research articles Journal of proteome research

Population Studies of Vitamin D Binding Protein Microheterogeneity by Mass Spectrometry Lead to Characterization of its Genotype-Dependent O-Glycosylation Patterns

Chad R. Borges,* Jason W. Jarvis, Paul E. Oran, and Randall W. Nelson

Molecular Biosignatures Analysis Unit, The Biodesign Institute at Arizona State University, Tempe, Arizona 85287

Received April 17, 2008

Mass spectrometric evidence presented here characterizes the genotype-dependent glycosylation patterns for each of the three major allele products of Vitamin D Binding Protein found in the general human population. Findings based on the analysis of over 100 individual plasma samples demonstrated that all DBP allele products, except GC*2, are modified (10−25 mol%) with a linear $(NeuNAc)_1$-$(Gal)_1(GalNAc)_1$ trisaccharide and, to a much lesser extent (1−5 mol%) with a trisaccharide-independent $(Gal)_1(GalNAc)_1$ dissaccharide. GC*2 protein contains the disaccharide but remains completely free of the trisaccharide, even in heterozygous individuals possessing a second gene product that is modified with the trisaccharide. Thus, all allelic forms of DBP except GC*2 possess two independent O-glycosylation sites occupied by separate, yet consistently isomass oligosaccharides and, despite a consensus sequence, lack N-glycosylation.

Keywords: Vitamin D binding protein • GcG • population proteomics • O-glycosylation • genotype • mass spectrometric immunoassay

Introduction

Modern genomic[1] and proteomic[2] science increasingly recognize the need to characterize individual biochemical diversity. Ultimately, thorough understanding of this diversity promises to form the framework around which personalized human medicine will operate.[1]

Group-specific component globulin, or vitamin-D binding protein (DBP) as it came to be known in 1975,[3] has been studied as a differentiating marker of inheritance since 1960.[4] Today, three common alleles are known to exist along with many rare variants.[5,6] Major biological roles for DBP include vitamin D metabolite transport,[3,7] fatty acid transport,[8,9] actin sequestration,[6] and macrophage activation.[10,11] The amino acid and gene sequences of the three common variants of DBP are known (UniProtKB/Swiss-Prot entry P02774) and numerous studies in the past few decades have solved one or more structural aspects of DBP glycosylation, but because the protein has not been systematically and, simultaneously, comprehensively characterized across a large sampling of the human population, the story has remained incomplete. Briefly, according to a 1983 study by Viau et al.[12] only the GC*1 gene product is (partially) O-glycosylated with the linear trisaccharide NeuNAc α(2 → 3) Gal β(1 → 3) GalNAc α(1 → 0) Ser or Thr. (However, in cases of liver cirrhosis or hepatitis, a branched tetrasaccharide containing two sialic acid residues may be present on some GC*1 proteins.[12,13]) Other studies have provided evidence to suggest that the GC*2 variant is also modified with small quantities of carbohydrate[14,15] (a disaccharide[16–18]) that the oligosaccharide structure and composition differs between the three common DBP protein variants[18,16,17] (and is not the linear structure asserted by Viau[12]), and that all circulating GC*1 DBP molecules are glycosylated (i.e., GC*1 DBP is 100% glycosylated) but only 10% of GC*2 DBP molecules are glycosylated.[16,17] A recent study of intact DBP by mass spectrometry found evidence for the nominal trisaccharide suggested by Viau et al.,[12] but was unable to confirm which allelic forms of the protein were glycosylated and where the glycosylation site might be located.[19]

Thus, some important features of DBP glycosylation have been elucidated, but several important questions regarding the diversity of DBP glycosylation in the human population remain unanswered. For example, If all three common allelic variants of DBP are glycosylated, what percentage of the expressed protein is glycosylated for each allele product? Is glycosylation homogeneous, that is, the same glycoform across all protein variants? If not, what are the different glycoforms and with what distribution are they attached to the common genetic variants? Is glycosylation qualitatively consistent from person to person within a single genetic variant? DBP contains an N-glycosylation consensus sequence. Does any N-glycosylation occur as suggested in the UniProtKB/Swiss-Prot database? Indirect evidence[14] suggests that the K420T mutation that defines the shift from the GC*2 to the GC*1 variants provides a site of O-glycosylation, but no direct structural evidence has been provided showing that T420 of the GC*1 variants is a precise

* Author to whom correspondence should be addressed: Chad R. Borges, Molecular Biosignatures Analysis Unit, The Biodesign Institute at Arizona State University, P.O. Box 876401, Tempe, AZ 85287. Tel, 480-727-9928; fax, 480-727-9464; e-mail, chad.borges@asu.edu.

10.1021/pr8002936 CCC: $40.75

Published on Web 08/08/2008

site of glycosylation. Can T420 be directly confirmed as a site of glycosylation? Is there evidence for trisaccharide independent disaccharidic O-glycosylation of GC*1 variants? If so, where might it be attached?

As part of an ongoing study of targeted, intact proteins across human populations, these studies began at the analysis of intact DBP from the plasma of over 100 individuals by high-throughput mass spectrometric immunoassay coupled with electrospray ionization time-of-flight mass spectrometry (MSIA ESI-TOF-MS). The resolving power of this technique led to the observation of over a dozen different forms of DBP, some of which appeared coupled with particular genotypes. This prompted a more detailed investigation upon specific Arg-C derived DBP peptides by MALDI-LIFT-TOF/TOF which, combined with the populational observations on intact DBP, led to definitive answers for the above questions.

Materials and Methods

Materials. Polyclonal rabbit anti-human DBP (GC-Globulin) antibodies (Cat. No. A0021) were obtained from DAKO (Carpinteria, CA). According to the manufacturer's specifications, this antibody is for *in vitro* diagnostic use and is intended for the determination of DBP in gel immunoprecipitation techniques and for phenotyping of DBP by immunofixation. Premixed MES-buffered saline powder packets were from Pierce (Rockford, IL). Isolation of DBP from plasma was carried out with proprietary MSIA pipet tips from Intrinsic Bioprobes, Inc. (Tempe, AZ) derivatized with the DBP antibodies via 1,1′-carbonyldiimidazole (CDI) chemistry as previously described.[20,21] EDTA plasma samples from African-American, Caucasian, and Hispanic individuals were purchased from ProMedDx (Norton, MA). Protein Captrap catridges for LC/MS were obtained from Michrom Bioresources, Inc. (Auburn, CA). Premade 10 mM Hepes-buffered saline (HBS) was bought from Biacore, Inc. (Piscataway, NJ). Endoproteinase Arg-C from *Clostridium histolyticum* was obtained from Roche Applied Science (Indianapolis, IN). All other chemicals were obtained from Sigma-Aldrich (St. Louis, MO).

Sample Preparation for the Analysis of Intact DBP by MSIA ESI-TOF-MS. One hundred twenty-five microliters of EDTA plasma was mixed with 1.25 μL of 10% Tween 20 then diluted to 250 μL total volume with HBS. Samples were frozen at −80 °C until use. With the aid of a Beckman Multimek 96-channel automated pipettor, MSIA pipet tips that had been preactivated with CDI[20,21] were derivatized with DBP antibodies by repetitively flowing (aspirating and dispensing 500 times) 50 μL volumes of antibody solution (150 μL/well; 0.05 g/L in 0.1 M MES buffered saline, pH 4.7) through the tips. Antibody-linked tips were stored in HBS at 4 °C until the day of use at which time they were prerinsed (400 μL/well; 150 μL aspirate and dispense cycles; 10 cycles) with HBS, then used to extract DBP from individual plasma samples at room temperature (250 μL/well; 85 μL aspirate and dispense cycles; 250 cycles). Pipette tips were then ejected from the robot and allowed to sit in their respective plasma samples at room temperature until they were individually (manually) washed (by drawing from a fresh reservoir of liquid and dispensing to waste) and eluted as follows: five cycles of 150 μL of HBS, five cycles of 150 μL distilled water, five cycles of 2 M ammonium acetate/acetonitrile (3:1 v/v), 10 cycles of distilled water. Elution was accomplished by air-drying the pipet frits, then drawing 5 μL of a mixture of 100% formic acid/acetonitrile/distilled water (9/5/1 v/v/v), mixing over the pipet affinity capture frit for 20−30 s, and dispensing into a 96-conical well polypropylene autosampler tray. Frits were then washed with an additional 5 μL of distilled water which was used to dilute the eluted sample. Five microliters was injected into the LC-TOF-MS within 10 min of elution.

ESI-TOF-MS. A trap-and-elute form of sample concentration/solvent exchange rather than traditional LC was used for these analyses. Five-microliter samples were injected by a Spark Holland Endurance autosampler in microliter pick-up mode and loaded by an Eksigent nanoLC*1D at 10 μL/min (90/10 water/acetonitrile containing 0.1% formic acid, Solvent A) onto a protein captrap (polymeric/reversed phase sorbent, Microm Bioresources, Auburn, CA) configured for unidirectional flow on a 6-port divert valve. After 2 min, the divert valve position was automatically toggled and flow over the cartridge changed to 1 μL/min Solvent A (running directly to the ESI inlet) which was immediately ramped over 8 min to 10/90 water/acetonitrile containing 0.1% formic acid. By 10.2 min, the run was completed and the flow was back to 100% solvent A.

