NIH Public Access
Author Manuscript
Anal Chem. Author manuscript; available in PMC 2012 November 15.

Published in final edited form as:
Anal Chem. 2011 November 15; 83(22): 8484–8491. doi:10.1021/ac2017037.

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

Abstract

Using recombinant DNA technology for expression of protein therapeutics is a maturing field of pharmaceutical research and development. As recombinant proteins are increasingly utilized as biotherapeutics, improved methodologies ensuring the characterization of post-translational modifications (PTMs) are needed. Typically, proteins prepared for PTM analysis are proteolytically digested and analyzed by mass spectrometry. To assure full coverage of the PTMs on a given protein, one must obtain complete sequence coverage of the protein, which is often quite challenging. The objective of the research described here is to design a protocol that maximizes protein sequence coverage and enables detection of post-translational modifications, specifically *N*-linked glycosylation. To achieve this objective, a highly efficient proteolytic digest protocol using trypsin was designed by comparing the relative merits of denaturing agents (urea and Rapigest™ SF), reducing agents (dithiothreitol, DTT, and tris(2-carboxyethyl)phophine, TCEP), and various concentrations of alkylating agent (iodoacetamide, IAM). After analysis of human apo-transferrin using various protease digestion protocols, ideal conditions were determined to contain 6 M urea for denaturation, 5 mM TCEP for reduction, 10 mM IAM for alkylation, and 10 mM DTT, to quench excess IAM before the addition of trypsin. This method was successfully applied to a novel recombinant protein, human lysyl oxidase-like 2 (hLOXL2). Furthermore, the glycosylation PTMs were readily detected at two glycosylation sites in the protein. These digestion conditions were specifically designed for PTM analysis of recombinant proteins and biotherapeutics, and the work described herein fills an unmet need in the growing field of biopharmaceutical analysis.

INTRODUCTION

Recombinant proteins are designed and produced for a variety of reasons, most notably for use as therapeutic agents[1–3] and vaccine candidates.[4–8] Utilizing recombinant DNA technology and genetic engineering to produce a wide-range of proteins has been shown to be beneficial in the development of various pharmaceuticals, as demonstrated by pharmacological studies involving interferons,[2] reproductive hormones,[9] and monoclonal antibodies.[1, 10] More recently, the use of recombinant proteins has focused on the development of potential biopharmaceutical protein drugs that contain novel post-translational modifications (PTMs), in an effort to alter the solubility, efficacy, half-life, and in-vivo clearance rate in comparison to the characteristics of the corresponding native

*To whom correspondence should be addressed. University of Kansas, 2030 Becker Drive, Lawrence, KS 66047. Phone: 785-864-3015 hdesaire@ku.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

protein sequences.[1, 10, 11] For potential protein pharmaceuticals, full characterization including PTMs is critical in the drug development process.

Mass spectrometry (MS) is an important tool for protein identification[10, 12] and quantification,[11, 13–15] and is especially powerful in the analysis of post-translationally modified proteins. MS is perhaps the most commonly utilized technique for primary sequence characterization of recombinant proteins, as well as for the detection of PTMs.[2, 16] A common preparatory step prior to MS of proteins (recombinant or native) includes a protease digestion procedure, where the primary protein sequence is cleaved into peptides. These proteolized peptides typically retain their PTMs, thereby allowing MS and tandem MS experiments to be used to detect the modifications, while maintaining information about the site that was modified.[17, 18] For detection of all the different PTMs present in proteins, it is advantageous to achieve full protein sequence coverage.[19] Therefore, efficient protease digestions are crucial in order to achieve accurate characterization and full detection for peptides containing PTMs.[3, 20]

In order to develop optimized methods for MS analysis of peptides and PTMs on proteins, previous work has focused on several different stages of the protein preparation process ranging from evaluation of different types of mass spectrometers[21, 22] or separation methods,[20] to comparing specific aspects of a protease digestion procedure.[23–26] Inefficient protease digestion procedures inevitably result in poor mass spectrometry data, no matter how efficient the separation method or mass spectrometer parameters.[23, 25–28] In many instances, if a protein is not properly unfolded prior to addition of protease, the protease will not efficiently cleave the protein; therefore, MS data interpretation suffers because several peptides, consisting of different degrees of enzymatic mis-cleavage, would be present and diluted over multiple *m/z* values. Those peptides that are difficult to ionize will not be detected, leading to lower protein sequence coverage.[10, 19, 23, 27]

Proteolytic digestion methods consist of several procedural steps prior to the addition of an enzyme to cleave a protein into peptides, including: denaturation, reduction of disulfide bonds, and subsequent alkylation, or “capping,” of reduced cysteine residues. Previous studies have focused on each of the individual steps in the protease digestion process, and much of the optimization research has concentrated on denaturation.[14, 24, 25, 27, 29] Additionally, most of the recent emphasis has centered on the MS analysis of membrane proteins[27, 29] and cellular proteome elucidation.[24, 25] Analysis of cellular proteomes incorporates numerous membrane-bound proteins; thus, it is not surprising that researchers found ideal denaturants for these types of proteins to include MS friendly detergents, such as Rapigest™ SF, since detergents are known to aid in the solubilization of membrane-bound proteins.[29] It is unknown whether these conditions would be optimal for recombinantly expressed proteins, which are typically secreted, post-translationally modified, and not membrane-bound.

Due to the common presence of disulfide bonds present in many proteins, researchers also have focused on optimizing conditions for reduction[30, 31] and alkylation.[23, 32, 33] When reduction and alkylation are incomplete, a lower signal to noise ratio is often observed, as peaks may be present in the MS data that correspond to both derivatized and underivatized peptides.[15] Thus, peptides with already low ionization efficiencies, such as glycosylated peptides, may not be detected because splitting peptide ions over multiple *m/z* values can result in ion abundance too low for MS detection.[15] Moreover, over alkylation of peptides, or alkylation of the N-terminus or other amino acid side chains besides cysteine, can also occur when the alkylating agent is allowed to incubate with the sample for long periods of time, such as when the alkylating agent is not removed during the protease digestion step.[34]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Therefore, optimizing reduction and alkylation conditions is also essential for maximizing protein sequence coverage by MS.

