Talanta 85 (2011) 70–75

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Two-step protease digestion and glycopeptide capture approach for accurate glycosite identification and glycoprotein sequence coverage improvement

Yaohan Chen[a,b], Jing Cao[b], Guoquan Yan[a], Haojie Lu[a,b,*], Pengyuan Yang[a,b,*]

[a] Department of Chemistry, Fudan University, Shanghai 200433, PR China
[b] Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, PR China

ARTICLE INFO

Article history:
Received 25 December 2010
Received in revised form 5 March 2011
Accepted 12 March 2011
Available online 23 March 2011

Keywords:
Glycoproteomics
Glycopeptides
Hydrazide chemistry
Two-step digestion
Sequence coverage
N-glycosite

ABSTRACT

A novel two-step protease digestion and glycopeptide capture approach has been developed. It is different from traditional tryptic digestion, glycopeptide enriching and identification approach in glycoproteomics. Here, proteins were first digested by Lys-C into relatively large peptides. Glycopeptides among them were selectively captured by hydrazide resin through oxidized glycans. After thorough washing steps, trypsin was used as a second protease to in situ release non-glycosylated part (named as LT-peptides) from glycopeptides. Subsequently, the remaining part of glycopeptides on resin was de-glycosylated by peptide-N-glycosidase F, and collected as DG-peptides. Finally, both LT- and DG-peptides could be analyzed by mass spectrometer, achieving glycoprotein and glycosite identification. The approach was applied to cell lysate after positive validation by a model glycoprotein: 143 N-glycoproteins identified from DG- and LT-fraction both. In those glycoproteins, 189 DG-peptide-revealed N-glycosites got further confirmation by neighboring LT-peptides, which, in the meantime, made 109 glycoproteins get improved sequence coverage with increase even up to 350% (averagely 79.4%). Through controllable release, separate identification and combined interpretation of non-glycopeptides (newly introduced LT-peptides here) and traditional de-glycopeptides, the approach could not only achieve routine N-glycosite identification, but also provide further proofs of N-glycosites and increase glycoprotein sequence coverage.

1. Introduction

Post-translational modifications (PTMs) are ubiquitous in cells, and can regulate the function of proteins often by modulating their biophysical characteristics [1]. Protein glycosylation is one of the most common and complex PTMs. Glycan chains have important functions in many biological processes [2,3]. Their alterations usually indicate pathologic mechanisms [4,5]. As a new section of proteomics [6], glycoproteomic research has been rapidly developing in recent years. Owing to the development of different glycoprotein/glycopeptide enrichment methods and progressing of mass spectrometer (MS) techniques, high throughput glycosite identification has been achieved, especially in the field of N-glycosylation.

Currently, for enrichment of glycoproteins/glycopeptides, solid phase extraction approaches such as lectin affinity [7–9], hydrazide chemistry [10–12], hydrophilic affinity [13–15], and boronate affinity methods [16–18] are powerful tools. In these approaches, except that lectins are usually preferable in targeted glycosylation studies [19–21], mechanisms of all the other three methods are based on glycan hydroxyl groups, so that their common use is in broad glycoproteomics with less selectivity in glycan structures. However, due to the complexity and heterogeneity of glycosylation, none of these methods has been able to capture all glycoproteins/glycopeptides in complex biological samples yet.

N-glycopeptide capture-and-release using hydrazide chemistry and peptide-N-glycosidase F (PNGase F) is a popular approach. To the best of our knowledge, for practical sample applications, hydrazide chemistry shows much higher specificity than other methods, *e.g.* lectin enrichment [22] and hydrophilic separation [23]. Outstanding specificity of hydrazide chemistry (usually above 90%) means the capability to reduce sample complexity and also reliable identification. However, due to the high specificity, usually glycoprotein identification can only rely on a limited number of de-glycopeptides. Those identified glycoproteins have low sequence coverage, and most of them are single-peptide-hit identification likely [24]. On the other hand, a 0.98 Da mass increase

Abbreviations: LT-peptides, peptides produced first by Lys-C then by trypsin; DG-peptides, formerly glycosylated peptides released by PNGase F, namely de-glycopeptides.

* Corresponding authors at: Department of Chemistry, Fudan University, Shanghai 200433, PR China. Tel.: +86 21 54237961; fax: +86 21 54237961.
E-mail addresses: luhaojie@fudan.edu.cn (H. Lu), pyyang@fudan.edu.cn (P. Yang).

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.03.029

on Asn generated by PNGase F when releasing N-glycans from peptide backbones [25], is utilized as indirect evidence to locate N-glycosites. Although accurate mass measurements can be provided by MS, *e.g.* Fourier transform ion cyclotron resonance (FTICR), the small mass difference is still a potential problem causing false positive. In addition, automatic hydrolysis occurring naturally or experimentally may be another factor bring false positive. Replacing H_2O^{16} by H_2O^{18} may be a solution, but it is still possible that automatic hydrolysis could happen in such condition [26]. Unambiguous glycoprotein and glycosite identification is indeed challenged.