The bulk of DBP eluted between 5.5 and 7.5 min into a Bruker MicrOTOF-Q (Q-TOF) mass spectrometer operating in positive ion, TOF-only mode, acquiring spectra in the m/z range of 50−3000 Th. ESI settings for the Agilent G1385A capillary nebulizer ion source were as follows: end plate offset, −500 V; capillary, −4500 V; nebulizer, nitrogen 2 bar; dry gas, nitrogen 3.0 L/min at 225 °C. Data were acquired in profile mode at a digitizer sampling rate of 2 GHz. Spectra rate control was by summation at 1 Hz.

Data Analysis for Intact DBP. Approximately 1.5 min of recorded spectra were averaged across the chromatographic peak apex of DBP elution. The ESI charge-state envelope was deconvoluted with Bruker DataAnalysis v3.4 software to a mass range of 1000 Da on either side of any identified peak. Deconvoluted spectra were baseline substracted and all peaks were integrated. Tabulated mass spectral peak areas were exported to a spreadsheet for further calculation and determination of the peak areas of interest relative to all other forms of DBP present in the mass spectrum.

Sample Preparation for DBP Glycopeptide Analysis by MALDI-LIFT-TOF/TOF. Five hundred micrograms of 90% pure, lyophilized DBP from pooled human blood plasma (Sigma-Aldrich, Cat. no. G8764, Lot no. 025K12751) was reconstituted in 500 μL of water to give a 0.9 g/L or 18 μM solution of DBP. (The mass spectrum of this purified/pooled DBP looks like one might expect given the spectra of several individuals with different genotypes; i.e., all the major peaks are present but broadened due to the contributions from the various mass shifted genotypes.) To 50 μL (45 μg or 900 pmol DBP) was added 50 μL of 0.1 M Tris buffer, pH 7.6 containing 10 mM $CaCl_2$, 6 μL of acetonitrile, 13 μL of 50 mM DTT dissolved in the Tris buffer, and 10 μL (1 μg) of reconstituted Arg-C. The sample was incubated with shaking at 750 rpm for 2 h at 37 °C. A 2-μL aliquot was mixed with 2 μL of acetonitrile and 2 μL of 3 mg/mL 2,4,6- trihydroxyacetophenone (THAP) in 1:3 acetonitrile/50 mM dibasic ammonium citrate (as described by Pouria et al.[22]) on a ground steel MALDI target plate, then quickly dried under vacuum. This THAP spot was used for the analysis of trisaccharide-containing peptides. A second spot was made by mixing a 2-μL sample aliquot with 2 μL of acetonitrile and 2 μL of 1:2 acetonitrile/water containing 0.4% TFA (v/v) which was altogether nearly saturated with α-cyano-4-hydroxycinnamic acid (CHCA). The spot was dried quickly

in a vacuum chamber. This CHCA spot was used for the analysis of unmodified and disaccharide-modified peptides.

Analysis of Glycosylated DBP-Derived Peptides by MALDI-TOF and MALDI-LIFT-TOF/TOF. Single stage and LIFT-TOF/TOF mass spectra of glycosylated peptides were acquired on a Bruker Ultraflex MALDI-TOF/TOF instrument. Mass spectra were acquired in positive ion mode with the reflector engaged. Single stage MALDI-TOF mass spectra were externally calibrated with a mixture of 7 peptides supplied by Bruker ranging in monoisotopic m/z from 1046.54 (Angiotensin II) to 3147.47 (Somatostatin), then internally calibrated at m/z 1589.77, corresponding to the V203−R216 Arg-C fragment of DBP. The LIFT-TOF/TOF instrument control method was calibrated according to the instrument manufacturer's protocol and was run without CID gas. Trisaccharide-containing peptides and their sodium adducts were only observed in spectra acquired from the THAP matrix, but disaccharide-containing and unmodified peptides were best observed in spectra acquired from the CHCA matrix.

Results

Figure 1 shows a deconvoluted mass spectrum for each of the three major human DBP allele products in homozygous form and the GC*1F/GC*2 heterozygous case. Several features are notable.

DBP Genotype. Classification of DBP according to the genotypes listed in the UniProtKB/Swiss-Prot database was readily accomplished by MSIA ESI-TOF-MS of the intact protein. The three most common DBP alleles arise from two point mutations, D416E and T420K, which generate: the GC*1F allele (containing D416 and T420), GC*1S allele (containing E416 and T420), and the GC*2 allele (containing D416 and K420). All diploid genotype results are unambiguous and produced in an average of about 15−20 min per sample (from raw plasma sample to deconvoluted mass spectrum) when run in batch mode. The molecular masses of the DBP variants as given by the deconvoluted mass spectra fit accurately and within a 2 Da standard deviation the sequences of the major DBP alleles as reported in the UniProtKB/Swiss-Prot database considering that all cysteine residues are involved in disulfide bonds. On the basis of this highly accurate and precise observation of the molecular weights of the intact proteins, the GC*1 allele products do not contain Glu at residue number 152 or Arg at residue number 311 in contrast to the GC*2 allele product which contains Gly at residue number 152 and Glu at residue number 311, as has been reported in some cases.[23,24] Table 1 shows the genotype distribution for the 226 allele products (113 people) analyzed in this study.

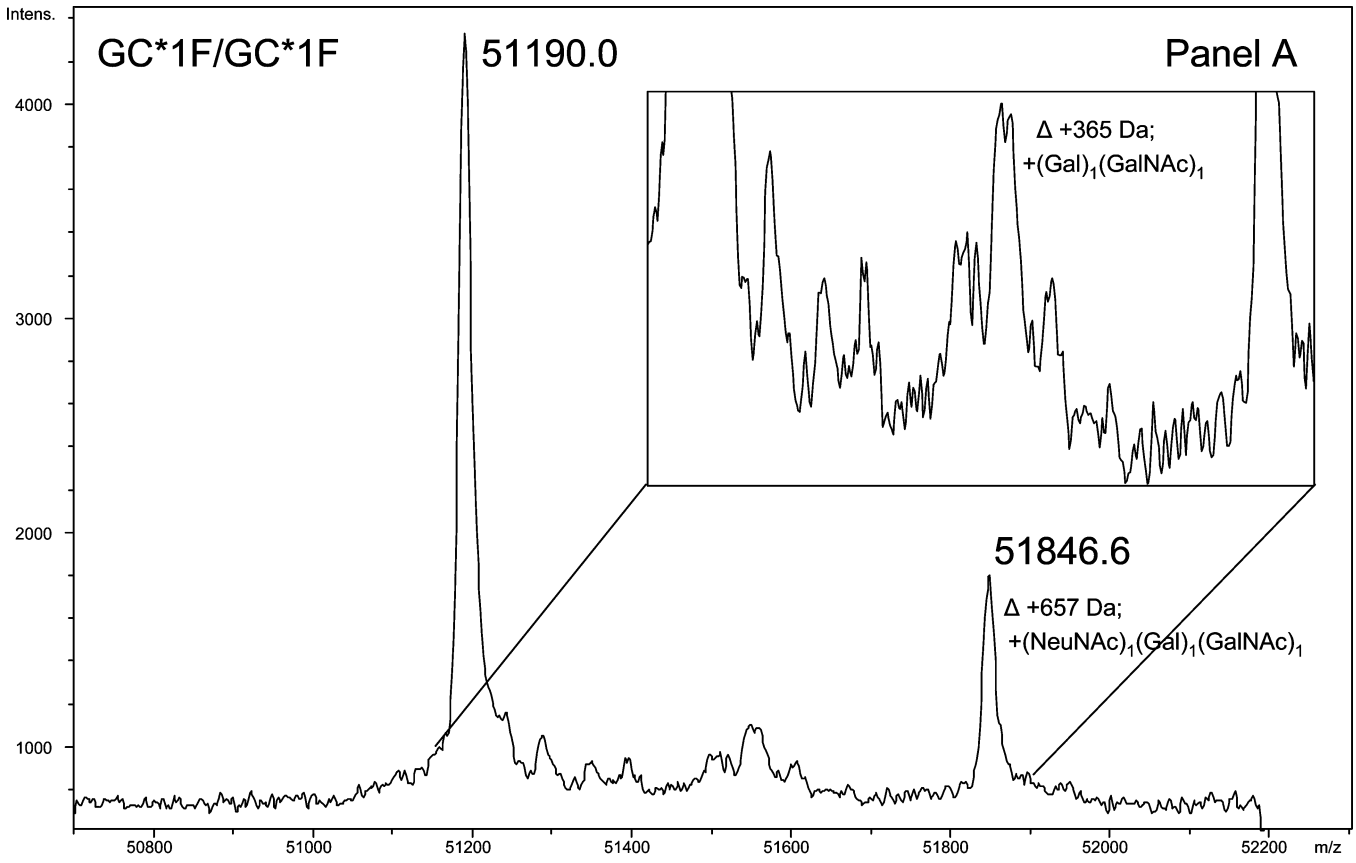
DBP Glycosylation. As seen in Figure 1, glycosylation of DBP by the $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide (Δm +656.6 Da)[12,16,17] is partial and is limited to the GC*1 gene products. (Based on previous literature reports involving enzyme-based study of the oligosaccharide structure of DBP,,[16,17] the specific designation of 'Gal' rather than 'Hex' is reported here.) However, a small, broad peak at about +365 Da from the unmodified GC*2 protein (Figure 1C) fits the expected mass shift of the $(Gal)_1(GalNAc)_1$ disaccharide proposed for this particular allele product by Yamamoto and Homma based on the generation of DBP-macrophage activating factor (GcMAF) upon treatment of GC*2 protein with only β-galactosidase without sialidase.[17] A similar peak with the same mass shift also appears in the deconvoluted mass spectra of intact GC*1F and GC*1S proteins (Figure 1A,B). (In some spectra, this wide peak is resolved into two peaks, confirming the presence of two species, the heaviest of which corresponds precisely with the expected mass shift of the predicted disaccharide. The identity of the lower mass peak remains unknown.) Additionally, a deconvoluted mass spectrum from some GC*1 samples such as that seen in Figure 1B (homozygous GC*1S/GC*1S) present trace evidence that an individual GC*1 protein molecule carries either the disaccharide, the trissacharide, or both oligosaccharides, as appears, for GC*1S, at m/z 52225. This evidence for the presence of both oligosaccharides on the same molecule (along with knowledge that the T420K point mutation (GC*2) completely eliminates the trisaccharide attachment site but retains the disaccharide site) suggests that the disaccharide of GC*1 proteins is not simply an asialo form of the $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide. Taken altogether, these data suggest that there are two distinct glycosylation sites on GC*1 proteins (one for a $(Gal)_1(GalNAc)_1$ disaccharide and one for a $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide) and a single glycosylation site for the $(Gal)_1(GalNAc)_1$ disaccharide on GC*2 proteins. Table 2 summarizes the relative degree of glycosylation for the three major DBP allele products in all combinations based upon the relative deconvoluted mass spectral peak area integrals of interest compared to (divided by) the summed areas of all mass spectral peaks present.