The work described herein focuses on designing an ideal protease digestion protocol for readily soluble proteins containing both disulfide bonds and *N*-linked glycosylation, with a more specific goal of identifying reaction conditions yielding high protein sequence coverage, as well as effective detection of *N*-linked glycosylation. Multiple parameters in the protease digestion process were assessed by developing several different reaction conditions on a model protein for determination of the optimal digestion strategy. These optimal digestion conditions were applied for the MS analysis of a recombinant form of human lysyl oxidase-like 2 (hLOXL2). hLOXL2 has been shown to be a very important protein in the progression of breast and ovarian cancers, as well as having the potential to be a glycoprotein therapeutic drug as the primary sequence of hLOXL2 contains potential *N*-linked glycosylation sites.[35–39] hLOXL2 has not been isolated, and its extent of PTMs, including glycosylation, has not been examined. Therefore, analysis of hLOXL2 by mass spectrometry is necessary prior to drug development, especially for the detection of its PTMs, most notably glycosylation.

EXPERIMENTAL SECTION

Materials and Reagents

All reagents, except for hLOXL2, were purchased from common commercial sources. More details are in the Supplemental setion.

Glycoprotein protease digestion denatured with RapiGest™ SF

Human apo-transferrin (10 mg/mL) was dissolved in 0.1% RapiGest™ SF containing 50 mM NH_4HCO_3, pH 7.8 buffer. For reduction of disulfide bonds, either dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP) was added. Table 1 shows the concentrations and type of reducing agent added for each of the 7 different reaction conditions. Samples were incubated for 45 min. at 60 °C. Iodoacetamide (IAM) was added as the alkylating agent for 60 min. at room temperature in the dark. As shown in Table 1, reaction condition 3 contained a step where DTT was added to quench the alkylation reaction after IAM had incubated with the protein samples for 1 hr. in the dark. Trypsin was added at a 1:30 (w/w) enzyme:protein ratio, and all samples were incubated at 37 °C for 18 hrs. HCl was added to a final concentration of 50 mM to stop the tryptic digestion, as well as provide an acidic solution for RapiGest™ SF precipitation. Samples were re-incubated at 37 °C for an additional 45 min., then centrifuged to pellet out RapiGest™ SF. The supernatant was removed and stored at −20 °C until analysis with mass spectrometry.

Glycoprotein protease digestion denatured with urea

Urea (6 M) was added to glycoproteins (> 2 mg/mL), which had been dissolved in 50 mM NH_4HCO_3, pH 7.8. DTT or TCEP was added to reduce the disulfide bonds (see Table 1), and samples were incubated at room temperature for 1 hr before IAM was added to alkylate Cys residues by incubation in the dark for 1 hr. Reaction conditions 5 and 7 contained an additional step where DTT was added to quench the alkylation reaction after IAM had been allowed to incubate in the dark for 1 hr. The NH_4HCO_3 buffer was added to dilute the urea concentration to 1 M before the addition of trypsin, at a 1:30 (w/w) enzyme:protein ratio. Samples were then incubated at 37 °C for 18 hrs. To stop the trypsin reaction, 1 μL acetic acid was added per 100 μL of solution before storing samples at −20 °C until analysis with mass spectrometry.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Liquid chromatography/mass spectrometry

All samples were analyzed using the same LC-MS parameters, where an ESI-LTQ-FTICR-MS instrument was used for detection. More details are available in Supplemental Materials.

Data analysis

MS and MS/MS data acquired on the hybrid LTQ FTICR mass spectrometer were searched using Mascot (Matrix Science, London, UK, version 2.2.04) against the SwissProt database (v2011x). The peak list from Xcalibur raw files were extracted using DTA SuperCharge (version 1.19, http://msquant.sourceforge.net). The mgf files were searched using the following parameters- (a) enzyme: trypsin, (b) missed cleavage: 2, (c) fixed modification: carbamidomethyl, (d) variable modification: deamidation (NQ), Gln->pyro-Glu (N-term Q), Glu->pyro-Glu (N-term E), methionine oxidation, carbamidomethyl (N-term), carbamyl, (e) peptide tolerance of 1.0 Da, and (f) MS/MS tolerance of 0.8 Da. Search results with Mascot ion score threshold of 44 for transferrin and 33 for hLOXL2, indicate peptide identification at a 95% confidence level. To evaluate underalkylation of Cys residues, a second pass Mascot error tolerant search was performed. A list of all possible combinations of Cys modifications for peptides with more than one Cys was generated. Significant peptide matches were confirmed manually to eliminate false positives. Glycopeptides, could not be detected using Mascot. More details regarding detection and analysis of glycopeptides can be found in the supporting information.

RESULTS AND DISCUSSION

The goal of the study described herein is to determine ideally suited reaction conditions for proteolytic digestion of recombinant proteins and glycoproteins. To achieve the objectives, various reaction conditions were tested on a model protein, so that the optimal protein digest condition could be identified. Human apo-transferrin (transferrin) was chosen as the model glycoprotein, because transferrin possesses a high number of co- and post-translational modifications, including 19 disulfide bonds and two *N*-linked glycosylation sites. As validation that the conditions are highly effective, they are used to characterize an important pharmaceutical target, hLOXL2. Figure 1 shows the amino acid sequence for transferrin and hLOXL2.

Survey of various protease digestion conditions

Transferrin was subjected to six different reaction conditions (Table 1), and the resulting MS and MS/MS data were assessed for the detection of peptides. The different reaction conditions chosen allowed for the comparison of two different denaturants, RapiGest™ SF and urea, as well as the reducing agents, DTT and TCEP, and various concentrations of IAM. Additionally, the necessity of adding an extra procedural step where DTT was added to quench the excess IAM, preventing unwanted side reactions and complications in MS data analysis, was investigated.

In addition to simply detecting transferrin peptides in the MS and MS/MS data, other factors governing digestion efficiency were also assessed, including: 1) complete alkylation (or under alkylation) of the Cys residues, 2) alkylation on the N-terminus of peptides (over alkylation), and 3) incomplete detection of glycans. For an efficacious digestion, one would expect to obtain complete alkylation of Cys residues, no overalkylation on the N-terminus, and complete detection of the glycans present. MS and MS/MS analysis was performed to evaluate these factors for each of the different reaction conditions in an effort to determine optimal protease digestion conditions for transferrin. All LC-MS/MS data files were submitted to a Mascot search. Using the parameters described in the Experimental section, the nonglycosylated peptides were identified. From the list of the best peptide matches, we

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

determined the overall coverage along with the species that were either under- or over-alkylated. The data were also manually inspected to identify the glycopeptides and to provide additional assurance that the automated assignments were correct.

