Journal of proteome research

pubs.acs.org/jpr

Article

Glycosylation Profiling of the Neoplastic Biomarker Alpha Fetoprotein through Intact Mass Protein Analysis

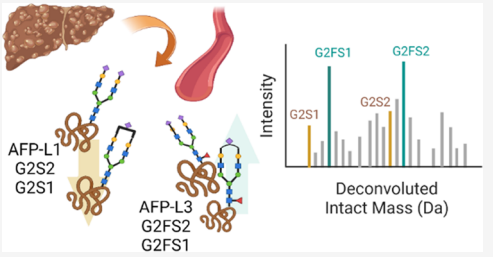
Carmen Dunbar,[‖] Mark M. Kushnir,[‖] and Yifei K. Yang*

Cite This: *J. Proteome Res.* 2023, 22, 226−234

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Elevated serum alpha-fetoprotein (AFP) can be observed in liver cirrhosis and hepatocellular carcinoma (HCC). The glycosylation patterns of AFP have been shown to differentiate these conditions, with AFP glycoforms with core fucosylation (AFP-L3) serving as a malignancy risk predictor for HCC. We have developed a method to detect endogenously present AFP proteoforms and to quantify the relative abundance of AFP-L3 glycoforms (AFP-L3%) in serum samples. This method consists of immune enrichment of endogenous AFP, followed by liquid chromatography coupled with high-resolution mass spectrometry (LC−HRMS) intact protein analysis of AFP. Data are available via ProteomeXchange with identifier PXD038606. Based on the AFP profiles in authentic patient serum samples, we have identified that the frequently observed AFP glycoforms without core fucosylation (AFP-L1) are G2S2 and G2S1, and common AFP-L3 glycoforms are G2FS1 and G2FS2. The intensities of glycoforms in the deconvoluted spectrum are used to quantify AFP-L3% in each sample. The method evaluation included reproducibility, specificity, dilution integrity, and comparison of AFP-L3% with a lectin-binding gel shift electrophoresis (GSE) assay. The AFP-L1 and AFP-L3 proteoforms were reproducibly identified in multiple patient serum samples, resulting in reproducible AFP-L3% quantification. There was considerable agreement between the developed LC−HRMS and commercial GSE methods when quantifying AFP-L3% (Pearson r = 0.63) with a proportional bias.



KEYWORDS: *core fucose, fucosylation, intact protein analysis, high-resolution mass spectrometry, AFP-L3 proteoforms*

INTRODUCTION

Glycosylation is an important post-translational modification (PTM) mechanism of proteins, which modulates the protein function and is implicated in protein folding, trafficking, and cell-to-cell communications.[1] Dynamic changes in glycosylation states have been associated with different disease processes, such as cancer,[2] viral infection responses,[3] and Alzheimer's disease;[4] the extent of the glycosylation could be associated with the disease progression[5] and prognosis.[6] These observed associations indicate that the specific glycosylation states of proteins can be utilized as biomarkers for clinical diagnosis[7] and disease monitoring.[8] Currently, measurement of clinically adopted cancer protein biomarkers is mostly performed using affinity-based techniques.[9] These methods can at times suffer from poor specificity and accuracy and cannot distinguish between closely related proteoforms corresponding to different glycosylation modifications; this may limit their diagnostic accuracy and value.

Alpha fetoprotein (AFP) is an established circulating cancer biomarker implicated in multiple neoplastic malignancies, including hepatocellular carcinoma (HCC) and nonseminomatous testicular tumors.[10,11] AFP is a glycoprotein predominantly expressed in the embryonic yolk sac and fetal liver, and its elevation as a circulating biomarker can be observed in liver cirrhosis, during pregnancy, and in neonates.[12] Mature AFP has a single-consensus N-linked glycosylation site at Asparagine 251 (N251); a high degree of structural variation in glycan structures has been identified in serum samples.[11,13−15] There was limited evidence suggesting a possible O-glycosylation site within AFP based on the cell culture study;[16] however, AFP O-glycosylation has not been confirmed in authentic serum samples. In addition to glycosylation, six putative phosphorylation sites in AFP have been reported.[17] The existence of different isoforms of post-translationally modified AFP was first described by Taketa and Hirari, who found different forms of glycosylated AFP in patients with benign and malignant diseases.[18] AFP glycoforms sharing a core fucosylation structure (AFP-L3) are used clinically as a biomarker to predict the risk of developing HCC[11] and to monitor its recurrence in patients with impaired liver function. Commonly used methodology to measure the relative abundance of AFP-L3 glycoforms is lectin-binding gel shift electrophoresis (GSE). Lens culinaris agglutinin (LCA) is a

Received: October 13, 2022
Published: December 21, 2022

type of lectin, which can bind specifically to the core fucosylation (α-linked fucose residue attached to N-acetylglucosamine) moiety. Electrophoresis-based methods utilizing LCA binding can identify AFP-L3; however, they do not provide insights into structural variations of AFP glycoforms beyond the core fucosylation site.

Several liquid chromatography with tandem mass spectrometry (LC−MS/MS) methods have been developed to profile and identify different glycoforms of AFP, which are based on quantification of AFP-specific glycopeptides through digestion of proteins enriched from plasma or serum samples.[14,19,20] These methods utilize glycopeptides representing the AFP glycoforms without core fucosylation (AFP-L1) and with core fucosylation (AFP-L3) to calculate the relative abundance of AFP-L3 glycoforms (AFP-L3%). Other studies also utilized enzymes and chemical treatment to release glycans from AFP and identified the glycan structure by electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) coupled with MS.[13,21,22] During MS analysis, glycans and glycopeptides require different collision energies for structure and sequence identification based on fragmentation. Certain glycan structures, such as sialic acids, require further derivatization or permethylation prior to MS analysis due to their heat labile nature.

Different from electrophoresis and glycan- or glycopeptide-based MS approaches, intact protein analysis using liquid chromatography coupled with high-resolution mass spectrometry (LC−HRMS) can directly profile different AFP proteoforms, including glycosylation. Potential advantages of intact protein methods include faster and less complex sample preparation (relative to released glycan- or glycopeptide-based MS), and more straightforward data analysis. Compared to the electrophoresis-based methods, masses of intact AFP proteoforms can be used to differentiate and identify post-translational modifications in their native states beyond the glycan core fucosylation site.[14,23]

Here, we describe an analytical method for AFP glycoform analysis that utilizes AFP-specific immune enrichment, followed by LC−HRMS analysis to differentiate and quantitate the relative abundance of AFP glycoforms. Based on the identified AFP proteoforms and the presence of core fucose, we further categorize different glycoforms to AFP-L1 and AFP-L3. We evaluated the performance of the analytical method, applied the method to analyze authentic patient samples, and compared AFP-L3% results quantified from the developed LC−HRMS approach and a commercial GSE method.