No data from the 113 samples analyzed support the presence of DBP N-glycosylation. This is in contrast to the potential for N-glycosylation reported on the UniProtKB/Swiss-Prot database and a report by Wang et al.,[25] but in corroboration of the findings of Viau et al.[12] Neither was a mass shift of +527 Da detected, indicative of the $(Gal)_1(Man)_1(GalNAc)_1$ trisaccharide suggested by Ohkura et al.[18] for the GC*1S protein.

Analysis of Glycosylated DBP-Derived Peptides by MALDI-LIFT-TOF/TOF. Figure 2 shows a MALDI-TOF mass spectrum over the m/z range expected for the unmodified Arg-C peptide L410-R429 of DBP obtained from a pooled human plasma source. The monoisotopic peaks at 2206.32 (calc. 2206.33) and 2220.34 (calc. 2220.34) represent the peptides (R)/LKAKLP-**D**ATPTELAKLVNKR/(S) and (R)/LKAKLP**E**ATPTELAKLVNKR/(S) of the GC*1F and GC*1S allele products, respectively. These peaks were among the most intense in the entire mass spectrum of the Arg-C digest of DBP. Only a trace quantity of the GC*2 allele product-derived peptide (R)/LKAKLP**D**ATP-**K**ELAKLVNKR/(S) at m/z 2233.4 is observed, likely due to a low fractional contribution of GC*2 allele products to the pooled plasma source used to purify this DBP sample.

Panel A of Figure 3 shows a MALDI-TOF mass spectrum over the m/z range expected for the same peptides depicted in Figure 2, but modified with a $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide. The monoisotopic peaks at m/z 2862.58 (calc. 2862.56) and 2884.56 represent the MH^+ and sodiated forms, respectively, of a $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide-glycosylated peptide with the sequence (R)/LKAKLP**D**ATPTELAKLVNKR/(S) derived from the GC*1F allele product. Likewise, the monoisotopic peaks at m/z 2876.58 (calc. 2876.57) and 2898.58 represent the MH^+ and sodiated forms, respectively, of a $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide-glycosylated peptide with the sequence (R)/LKAKLP**E**ATPTELAKLVNKR/(S) derived from the GC*1S allele product.

MALDI-LIFT-TOF/TOF analysis with precursor ion selection for the cluster of peaks from m/z 2862 to 2898 is shown in Panels B and C of Figure 3. The base peak in the TOF/TOF spectrum is due to the facile loss of sialic acid from the trisaccharide. Panel C of Figure 3 shows all peaks in the TOF/



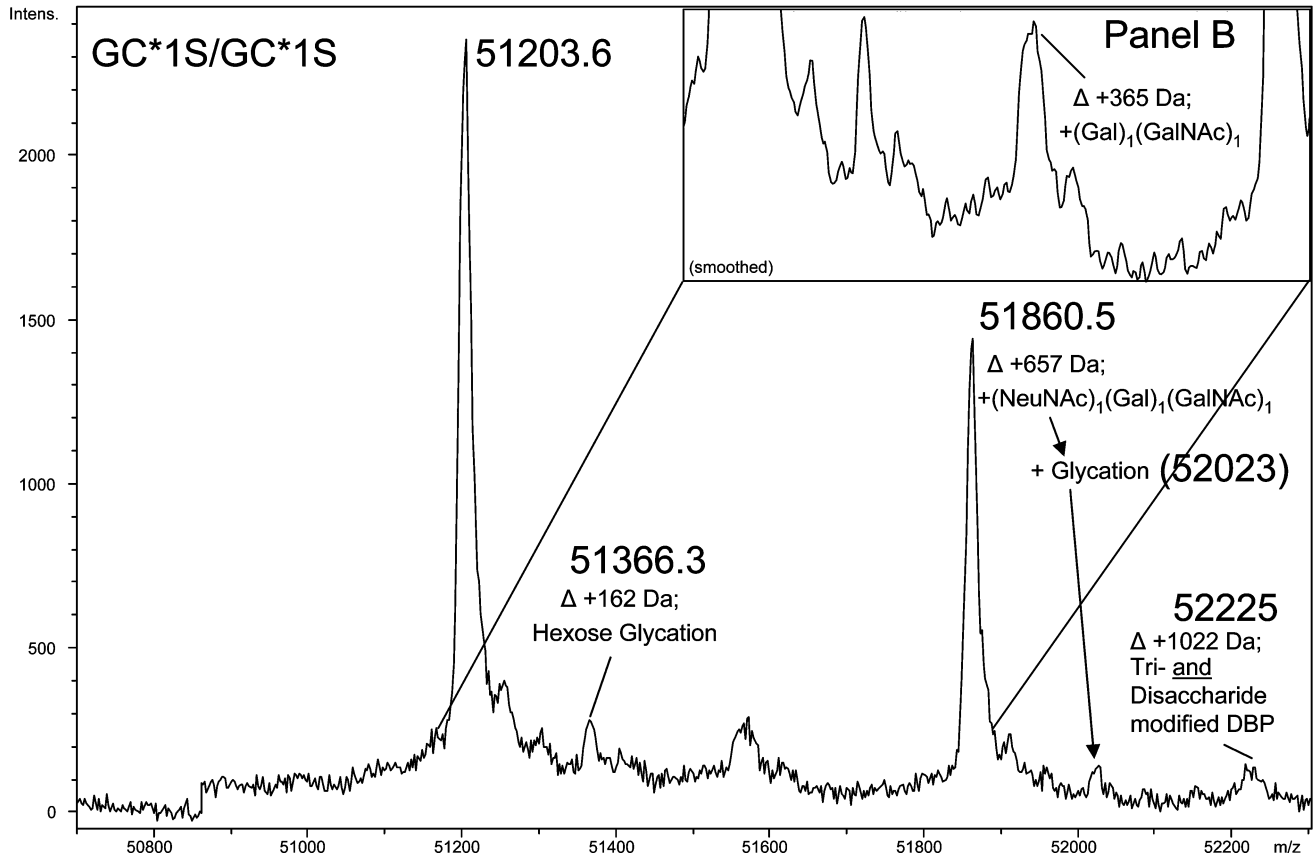


Figure 1. (Continued)