Figure 2 shows an example of some of the data. In Figure 2A, an HPLC chromatogram from condition 3 in Table 1 is shown. The highlighted region from 41–42 min corresponds to the retention time averaged for the high resolution mass spectrum shown in Figure 2B. The MS/MS data in Figure 2C resulted from the circled peak labeled in the mass spectrum (Figure 2B). These particular MS and MS/MS data correspond to a peptide that was not alkylated, indicating that the alkylation reaction was incomplete in this case. Thus, as shown in Figure 2B, the peptide is diluted between the fully alkylated peak (at *m/z* 1086) and the non-alkylated form (at *m/z* 1057). Fortunately, this particular peptide happened to ionize efficiently, so both the alkylated and non-alkylated forms were detected. However, if the ionization efficiency of this species had been lower, it would be likely that neither of these peaks would be detected in the MS/MS data, leading to lower protein sequence coverage.

Comparison of denaturing and reducing agents

The overall sequence coverage for each of the different conditions tested is shown in Table 2. Based on these data, urea is clearly the optimal denaturing agent for the model protein used in this study. In three out of the four reaction conditions where transferrin was subjected to denaturing by urea, high sequence coverage was obtained. The only reaction condition where a urea denatured sample did not achieve greater sequence coverage than samples utilizing Rapigest™ SF for denaturation was condition 2. The analogous reaction condition for condition 2 was condition 1, where the type and concentration of reducing agent, as well as concentration of alkylating agent were the same, and only the type of denaturing agent was different. Both of these analogous reaction conditions performed poorly and had the lowest sequence coverage compared to all other the reaction conditions tested. Therefore, it was not necessarily that urea did not perform well in condition 2 (the data were still better than condition 1 where Rapigest™ SF was utilized for denaturation), but most likely a poor combination of reagents and concentrations of reducing and alkylated agents led to the poor sequence coverage result in condition 2.

DTT and TCEP were chosen as competing reducing agents for the optimization study. As shown in Table 2, when DTT was used as the reducing agent, and the excess IAA was not quenched with additional DTT (such as in reactions 1 and 2) the coverage was poor. However, in reactions 3 and 5, where this quenching step was added to the DTT reactions, then DTT slightly out-performed TCEP, where no quenching step was employed. These results suggest two things: First, the second addition of DTT is absolutely necessary, when using DTT as the reducing agent. Second, as long as that quenching step is done, DTT performs comparably to TCEP. The data from the first six reactions were acquired initially, and based on the findings of these reactions, we hypothesized that the best set of conditions would involve using TCEP *and* an additional DTT quenching step. These results prompted us to develop a seventh set of conditions, where TCEP is used and quenched with additional DTT. This set of conditions led to the seventh set of data in Table 2, and it is described in more detail below. In short, the additional quenching step did provide a slight improvement in overall sequence coverage.

Under alkylation of Cys residues

Incomplete alkylation of Cys residues may occur when a protein is not entirely unfolded or when disulfide bonds are inefficiently reduced, thereby rendering the alkylating agent inaccessible to those residues.[15] The only protease digestion procedure where more than 10% of the Cys-containing peptides were detected as being underalkylated was condition 3,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

as shown in Table 1. Aside from condition 3, which is obviously not an optimal protocol, underalkylation was not a significant issue.

Over alkylation

Overalkylation occurs when alkylation is detected on the N-terminus of peptides, as opposed to being limited to Cys residues. Although alkylating agents are most selective to thiol groups, amines can also become reactive when given enough incubation time.[34] As no reagents were removed from the reaction mixtures during the protease digestion, there was ample time for IAM to react with the N-terminus of newly-formed peptides that were generated after cleavage by trypsin. Indeed, the results from Table 2 indicate that over alkylation was detected in all reaction conditions that lacked an additional step of adding DTT to quench alkylation of IAM, regardless of the identity of reducing or denaturing agent used. Therefore, the optimal protease digestion condition should incorporate a step to quench IAM after alkylation.

Evaluation of cysteine-containing peptides detected

Due to the many peptides that contained one or more Cys residues, and because under and over alkylation was being evaluated, the percentage of Cys-containing peptides detected was also evaluated. As shown in Table 2, conditions 5 and 7 led to the highest number of these peptides detected. In both these cases, there were no under alkylation or over alkylation products detected. This point further demonstrates that under and over alkylation leads to lower sequence coverage, and efforts should be taken to minimize these products.

Determination of the optimal conditions

As described above, it was determined that urea outperformed Rapigest™ SF as a denaturant, while TCEP and DTT performed comparably as a reducing agents (when an “IAM quenching step” was used with the DTT reactions). Of the initial six trials, reaction 5 had the highest overall sequence coverage, at nearly 75%, yet it did not appear that this reaction could be optimized any further, since no under alkylation or over alkylation products were detected. Condition 6, which contained both urea and TCEP for denaturation and reduction, respectively, was nearly as good as condition 5, in terms of its sequence coverage; yet the results from Condition 6 included several overalkylated peptides (and a few underalkylated ones). Based on these findings, the seventh reaction condition was developed with the same protease digestion procedure from Condition 6, but with an additional step after alkylation, where DTT was added to quench unreacted IAM (see Table 1 for full details). The transferrin peptides identified in the MS data from the Condition 7 digestion protocol are highlighted in Figure 1A. As described in Table 2, condition 7 had the greatest sequence coverage, and no under or over alkylation was detected. Thus, condition 7 was determined to be optimal for protease digestion.

