1	Complement C8 beta chain	R.FRK#PYNVESYTPQTQGK.Y	5	0.5–1
C8B	IPI:IPI00294395.1	Complement C8 beta chain	R.GILNEIK#DR.N	1	0.5–1
C8B	IPI:IPI00294395.1	Complement C8 beta chain	R.K#PYNVESYTPQTQGK.Y	1	0.5–1
C8B	IPI:IPI00294395.1	Complement C8 beta chain	R.GDYTLNNVHACAK#NDFK.I	5	0.5–1
C8B	IPI:IPI00294395.1	Complement C8 beta chain	R.SDLEVAHYK#LKPR.S	5	0.5–1
C9	IPI:IPI00022395.1	Complement C9	K.LSPIYNLVPVK#M[*] K.N	1	0.4–1
C9	IPI:IPI00022395.1	Complement C9	R.RPWNVASLIYETK#GEK.N	4	0.4–1
C9	IPI:IPI00022395.1	Complement C9	R.TEHYEEQIEAFK#SIIQEK.T	2	0.4–1
C9	IPI:IPI00022395.1	Complement C9	K.SIIQEK#TSNFNAAISLK.F	5	0.4–1
C9	IPI:IPI00022395.1	Complement C9	R.K#YAFELKEK.L	3	0.4–1
CFHR3	IPI:IPI00027507.1	Complement factor H-related protein 3	R.K#CYFPYLENGYNQNYGRK.F	3	0.2–1
CFHR3	IPI:IPI00027507.1	Complement factor H-related protein 3	R.K#CYFPYLENGYNQNYGR.K	2	0.2–1
CFI	IPI:IPI00291867.3	Complement factor I	R.GLETSLAECTFTK#R.R	1	0.5 for alpha or beta chain
CFI	IPI:IPI00291867.3	Complement factor I	K.YISQM[*] K#ACDGINDCGDQSDELCCK.A	2	0.5 for alpha or beta chain
CFI	IPI:IPI00291867.3	Complement factor I	K.LVDQDK#TM[*] FICK.S	3	0.5 for alpha or beta chain
GPX3	IPI:IPI00026199.1	Glutathione peroxidase 3	R.TTVSNVK#MDILSYM[*] R.R	1	0.8
GPX3	IPI:IPI00026199.1	Glutathione peroxidase 3	R.TTVSNVK#M[*] DILSYM[*] R.R	2	0.8
GPX3	IPI:IPI00026199.1	Glutathione peroxidase 3	K.GDVNGEKEQK#FYTFLK.N	4	0.8
GPX3	IPI:IPI00026199.1	Glutathione peroxidase 3	K.EQK#FYTFLK.N	1	0.8
HABP2	IPI:IPI00746623.2	Hyaluronan-binding protein 2	K.LK#PVDGHCALESK.Y	3	0.07–0.2
LRG1	IPI:IPI00022417.4	Leucine-rich alpha-2-glycoprotein	R.LHLEGNKLQVLGK#DLLLPQPDLR.Y	2	0.4
PROS1	IPI:IPI00294004.1	Vitamin K-dependent protein S	K.IAK#EAVMDINKPGPLFKPENGLLETK.V	1	0.3

Gene Symbol	IPI No	Protein Name	Peptide Sequence	Peptide Observation Count	Protein concentration (μM)[¥]
PROS1	IPI:IPI00294004.1	Vitamin K-dependent protein S	K.QGASGIK#EIIQEK.Q	2	0.3
PROS1	IPI:IPI00294004.1	Vitamin K-dependent protein S	K.IAK#EAVM* DINKPGPLFKPENGLLETK.V	1	0.3
PROS1	IPI:IPI00294004.1	Vitamin K-dependent protein S	K.DGK#ASFTCTCKPGWQGEK.C	3	0.3
PROS1	IPI:IPI00294004.1	Vitamin K-dependent protein S	R.NNLELSTPLK#IETISHEDLQR.Q	1	0.3
PROS1	IPI:IPI00294004.1	Vitamin K-dependent protein S	R.K#VESELIKPINPR.L	4	0.3
SERPINA4	IPI:IPI00328609.3	Kallistatin	R.VGSALFLSHNLK#FLAK.F	2	0.2
SERPINA4	IPI:IPI00328609.3	Kallistatin	R.M* DYKGDATVFFILPNQGK#M* R.E	1	0.2
SERPINA7	IPI:IPI00292946.1	Thyroxine-binding globulin	K.MGIQHAYSENADFSGLTEDNGLK#LSNAAHK.A	1	0.2–0.5
SERPINA7	IPI:IPI00292946.1	Thyroxine-binding globulin	K.EGQM* ESVEAAM* SSK#TLK.K	2	0.2–0.5
SERPINA7	IPI:IPI00292946.1	Thyroxine-binding globulin	K.M* GIQHAYSENADFSGLTEDNGLK#LSNAAHK.A	1	0.2–0.5
SERPINF1	IPI:IPI00006114.4	Pigment epithelium-derived factor	R.K#TSLEDFYLDEER.T	3	0.1
SERPINF1	IPI:IPI00006114.4	Pigment epithelium-derived factor	K.LSYEGEVTK#SLQEM* K.L	1	0.1
SERPINF1	IPI:IPI00006114.4	Pigment epithelium-derived factor	K.LQSLFDSPDFSK#ITGKPIK.L	2	0.1
SERPINF1	IPI:IPI00006114.4	Pigment epithelium-derived factor	R.IK#SSFVAPLEK.S	1	0.1
SERPINF1	IPI:IPI00006114.4	Pigment epithelium-derived factor	K.LK#LSYEGEVTK.S	1	0.1
SERPINF1	IPI:IPI00006114.4	Pigment epithelium-derived factor	R.VPM* M* SDPK#AVLR.Y	1	0.1
SERPINF1	IPI:IPI00006114.4	Pigment epithelium-derived factor	K.ITGK#PIKLTQVEHR.A	2	0.1
SHBG	IPI:IPI00023019.1	Isoform 1 of Sex hormone-binding globulin	R.RDSWLDK#QAEISASAPTSLR.S	2	0.03–0.2