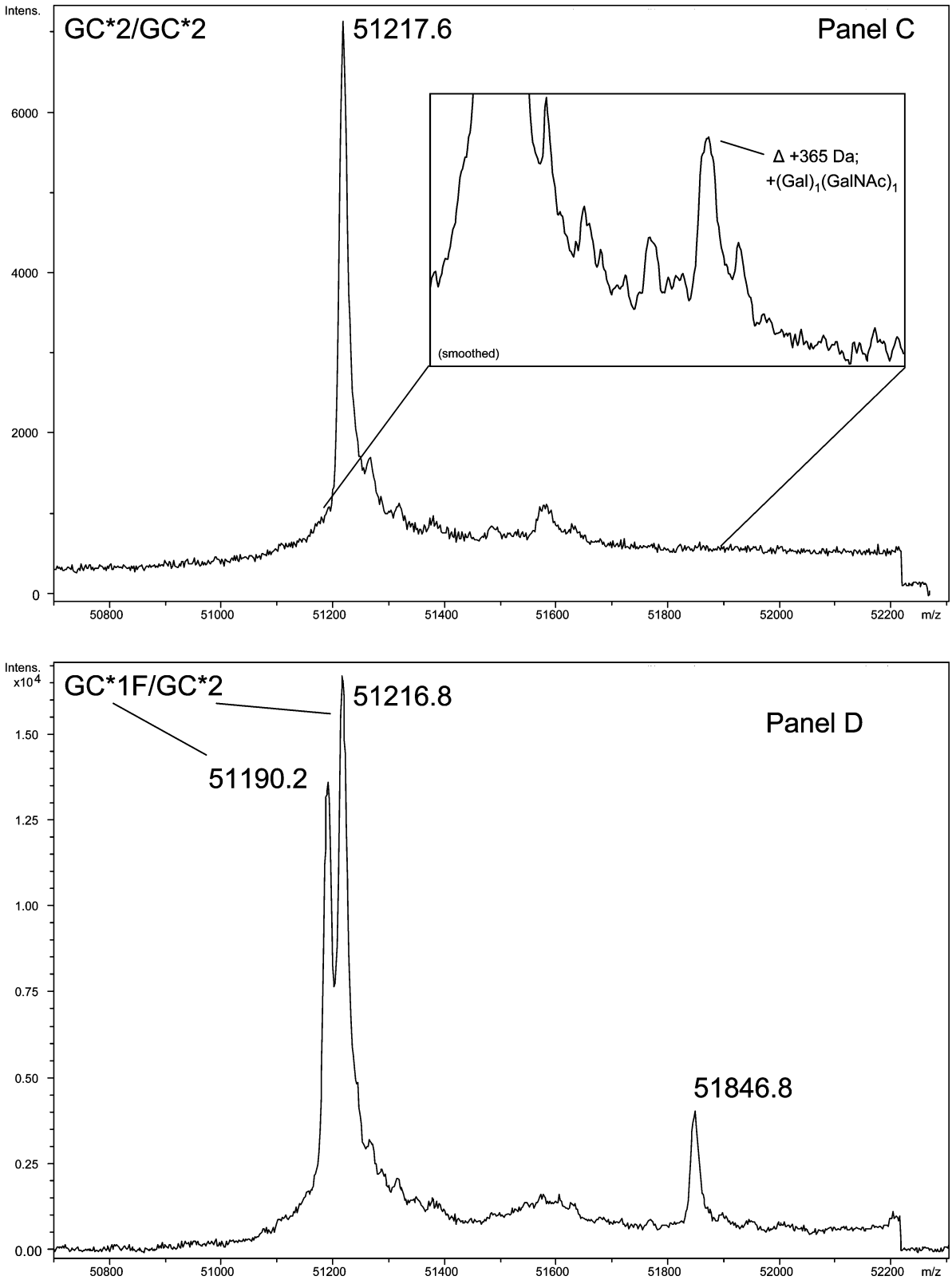


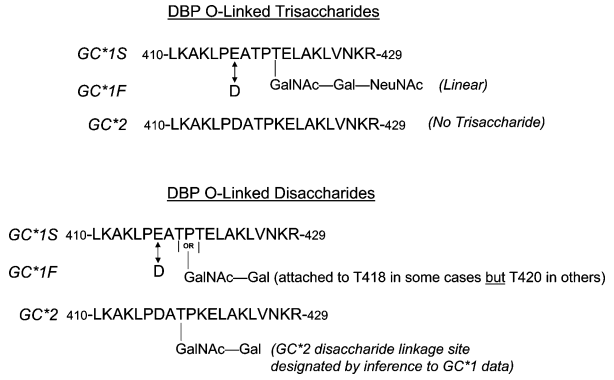
Figure 1. ESI-TOF mass spectra of intact DBP directly isolated from the blood plasma of 4 different people, the first three of which are homozygous for one of the three major DBP alleles. The base peaks at approximately 51 200 Da represent unmodified DBP (which varies in exact mass according to DBP genotype), a peak at Δm +656.6 represents an O-linked glycoform (a $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide),[12,16,17] and a small peak at Δm +162 Da (if present) represents glycated DBP. Displayed masses correspond to the observed *m*/*z* value of an "MH^+" species. Insets are magnifications of the indicated region of the spectrum. (A) Homozygous GC*1F/GC*1F DBP (calculated unmodified DBP *m*/*z* = 51189.2); (B) homozygous GC*1S/GC*1S DBP (calculated unmodified DBP *m*/*z* = 51203.2); (C) homozygous GC*2/GC*2 DBP (calculated unmodified DBP *m*/*z* = 51216.3); (D) heterozygous GC*1F/GC*2 showing resolved allele products and association of trisaccharide glycoform with only the GC*1F gene product. Though not shown directly, the mass resolution is sufficient to readily classify all heterozygous combinations according to diploid genotype.

Table 1. Frequency of DBP Alleles Observed in 113 Individual Samples (226 Alleles)[a]

	Not in DB	GC*1F	GC*1S	GC*2
Number of Alleles Observed	5	121	61	39
Percentage of All Alleles Observed	2%	54%	27%	17%

[a] Samples analyzed were of African-American, Caucasian, or Hispanic origin. 'Not in DB' indicates not in UniProtKB/Swiss-Prot database.

Scheme 1. Genotype-Linked O-Glycoform Summary for the Three Most Common DBP Allele Products



TOF spectrum with masses less than that of the base peak. Of interest are peaks at m/z 2423.96 (calc. 2423.42) and 2220.73 (calc. 2220.34) which indicate loss of a $NeuNAc_1Gal_1$ disaccharide and loss of the entire $NeuNAc_1Gal_1GalNAc_1$ trisaccharide, respectively, to generate the unmodified L410−R429 peptide of GC*1S. (Analogous peaks for the GC*1F peptide are present but not annotated.) As shown in the inset of Panel B of Figure 3, there is no evidence for the loss of galactose alone from the trisaccharide. Taken together, these data support a linear (rather than branched) trisaccharide structure.

Also of interest in Panel C of Figure 3 are peaks at m/z 1828.01 (calc. 1827.94) and 1853.98 (calc. 1853.92) which correspond to y10 and x10 ions, respectively, containing the trisaccharide located at the second threonine residue (i.e., T420). (If, by *in silico* simulation, the trisaccharide is placed at T418, no assignments can be made to the peaks at 1828.01 and 1853.98.)

Panel A of Figure 4 shows a MALDI-TOF spectrum over the m/z range expected for the $(Gal)_1(GalNAc)_1$ disaccharide-modified peptides discussed above. The monoisotopic peaks at 2571.46 (calc. 2571.45) and 2585.47 (calc. 2585.47) represent the disaccharide-modified peptides (R)/LKAKLP**D**ATPTELAKLVNKR/(S) and (R)/LKAKLP**E**ATPTELAKLVNKR/(S) of the GC*1F and GC*1S allele products, respectively. To determine the site of disaccharide modification of the GC*1F and GC*1S allele products, these peptides were simultaneously subjected to MALDI-LIFT-TOF/TOF as shown in Panel B of Figure 4. A y11 ion at monoisotopic m/z 1268.18 (calc. 1268.77) indicates that, for at least a portion of the peptide population subjected to MS/MS, the $(Gal)_1(GalNAc)_1$ disaccharide is attached to T418. However, a peak at monoisotopic m/z 1633.04 (Panel B of Figure 4) (calc. 1633.90) corresponds to a different y11 ion, the mass of which indicates that, for a separate subpopulation of the ions, the disaccharide is attached to T420. Thus, while all disaccharide-containing peptides subjected to MS/MS contained only one disaccharide moiety (as dictated by the precursor ion mass), the disaccharide was attached to T418 in a portion of the peptide population, but was attached to T420 in the remaining portion of the peptide population.

Scheme 1 provides the structural summary from these studies of the genotype-specific O-glycoforms of DBP as linked to the three most common allele products.

Beyond these modified DBP glycoforms, trace quantities of Arg-C peptides representing GalNAc-only modified GC*1 (i.e., GcMAF derived from GC*1 allele products) were detected in single stage mass spectra at monisotopic m/z 2409.40 (calc. 2409.41) and monisotopic m/z 2423.40 (calc. 2423.42) (data not shown). Finally, despite trace evidence in mass spectra of intact DBP, no evidence was found in Arg-C digests of DBP for doubly (disaccharide and trisaccharide)-modified DBP. (The trace evidence shown in Figure 1, Panel B for such a diglycosylated species was only rarely observed for the >100 samples analyzed.) This suggests that not only is the modification of low relative abundance, but that it may vary in abundance from person to person. Given this information, it cannot be expected that the pooled plasma source of DBP would necessarily contain a sufficiently large quantity of this uniquely modified protein to ensure its detection. The major importance of this unique peak in the mass spectrum of the intact protein lies in its suggestion that there are two independent sites of glycosylation on GC*1F and GC*1S allele products—a notion which is confirmed by MS/MS data.