Detection and analysis of glycopeptides

In addition to the criteria above, one key feature of an ideal digestion protocol is that it produces high coverage of the PTMs on the protein being analyzed. Therefore, in addition to checking for sequence coverage and alkylation state, the seven data sets were also searched for the known PTMs on transferrin, which contain two *N*-linked glycosylation sites. For a set of reaction conditions to be considered optimal, the glycopeptides detected in the transferrin MS data needed to encompass all the glycoforms described in the literature for this protein sample. The major glycoform known to be present on transferrin is an *N*-linked biantennary sialylated complex type glycan.[40] However, transferrin also has an *N*-linked triantennary sialylated complex type glycan present in lower abundance.[40] Therefore, the MS data were searched for both biantennary and triantennary sialylated complex type

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

glycopeptides in the data sets from all 7 reaction conditions. As shown in Table 2, biantennary *N*-linked sialylated glycopeptides were detected at both glycosylation sites in transferrin in data sets from all 7 reaction conditions. This was expected, since the biantennary glycans are the most abundant glycoforms in transferrin.[40] The only data set where the lesser abundant triantennary glycans were detected was from condition 7. The triantennary *N*-linked glycan details are listed in Supplementary Table 1. The presence of the less common glycoform detected only in the MS data for condition 7 was further confirmation that condition 7 was indeed the most optimal reaction protocol.

Application of an optimized proteolytic digest to hLOXL2

After identifying and validating an ideal set of reaction conditions for the model protein, transferrin, the final objective of this work was to demonstrate that these conditions were also effective in the analysis of a biologically important protein, highlighting the general utility of the method. Lysyl oxidase (LOX) is a secreted copper-containing amine oxidase, which forms reactive aldehydes by oxidizing the ε-amino group of lysine side chains in collagen and elastin. LOX contains a cross-linked quinone cofactor arising from the PTM of its lysyl and tyrosol residues, which are conserved across all LOX and lysyl oxidase-like (LOXL) proteins.[35, 41, 42] Research involving lysyl oxidase and lysyl oxidase-like proteins (LOXL, LOXL2, LOXL3 and LOXL4) has implicated that these enzymes participate in a variety of biological processes, including extracellular matrix stabilization, cellular growth and homeostasis.[41, 43, 44] Moreover, LOX family participants are attractive pharmacological targets, since dys-regulation of LOX has been found to correlate with numerous diseases and adverse physiological states, including cancer formation and metastasis, connective tissue disorders, neurodegenerative pathologies, and cardiovascular abnormalities.[36, 38, 39, 44] Specifically, LOXL2 has been shown to be involved in abnormal collagen deposition, tumor invasion, lymph node metastasis, and cancer progression in breast and ovarian cancers.[35–39] As such, characterization of LOXL2 is as essential step in assessing its viability as a pharmaceutical candidate of interest.

To confirm the sequence of the recombinant hLOXL2, as well as to detect its PTMs, specifically *N*-linked glycosylation, proteolytic digestion followed by mass spectrometry was performed. As described above, the optimal set of digestion conditions on the model protein transferrin was protocol 7 from Table 1. The protein, hLOXL2, was subjected to these conditions, followed by LC-MS and MS/MS analysis of the rendered peptides. A summary of the hLOXL2 amino acid residues detected is illustrated in Figure 1B. The overall results from the MS data analysis of hLOXL2 are described in Table 3, and high resolution MS data for each detected peptide are shown in Supplemental Table 1. The percent protein sequence coverage was 71 % and the percent Cys-containing peptides were 78 %. In fact, only two peptides containing Cys residues were not detected in the analysis of hLOXL2. The non-detected Cys-containing peptides are highlighted in red in Figure 1B. The first such peptide is very large, 72 amino acids in length. Additionally, the peptide has eleven acidic residues (D or E) and only one basic residue. With that in mind, it is unlikely that this peptide would ionize well in the mass range of the experiment (up *to m*/z 2000), as the lowest detectable charge state for this ion would be +4, producing an *m*/z of 1980. Even upon manual inspection of the data, an ion at this mass and charge state was not detected. Likewise, a +5 charged ion for this peptide, at *m/z* 1584 was also searched for manually, but could not be conclusively identified. Therefore, it seems most probable that the inability to detect this peptide was not due to the reaction conditions, but rather due to the fact that the peptide does not ionize in the mass range of the experiment. The second Cys-containing peptide, which was not detected, also was not expected to be observable by this MS experiment, because its tryptic peptide is only CR, which would have an *m/z* of 335. This value is outside the scan range used for this experiment.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

A few additional peptides (which do not contain Cys) were also not detected. These peptides include: LNGGR, NPYEGR, LLR, SR, VAEGHK, and YDGHR. All these undetected peptides are short, with 6 amino acid residues or fewer, as highlighted in red in Figure 1B. Of these peptides LLR and SR were not expected to be detected because their masses fall below the scan range (ie. < 500 Da). The other four peptides are five to six amino acid residues in length and could potentially be detected by MS with the scan range utilized. However, we expect that these peptides have low ionization efficiencies because of the presence of acidic amino acid residues (D and E) in three of the four undetected peptides. In summary, this work illustrates that the digestion conditions were clearly optimized in that most peptides were detected, and those that were not detected could reasonably be expected to suffer from poor ionization efficiency. Additionally, no under alkylated or over alkylated peptides were detected in the hLOXL2 sample, further supporting the conclusion that the digestion conditions were optimal for this sample.

Glycopeptide analysis of hLOXL2

The CID MS/MS data of glycopeptides, including the glycopeptides from hLOXL2, are distinct from data for peptides in that, unlike peptides, fully glycosylated peptides cannot be identified in an automated fashion using a Mascot search, but instead are typically characterized using other search tools. A full description of the search strategy for detecting the hLOX glycopeptides is available in the Supplemental Materials.