EXPERIMENTAL SECTION

Standards and Reagents

AFP purified from human umbilical cord blood was purchased from Lee Biosolutions (Maryland Heights, MO); monoclonal rabbit anti-AFP antibody and magnetic beads coated with anti-rabbit IgG antibodies were purchased from Thermo Fisher Scientific (Waltham, MA) and Eurofins Abraxis (Warminster, PA) respectively. Formic acid, glycine, and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) were purchased from MilliporeSigma (St. Louis, MO); acetonitrile was purchased from Fisher Scientific (Waltham, MA). All other reagents were of the highest purity grade commercially available.

AFP-Specific Immune Enrichment from Serum Samples

Serum samples (500 μL) were aliquoted in 1.5 mL LoBind tubes (Eppendorf, Hamburg, Germany); 500 μL of HEPES buffer (pH 7.4) and 5 μL of anti-AFP antibody (rabbit monoclonal antibody (3)) were added to the samples and the samples were incubated at 10 °C for 1 h (with mixing, at 1100 rpm) on a thermomixer (Eppendorf). After incubation, 5 μL of magnetic beads coated with anti-rabbit IgG antibody were added to the samples, and the samples were incubated at 10 °C for 1 h (with mixing at 1100 rpm). After incubation, the tubes were placed in a magnetic rack for 20 min to separate the magnetic beads from the supernatants, and the supernatants were removed as waste. The magnetic beads were washed 5 times with 300 μL of phosphate-buffered saline (PBS, pH 7.4, 20 mM). The washes included addition of 300 μL of PBS to the beads, vortexing the beads in LoBind tubes for 2 min at 1100 rpm, followed by separation of the beads in a magnetic rack (tubes were held in the rack for 5 min), and removal of the wash solution. After the last wash cycle, AFP was eluted from the magnetic beads with 75 μL of glycine buffer (pH 2); the eluents were transferred into high-performance liquid chromatography (HPLC) vials, and the samples were injected (20 μL) on LC−HRMS for intact protein analysis.

Instrumental Analysis

Instrumental analysis was performed using a quadrupole time-of-flight (Q-TOF) 6550 mass spectrometer equipped with a 1290 infinity I HPLC stack (Agilent Technologies, Santa Clara, CA). HPLC separation was performed using a 300 Å, 3 μm, 2.1 mm × 50 mm PLRP column (Agilent Technologies). Mobile phase A was 0.1% formic acid in water, and mobile phase B was 0.1% formic acid in acetonitrile. The flow rate was 300 μL/min, and the mobile phase gradient was 80 to 40% A over 12 min, followed by column wash, conditioning, and re-equilibration. The column temperature was kept at 80 °C.

Data acquisition on the Q-TOF instrument was performed using high-resolution profile mode (4 GHz); the mass range was 1000−3200 m/z. Source conditions and acquisition parameters are listed in Supporting Information Table S1. The mass spectrometry intact proteim data have been deposited to the ProteomeXchange Consortium via the PRIDE[24] partner repository with the dataset identifier PXD038606.

Intact Mass Data Analysis

Multiple charge envelopes corresponding to chromatographic peaks at the retention time (RT) of AFP and human albumin were selected for deconvolution using the maximum entropy algorithm in BioConfirm software 10.1 (Agilent Technologies). For AFP, the deconvolution was performed with a target mass range between 68,000 and 80,000 Da using a 7 Da mass step size. Deconvoluted masses at the expected AFP retention time were then matched to AFP full-length sequence without its signal peptide (P02771|19-609) and with 16 disulfide bonds allowing variable oxidation and deamidation. The variable modifications included known AFP glycan modifications and a maximum of six sites of phosphorylation.[17] A small fraction of albumin was coenriched from patient serum samples along with AFP (because of the cross-reactivity of the antibody); the presence of the albumin peak and agreement of the deconvoluted mass with the expected mass of albumin were used as an internal control to assess performance of the assay in every sample. Deconvolution of the peak at the retention

https://doi.org/10.1021/acs.jproteome.2c00656
J. Proteome Res. 2023, 22, 226−234

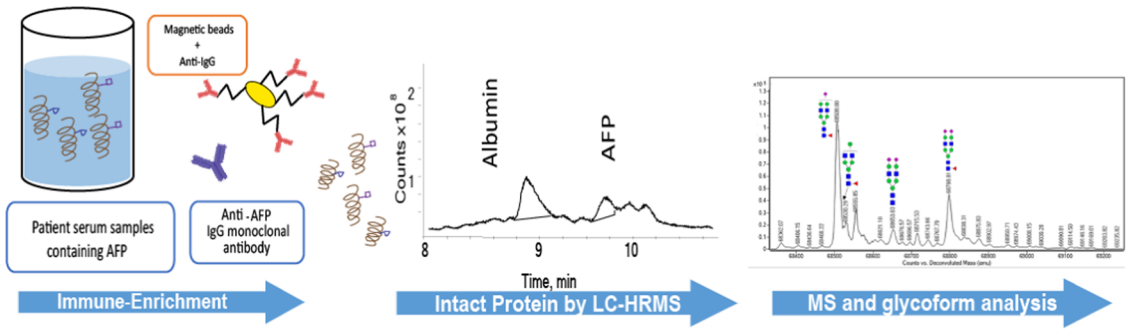


Figure 1. Workflow of AFP glycoform profiling using intact mass LC−HRMS. AFP is affinity-enriched from serum samples, followed by instrumental analysis using analytical flow rate LC separation and intact protein analysis by Q-TOF HRMS.

time of albumin was performed using a mass range of 66,000−67,500 Da with a 7 Da step size.