Glycation of DBP. Protein glycation is defined as the nonenzymatically mediated covalent attachment of glucose to protein amino groups via the Maillard reaction with Amadori rearrangement resulting in a permanent 1-deoxyfructosyl adduct. Direct evidence for DBP glycation in the form of the signature +162 Da mass shift is readily seen in the deconvoluted mass spectra of some samples (e.g., as in Figure 1, Panel B). Indirect evidence that DBP is a likely target for glycation was first established by Jaleel et al. in 2005.[26] Details regarding the degree of DBP glycation within the populations studied will be published elsewhere as part of a larger study involving additional proteins.

Discussion

In the data presented here, the qualitative identification of genotype-specific DBP glycosidic structures, their location, branching pattern (i.e., lack thereof), and relative quantity of DBP molecules glycosylated for each genotype are presented. Research on the structure and attachment site of carbohydrate-modified DBP, including the influence of polymorphisms on DBP glycosylation has been going on since the late 1970s. The analytical techniques employed here that provide a high resolution and systematic-yet-comprehensive analysis of DBP across the human population were not available for these foundational studies and likely account for some of the differential observations.

In 2007, Christiansen et al[19] suggested that, "except for one study[12] the location and structure of the O-linked [DBP] glycan has not been thoroughly characterized and its existence is mainly based on indirect evidence." In 1978, Svasti and Bowman[15] showed that glycosylation differences involving sialic acid (NeuNAc) were responsible for the "slow" and "fast" difference between two isoelectrically focused bands of GC*1 protein. In 1979, Svasti[27] reported that there appeared to be carbohydrate content differences between the GC*1 and GC*2 forms of DBP, concluding upon "indirect evidence"[14] that an O-glycosidically linked oligosaccharide was present in GC*1 which was absent in GC*2. Nevertheless, the reports by Svasti et al. in 1978 and 1979 suggest that the GC*2 protein carries small quantities of sialic acid, a feature not supported by the

Table 2. Relative Degree of DBP Glycosylation by Genotype for the Six Major DBP Diploid Genotypes[a]

	Homozygous Genotypes			Heterozygous Genotypes[b]		
	GC*1F	GC*1S	GC*2	GC*1F/GC*1S	GC*1S/GC*2	GC*1F/GC*2
Average Relative Percent Glycosylation ± SD	12.5 ± 2.6%	19.4 ± 3.1%	2.0 ± 1.1%	17.1 ± 4.8%	10.5 ± 2.1%	6.1 ± 2.3%
n	44	11	5	20	18	10

[a] The five genotypes observed which do not correspond to any known database entries are not included. [b] Glycosylation by the $(Gal)_1(GalNAc)_1$ disaccharide only considered if sample was homozygous for GC*2.

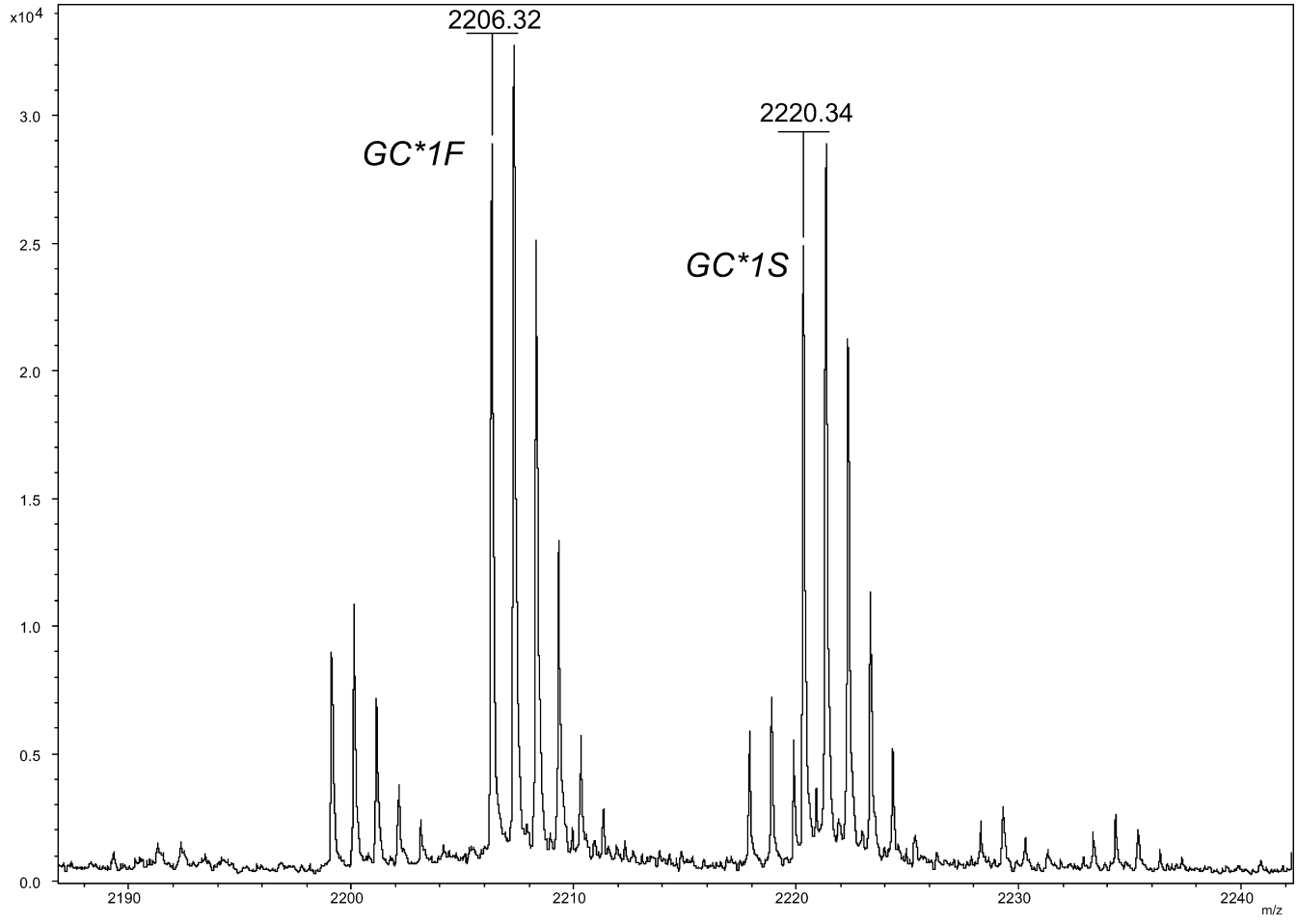


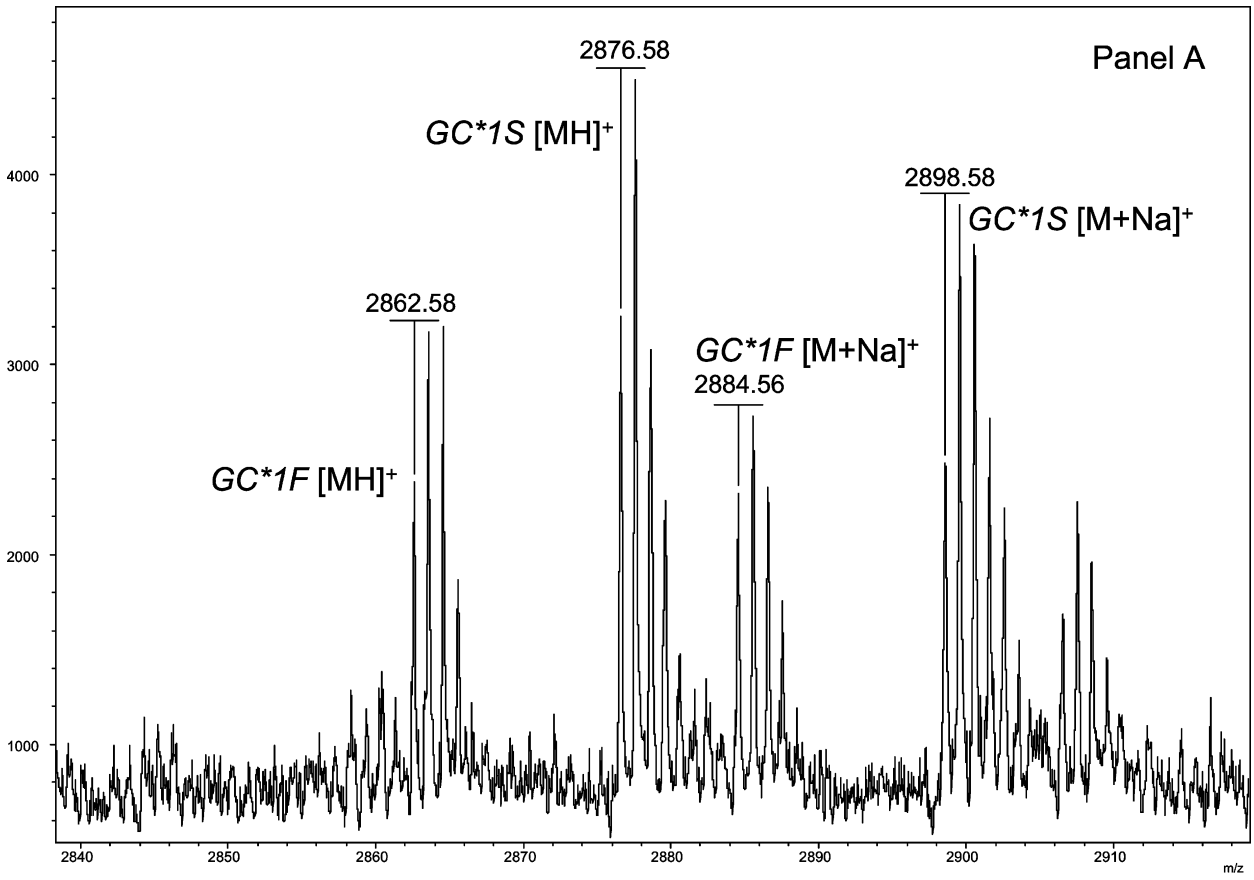
Figure 2. MALDI-TOF mass spectrum of the unmodified Arg-C peptides L410-R429 of DBP obtained from a pooled human plasma source. The monoisotopic peaks at 2206.32 (calc. 2206.33) and 2220.34 (calc. 2220.34) represent the peptides (R)/LKAKLP**D**ATPTELAKLVNKR/(S) and (R)/LKAKLP**E**ATPTELAKLVNKR/(S) of the GC*1F and GC*1S allele products, respectively. Only a trace quantity of the GC*2 allele product-derived peptide (R)/LKAKLP**D**ATP**K**ELAKLVNKR/(S) at *m/z* 2233.4 is observed.

