Optimization of Sample Preparation Methods for Building Human Glycopeptide Libraries

Yi Liu, Meghan C. Burke, Zachary C. Goecker, Sergey L. Sheetlin, Guanghui Wang, Yuri A. Mirokhin, Zheng Zhang, Yuxue Liang, Xiaoyu Yang, Dmitrii V. Tchekhovskoi, Stephen E. Stein

Mass Spectrometry Data Center, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899, United States

PIPELINE_GUI SOFTWARE
GLYCOPEPTIDE_LIBRARY

INTRODUCTION

Since glycoproteins are ubiquitous in human biology and the specific glycans attached to each glycosylated site depend on cellular kinetics, not genomics, determination of these glycan structures requires special methods. In this report we described work being done at NIST to determine individual N-linked glycans and record their spectra and abundance distributions at each site for over 30 native and/or recombinant human glycoproteins. Experiments were designed to optimize glycoprotein sample processing to achieve reliable and reproducible results. By applying NIST developed validation methods to tentative identifications made by MSFragger, we derived a comprehensive library of proteolytic glycopeptide data for commonly observed plasma and related glycoproteins.

MATERIALS AND METHODS

More than thirty native and/or recombinant human plasma glycoproteins, such as Alpha-1-Glycoprotein (A1AG), Apolipoprotein-B100 (APOB), Alpha-1-Antitrypsin (A1AT) and other selected human glycoproteins, such as Thyroglobulin (TG), Prostate-Specific-Antigen (PSA) and Erythropoietin (EPO), were processed for building human glycopeptide libraries. The workflow of sample processing, before and after digestion, typically entailed protein sample denaturation, reduction and alkylation, termination of proteolytic activity, peptide desalting and clean-up. These peptides were then analyzed by nanoflow liquid chromatography with tandem spectra acquired by DDA in a high resolution/high mass accuracy mass spectrometer. Library creation was done using a multistep NIST developed analysis platform, Pipeline_GUI, leading to a GADS library, consisting of Glycopeptide Abundance Distribution Spectra (GADS) for each N-glycosylated site, along with individual MS/MS spectra.

RESULTS

1. Digestion buffer– Two buffer systems were evaluated: 50mM Tris-HCl, pH8.5 and 50mM Ammonia Bicarbonate (ABC), pH 8.5. ABC was found to be optimal for most of protease digestions based on its greatly improved pH stabilization. However, Tri-HCL is ideal for certain proteases, such as Subtilisin (Table 1).

Table 1. Thyroglobulin Digested in Different Buffers

Protease(s)	Ammonia Bicarbonate (ABC)				Tris-HCl			
	NoGpep	TotGpep	DiffGpep	UniGlycan	NoGpep	TotGpep	DiffGpep	UniGlycan
AlphaLytic	5265	500	249	29	2565	84	59	12
Chymotrypsin	6573	684	316	43	2631	140	83	16
Glu-C	5157	306	167	40	3050	161	107	29
Trypsin	6415	782	335	60	2870	150	84	22
Subtilinsin	14223	1224	372	37	12933	1398	410	34
Subtilinsin/Trypsin	9430	1028	283	38	10327	1866	433	45

2. Denaturants– Eight denaturant conditions were examined (Table 2a). Among of them, an acid labile surfactant (ASF), SDS and Urea were further evaluated for several other glycoproteins based on their denaturing efficiency. ASF was generally the most effective and easy-to-use, so used for glycoproteins (Table 2b). As to those without disulfide bonds, glycoproteins such as Alpha-1-Antitrypsin, we tested without denaturants and heat alone methods. These were found less effective than digestions with chemical denaturants (Table 2c).

Table 2a. Comparing of Protein Denaturants for APOB

Denaturants	Temp/Time	NoGpep	TotGpep	DiffGpep	UniGlycan
80% ACN	25°C/60mins	6412	106	51	27
6M Guanidine	40°C/60mins	8923	175	69	15
20% Methanol	25°C/60mins	7014	108	53	21
0.1% RapiGest	60°C/30mins	12068	552	130	24
0.4% Sodium Dodecyl Sulfate	95°C/10mins	11989	570	164	23
50% Trifluoroethanol	25°C/60mins	5762	111	54	23
8M Urea	25°C/60mins	14609	684	146	24
No Denaturant	95°C/10mins	5161	123	47	17

Table 2b. Comparing of Protein Denaturants on Fibrinogen

Denaturants	NoGpep	TotGpep	DiffGpep	UniGlycan
0.1% ASF	9863	411	96	15
0.4% SDS	7353	299	72	14
8M Urea	7484	239	57	15

Table 2c. Comparing with/without Denaturants for Alpha-1-Antitrypsin

Protease(s)	Denaturant				No-Denaturant			
	NoGpep	TotGpep	DiffGpep	UniGlycan	NoGpep	TotGpep	DiffGpep	UniGlycan
Trypsin/Lys-C	15540	1104	225	44	12582	588	175	18
Lys-C/Glu-C	13106	1019	200	31	13035	1259	255	35
Lys-C/Arg-C	15585	506	145	25	9825	512	126	17
Chymotrypsin	12927	1368	265	44	6897	154	54	7

3. Termination– Three protease termination methods were evaluated: 1) 0.1% Formic Acid: Acidic protease reaction buffer till final concentration of 0.1% FA to inhibit the protease activity; 2) 95°C heat: Inactive protease with high temperature; and 3) Speed Vacuum drying: High temperature inactivation, specially used for sequential digestion with different buffer. Relative abundances of site specific glycans were not affected by the methods (Fig 1), but 0.1% FA and 95°C heat inactivated proteases generated the larger glycopeptide abundances as showed in Table 3.