¥ Protein concentrations as shown in Reference 39.

* and # in peptide sequence represent oxidized methionine and the Amadori compound modification, respectively

Table 3

Major glycation sites identified from the most abundant proteins in human plasma, erythrocyte membrane and cytosol. Only individual peptides with spectra count of ≥5 are included. To be consistent with previous literature reports, the glycation sites were assigned after having the signal peptides, propeptides or initiator methionine removed from the protein precursor sequences

Protein Entry	Gene Symbol	Protein Name	Glycation Sites
P02768	ALB	Serum Albumin	12, 51, 64, 73, 93, 136*, 137*, 159, 162, 174, 181, 190, 195, 199, 205, 225, 233, 240, 262, 274, 276, 281, 313, 323, 351, 359, 378, 389, 402, 413*, 414, 444, 466, 500*, 519, 524, 525, 534, 541, 545, 557
P69905	HBA1, HBA2	Hemoglobin subunit alpha	11, 16, 40, 56, 60, 61, 99, 127,139
P68871	HBB	Hemoglobin subunit beta	8, 59, 66, 82, 120, 132
P02042	HBD	Hemoglobin subunit delta	8, 65, 66, 132, 144
P02549	SPTA1	Spectrin alpha chain	59, 67, 72, 104, 130, 210, 281, 283, 319, 345, 554, 579, 586, 595, 634, 728, 779, 858, 918, 940, 1102, 1374, 1468, 1486, 1503, 1579, 1586, 1708, 1732, 1947, 1992, 1994, 2051, 2061, 2132, 2200
P11277	SPTB	Spectrin beta chain	79, 89, 213, 291, 301, 358, 461, 547, 568, 640, 653, 655, 728, 755, 785, 1247, 1261, 1342, 1518, 1675, 1684, 1695, 1771, 1815, 1869, 1876, 1904, 1982, 1996

* denotes identified from *in vitro* glycated plasma only

Table 4

Glycated proteins identified with the majority of their peptides having consistent up-regulation in AGT vs NGT blood plasma