deconvoluted spectra of intact GC*2 protein reported here. Neither is this feature of GC*2 supported by activity assays, that is, activation of DBP to GcMAF, performed by Yamamoto et al.,[11,16,17,28] which suggest a branched $(NeuNAc)_1(Gal)_1$-$(GalNAc)_1$ trisaccharide for GC*1 proteins and a $(Gal)_1(GalNAc)_1$ disaccharide for GC*2 allele products, but do not mention a separate, trisaccharide-unrelated $(Gal)_1(GalNAc)_1$ disaccharide for GC*1 proteins, which, to our knowledge, has not been described until now. In contrast to the activity based studies of Yamamoto et al., but in agreement with the data presented here, the 1983 paper by Viau et al.[12] cited by Christiansen in 2007[19] presents evidence that the GC*1S (and, by analogy, GC*1F) trisaccharide is the linear NeuNAc α(2 → 3) Gal β(1 → 3) GalNAc α(1 → 0) Ser or Thr structure.

Regarding the role of genotype, many experts in DBP biochemistry have deduced that the T420K mutation that delineates the amino acid sequence difference between the GC*1F and GC*2 allele products is responsible for the almost complete lack of carbohydrate content of the GC*2 protein,[6] but direct structural evidence for T420 as the site of DBP trisaccharide glycosylation has not been provided until now.

The data presented in Table 2 reflect relative deconvoluted mass spectral peak area integrals. Because of possible differences in ionization efficiency between DBP structural forms, these integrals may not represent the exact molar percentage of each form of DBP catalogued. But considering the overall size of DBP, the possible effect of a single difference in ionizable functional group on whole protein ionization, and the sheer number of charges on each protein molecule (+20−40 according to the raw nondeconvoluted mass spectra), this effect of ionization efficiency on apparent molar abundance is likely to be minimal.

A few reports have suggested DBP features such as a (branched) $(Man)_1(Gal)_1(GalNAc)_1$ trisaccharide for GC*1S,[18,24]



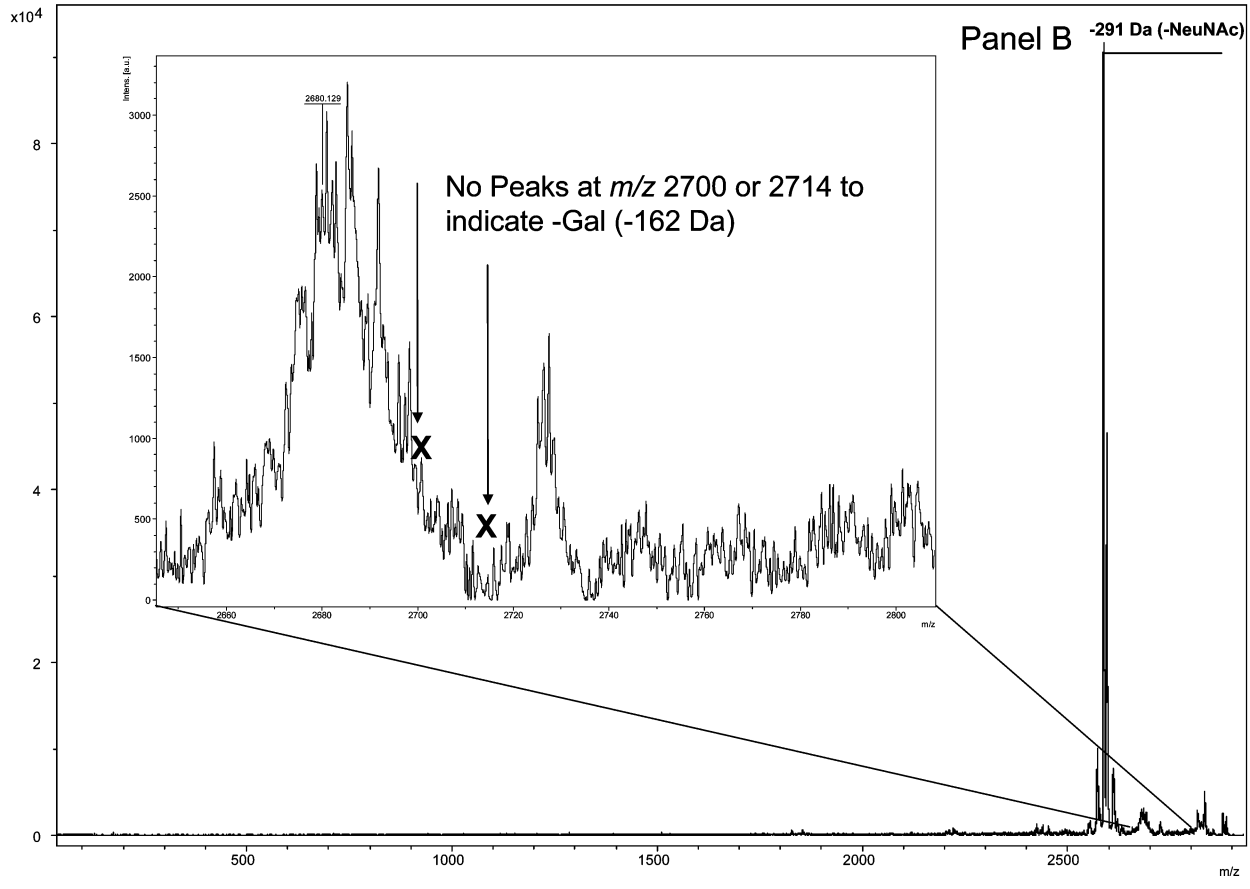


Figure 3. (Continued)