Based on consensus results reported in the literature,[13−15,19,21] the allowed AFP glycoforms were G0, G0F, G1, G1F, G1S1, G1FS1, G2, G2F, G2S2, G2S1, G2FS1, and G2FS2 (Figure S1).[13,14,21] These glycoforms also represent an unbiased and equal number of potential proteoforms representing AFP-L1 and AFP-L3 isoforms. Match tolerance for the difference between the observed and the expected intact masses of the deconvoluted peaks was ±15 Da. For glycoforms G2FS1 and G2FS2, the mass matching tolerance was narrowed to ±5 Da due to the presence of interferences from unknown serum proteins observed in the deconvoluted mass spectra of AFP-negative serum samples (see the Results and the Discussion sections). Multiple deconvoluted masses within a ±1 Da window were summed up using their intensities as the abundance of the matched glycoform. The deconvoluted mass with the smallest delta relative to the theoretical target mass of each AFP proteoform was ruled in as the putative match. Based on the results observed in the analysis of multiple patient samples and their abundances, the reproducibly observed glycoforms (AFP-L1: G1, G1S1, G2, G2S1, G2S2 and AFP-L3: G1F, G1FS1, G2F, G2FS1, G2FS2) were used to quantify the relative AFP-L3%. The formula below was used for calculating the relative AFP-L3%.

$$\text{AFP-L3\%} = \frac{\sum I_{\text{identified L3 isoforms}}}{\sum I_{\text{identified L3 isoforms}} + \sum I_{\text{identified L1 isoforms}}}$$

where "I" is the intensity height of the isoforms.

Immunoassay and GSE to Quantify Total AFP and AFP-L3%

In the study, we analyzed residual deidentified aliquots of patient serum samples (n = 40) submitted for total AFP and AFP-L3% testing to ARUP Laboratories (Salt Lake City, UT). The AFP concentration was measured with an Access AFP quantitative chemiluminescent immunoassay (Beckman Coulter, IN). The AFP-L3% was measured by a GSE method using the μTASWako AFP-L3 Immunological Test System (Wako Pure Chemical Industries, Osaka, Japan). In this method, LCA first recognizes and binds to the core fucosylation structure shared by AFP-L3 glycoforms; LCA and AFP-L3 complex is further separated from AFP-L1 by gel electrophoresis,[25] and lastly an AFP-specific antibody with fluorescence label identifies and quantifies the relative abundance of AFP-L3 isoforms. The electrophoretic separation and fluorescence detection are both performed on a microfluidic chip.

All studies with human samples were approved by the Institutional Review Board of the University of Utah (Salt Lake City, UT). Patient serum specimens were measured by the AFP immunoassay and GSE assay within one week of sample collection. The residual samples were stored at −20 °C prior to analysis by LC−HRMS within 6 months of the sample collection.

Method Evaluation

The method performance evaluation consisted of the following experiments: analytical reproducibility, analytical specificity, dilution integrity, and comparison with the lectin-binding GSE assay using authentic patient serum samples. The analytical reproducibility was evaluated by the extraction and analysis of three AFP serum sample pools in four replicates; the samples contained different total AFP concentrations and different AFP-L3%. The specificity of the method was evaluated by analyzing five individual serum samples containing total AFP ≤0.01 μg/mL. The dilution integrity and sensitivity of the method for detecting relative AFP-L3% were assessed by analysis of a dilution series of serum patient samples containing highly elevated AFP, diluted with AFP-negative serum samples, to create four samples containing progressively lower AFP concentrations, while maintaining constant AFP-L3%.

Statistical Analysis

Statistical analyses were performed using Excel (Microsoft) with additional modules from Analyse-IT and JMP12 software (SAS Institute).

RESULTS

Immune-Enrichment Workflow and Intact Protein Analysis of AFP

We have developed a method for the detection of AFP glycoforms and quantification of the relative AFP-L3% in serum samples. The method is based on enrichment of endogenous AFP, followed by LC−HRMS analysis of the intact AFP (Figure 1). The assay consisted of the following steps: (i) enrichment of endogenous AFP using rabbit anti-AFP monoclonal antibody and magnetic beads with conjugated anti-rabbit IgG, (ii) removal of nonspecifically bound proteins by washing the magnetic beads, (iii) elution of AFP from the beads, (iv) LC−HRMS analysis of the intact AFP, followed by (v) deconvolution of AFP and albumin from multiple charge envelopes, and (vi) AFP proteoform detection and AFP-L3% quantification.

Following chromatographic separation, intact AFP was eluted at a retention time (RT) of 9.9 ± 0.2 min. From the deconvoluted intact protein mass spectrum, we can see a

https://doi.org/10.1021/acs.jproteome.2c00656
J. Proteome Res. 2023, 22, 226−234

collection of deconvoluted masses matched to full-length AFP with different post-translational modifications (Figure 2). The

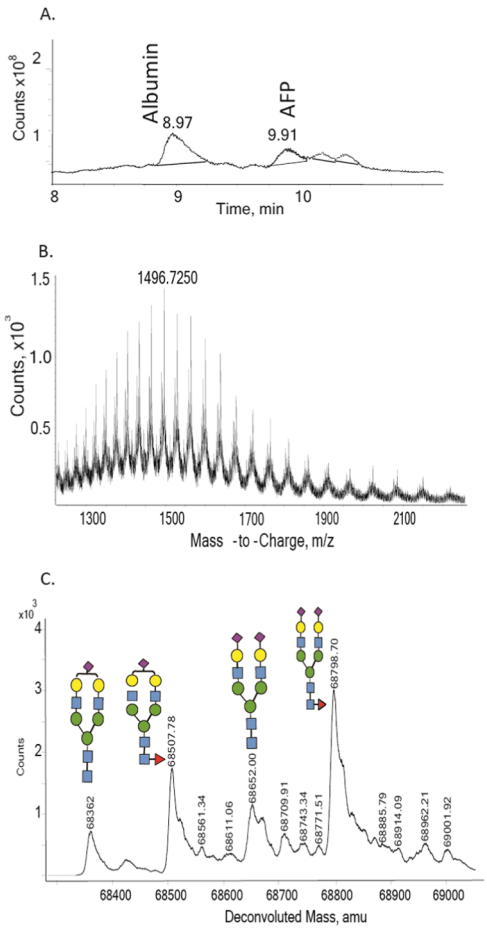


Figure 2. (A) Total ion chromatogram of a patient serum sample containing 16.7 μg/mL AFP. (B) Multiple charge states detected within the AFP peak eluted at RT 9.9 min. (C) Deconvoluted intact mass spectrum and the assigned glycoforms of AFP at N251 based on the intact mass match.