Combinations of different enzymes are helpful in glycoproteomic research, either in parallel [27–30] or sequential [31–33] ways. For example, Chen et al. [30] applied three different proteinases (trypsin, pepsin and thermolysin) in a parallel way, to get a more comprehensive result of liver tissue N-glycosite profile. Another work of ours had combined two proteinases, through two levels of enrichment, to achieve intact glycopeptide identification [31]. Bunkenborg et al. [32] separated Lys-C produced glycopeptides by lectin, and then obtained de-glycopeptides by PNGase F treatment; finally de-glycopeptides were digested by trypsin to generate appropriate size of peptides and achieve glycoprotein and glycosite identification. Nevertheless, resulting analyte itself was a mixture of de-glycopeptides, non-glycopeptides and non-specifically adsorbed peptides by lectin enriching method, so probably the analyte contents would interfere with each other in following analysis. Unlike lectins which are proteins themselves and cannot sustain in protease and glycosidase treatment, chemical capture approach especially the one with satisfactory specificity makes it possible to achieve in situ enzyme treatment, controllable release and efficient fractionation of non-glycopeptides and de-glycopeptides.

Herein, based on the high specificity of hydrazide chemistry and different specificities of endoproteinases, we have developed a novel two-step protease digestion and glycopeptide chemical capture approach. By applying the specificity of endoproteinase Lys-C (hydrolyzing proteins only at K), larger glycopeptides were provided for hydrazide chemistry to capture. From those resin-immobilized glycopeptides, LT-peptides (non-glycopeptides produced by Lys-C and then by trypsin) and DG-peptides (de-glycopeptides) could be controllably released by a second endoproteinase trypsin and glycosidase PNGase F respectively, and separately collected, so to avoid their interference and competition with each other in following liquid chromatography (LC)–MS/MS analysis. Besides routine DG-peptides, LT-peptides were informative and supplementary, either for increasing sequence coverage or obtaining more confirmed N-glycosite information by appearing neighboringly.

2. Material and methods

2.1. Chemicals and instrumentations

Water used in this experiment was prepared from a Milli-Q system (Millipore, Bedford, MA, USA). Bradford assay reagent, Readyprep™ 2-D cleanup kit, sodium periodate and hydrazide resin were obtained from Bio-Rad (Hercules, CA, USA). Endoproteinase Lys-C was from Wako (Osaka, Japan); sequencing grade modified trypsin was from Promega (Madison, WI, USA); peptide-N-glycosidase F (PNGase F, 500 U/μL) was from New England Biolabs (Ipswich, MA, USA); protease inhibitor cocktail used during cell lysis was from Pierce (Rockford, IL, USA). Iodoacetamide (IAA), dithiothreitol (DTT), ammonium bicarbonate (NH_4HCO_3), sodium sulphite, urea, thiourea, methanol, formic acid (FA) and acetonitrile (ACN) were purchased from Aldrich (Milwaukee, WI, USA).

4700 Proteomics Analyzer was from Applied Biosystems (Foster City, CA, USA). LTQ-Orbitrap XL system was from Thermo Fisher Scientific (Bremen, Germany).

2.2. Lys-C digestion of samples

Cell lysate (293 T) was collected as supernatant through adding lysis buffer (7 M urea, 2 M thiourea with cocktail) to cell pellet, disrupting by ultrasonic wave and centrifuging at 15,000 × g. Protein concentration of the lysate was measured by the Bradford assay. Reduction and alkylation of cysteine was carried out under 10 mM DTT (37 °C for 1 h) and 30 mM IAA (room temperature for 1 h in the dark), respectively. Then Readyprep™ 2-D cleanup kit was used for protein desalting. Final protein precipitate was resuspended in 50 mM NH_4HCO_3, and Lys-C was added as 1/50 of the protein by weight, incubated at 37 °C overnight. Resulting peptide mixture was dried in a vacuum centrifuge repeatedly to remove NH_4HCO_3.

2.3. Glycopeptide capture using hydrazide resin

Glycopeptides were captured using the method previously described [24] with minor modifications. Briefly, Lyophilized Lys-C digested peptides (~200 μg) were dissolved in a coupling buffer (100 mM sodium acetate, 150 mM NaCl, pH 5.5) at a concentration of 1 mg/100 μL. Sodium periodate at 10 mM final concentration was introduced into the peptide solution and agitated in the dark at room temperature for 1 h, to oxidize the *cis*-diol groups on glycans to aldehydes. Then sodium sulphite was added to the solution at a final concentration of 20 mM and incubated for another 10 min to quench the excess oxidant. The coupling reaction was initiated by introducing hydrazide resin (prepared by washing five times with coupling buffer) into the quenched peptide solution at about 10 mg/mL of resin, and extra coupling buffer was added to reach a solid to liquid ratio of 1:5. The coupling reaction was performed at 37 °C overnight under agitation. After that, the resin was washed three times thoroughly and successively with 1.5 M NaCl, methanol, and 50 mM NH_4HCO_3. Then resin was resuspended in 50 mM NH_4HCO_3 (a solid to liquid ratio of 1:3) for the second protease digestion.