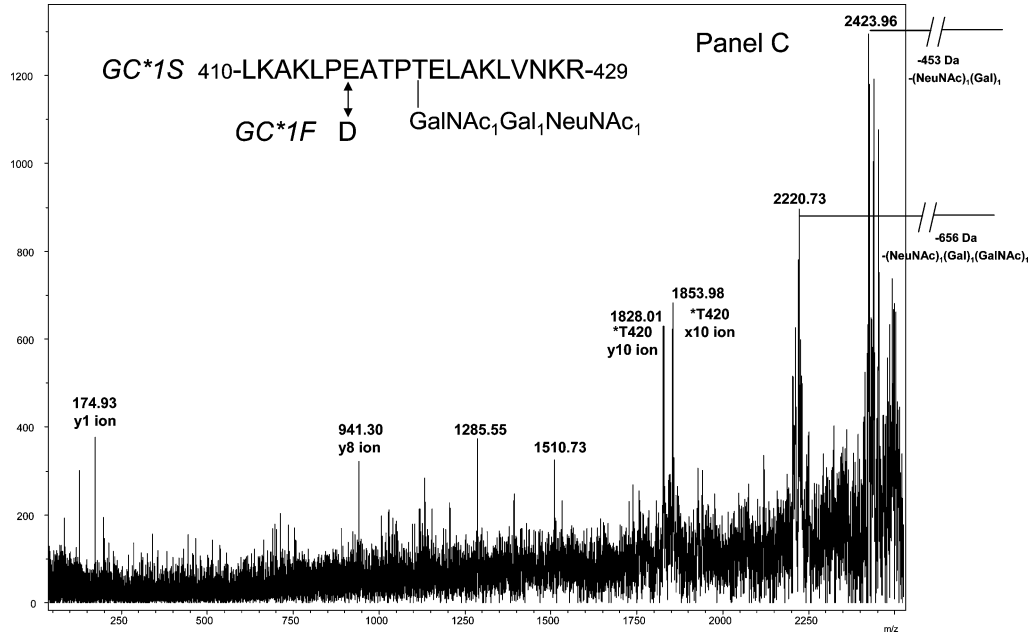


Figure 3. (A) MALDI-TOF mass spectrum of $(NeuNAc)_1(Gal)_1$ $(GalNAc)_1$ trisaccharide-modified L410-R429 GC*1F and GC*1S peptides of DBP. The peaks at *m*/*z* 2862.58 (calc. 2862.56) and 2884.56 represent the MH^+ and sodiated forms, respectively, of a $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide-glycosylated peptide with the sequence (R)/LKAKLP**D**ATPTELAKLVNKR/(S) derived from the GC*1F allele product. Analogous GS*1S products (corresponding to a D416E mutation) are shifted up in mass by 14 Da. (B) MALDI-LIFT-TOF/TOF spectrum with precursor ion selection for the cluster of peaks from *m*/*z* 2862 to 2898 that is shown in Panel A. The base peak in the TOF/TOF spectrum is due to the facile loss of sialic acid from the trisaccharide. The inset shows that there is no evidence for the loss of galactose alone from the trisaccharide. (C) Peaks in the TOF/TOF spectrum with masses less than that of the base peak: *m*/*z* 2423.96 (calc. 2423.42) and 2220.73 (calc. 2220.34) indicate loss of a $NeuNAc_1Gal_1$ disaccharide and loss of the entire $NeuNAc_1Gal_1GalNAc_1$ trisaccharide, respectively, to generate the unmodified L410−R429 peptide of GC*1S. (Analogous peaks for the GC*1F peptide are present but not annotated.) Peaks at *m*/*z* 1828.01 (calc. 1827.94) and 1853.98 (calc. 1853.92) correspond to y10 and x10 ions, respectively, containing the trisaccharide located at the second threonine residue (i.e., T420) (Asterisks indicate threonine residues that are supported by the data as sites of trisaccharide glycosylation.)

Glu at residue number 152 and Arg at residue number 311,[23,24] and N-glycosylation of DBP.[25] These are features which would have been apparent in at least some of the 113 mass spectra of intact DBP from separate individuals gleaned in this study. The evidence presented here does not support these structural features of DBP.

As alluded to above, the most unanticipated finding presented here lies in the evidence for a separate, trisaccharide-unrelated $(Gal)_1(GalNAc)_1$ disaccharide for GC*1 proteins. Given, however, that only one mutation affecting an amino acid residue capable of serving as a carbohydrate attachment point (K420T) differs between GC*2 and the GC*1 proteins, it is reasonable to deduce that the disaccharide attachment point for GC*2 would remain intact in the GC*1 proteins. Whether the T420 disaccharide may be metabolically related to the T420 trisaccharide and/or simply exists as an analytical artifact (e.g., from prompt fragmentation/loss of the sialic acid residue prior to ion acceleration in the MALDI-TOF instrument) of the T420 trisaccharide remains unresolved.

The evidence for a linear trisaccharide structure as shown here and as described by Viau et al.,[12] makes it difficult to explain the GcMAF activity studies of Yamamoto et al.[16] where immobilized β-galactosidase and immobilized sialidase can be used in any order upon GC*1 proteins to produce GcMAF (which has been reported to be $(GalNAc)_1$ monosaccharide-modified DBP.[16] In combination, however, with the evidence presented here for a separate, trisaccharide-unrelated $(Gal)_1(GalNAc)_1$ disaccharide for GC*1 proteins, it becomes feasible to hypothesize that the terminal sialic acid residue of the GC*1 trisaccharide acts in an inhibitory way toward GcMAF, that is, $(GalNAc)_1$ monosaccharide-modified DBP. Thus, an initial action of β-galactosidase upon GC*1 protein would produce an inhibited population of GcMAF molecules, which would become uninhibited by the action of sialidase. Likewise, an initial action of sialidase upon GC*1 proteins would produce a population of DBP molecules which would be primed and ready for formation of active GcMAF upon exposure to β-galactosidase.

An experiment in which unmodified GC*1 protein is added to pregenerated GcMAF and macrophage activation measured before and after addition of the unmodified GC*1 protein would begin to address this hypothesis.

Conclusions

GC*1 DBP allele products are modified with a linear $(NeuNAc)_1(Gal)_1(GalNAc)_1$ trisaccharide (at 10−25 mol %) at T420 and, to a much lesser extent (<5 mol %), with a $(Gal)_1(GalNAc)_1$ dissaccharide at T418. GC*2 protein contains the disaccharide but remains completely free of the trisaccharide, even in heterozygous individuals possessing a second gene product that is modified with the trisaccharide. Direct structural evidence is provided to show that the lack of trisaccharidic glycosylation of GC*2 is due to substitution of the T420 trisaccharide attachment site to K420. No evidence was found to suggest that any of the three major DBP allele products are N-glycosylated, contain mannose, or possess Glu at residue number 152 or Arg at residue number 311.

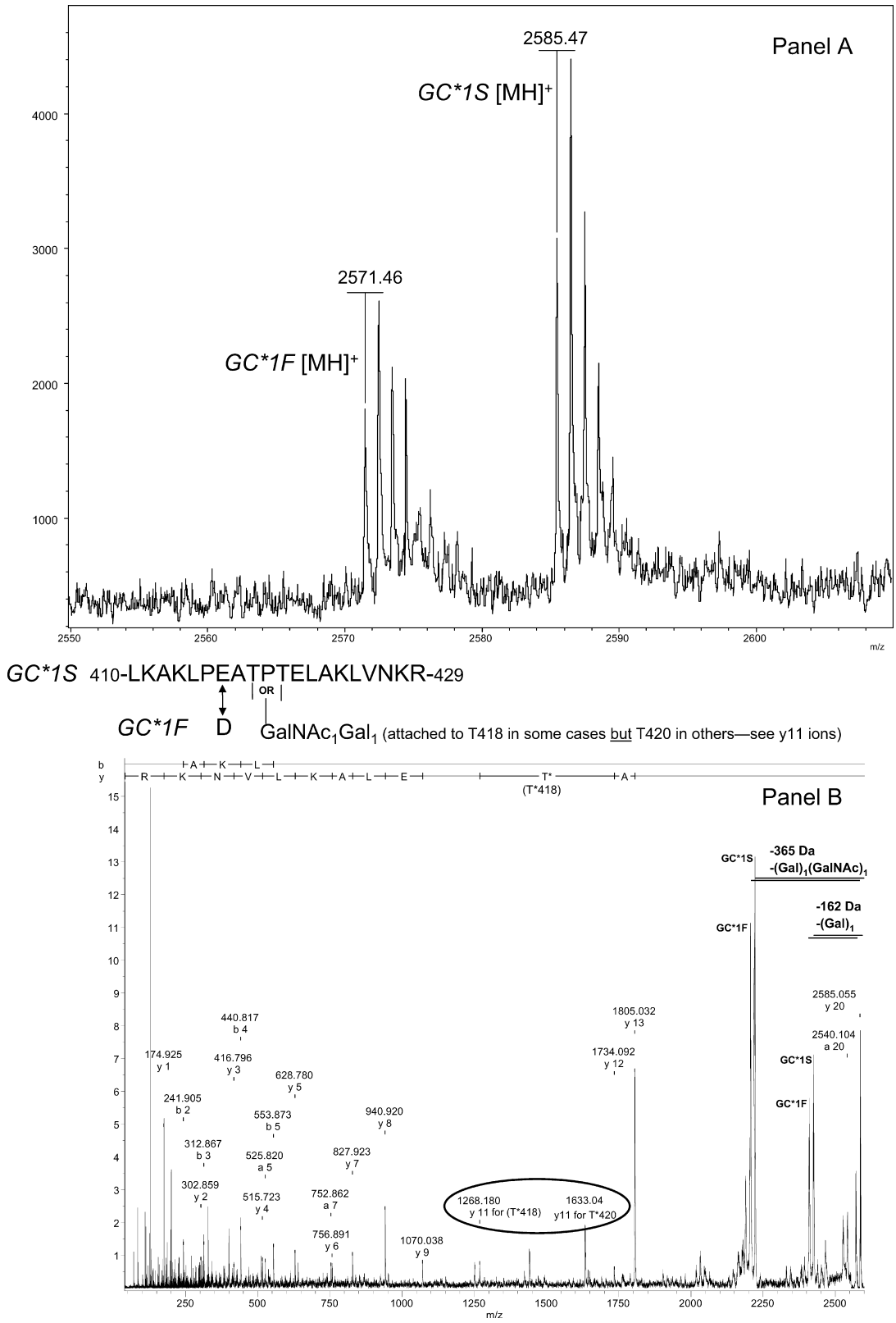


Figure 4. (A) MALDI-TOF mass spectrum of $(Gal)_1(GalNAc)_1$ disaccharide-modified L410−R429 GC*1F and GC*1S peptides of DBP. The peaks at 2571.46 (calc. 2571.45) and 2585.47 (calc. 2585.47) represent the disaccharide-modified peptides (R)/LKAKLP**D**ATPTELAKLVNKR/(S) and (R)/LKAKLP**E**ATPTELAKLVNKR/(S) of the GC*1F and GC*1S allele products, respectively. (B) MALDI-LIFT-TOF/TOF spectrum of the peptides shown in Panel A. A y11 ion at *m*/*z* 1268.18 (calc. 1268.77) indicates that for at least a portion of the peptide population subjected to MS/MS, the $(Gal)_1(GalNAc)_1$ disaccharide is attached to T418. However, a peak at *m*/*z* 1633.04 (calc. 1633.90) corresponds to a different y11 ion, the mass of which indicates that, for a separate subpopulation of the ions, the disaccharide is attached to T420. Asterisks indicate threonine residues that are supported by the data as sites of disaccharide glycosylation.