Figure 3 shows example data outlining the detection of a glycopeptide from one of the two glycosylation sites in hLOXL2. Figure 3A shows a high resolution mass spectrum, where the circled peak corresponded to the glycopeptide of interest. By zooming in, as shown in Figure 3A, the isotopic distribution can be seen and the monoisotopic *m/z* is determined and compared to the calculated *m/z*. In this case, the experimental mass error is 8.2 ppm, as shown in Supplemental Table 2. Figure 3B shows the MS/MS data from the precursor ion circled in Figure 3A. As can be elucidated in Figure 3B, there are losses of monosaccharide sugar residues present in glycopeptide data. These sugar losses help to identify the *N*-linked glycan present on the glycopeptide as a fucosylated *N*-linked glycan core ([Hex]3[HexNAc]2[Fuc]1, where Hex = hexose, HexNAc = *N*-acetylhexosamine, and Fuc = fucose). See Figure 3B. To further confirm this assignment, the MS/MS data are searched for a peak corresponding to the potential peptide plus one HexNAc residue, called the Y_1 ion. This ion is common to MS/MS data of glycopeptides acquired in positive ion mode.[45] When a peak corresponding to the Y_1 ion is detected in the data, and it correlates to the monosaccharide sugar losses, the candidate glycopeptide is further supported, as shown in Figure 3B. The predominant glycoform detected in the MS/MS data of hLOXL2 at both glycosylation sites was a fucosylated *N*-linked glycan core. This assignment is consistent with the fact that a fucosylated *N*-linked glycan core is known to be one of the most common glycoforms in insect cells and in proteins expressed in insect cell lines.[46] In addition to this glycoform, eight additional glycoforms, of lower abundance, were also detected for the peptide, HYHSMEVFTHYDLLNLNGTK; they are shown in Supplemental Table 2. Finally, a very small fraction of this peptide was detected as completely nonglycosylated, and this result was verified by an automated Mascot search. The other glycosylation site on the protein resides on the peptide, NGSLVWGMVCGQNWGIVEAMVVCR. For this peptide, only the single glycoform, containing Hex]3[HexNAc]2[Fuc]1 was detected. However, fact that more glycoforms were not detected is possibly because this site is also nonglycosylated. In one LC-MS run, the peptide was detected, in large abundance, as a nonglycosylated species, concomitant with an N-to-D conversion (deamidation) at the glycosylation site. This PTM was manually verified both by high resolution MS and MS/MS experiments and a Mascot search. The Mascot ion score for this assignment was: 84, with an EValue of $4*10^{-10}$. This PTM is also shown in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Supplemental Table 2. While this peptide was detected with high confidence in one LC-MS file, replicate experiments indicated that the species was not always detectable. Additional biological experiments are ongoing to confirm the relevance of this species.

CONCLUSIONS

Efficient proteolytic digestion of proteins for mass spectrometric analysis is a critical step in the process of protein characterization. The optimal protease digestion reaction contained 6 M urea for denaturation, 5 mM TCEP for reduction, 10 mM IAM for alkylation, and 10 mM DTT for quenching the alkylation reaction. As described in Table 2, these conditions illustrated the highest protein sequence coverage with no under or over alkylation detected in the MS data. These conditions were also the only conditions where triantennary *N*-linked glycopeptides were detected in the transferrin data. This is significant because other researchers have been able to detect the triantennary *N*-linked glycans in deglycosylated transferrin data.[40] Therefore, these glycoforms are expected to be detected as glycopeptides as well.

Using the optimal digestion protocol identified above, we have successfully analyzed a recombinant form of hLOXL2 expressed in insect cells by mass spectrometry. Protein sequence coverage was high, and all undetected peptides were either very large or very small and likely suffered from poor ionization. In addition, the major PTM on this protein, glycosylation, was detected at both putative glycosylation sites. The glycoforms identified were consistent with the glycoforms typically present from the cell line used.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge financial support from an NSF CAREER award (project number 0645120) to HD, an NSF Fellowship to KR and CW (DGE-0742523), and NSF CAREER award (MCB-0747377) and NIH (5R01GM079446-02) to MM.

REFERENCES

1. Beck A, Wagner-Rousset E, Bussat MC, Lokteff M, Klinguer-Hamour C, Haeuw JF, Goetsch L, Wurch T, Dorsselaer AV, Corvaia N. Current Pharmaceutical Biotechnology. 2008; 9:482–501. [PubMed: 19075687]
2. Liu YH, Wylie D, Zhao J, Cure R, Cutler C, Cannon-Carlson S, Yang X, Nagabhushan TL, Pramanik BN. Anal. Biochem. 2011; 408:105–117. [PubMed: 20807495]
3. Walsh G, Jefferis R. Nature Biotechnol. 2006; 24:1241–1252. [PubMed: 17033665]
4. Koff R. J. Parasitology. 2003; 33:517–523.
5. Madrid-Marina V, Torres-Poveda K, Lopez-Toledo G, Garcia-Carranca G. Achives of Medical Research. 2009; 40:471–477.
6. Go EP, Irungu J, Zhang Y, Dalpathado DS, Liao HX, Sutherland LL, Alam SM, Haynes BF, Desaire H. J. Proteome Res. 2008; 7:1660–1674. [PubMed: 18330979]
7. Go EP, Chang Q, Liao HL, Sutherland LL, Alam SM, Baynes BF, Desaire H. J. Proteome Res. 2009; 8:4231–4242. [PubMed: 19610667]
8. Barouch DH. Nature. 2008; 455:613–619. [PubMed: 18833271]
9. Thakur D, Rejtar T, Karger B, Washburn NJ, Bosques CJ, Gunay NS, Shriver Z, Venkataraman G. Anal. Chem. 2009; 81:8900–8907. [PubMed: 19817480]
10. Barnes CAS, Lim A. Mass Spectrometry Reviews. 2007; 26:370–388. [PubMed: 17410555]
11. Ezan E, Dubois M, Becher F. Analyst. 2009; 134:825–834. [PubMed: 19381370]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