Gene Symbol	IPI No	Protein Name	Peptide Sequence	Peptide Count AGT	Peptide Count NGT
APOH	IPI:IPI00298828.3	Beta-2-glycoprotein 1	K.NGM* LHGDKVSFFCK#NK.E	5	2
APOH	IPI:IPI00298828.3	Beta-2-glycoprotein 1	K.NGMLHGDKVSFFCK#NK.E	4	NA
C4BPA	IPI:IPI00021727.1	C4b-binding protein alpha chain	R.TPSCGDICNFPPK#IAHGHYK.Q	9	5
C4BPA	IPI:IPI00021727.1	C4b-binding protein alpha chain	R.KPDVSHGEMVSGFGPIYNYK#DTIVFK.C	8	NA
C4BPA	IPI:IPI00021727.1	C4b-binding protein alpha chain	R.KPDVSHGEM* VSGFGPIYNYK#DTIVFK.C	3	NA
C8B	IPI:IPI00294395.1	Complement C8 beta chain	R.FRK#PYNVESYTPQTQGK.Y	4	1
C8B	IPI:IPI00294395.1	Complement C8 beta chain	R.SDLEVAHYK#LKPR.S	4	1
CLU	IPI:IPI00291262.3	Clusterin	R.KNPK#FM* ETVAEK.A	6	3
CLU	IPI:IPI00291262.3	Clusterin	R.KNPK#FMETVAEK.A	5	3
F13A1	IPI:IPI00297550.8	Coagulation factor XIII A chain	K.KDGTHVVENVDATHIGK#LIVTK.Q	7	1
HPX	IPI:IPI00022488.1	Hemopexin	K.EKGYPK#LLQDEFPGIPSPLDAAVECHR.G	33	1
HPX	IPI:IPI00022488.1	Hemopexin	K.GGYTLVSGYPK#RLEK.E	18	9
HPX	IPI:IPI00022488.1	Hemopexin	K.GDKVWVYPPEKK#EK.G	11	4
HPX	IPI:IPI00022488.1	Hemopexin	K.GEFVWK#SHKWDR.E	9	3
HPX	IPI:IPI00022488.1	Hemopexin	K.SLGPNSCSANGPGLYLIHGPNLYCYSDVEK#LNAAK.A	8	3
HPX	IPI:IPI00022488.1	Hemopexin	K.GDK#VWVYPPEKK.E	7	3
HPX	IPI:IPI00022488.1	Hemopexin	K.GEFVWK#SHK.W	5	NA
HRG	IPI:IPI00022371.1	Histidine-rich glycoprotein	K.ALEK#YKEENDDFASFR.V	7	2
HRG	IPI:IPI00022371.1	Histidine-rich glycoprotein	K.FK#SGFPQVSM* FFTHTFPK.-	5	2
IGL@	IPI:IPI00154742.6	IGL@ protein	K.ADSSPVK#AGVETTTPSK.Q	6	2
IGL@	IPI:IPI00154742.6	IGL@ protein	K.LTVLGQPK#AAPSVTLFPPSSEELQANK.A	6	3

* and # in peptide sequence represent oxidized methionine and Amadori compound modification, respectively. NA = not observed

Table 5

A selected list of glycated peptides identified from spectrin that show up-regulation in AGT vs NGT erythrocytes

Gene Symbol	IPI No	Protein Name	Glycation Site	Peptide Sequence	Peptide Count AGT	Peptide Count NGT
SPTA1	IPI:IPI00220741.6	spectrin alpha chain	210	K.FEDFQVELVAK#EGR.V	7	4
			210	K.KFEDFQVELVAK#EGR.V	50	22
			281	R.DVTEAIQWIK#EK.E	14	9
			319	R.NLAVMSDK#VK.E	8	3
			579	R.RLLK#ESLLLQK.L	4	1
			595	K.LYEDSDDLK#NWINK.K	5	2
			779	R.FEALK#EPLATR.K	6	3
			940	K.K#HEAFLLDLNSFGDSM* K.A	8	3
			1102	K.K#AENTGVELDDVWELQK.K	21	12
			1468	K.ITDLEHFAESLIADEHYAK#EEIATR.L	9	7
			1486	K.ALK#AQLIDER.T	29	15
			1992	R.LPEITDLK#DK.L	30	5
			2132	R.TWK#HLSDIIEER.E	10	6
SPTB	IPI:IPI00216704.8	Spectrin beta chain	79	R.ITDLYK#DLR.D	6	4
			291	R.VGK#VIDHAIETEK.M	6	2
			354	R.TVEKPPK#FQEK.G	3	NA
			568	K.HLLEVEDLLQK#HK.L	5	NA
			640	R.LWK#FFWEM* DEAESWIK.E	5	NA
			653	K.FFWEM* DEAESWIK#EK.E	6	2
			653	K.FFWEMDEAESWIK#EK.E	7	2
			728	K.EVSAQWDQLK#DLAAFCK.K	6	3
			728	R.IKEVSAQWDQLK#DLAAFCK.K	18	5
			1174	R.SHTLAQCLGFQEFQK#DAK.Q	3	NA
			1675	R.LQGQVDKHYAGLK#DVAEER.K	12	9
			1684	K.RK#LENM* YHLFQLK.R	3	NA
			1684	K.RK#LENMYHLFQLK.R	3	NA
			1684	R.K#LENM* YHLFQLK.R	9	3

Gene Symbol	IPI No	Protein Name	Glycation Site	Peptide Sequence	Peptide Count AGT	Peptide Count NGT
			1684	R.K#LENMYHLFQLK.R	12	7
			1695	K.LENM* YHLFQLK#R.E	9	7
			1695	K.LENMYHLFQLK#R.E	7	4
			1695	K.RKLENM* YHLFQLK#R.E	15	4
			1695	R.KLENM* YHLFQLK#R.E	69	39
			1695	R.KLENMYHLFQLK#R.E	26	13
			1815	R.YFYTGAEILGLIDEK#HR.E	37	32
			1869	R.LQTAYAGEK#AEAIQNK.E	10	8
			1876	K.AEAIQNK#EQEVSAAWQALLDACAGR.R	35	19
			1996	R.RKEMNEK#WEAR.W	4	1

* and # in peptide sequence represent oxidized methionine and Amadori compound modification, respectively. NA = not observed