Abbreviations: DBP, Vitamin D Binding Protein; MSIA ESI-TOF-MS, mass spectrometric immunoassay coupled with electrospray ionization time-of-flight mass spectrometry; MALDI-LIFT-TOF/TOF, matrix-assisted laser desorption ionization tandem time-of-flight/time-of-flight mass spectrometry; NeuNAc, *N*-acetylneuraminic acid (aka sialic acid); Gal, galactose; GalNAc, *N*-acetylgalactosamine; GC*, DBP allelic designation; GcMAF, DBP-derived macrophage activating factor; CDI, 1,1′-carbonyldiimidazole; HBS, Hepes-buffered saline; Arg-C, endoproteinase Arg-C from *C. histolyticum*; DTT, dithiothreitol; THAP, 2,4,6- trihydroxyacetophenone; CHCA, α-cyano-4-hydroxycinnamic acid.

Acknowledgment. The authors are grateful to Intrinsic Bioprobes, Inc. team members for their helpful discussions. Funding was provided by Arizona State University's Technology and Research Initiative Fund and by the Agilent Technologies Foundation in the form of Research Project Gift 08US-422UR.

References

(1) Pennisi, E.; Kaiser, J. Breakthrough of the year: human genetic variation. *Science* **2007**, *318*, 1842–3.
(2) Nedelkov, D.; Kiernan, U. A.; Niederkofler, E. E.; Tubbs, K. A.; Nelson, R. W. Investigating diversity in human plasma proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (31), 10852–7.
(3) Daiger, S. P.; Schanfield, M. S.; Cavalli-Sforza, L. L. Group-specific component (Gc) proteins bind vitamin D and 25-hydroxyvitamin D. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72* (6), 2076–80.
(4) Hirschfeld, J.; Jonsson, B.; Rasmuson, M. Inheritance of a new group-specific system demonstrated in normal human sera by means of an immuno-electrophoretic technique. *Nature* **1960**, *185*, 931–2.
(5) Cleve, H.; Constans, J. The mutants of the vitamin-D-binding protein: more than 120 variants of the GC/DBP system. *Vox Sang.* **1988**, *54* (4), 215–25.
(6) Speeckaert, M.; Huang, G.; Delanghe, J. R.; Taes, Y. E. Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. *Clin. Chim. Acta* **2006**, *372* (1−2), 33–42.
(7) Cooke, N. E.; Haddad, J. G. Vitamin D binding protein (Gc-globulin). *Endocr. Rev.* **1989**, *10* (3), 294–307.
(8) Ena, J. M.; Esteban, C.; Perez, M. D.; Uriel, J.; Calvo, M. Fatty acids bound to vitamin D-binding protein (DBP) from human and bovine sera. *Biochem. Int.* **1989**, *19* (1), 1–7.
(9) Williams, M. H.; Van Alstyne, E. L.; Galbraith, R. M. Evidence of a novel association of unsaturated fatty acids with Gc (vitamin D-binding protein). *Biochem. Biophys. Res. Commun.* **1988**, *153* (3), 1019–24.
(10) Homma, S.; Yamamoto, M.; Yamamoto, N. Vitamin D-binding protein (group-specific component) is the sole serum protein required for macrophage activation after treatment of peritoneal cells with lysophosphatidylcholine. *Immunol. Cell Biol.* **1993**, *71* (Pt. 4), 249–57.
(11) Yamamoto, N.; Kumashiro, R. Conversion of vitamin D3 binding protein (group- specific component) to a macrophage activating factor by the stepwise action of beta- galactosidase of B cells and sialidase of T cells. *J. Immunol.* **1993**, *151* (5), 2794–802.
(12) Viau, M.; Constans, J.; Debray, H.; Montreuil, J. Isolation and characterization of the O-glycan chain of the human vitamin-D binding protein. *Biochem. Biophys. Res. Commun.* **1983**, *117* (1), 324–31.
(13) Constans, J.; Arlet, P.; Viau, M.; Bouissou, C. Unusual sialilation of the serum DBP associated with the Gc 1 allele in alcoholic cirrhosis of the liver. *Clin. Chim. Acta* **1983**, *130* (2), 219–30.
(14) Coppenhaver, D. H.; Sollenne, N. P.; Bowman, B. H. Post-translational heterogeneity of the human vitamin D-binding protein (group-specific component). *Arch. Biochem. Biophys.* **1983**, *226* (1), 218–23.
(15) Svasti, J.; Bowman, B. H. Human group-specific component. Changes in electrophoretic mobility resulting from vitamin D binding and from neuraminidase digestion. *J. Biol. Chem.* **1978**, *253* (12), 4188–94.
(16) Yamamoto, N. Structural definition of a potent macrophage activating factor derived from vitamin D3-binding protein with adjuvant activity for antibody production. *Mol. Immunol.* **1996**, *33* (15), 1157–64.
(17) Yamamoto, N.; Homma, S. Vitamin D3 binding protein (group-specific component) is a precursor for the macrophage-activating signal factor from lysophosphatidylcholine-treated lymphocytes. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88* (19), 8539–43.
(18) Ohkura, K.; Nagasawa, H.; Uto, Y.; Okamura, N.; Murakami, A.; Hori, H. The role of Gc protein oligosaccharide structure as a risk factor for COPD. *Anticancer Res.* **2006**, *26* (6A), 4073–8.
(19) Christiansen, M.; Jorgensen, C. S.; Laursen, I.; Hirschberg, D.; Hojrup, P.; Houen, G. Protein chemical characterization of Gc globulin (vitamin D-binding protein) isoforms; Gc-1f, Gc-1s and Gc-2. *Biochim. Biophys. Acta* **2007**, *1774* (4), 481–92.
(20) Hermanson, G. T.; Krishna Mallia, A.; Smith, P. K., *Immobilized Affinity Ligand Techniques*; Academic Press: San Diego, CA, 1992.
(21) Niederkofler, E. E.; Tubbs, K. A.; Kiernan, U. A.; Nedelkov, D.; Nelson, R. W. Novel mass spectrometric immunoassays for the rapid structural characterization of plasma apolipoproteins. *J. Lipid Res.* **2003**, *44* (3), 630–9.
(22) Pouria, S.; Corran, P. H.; Smith, A. C.; Smith, H. W.; Hendry, B. M.; Challacombe, S. J.; Tarelli, E. Glycoform composition profiling of O-glycopeptides derived from human serum IgA1 by matrix-assisted laser desorption ionization-time of flight-mass spectrometry. *Anal. Biochem.* **2004**, *330* (2), 257–63.
(23) Cooke, N. E.; David, E. V. Serum vitamin D-binding protein is a third member of the albumin and alpha fetoprotein gene family. *J. Clin. Invest.* **1985**, *76* (6), 2420–4.
(24) Nagasawa, H.; Uto, Y.; Sasaki, H.; Okamura, N.; Murakami, A.; Kubo, S.; Kirk, K. L.; Hori, H. Gc protein (vitamin D-binding protein): Gc genotyping and GcMAF precursor activity. *Anticancer Res.* **2005**, *25* (6A), 3689–95.
(25) Wang, Y.; Wu, S. L.; Hancock, W. S. Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry. *Glycobiology* **2006**, *16* (6), 514–23.
(26) Jaleel, A.; Halvatsiotis, P.; Williamson, B.; Juhasz, P.; Martin, S.; Nair, K. S. Identification of Amadori-modified plasma proteins in type 2 diabetes and the effect of short- term intensive insulin treatment. *Diabetes Care* **2005**, *28* (3), 645–52.
(27) Svasti, J.; Kurosky, A.; Bennett, A.; Bowman, B. H. Molecular basis for the three major forms of human serum vitamin D binding protein (group-specific component). *Biochemistry* **1979**, *18* (8), 1611–7.
(28) Yamamoto, N.; Homma, S.; Millman, I. Identification of the serum factor required for in vitro activation of macrophages. Role of vitamin D3-binding protein (group specific component, Gc) in lysophospholipid activation of mouse peritoneal macrophages. *J. Immunol.* **1991**, *147* (1), 273–80.

PR8002936