12. Chen G, Mirza UA, Pramanik BN. Advances in Chromatography. 2009; 47:1–29.
13. Rebecchi KR, Wenke JL, Go EP, Desaire H. J. Am. Soc. Mass Spectrom. 2009; 20:1048–1059. [PubMed: 19278867]
14. Norrgran J, Williams TL, Woolfitt AR, Solano MI, Pirkle JL, Barr JR. Anal. Biochem. 2009; 393:48–55. [PubMed: 19501563]
15. Hamdan M, Righetti PG. Mass Spectrom. Reviews. 2002; 21:287–302.
16. Itoh S, Kawasaki N, Ohta M, Hayakawa T. J. Chromatography A. 2002; 978:141–152.
17. Dalpathado D, Desaire H. Analyst. 2008; 133:731–738. [PubMed: 18493671]
18. Wuhrer M, Catalina MI, Deelder AM, Hokke CH. J. Chromatography B. 2007; 849:115–128.
19. Meyer B, Papasotiriou DG, Karas M. Amino Acids. 2010; 41:291–310. [PubMed: 20625782]
20. Zhang Y, Go EP, Desaire H. Anal. Chem. 2008; 80:3144–3158. [PubMed: 18370425]
21. Second TP, Blethrow JD, Schwartz JC, Merrihew GE, MacCoss MJ, Swaney DL, Russell JD, Coon JJ, Zabrouskov V. Anal. Chem. 2009; 81:7757–7765. [PubMed: 19689114]
22. Garza S, Moini M. Anal. Chem. 2006; 78:7309–7316. [PubMed: 17037937]
23. Ren D, Julka S, Inerowicz HD, Regnier FE. Anal. Chem. 2004; 76:4522–4530. [PubMed: 15283597]
24. Chen EI, Cociorva D, Norris JL, Yates JR III. J. Proteome Res. 2007; 6:2529–2538. [PubMed: 17530876]
25. Arnold RJ, Hrncirova P, Annaiah K, Novotny MV. J. Proteome Res. 2004; 3:653–657. [PubMed: 15253449]
26. Strader MB, Tabb DL, Hervey WJ, Pan C, Hurst GB. Anal. Chem. 2006; 78:125–134. [PubMed: 16383319]
27. Yu YQ, Gilar M, Lee PJ, Bouvier ESP, Gebler JC. Anal. Chem. 2003; 75:6023–6028. [PubMed: 14588046]
28. Hervey WJ IV, Strader MB, Hurst GB. J. Proteome Res. 2007; 6:3054–3061. [PubMed: 17616116]
29. Masuda T, Tomita M, Ishihama Y. J. Proteome Res. 2008; 7:731–740. [PubMed: 18183947]
30. Getz EB, Xiao M, Chakrabarty T, Cooke R, Selvin PR. Anal. Biochem. 1999; 273:73–80. [PubMed: 10452801]
31. Cline DJ, Redding SE, Brohawn SG, Psathas JN, Schneider JP, Thorpe C. Biochemistry. 2004; 43:15195–15203. [PubMed: 15568811]
32. Jacobs JM, Mottaz HM, Yu LR, Anderson DJ, et al. J. Proteome Res. 2004; 3:68–75. [PubMed: 14998165]
33. Sechi S, Chait BT. Anal. Chem. 1998; 70:5150–5158. [PubMed: 9868912]
34. Boja ES, Fales HM. Anal. Chem. 2001; 73:3576–3582. [PubMed: 11510821]
35. Vadasz Z, Kessler O, Akiri G, Gengrinovitch S, Kagan HM, Baruch Y, Izhak OB, Neufeld G. J. Hepatology. 2005; 43:499–507.
36. Kirschmann DA, Seftor EA, Fong SFT, Nieva DRC, Sullivan CM, Edwards EM, Sommer P, Csiszar K, Hendrix MJC. Cancer Res. 2002; 62:4478–4483. [PubMed: 12154058]
37. Peng L, Ran YL, Hu H, Yu L, Liu Q, Zhou Z, Sun YM, Sun LC, Pan J, Sun LX, Zhao P, Yang ZH. Carcinogenesis. 2009; 30:1660–1669. [PubMed: 19625348]
38. Barker HE, Chang J, Cox TR, Lang G, Bird D, Nicolau M, Evans HR, Gartland A, Erler JT. Cancer Res. 2011; 71:1561–1572. [PubMed: 21233336]
39. Fong SFT, Dietzsch E, Fong KSK, Hollosi P, Asuncion L, He QP, Parker MI, Csiszar K. Genes, Chromosomes & Cancer. 2007; 46:644–655. [PubMed: 17394133]
40. Wada Y, Azadi P, Costello CE, Dell A, et al. Glycobiology. 2007; 17:411–422. [PubMed: 17223647]
41. Molnar J, Fong KSK, He QP, Hayashi K, Kim Y, Fong SFT, Fogelgren B, Szauter KM, Mink M, Csiszar K. Biochim et Biophys Acta. 2003; 1647:220, 224.
42. Seve S, Decitre M, Gleyzal C, Farjanel J, Sergeant A, Ricard-Blum S, Sommer P. Conn. Tiss. Res. 2002; 43:613–619.
43. Lucero HA, Kagan HM. Cell. Mol. Life Sci. 2006; 63:2304–2316. [PubMed: 16909208]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

44. Rodriguez C, Rodriguez-Sinovas A, Martinez-Gonzalez J. Drug News Perspect. 2008; 21:218–224. [PubMed: 18560621]

45. Segu ZM, Mechref Y. Rapid Comm. Mass Spectrom. 2010; 24:1217–1225.

46. Fabini G, Freilinger A, Altmann F, Wilson IBH. J. Biol. Chem. 2001; 276:28058–28067. [PubMed: 11382750]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(A) Transferrin

```
  1 MRLAVGALLV CAVLGLCLAV PDKTVRWCAV SEHEATKCQS FRDHMKSVIP
 51 SDGPSVACVK KASYLDCIRA IAANEADAVT LDAGLVYDAY LAPNNLKPVV
101 AEFYGSKEDP QTFYYAVAVV KKDSGFQMNQ LRGKKSCHTG LGRSAGWNIP
151 IGLLYCDLPE PRKPLEKAVA NFFSGSCAPC ADGTDFPQLC QLCPGCGCST
201 LNQYFGYSGA FKCLKDGAGD VAFVKHSTIF ENLANKADRD QYELLCLDNT
251 RKPVDEYKDC HLAQVPSHTV VARSMGGKED LIWELLNQAQ EHFGKDKSKE
301 FQLFSSPHGK DLLFKDSAHG FLKVPPRMDA KMYLGYEYVT AIRNLREGTC
351 PEAPTDECKP VKWCALSHHE RLKCDEWSVN SVGKIECVSA ETTEDCIAKI
401 MNGEADAMSL DGGFVYIAGK CGLVPVLAEN YNKSDNCEDT PEAGYFAIAV
451 VKKSASDLTW DNLKGKKSCH TAVGRTAGWN IPMGLLYNKI NHCRFDEFFS
501 EGCAPGSKKD SSLCKLCMGS GLNLCEPNNK EGYYGYTGAF RCLVEKGDVA
551 FVKHQTVPQN TGGKNPDPWA KNLNEKDYEL LCLDGTRKPV EEYANCHLAR
601 APNHAVVTRK DKEACVHKIL RQQQHLFGSN VTDCSGNFCL FRSETKDLLF
651 RDDTVCLAKL HDRNTYEKYL GEEYVKAVGN LRKCSTSSLL EACTFRRP
```