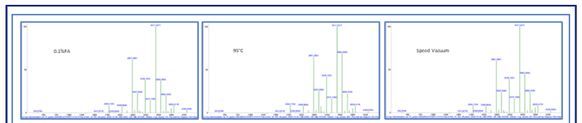


Fig.1 A1AG Abundance with Three Different Termination Methods

Table 3. Comparing Terminating Methods

Terminating Methods	A1AG Denatured with RapiGest				A1AG Denatured with Methanol			
	NoGpep	TotGpep	DiffGpep	UniGlycan	NoGpep	TotGpep	DiffGpep	UniGlycan
0.1% FA Acidification	2670	1092	421	57	2143	687	265	41
95°C Heating	2231	1072	430	64	1731	780	293	47
SpeedVac Dry	1971	995	406	62	1605	584	242	40

4. Desalting– Two desalting methods, C18 Zip-Tip Peptide Clean-up tips and MonoSpin SPE centrifugal spin columns, were applied for the purpose of proteolytic peptides clean-up. MonoSpin SPE centrifugal spin columns (MS) with the low-pressure, high-flow, and low-liquid-retention properties, as well using stepped elution buffers, yielded the largest number of identifications (Table 4a).

Table 4a. Comparing Desalt Methods

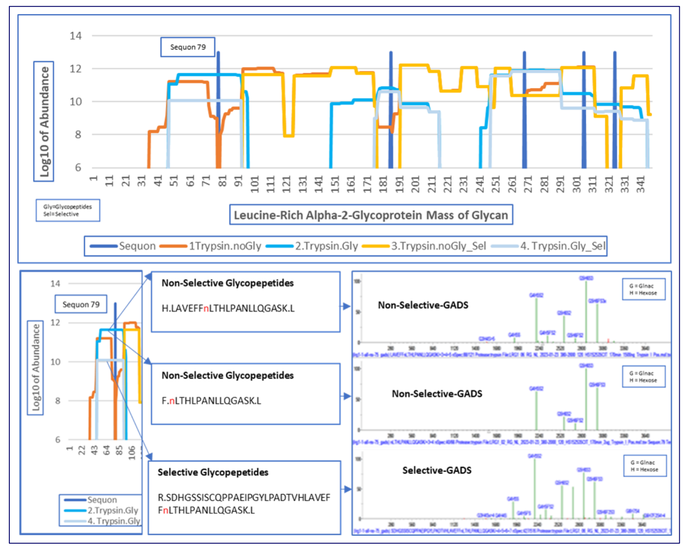
Methods	NoGpep	TotGpep	DiffGpep	UniGlycan
MS-1 (100% ACN)	9750	313	107	30
MS-2 (50%, 80%ACN/0.1%FA)	20994	532	167	33
ZipTip-1 (3xZip-Tips)	16568	378	124	30
ZipTip-2 (1xZip-Tip)	16580	369	120	30

During desalting with MonoSpin columns, a washing solution was used for removing salts. Two solutions were compared. One is vendor recommended LC-H2O, and another is 0.1%FA solution. Our results found 0.1%FA clearly optimal (Table 4b).

Table 4b. Comparing Desalt Washing Solutions

Protease(s)	Washing Column with 0.1% Formid Acid				Washing Column with LC-H2O			
	NoGpep	TotGpep	DiffGpep	UniGlycan	NoGpep	TotGpep	DiffGpep	UniGlycan
Lys-C/Arg-C	7868	1225	311	64	6065	883	207	56
Trypsin	7630	1536	343	68	6652	873	200	54
Trypsin/Lys-C	8747	1980	438	83	6308	1269	328	67

5. Data analysis– Using the NIST developed Pipeline, surprisingly, we found the MSFragger non-specific cleavage settings yielded the best results, regardless of the protease employed. For the tested proteases, such as Trypsin, Chymotrypsin, Lys-C, Glu-C, Asp-N, Arg-C, LysargiNase and Alpha Lytic, this method generated peptides from unexpected cleavage sites, generating additional high-quality glyco- and non-glycopeptides. Fig 2. presents an example, Leucine-Rich Alpha-2-Glycoprotein Sequon 79, in which two of high-quality tryptic unexpected unique Glycopeptide Abundance Distribution Spectra (GADS) were detected with non-selective search while with selective, only one tryptic glycopeptide was found. This provides important confirmation of glycan distributions. We find that non-selective search is a very powerful calculation method capable of reliable extracting all usable information derivable from protein digestions.



CONCLUSION

In this report, more than thirty Human Glycoproteins were studied. The preparation methods of glycopeptide were optimized and evaluated using a NIST developed Pipeline software for building Glycopeptide Libraries.