chromatographic peak of human albumin (RT 9.0 ± 0.2 min) was eluted ~1 min earlier than the AFP peak (Figures 2 and S2), and the albumin peak was present in all analyzed patient serum samples. Multiple charge envelopes at the respective retention time of AFP and albumin were selected and deconvoluted (Figures 2 and S2); the observed intact masses were consistent with the expected masses of human albumin and AFP containing previously reported PTMs or other nonspecific modifications.[8−10,17,18]

We assessed the efficiency of AFP immune enrichment with authentic serum samples with both high and low AFP-L3% and observed that the enrichment recovery (absolute recovery of 30%) was independent of AFP isoform compositions. Based on the protein electrophoresis evaluation of postenrichment serum samples, AFP was the prominent protein species (Figure S3). The specificity of the intact AFP enrichment was further confirmed by bottom-up analysis of AFP in a workflow utilizing the same intact AFP enrichment method, followed by protein reduction, alkylation, tryptic digestion, and LC-MS/MS analysis. Nine patient serum samples (mean/median AFP concentrations of 60.8/18.1 μg/mL) and a commercial AFP standard (Lee Biosolutions) were prepared and analyzed using the data-dependent acquisition method on Q-TOF instrument; in these samples, we detected 32 unique peptides from AFP, representing 69.9% AFP sequence coverage (Supporting Information Table S2).

AFP Glycoform Profiles in Authentic Patient Serum Samples

The intact AFP peaks observed in the deconvoluted mass spectra of authentic patient serum samples were matched to theoretical intact masses of the full-length AFP sequence with various glycosylation isoforms (Figures 2 and S1). In the initial proteoform profiling and assignment process, we allowed all common glycan modifications available in the BioConfirm™ library. We identified several glycoforms corresponding to AFP-L1 and AFP-L3, which were frequently observed in authentic serum samples (Figure 3). The list of these most

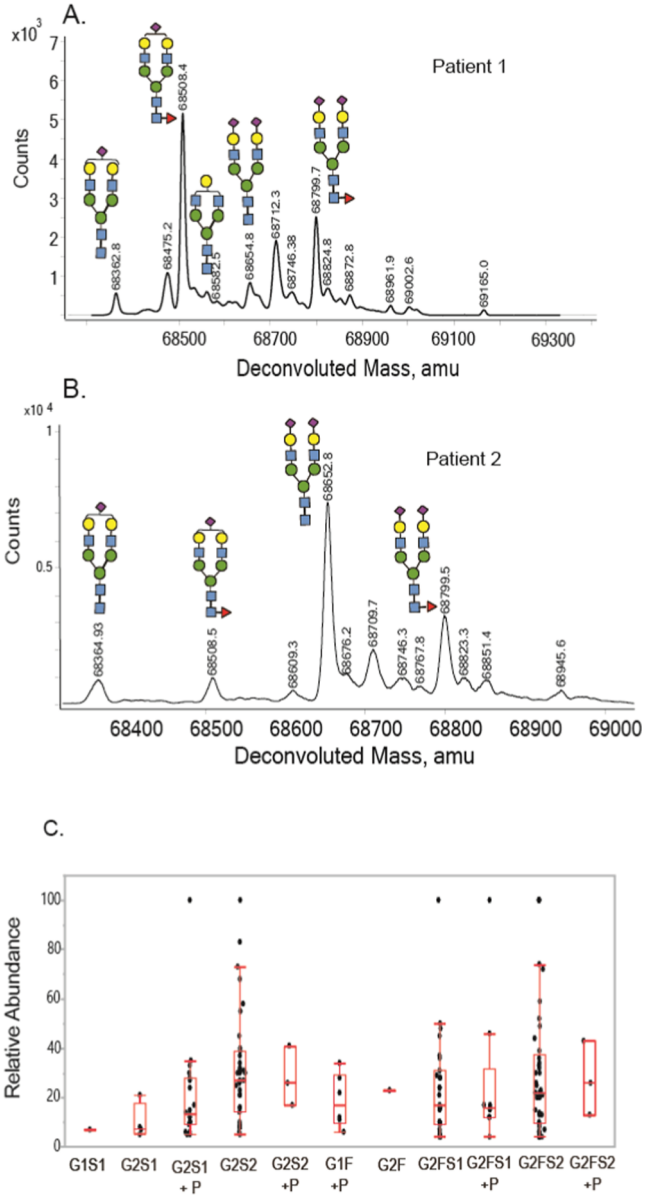


Figure 3. (A) AFP glycoforms identified in a patient sample with relatively high AFP-L3%. (B) AFP glycoforms identified in a patient sample with relatively low AFP-L3%. (C) Distribution of relative abundances of various AFP proteoforms observed in analyzed authentic patient serum samples (n = 40).

https://doi.org/10.1021/acs.jproteome.2c00656
J. Proteome Res. 2023, 22, 226−234

Table 1. Frequently Observed AFP Proteoforms and Their Detection Frequencies in 40 Analyzed Patient Samples

intact mass (Da)	occurrence frequencies (number of patients observed)	glycan modifications	phosphorylation	glycoform classification
68,361 ± 15	4	G2S1	none	L1 isoform
68,508 ± 5	17	G2FS1	none	L3 isoform
68,534 ± 15	6	G1F	6 phosphorylation	L3 isoform
68,799 ± 5	25	G2FS2	none	L3 isoform
68,652 ± 15	21	G2S2	none	L1 isoform
68,841 ± 15	18	G2S1	6 phosphorylation	L1 isoform
68,987 ± 15	9	G2FS1	6 phosphorylation	L3 isoform

prominent glycoforms (Figure S4) was further expanded to include equal numbers of glycoforms corresponding to the AFP-L1 and the AFP-L3 isoforms. The final glycoform list consisted of 12 glycan modifications: G0, G0F, G1, G1F, G1S1, G1FS1, G2, G2F, G2S2, G2S1, G2FS1, and G2FS2. Using the developed method, we analyzed commercial AFP standard purified from human umbilical cord blood.[26] The deconvoluted mass spectrum of AFP from the commercial standard showed distinctively different intact mass spectra (Figure S5), suggesting different features of proteoforms in umbilical cord blood compared to pathologic serum samples from adults (Figure 3A,B).