(B) hLOXL2

```
  1 RSPWPGVPTS MRLNGGRNPY EGRVEVLVER NGSLVWGMVC GQNWGIVEAM
 51 VVCRQLGLGF ASNAFQETWY WHGDVNSNKV VMSGVKCSGT ELSLAHCRHD
101 GEDVACPQGG VQYGAGVACS ETAPDLVLNA EMVQQTTYLE DRPMFMLQCA
151 MEENCLSASA AQTDPTTGYR RLLRFSSQIH NNGQSDFRPK NGRHAWIWHD
201 CHRHYHSMEV FTHYDLLNLN GTKVAEGHKA SFCLEDTECE GDIQKNYECA
251 NFGDQGITMG CWDMYRHDID CQWVDITDVP PGDYLFQVVI NPNFEVAESD
301 YSNNIMKCRS RYDGHRIWMY NCHIGGSFSE ETEKKFEHFS GLLNNQLSPQ
351 SAWSHPQFE
```

Figure 1.

(A) Amino acid sequence of human apo-transferrin. (B) Amino acid sequence of a recombinant hLOXL2. Black text indicates the signal peptide. Blue text signifies amino acids that were detected using the seventh set of digestion conditions (See Table 1). Red text illustrates amino acids that were not detected using these conditions.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

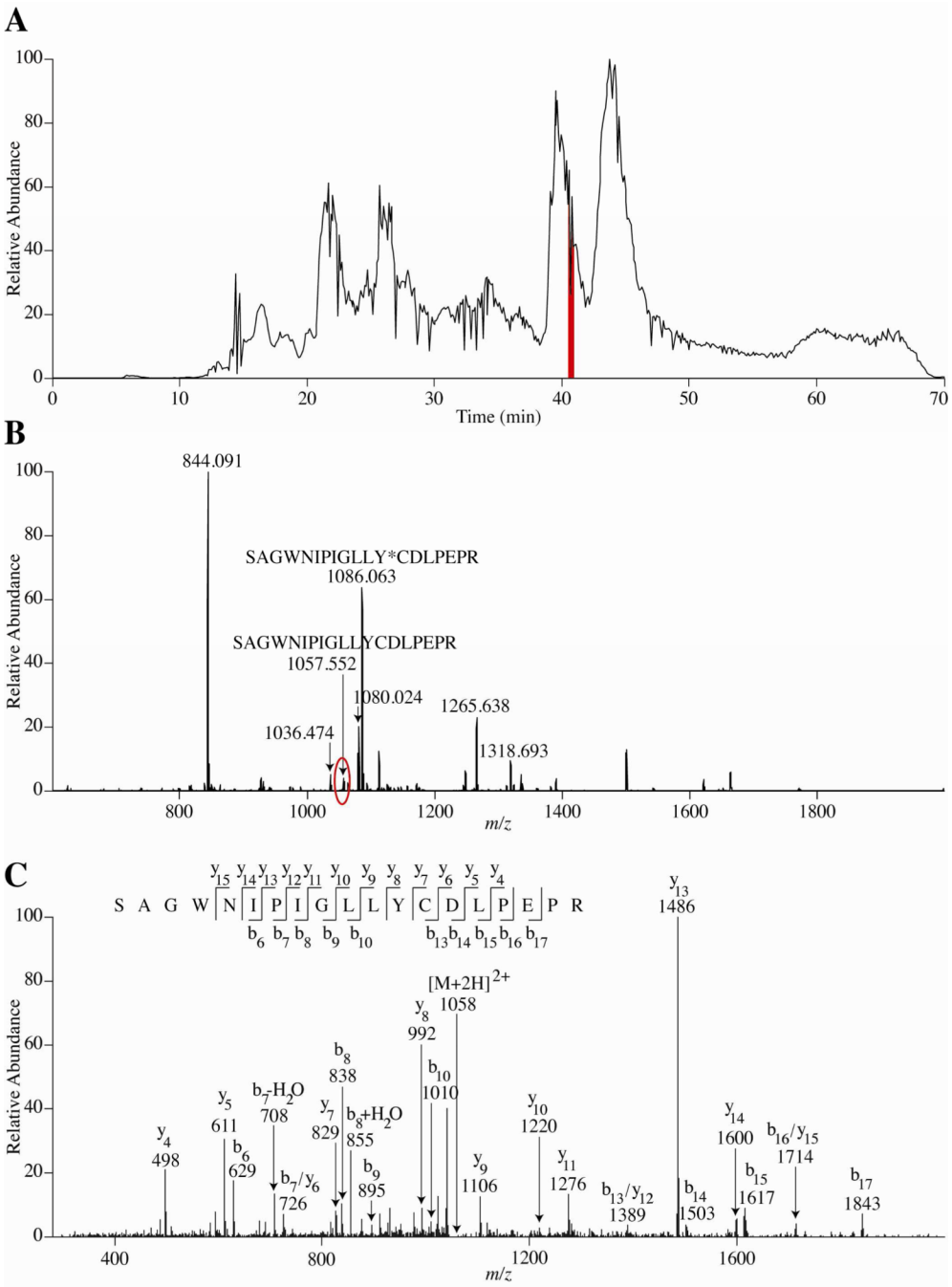

Figure 2.

