A general protease digestion procedure for optimal protein sequence coverage and PTM analysis of recombinant glycoproteins: Application to the characterization of hLOXL2 glycosylation.

*Kathryn R. Rebecchi, Eden P. Go, Li Xu, Carrie L. Woodin, Minae Mure, and Heather Desaire**

Department of Chemistry, University of Kansas, Lawrence, KS

SUPPLEMENTAL INFORMATION

Additional Experimental Details.

Materials and Reagents.

Human apo-transferrin (transferrin), urea, tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), iodoacetamide (IAM), hydrochloric acid, acetic acid, and formic acid were purchased from Sigma (St. Louis, MO). Ammonium bicarbonate (NH_4HCO_3) was purchased from Fluka (Milwaukee, WI). HPLC grade acetonitrile (CH_3CN) was purchased from Fisher Scientific (Fair Lawn, NJ). RapiGest™ SF was purchased from Waters Corporation (Milford, MA). PNGase F from *Flavobacterium meningosepticum* was purchased from New England BioLabs (Ipswich, MA). Sequencing Grade Modified Trypsin was purchased from Promega (Madison, WI). Ultrapure water was obtained from an in-house Millipore Direct-Q® UV 3 system (Billerica, MA) with a resistance greater than 18 M.

HLOXL2 expression and purification: The cDNA containing the complete open reading frame of HLOXL2 (Origene Technologies, Rockville, MD) was used as a template and a 1 kb cDNA (+1306 to +2323 of HLOXL2 gene, GenBank accession number: NM_002318.2) was amplified by PCR and cloned into a pMT/Bip/V5-HisA (Invitrogen Carlsbad, CA) expression vector. The expression plasmid containing the 1 kb HLOXL2 gene was transfected into *Drosophila S*2 cells using SuperFect transfection reagent (Qiagen, Valencia, CA). The *S*2 cell expressed protein was purified by affinity chromatography, and the protein concentration was determined by BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA).[1]

Liquid chromatography/mass spectrometry.

The tryptic digest samples (5 μL, ~ 15 μg) were injected onto a reversed phase (C_{18}) column (300 μm i.d. x 5 cm, 3 μm particle size, CVC MicroTech, Fontana, CA) using a Dionex UltiMate capillary HPLC system (Sunnyvale, CA) containing a FAMOS well plate autosampler.

The HPLC system was connected to an ESI-LIT-FTICRMS (electrospray ionization – linear ion trap – Fourier transform ion cyclotron resonance mass spectrometer), ThermoScientific (San Jose, CA) containing a 7 Tesla actively shielded magnet. Mobile phase solvents A and B were composed of H_2O and CH_3CN, respectively, where both solvents contained 0.1 % formic acid. To elute peptides from the column, the solvent conditions were held initially for 5 min at 5 % B, linearly increased to 40 % B over 50 min, further increased to 90 % B in 10 min, then held at 90 % B for 10 min, and lastly the column was re-equilibrated before the next injection. To ensure that no sample carryover was detected in the MS data, a 30 min wash followed by a blank injection was implemented between each sample run. The mass spectrometer was set to the following parameters for all samples: ESI source voltage was 2.8 kV, capillary voltage offset was 47 V, capillary temperature was 200 °C, FT-ICR resolution was set to 25,000 for *m/z* 400, and MS/MS data were collected in a data dependent manner by selecting the five most intense ions in an FT-ICR MS scan for collision induced dissociation (CID), where a relative collision energy of 30 % and a dynamic exclusion window of 3 min were utilized.

Detection of glycopeptides

To identify glycopeptides, manual interpretation of the high resolution MS and MS/MS data was performed. To confirm a glycopeptide candidate composition, the monoisotopic mass must be present and within 20 ppm mass error of the theoretical mass. Additionally, the MS/MS data must support the assignment by containing an expected characteristic fragmentation pattern (losses of expected monosaccharide residues). For transferrin, the glycopeptides to be detected were known, since the glycosylation for this protein has been characterized previously. Therefore, the specific glycopeptide masses of interest were searched for and, when found, the composition was verified by MS/MS.