Out of all analyzed patient samples (n = 40), the AFP-L3 isoform G2FS2 (mass 68,799 Da) appeared in 62.5% of the patient samples (Table 1 and Figure S4). The most frequently observed AFP-L1 isoform was G2S2 (mass 68,652 Da, observed in 52.5% of all samples). Both G2FS2 and G2S2 are complex glycan structures with two terminal sialic acids, which have been reported as glycan modifications in other studies based on bottom-up proteomics methods for AFP.[19] Another frequently observed AFP-L3 isoform, G2FS1 (mass 68,508 Da), was identified in 42.5% of patient samples (Table 1 and Figure S4). The presence of G2FS2 and G2FS1 is strongly associated with each other in patient samples (Figures 3 and S6). G2S1 glycosylation (mass 68,841 Da) was another frequently observed AFP-L1 isoform (45% of patient samples; Table 1 and Figure S4).

AFP Phosphorylation in Authentic Patient Serum Samples

In addition to glycosylation, phosphorylation was an additional PTM allowed in intact mass analysis and proteoform identification. Six putative phosphorylation sites in the AFP sequence have been reported;[17] however, there is no published literature to suggest the occupancy of these phosphorylation sites. Variable phosphorylation sites combined with five possible glycan modifications can result in numerous closely spaced theoretical target masses (Figure S7), hence creating large numbers of matches for some of the deconvoluted masses. To minimize these potential matches, we narrowed the identification criteria to allow two occupancy states: 0 or 6 phosphorylation. In addition to nonphosphorylated forms (glycans G2S1, G2S2, G1F, G2FS1, and G2FS2), corresponding proteoforms with six phosphorylation sites (mass 68,841, 69,132, 68,534, 68,987, and 69,278 Da, respectively) were used to quantify AFP-L3%. The identified proteoforms exhibited variable relative abundance among the analyzed patient samples (Figure 3C).

Method Performance for Determining AFP-L3%

Specificity. Serum samples containing less than 0.01 μg/mL of AFP (determined by an immunoassay) were analyzed using the developed LC−HRMS method to assess the specificity of the analysis. The absence of AFP in the serum samples was corroborated by the absence of the AFP chromatographic peak at the expected RT (Figure S8). To assess method specificity, background MS intensity within ±0.5 min of the expected RT window of AFP was deconvoluted and the intact mass spectra were inspected for the presence of the expected intact masses corresponding to the monitored AFP proteoforms. Deconvoluted intact masses in the vicinity of 68,550 and 68,679 Da were observed in few of AFP-negative serum samples. Based on the above observations, the two possible proteoforms, phosphorylated AFP-G2 (mass 68,550 Da) and AFP-G1S1 with (mass 68,679 Da), were excluded as proteoform target matches due to potential false identification. Since both proteoforms belong to the AFP-L1 isoform, the corresponding AFP-L3 glycoforms (G2F, G1FS1) with phosphorylation (mass 68,696 and 68,825 Da) were also excluded from the potential target list for AFP-L3% calculation.

Reproducibility. To evaluate assay reproducibility in quantifying AFP-L3%, we prepared and analyzed three serum sample pools at different AFP concentrations (Table 2) by mixing patient serum samples containing high concentrations of AFP with AFP-negative serum samples. Four replicate extractions and subsequent analysis were performed for each pool, and AFP-L3% was calculated using the height of deconvoluted masses of the detected glycoforms (Table 2). The variation in AFP-L3% among the replicate analyses was calculated and expressed as the coefficient of variation (CV%). In two serum pools with similar total AFP concentrations, but different AFP-L3% (pools 1 and 2), the imprecision was 1.6 and 17.7% (Table 2). In sample pools 1 and 2, different glycoforms of AFP were detected, potentially impacting the reproducibility in quantifying AFP-L3% in these samples. The highest AFP concentration of pool 3 (66.4 μg/mL) exhibited consistent glycoform intensities across all four replicates (Table 2). The technical replicates (reinjection of the final prepared patient serum samples) showed reproducible performance for quantifying AFP-L3% (Table S4).

Table 2. Reproducibility in Quantifying AFP-L3% Based on Replicate Analysis of Serum Sample Pools

serum pools	AFP concentration (μg/mL)	average AFP-L3%	CV%
pool 1	16.7	70.8%	1.6%
pool 2	17.0	55.8%	17.7%
pool 3	66.4	55.3%	6.9%

Dilution Integrity. Samples containing high and low concentrations of AFP were mixed in different proportions to create four AFP pools with varying total AFP concentrations (115−450 μg/mL), but the same AFP-L3% (Table 3). Based on the LC−HRMS intact protein analysis, the most prominent glycoforms identified in these samples were G2FS2 and G2FS1 (both AFP-L3 isoforms). The mean measured AFP-L3% was

https://doi.org/10.1021/acs.jproteome.2c00656
J. Proteome Res. 2023, 22, 226−234

Table 3. Dilution Integrity of AFP-L3%

serum sample pools (varying AFP proportions)	AFP concentration (μg/mL)	AFP-L3%
100%	456.5	81.4%
75% pool	342.4	88.6%
50% pool	228.3	92.2%
25% pool	114.1	86.3%

consistent across the four pools (mean 87.1%, range 81.4−92.2%, Table 3).

Comparison with the Lectin-Based GSE Method. As part of the method evaluation, we analyzed 40 individual authentic serum samples using the LC−HRMS workflow. The total AFP concentrations in these serum samples (determined by immunoassay) ranged between 6 and 450 μg/mL (mean/median of 87.2/37.2 μg/mL). The age of the participants ranged from 48 to 91 years (mean/median of 65/64 years). Based on the LC−HRMS method, the relative AFP-L3% isoforms in these specimens ranged from 0% (only AFP-L1 isoforms identified) to 100% (only AFP-L3 isoforms identified) (Figure 4). The majority of the samples contained

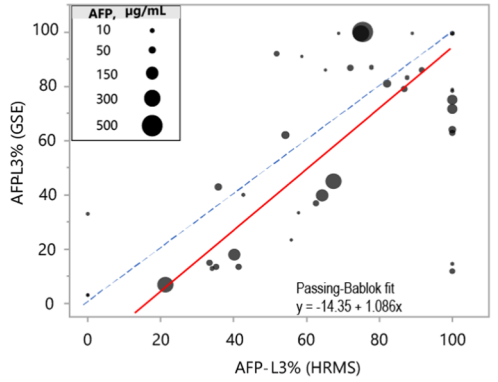


Figure 4. Comparison and correlation of AFP-L3% results based on LC−HRMS and GSE (Wako) in authentic serum samples. Pearson's r: 0.645.

both AFP-L1 and AFP-L3 glycoforms (Figure S6). While the AFP-L3% results determined by the LC−HRMS and GSE methods are comparable (Pearson's r: 0.645, Figure 4), the GSE method quantifies AFP-L3% with a noticeable positive proportional bias (8.6%, Figure 4). The greater discrepancies were observed in patient specimens with predominant AFP-L1 isoforms and lower total AFP concentrations. It is likely that these AFP-L1 isoforms are not detected as effectively as AFP-L3 isoforms based on the LC−HRMS method.