Representative human apo-transferrin data from condition 3 (See Table 1). (A) Total ion chromatogram. Highlighted in red is the region (41–42 min.) where the MS data in (B) are shown. (B) The two labeled peptides illustrate both the fully and non-alkylated peptide SAGWNIPIGLLYCDLPEPR. The peak that is circled in red is the *m/z* where MS/MS was acquired as shown in (C). Additional labeled peaks indicate other detected human apo-transferrin peptides. The * indicates the alkylated Cys residue. (C) MS/MS data for the non-alkylated SAGWNIPIGLLYCDLPEPR with a retention time of 41.55 min.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

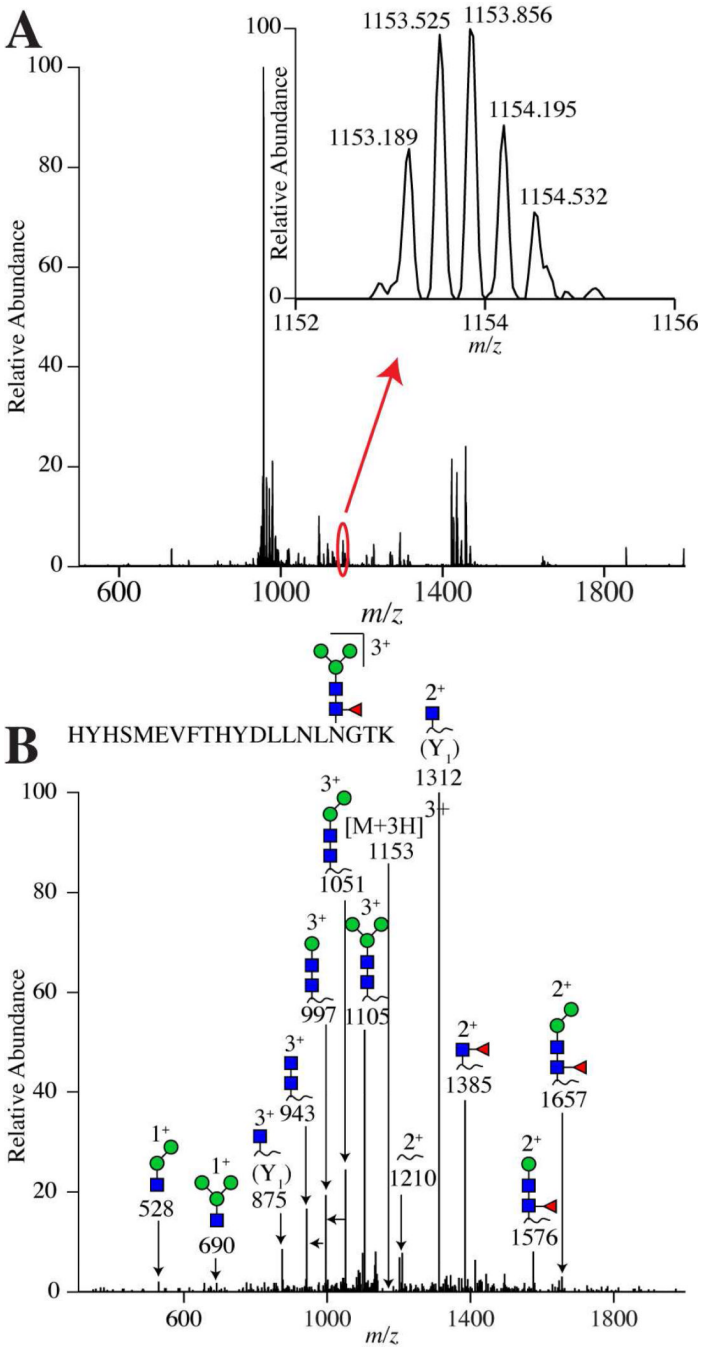

Figure 3.

(A) High resolution mass spectrum from hLOXL2 from the retention time 38–39 min. The peak circled in red illustrates where the hLOXL2 glycopeptide ion in (B) is located in the spectrum and the zoomed in region shows the isotopic distribution for the hLOXL2 glycopeptide ion, where the mass error can be calculated from the monoisotopic *m/z* value. (B) hLOXL2 glycopeptide MS/MS data at *m/z* 1153. The blue squares are *N*-acetylhexosamine, green circles are hexose, and the red triangle is fucose. The MS/MS data show losses of glycan residues that aid in determining the glycan composition.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Protease digestion preparation conditions for human apo-transferrin.

Reaction Conditions	Denaturing Agent	Reducing Agent	Alkylating Agent	Quenching Agent
#1[a]	0.1% RapiGest™	5 mM DTT	15 mM IAA	
#2[a]	6 M Urea	5 mM DTT	15 mM IAA	
#3[a]	0.1% RapiGest™	10 mM DTT	15 mM IAA	20 mM DTT
#4[a]	0.1% RapiGest™	5 mM TCEP	10 mM IAA	
#5[a]	6 M Urea	10 mM DTT	15 mM IAA	20 mM DTT
#6[a]	6 M Urea	5 mM TCEP	10 mM IAA	
#7[b]	6 M Urea	5 mM TCEP	10 mM IAA	10 mM DTT

[a]Initial reaction conditions tested.

[b]Reaction condition predicted as optimal after analysis of conditions 1 through 6.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Transferrin peptides detected for seven different protein digestion reaction conditions.

Reaction Conditions	[a]#1	[a]#2	[a]#3	[a]#4	[a]#5	[a]#6	[b]#7
Protein Coverage	51.1 %	62.0 %	69.8 %	63.8 %	74.6 %	70.8 %	75.9 %
Cys Containing Peptides Detected	47.4 %	63.2 %	66.7 %	69.2 %	76.3 %	68.4 %	76.3 %
Under Alkylation Detected	0.0 %	0.0 %	13.8 %	0.0 %	0.0 %	6.5 %	0.0 %
Over Alkylation Detected	34.8 %	38.5 %	3.4 %	33.3 %	0.0 %	9.7 %	0.0 %
Type of Glycans Detected at N^{432}	[c]Bi	Bi	Bi	Bi	Bi	Bi	Bi
Type of Glycans Detected at N^{630}	Bi	Bi	Bi	Bi	Bi	Bi	Bi/[d]Tri

[a]Initial reaction conditions tested.

[b]Reaction condition predicted as optimal after analysis of conditions 1 through 6.

[c]Bi = Biantennary *N*-linked Glycans.

[d]Tri = Triantennary *N*-linked Glycans.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

hLOXL2 peptides detected using optimal protease digest conditions

	hLOXL2
Protein Coverage	71.3%
Cys Containing Peptides Detected	77.8 %
Under Alkylation Detected	0.0 %
Over Alkylation Detected	0.0 %
Major Glycorform at N^{31}	[a][Hex]3[HexNAc]2[Fuc]1
Major Glycorform at N^{220}	[a][Hex]3[HexNAc]2[Fuc]1

[a]Hex = Hexose, HexNAc = N-acetyl hexosamine, Fuc = Fucose.