For hLOXL2, the glycan compositions were not known, as the glycosylation on this protein had not been characterized previously. Therefore, two complementary strategies were used to detect the glycopeptides present in the hLOXL2 MS data. The first strategy was to construct a prediction table of possible glycopeptide *m/z* values and then search the MS data for these *m/z* values. To construct the prediction table, the mass of either peptide containing the two *N*-linked glycosylation sites was added to possible glycan masses determined from glycans expected in glycoproteins from insect cells.[2] After searching the MS data for all the theoretical glycopeptide *m/z* values, the second strategy utilized to identify hLOXL2 glycopeptides was to scan the entire LC chromatogram for glycopeptides in the MS/MS data. Characteristic glycan product ions are typically present in all glycopeptide CID spectra, therefore when ions at *m/z* 366 (a hexose plus an *N*-acetylhexosamine) and/or *m/z* 528 (two hexose residues plus an *N*-acetylhexosamine) were detected in a CID spectrum, the spectrum was further searched for glycan losses characteristic for a glycopeptide spectrum. Before a glycopeptide could be considered identified, two criteria were required: An MS/MS data file was needed to support the assignment, and the monoisotopic peak must have an *m/z* value of less than 20 ppm mass error, compared to the calculated *m/z,* in the high resolution MS data.

References:

1) Xu, L.; Rebecchi, K. R.; Nightengale, R.; Taylor, M.; Calvert, W.; Limburg, J.; Desaire, H.; Mure, M. *J. Biol. Chem.* Submitted.
2) Fabini, G.; Freilinger, A.; Altmann, F.; Wilson, I. B. H. *J. Biol. Chem.* **2001**, *276*, 28058-28067.

Table S1. hLOXL2 nonglycosylated peptides detected using protease digestion condition 7

Peptide	Charge State	Theoretical *m/z*	Experimental *m/z*	Mass Error (ppm)
SPWPGVPTSMR	2	607.803	607.807	7
VEVLVER	1	843.494	843.498	5
QLGLGFASNAFQETWYWHGDVNSNK	2	1435.170	1435.188	13
VVMSGVK	1	719.412	719.413	2
CSGTELSLAHCR	2	695.803	695.806	4
FSSQIHNNGQSDFRPK	2	931.451	931.456	6
FSSQIHNNGQSDFRPKNGR	2	1095.015	1095.033	13
ASFCLEDTECEGDIQK	2	951.398	951.385	14
NYECANFGDQGITMGCWDMYR	2	1294.509	1294.527	15
HDIDCQWVDITDVPPGDYLFQVVINPNFE VAESDYSNNIMK	3	1599.742	1599.762	12
IWMYNCHIGGSFSEETEK	2	1094.477	1094.490	12
FEHFSGLLNNQLSPQSAWSHPQFE	2	1400.659	1400.669	7

Table S2: PTM's detected for hLOXL2 at the two putative glycosylation sites.

Peptide Composition	Glycan Composition[a]	Theoretical m/z[b]	Experimental m/z[b]	Mass Error (ppm)
NGSLVWGMVCGQNWGIVEAMVVCR	[Hex]3[HexNAc]2 [Fuc]1 + [Na]1	1261.548	1261.554	5
DGSLVWGMVCGQNWGIVEAMVVCR	No Glycan	1362.132	1362.153	15
HYHSMEVFTHYDLLNLNGTK	[Hex]0[HexNAc]1 [Fuc]1	923.434	991.435	1
HYHSMEVFTHYDLLNLNGTK	[Hex]0[HexNAc]2	992.441	992.443	2
HYHSMEVFTHYDLLNLNGTK	[Hex]1[HexNAc]2	996.460	996.460	0
HYHSMEVFTHYDLLNLNGTK	[Hex]0[HexNAc]2 [Fuc]1	991.127	991.134	7
HYHSMEVFTHYDLLNLNGTK	[Hex]1[HexNAc]2 [Fuc]1	1045.145	1045.140	5
HYHSMEVFTHYDLLNLNGTK	[Hex]2[HexNAc]2	1050.476	1050.480	4
HYHSMEVFTHYDLLNLNGTK	[Hex]2[HexNAc]2 [Fuc]1	1099.162	1099.178	15
HYHSMEVFTHYDLLNLNGTK	[Hex]3[HexNAc]2	1104.494	1104.514	18
HYHSMEVFTHYDLLNLNGTK	[Hex]3[HexNAc]2 [Fuc]1	1153.180	1153.172	7

[a]Hex = Hexose, HexNAc = *N*-acetyl hexosamine, Fuc = Fucose, Na = Sodium
[b]All glycopeptides were detected in the 3+ charge state.