DISCUSSION

Here, we report a rapid and reproducible LC−HRMS method to detect and identify glycoforms of AFP, a clinical biomarker to monitor impaired liver function. We also selected the AFP proteoforms of reproducible and specific signals to quantify AFP-L3%, a malignancy risk differentiator. To measure AFP-L3%, we established an intact protein analysis workflow to determine the relative abundance of a subset of AFP glycoforms with and without core fucosylation. While developing the method, we designed and optimized several aspects of the workflow, which makes the method feasible to be further developed into a clinical assay for HCC biomarker, AFP-L3%. Preliminary performance evaluation suggests that this method is adequately specific and reproducible, and AFP-L3% determined by the developed LC−HRMS method reasonably agrees with the relative abundances measured by the GSE method.[25]

Optimization of the AFP Enrichment Process for Intact Protein Analysis

Due to the highly complex composition of human serum, antibody-based sample preparation was developed to selectively enrich AFP and to remove the majority of unrelated proteins prior to LC−HRMS analysis. We evaluated multiple commercially available anti-AFP antibodies for AFP recovery and unbiased enrichment of different AFP glycoforms from serum samples. In addition to AFP, the immune enrichment recovered a small fraction of albumin, which can be seen by intact mass analysis in all analyzed serum samples after immune enrichment. A likely explanation for the albumin presence is the partial homology between AFP and albumin, resulting in a certain degree of cross reactivity with the anti-AFP antibody and a small fraction (~0.05%) of the endogenously present albumin coenriched with AFP. In addition to albumin, there are other serum proteins present in post-immune-enrichment samples (Figure S3 and Table S3). Nevertheless, these proteins are of significant different molecular weights relative to AFP and do not interfere during MS analysis.

Compared to glycopeptide-based bottom-up methods to profile AFP glycoforms,[14,19] advantages of our intact protein workflow include the following: (i) the immune-enrichment process is performed at physiologic conditions, which should preserve native AFP proteoforms; (ii) the sample preparation is faster and less complex without enzyme digestions; and (iii) comparable ionization efficiency of different AFP proteoforms enables AFP-L3% determination without the need to quantify the absolute concentrations of targeted glycoforms. In agreement with glycopeptide-based methods,[14] the most abundant and frequently observed AFP-L1 and AFP-L3 proteoforms contain complex glycan structures with one or two terminal sialic acids.[14] Although the GSE-based methods can detect the core fucosylation motif of AFP-L3 glycoforms based on lectin binding, they cannot provide additional glycan structure information such as terminal sialic acids, identifiable based on the LC−HRMS method. With further optimization and automation, the LC−HRMS workflow can be more readily scaled up to accommodate high-volume clinical testing.

Since there are several heavily glycosylated serum proteins (IgG, α1-antitrypsin, α2 macroglobulin, etc.) present in the postenrichment samples (Table S3), released glycan profiling can include glycans from other proteins rather than specific to AFP glycosylation.

Optimization of LC−HRMS and Data Analysis Workflow

Chromatographic separation utilizing the 300 Å PLRP column was optimized to achieve AFP peak separation from other coenriched proteins such as albumin and to ensure the reproducible detection of AFP using deconvolution of the multiple charged envelopes within the AFP RT window. For MS detection, the mass range of 1000−3000 m/z was sufficient to capture multiple charged envelopes of AFP and albumin and to minimize interferences from other proteins or small molecules.

The data analysis workflow was optimized to reduce the computation time and to simplify the glycoform identification process. We selected a moderate step size of 7 Da to provide

https://doi.org/10.1021/acs.jproteome.2c00656
J. Proteome Res. 2023, 22, 226−234

reliable mass resolution in the deconvoluted mass spectra and high confidence matches to the expected AFP proteoforms with defined glycan and phosphorylation modifications. Based on the intact mass differences between AFP and human albumin, we devise two discrete mass ranges, 68,000−80,000 and 66,000−67,500 Da, respectively, for deconvolution to obtain adequate quality mass spectra.

While developing the method, we optimized the mass matching tolerance using acquired data from enriched AFP-positive patient samples. We initially utilized a narrow mass matching window of ±5 Da and found that several deconvoluted masses of significant intensity were not identified as AFP glycoforms by BioConfirm software, while they were clearly in the vicinity of the masses of the expected AFP with various glycan modifications. Evaluation of the larger tolerance window demonstrated that the mass tolerance of ±15 Da enabled adequate identification of the targeted AFP glycoforms without introducing additional false-positive matches.

Altered Glycosylation and Other Post-Translational Modifications in Cancers

Glycan synthesis and conjugation to proteins is a dynamic process that is impacted by changes in available sugar precursors, enzymes, and intracellular signal processes. In neoplasm conditions, the associated glycosylation changes have been shown to alter core fucosylation,[8] increase α-2-6-sialylation,[27] and increase N-glycan branching.[28] When using LC−HRMS to profile intact AFP protein glycoforms, we observed diverse AFP glycoform distributions in majority of the analyzed patient samples (Table 1 and Figures 3B and S6); these glycoforms were also observed in previous glycomics studies that profile AFP isoforms.[13,15,21] In addition to the core fucosylation, our study also showed that the most abundant AFP glycoforms detected are heavily sialylated, regardless of the core fucosylation status in the patient samples (Figures 2 and S6). Because clinical diagnoses associated with the analyzed patient specimens were not available, we cannot speculate on the association between the degree of sialylation and hepatic malfunctions related to different liver pathological conditions. In the previous bottom-up study, Kim et al reported that approximately half of the glycans identified in samples of HCC patients were sialylated, suggesting that sialylation in HCC is common as sialylated carbohydrates play a role in cellular recognition, cell adhesion, and cell signaling.[23]

Although AFP phosphorylation has not been investigated extensively at a secretory protein level, our study revealed that a significant proportion of the AFP glycoforms is phosphorylated. A previous proteomics study has identified six putative phosphorylation sites on the AFP protein.[17] Phosphorylation is involved in the regulation of cell functions, including cell growth, differentiation, and apoptosis with multiple signaling pathways participating in phosphorylation cascade; alterations in phosphorylation pathways have been shown to be associated with cancers.[29]

Limitations of the Study

One of the limitations of this method is inadequate sensitivity to reliably detect AFP and its different proteoforms in patient samples with total AFP concentrations below 1 μg/mL. When analyzing such low concentration samples, our method cannot reliably detect the glycoforms of interest; therefore, additional work is needed to enhance the method's sensitivity.

Another limitation of the study is the lack of detailed clinical information related to the analyzed samples; future studies using patient samples with available clinical information are needed to assess the diagnostic utility of the various AFP proteoforms with different degrees of sialylation and phosphorylation.

CONCLUSIONS

In summary, we developed a method to analyze AFP glycoforms that utilizes AFP-specific immune enrichment, followed by LC−HRMS intact protein analysis. The method allows to profile different AFP proteoforms, which are further categorized to AFP-L1 and AFP-L3 based on their core fucosylation features. The combined relative intensity of detected glycoforms can be utilized to determine AFP-L3% in analyzed serum samples. Different from bottom-up workflows, the intact AFP analysis method reduces the complexity of the sample preparation and chromatographic separation and allows the determination of AFP-L3% based on the relative abundances of the identified AFP proteoforms.

As LC−HRMS instruments become more robust, allowing reliable detection of intact proteins, analysis of intact proteins which have low proteoform diversity could become more common in clinical laboratories.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00656.

Instrument source settings for HRMS data acquisition (Table S1); list of identified AFP tryptic peptides (Table S2); list of proteins identified in authentic patient serum samples post AFP enrichment (Table S3); reproducibility of AFP-L3% based on replicate injections (N = 2) (Table S4); glycan modifications of AFP (Figure S1); deconvoluted mass spectrum of human serum albumin post AFP enrichment (Figure S2); protein electrophoresis of serum AFP after immune enrichment and AFP standard (Figure S3); histogram of the observed AFP proteoforms (Figure S4); deconvoluted mass spectrum of AFP derived from human umbilical cord blood (Figure S5); different AFP glycoforms in authentic patient serum samples (Figure S6); possible proteoforms with variable phosphorylations and their expected intact masses (Figure S7), and chromatogram of a patient serum sample with AFP < 0.001 μg/mL (Figure S8) (PDF)

AUTHOR INFORMATION

Corresponding Author

Yifei K. Yang − *ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah 84108, United States; Department of Pathology, University of Utah, Salt Lake City, Utah 84108, United States;* Present Address: Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota 55902, United States; orcid.org/0000-0001-9817-484X; Email: yang.yifei@mayo.edu

Authors

Carmen Dunbar − *ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah 84108, United States*

https://doi.org/10.1021/acs.jproteome.2c00656
J. Proteome Res. 2023, 22, 226−234

Mark M. Kushnir − *ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah 84108, United States; Department of Pathology, University of Utah, Salt Lake City, Utah 84108, United States;* orcid.org/0000-0003-4480-0854

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.2c00656

Author Contributions

[‖]C.D. and M.M.K. contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

The work was supported by the William L. Roberts Memorial Fund from ARUP Institute for Clinical and Experimental Pathology.

Notes

The authors declare the following competing financial interest(s): The method disclosed herein is the subject of a United States patent application Serial No. 17/951,072.

■ ACKNOWLEDGMENTS

The authors thank Rachel Law for her help in retrieving and deidentifying residual patient samples. They also thank Jon Seiter and Cindy Gin for their technical support regarding the AFP immunoassay and gel shift electrophoresis assay.

■ ABBREVIATIONS

AFP, alpha fetoprotein; AFP-L1, alpha fetoprotein glycoforms without core fucosylation modification; AFP-L3, alpha fetoprotein glycoforms with core fucosylation modification; AFP-L3%, relative abundance of AFP-L3 glycoforms; LCA, lens culinaris agglutinin; GSE, gel shift electrophoresis; LC−HRMS, liquid chromatography coupled with high-resolution mass spectrometry; LC−MS/MS, liquid chromatography coupled with tandem mass spectrometry; MALDI, matrix-assisted laser desorption/ionization; RT, retention time; HCC, hepatocellular carcinoma; PTM, post-translational modification

■ REFERENCES

(1) Schjoldager, K. T.; Narimatsu, Y.; Joshi, H. J.; Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 729−749.

(2) Pinho, S. S.; Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **2015**, *15*, 540−555.

(3) Larsen, M. D.; de Graaf, E. L.; Sonneveld, M. E.; Plomp, H. R.; Nouta, J.; Hoepel, W.; Chen, H. J.; Linty, F.; Visser, R.; Brinkhaus, M.; et al. Afucosylated IgG characterizes enveloped viral responses and correlates with COVID-19 severity. *Science* **2021**, *371*, No. eabc8378.

(4) Haukedal, H.; Freude, K. K. Implications of Glycosylation in Alzheimer's Disease. *Front. Neurosci.* **2020**, *14*, No. 625348.

(5) Nevone, A.; Girelli, M.; Mangiacavalli, S.; Paiva, B.; Milani, P.; Cascino, P.; Piscitelli, M.; Speranzini, V.; Cartia, C. S.; Benvenuti, P.; et al. An N-glycosylation hotspot in immunoglobulin kappa light chains is associated with AL amyloidosis. *Leukemia* **2022**, *36*, 2076−2085.

(6) Xu, M.; Jin, H.; Wu, Z.; Han, Y.; Chen, J.; Mao, C.; Hao, P.; Zhang, X.; Liu, C. F.; Yang, S. Mass Spectrometry-Based Analysis of Serum N-Glycosylation Changes in Patients with Parkinson's Disease. *ACS Chem. Neurosci.* **2022**, *13*, 1719−1726.

(7) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **2019**, *15*, 346−366.

(8) Ramachandran, P.; Xu, G.; Huang, H. H.; Rice, R.; Zhou, B.; Lindpaintner, K.; Serie, D. Serum Glycoprotein Markers in Non-alcoholic Steatohepatitis and Hepatocellular Carcinoma. *J. Proteome Res.* **2022**, *21*, 1083−1094.

(9) Andre, F.; Ismaila, N.; Allison, K. H.; Barlow, W. E.; Collyar, D. E.; Damodaran, S.; Henry, N. L.; Jhaveri, K.; Kalinsky, K.; Kuderer, N. M.; et al. Biomarkers for Adjuvant Endocrine and Chemotherapy in Early-Stage Breast Cancer: ASCO Guideline Update. *J. Clin. Oncol.* **2022**, *40*, 1816−1837.

(10) Milose, J. C.; Filson, C. P.; Weizer, A. Z.; Hafez, K. S.; Montgomery, J. S. Role of biochemical markers in testicular cancer: diagnosis, staging, and surveillance. *Open Access J. Urol.* **2011**, *4*, 1−8.

(11) Sato, Y.; Nakata, K.; Kato, Y.; Shima, M.; Ishii, N.; Koji, T.; Taketa, K.; Endo, Y.; Nagataki, S. Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. *N. Engl. J. Med.* **1993**, *328*, 1802−1806.

(12) Benn, P.; Cuckle, H.; Pergament, E. Non-invasive prenatal testing for aneuploidy: current status and future prospects. *Ultrasound Obstet. Gynecol.* **2013**, *42*, 15−33.

(13) Johnson, P. J.; Poon, T. C.; Hjelm, N. M.; Ho, C. S.; Ho, S. K.; Welby, C.; Stevenson, D.; Patel, T.; Parekh, R.; Townsend, R. R. Glycan composition of serum alpha-fetoprotein in patients with hepatocellular carcinoma and non-seminomatous germ cell tumour. *Br. J. Cancer* **1999**, *81*, 1188−1195.

(14) Kim, K. H.; Lee, S. Y.; Hwang, H.; Lee, J. Y.; Ji, E. S.; An, H. J.; Kim, J. Y.; Yoo, J. S. Direct Monitoring of Fucosylated Glycopeptides of Alpha-Fetoprotein in Human Serum for Early Hepatocellular Carcinoma by Liquid Chromatography-Tandem Mass Spectrometry with Immunoprecipitation. *Proteomics - Clin. Appl.* **2018**, *12*, No. 1800062.

(15) Johnson, P. J.; Poon, T. C.; Hjelm, N. M.; Ho, C. S.; Blake, C.; Ho, S. K. Structures of disease-specific serum alpha-fetoprotein isoforms. *Br. J. Cancer* **2000**, *83*, 1330−1337.

(16) Steentoft, C.; Vakhrushev, S. Y.; Joshi, H. J.; Kong, Y.; Vester-Christensen, M. B.; Schjoldager, K. T.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N. B.; Marcos-Silva, L.; et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **2013**, *32*, 1478−1488.

(17) Tagliabracci, V. S.; W, S. E.; Guo, X.; Kinch, L. N.; Durrant, E.; Wen, J.; Xiao, J.; Cui, J.; Nguyen, K. B.; Engel, J. L.; Coon, J. J.; Grishin, N.; Pinna, L. A.; Pagliarini, D. J.; Dixon, J. E. A single kinase generates the majority of the secreted phosphoproteome. *Cell* **2015**, *161*, 1619−1632.

(18) Taketa, K.; Hirai, H. Lectin affinity electrophoresis of alpha-fetoprotein in cancer diagnosis. *Electrophoresis* **1989**, *10*, 562−567.

(19) Kim, H.; Sohn, A.; Yeo, I.; Yu, S. J.; Yoon, J. H.; Kim, Y. Clinical Assay for AFP-L3 by Using Multiple Reaction Monitoring-Mass Spectrometry for Diagnosing Hepatocellular Carcinoma. *Clin. Chem.* **2018**, *64*, 1230−1238.

(20) Lee, J.; Yeo, I.; Kim, Y.; Shin, D.; Kim, J.; Kim, Y.; Lim, Y. S.; Kim, Y. Comparison of Fucose-Specific Lectins to Improve Quantitative AFP-L3 Assay for Diagnosing Hepatocellular Carcinoma Using Mass Spectrometry. *J. Proteome Res.* **2022**, *21*, 1548−1557.

(21) Nakagawa, T.; Miyoshi, E.; Yakushijin, T.; Hiramatsu, N.; Igura, T.; Hayashi, N.; Taniguchi, N.; Kondo, A. Glycomic analysis of alpha-fetoprotein L3 in hepatoma cell lines and hepatocellular carcinoma patients. *J. Proteome Res.* **2008**, *7*, 2222−2233.

(22) Nakagawa, T.; Takeishi, S.; Kameyama, A.; Yagi, H.; Yoshioka, T.; Moriwaki, K.; Masuda, T.; Matsumoto, H.; Kato, K.; Narimatsu, H.; et al. Glycomic analyses of glycoproteins in bile and serum during rat hepatocarcinogenesis. *J. Proteome Res.* **2010**, *9*, 4888−4896.

(23) Kim, Y. J.; Varki, A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconjugate J.* **1997**, *14*, 569−576.

(24) Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; et al. The PRIDE database

resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **2022**, *50*, D543−D552.
(25) *TASWako AFP-L3 Test*. https://www.wako-chemicals.de/en/products/diagnostics/hcc-biomarkers/taswako-afp-l3-test (accessed March 24, 2022).
(26) Biosolutions, L.*AFP Standard*. https://www.leebio.com/product/58/alpha-fetoprotein-afp-human-cord-serum-105-11 (accessed May 10, 2022).
(27) Becker, V.; Timmer, J.; Klingmüller, U.Receptor Dynamics in Signaling. In *Advances in Systems Biology*; Goryanin, I. I.; Goryachev, A. B., Eds.; Springer: New York, 2012; pp 313−323.
(28) Ščupáková, K.; Adelaja, O. T.; Balluff, B.; Ayyappan, V.; Tressler, C. M.; Jenkinson, N. M.; Claes, B. S.; Bowman, A. P.; Cimino-Mathews, A. M.; White, M. J.; et al. Clinical importance of high-mannose, fucosylated, and complex N-glycans in breast cancer metastasis. *JCI Insight* **2021**, *6*, No. e146945.
(29) Singh, V.; Ram, M.; Kumar, R.; Prasad, R.; Roy, B. K.; Singh, K. K. Phosphorylation: Implications in Cancer. *Protein J.* **2017**, *36*, 1−6.

https://doi.org/10.1021/acs.jproteome.2c00656
J. Proteome Res. 2023, 22, 226−234