2.4. In-situ tryptic digestion and PNGase F release of resin-immobilized glycopeptides

Trypsin (1:100, w/w) was added to the resin, and digestion was performed at 37 °C overnight. Supernatant (LT-peptides) was collected and dried in a vacuum centrifuge for later use.

After tryptic digestion and LT-peptide collection, resin was washed three times thoroughly and successively with 1.5 M NaCl, methanol, and 100 mM NH_4HCO_3. Finally, de-glycosylation was carried out at 37 °C overnight by PNGase F at a concentration of 1 μL of PNGase F per 1 mg of crude proteins in 100 mM NH_4HCO_3. The supernatant containing DG-peptides was collected, combined with the supernatant of an 80% ACN wash and dried in a vacuum centrifuge for later use.

2.5. Mass spectrometry analysis

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometric analysis was performed on 4700 Proteomics Analyzer in positive ion detection, equipped with an ND-YAG 355 nm laser. 8 mg/mL CHCA (α-cyano-p-hydroxycinnamic acid) dissolved in 70% ACN (v/v) containing 0.1% TFA was used as matrix. MS/MS fragmentation was achieved by collision induced dissociation (CID), using air as the collision gas. For automated acquirement of 4700 MALDI mass spectra (100 shots/spot, 100 shots/sub-spectrum), 1000 laser shots were averaged to acquire one MALDI mass

spectrum and 3500 laser shots were averaged to acquire one MALDI-MS/MS spectrum.

Electrospray ionization (ESI) mass spectrometric analysis was carried out by LTQ-Orbitrap XL Systems. Lyophilized tryptic peptides were resuspended in 5% ACN IN 0.1% FA, separated by reverse-phase analytical column (200 mm × 75 μm) packed with a 1.7 μm of bridged-ethyl-hybrid (BEH) C_{18} material (Waters Corporation, Milford, USA) and analyzed by on-line electrospray tandem mass spectrometry in positive mode. Samples were injected onto the trap-column with a flow of 20 μL/min and subsequently eluted with a gradient of 5–45% solvent B (95% ACN in 0.1% FA) over 100 min, and then injected into the mass spectrometer at a constant column-tip flow rate of 500 nL/min. Eluted peptides were analyzed by MS and data-dependent CID MS/MS acquisition, selecting the eight most abundant precursor ions for MS/MS with a dynamic exclusion duration of 60 s. All tandem mass spectra were collected using normalized collision energy of 35%, using helium as the collision gas.

2.6. Data exploration

Data from MALDI and ESI MS/MS analysis was searched against Swiss-Prot database by MASCOT and SEQUEST, respectively. Parameters were as follows: enzyme, trypsin (partially enzymatic); maximum missed cleavages (MCs), two; fixed modification, cysteine (C, 57.02150); variable modifications, oxidation (M, 15.99492) and asparagine de-glycosylation (N, 0.98402); mass values, monoisotopic. Precursor mass and fragment mass tolerance was 150 ppm and ±0.6 Da for MASCOT search, 10 ppm and ±1 Da for SEQUEST search. To further validate results obtained from SEQUEST, Trans-Proteomic Pipeline (TPP) was used [34]. Database search results were statistically analyzed using PeptideProphet [35]. By building a well-established non-parametric model using scores and other properties in the search results, PeptideProphet would give high-confidence spectrum-to-peptide interpretation (score > 0.9) [36]. Here only those peptides passed the peptide probability threshold 0.9 were accepted for further data interpretation.

Although the majority of N-linked glycosylation occurs at a consensus NXS/T sequon (X ≠ P) [25], about 3.3% of our identified Asn modified DG-peptides did not contain such a sequon. It is likely resulted from false positive database search, nonspecific adsorption by the resin, and isolation of atypical N-linked glycosites (*i.e.* not the NXS/T motif) which needs further validation. To focus on those N-linked glycosites we could be most confident about, only those sequon-containing DG-peptides were preserved. Because hydrazide enrichment could be applied to all glycan containing hydroxyl groups suitably oxidized, it is predictable that LT-peptides could be released from other glycopeptides rather than N-glycopeptides. Since PNGase F is only suitable for N-glycosylation, to clearly and reliably interpret the advantages of the approach, analysis will presently major focus on N-glycoproteins identified by DG-peptides and also identified from LT-peptides. Recognition of LT-peptides which were related to DG-peptides in the same large glycopeptides was carried out allowing two missed-cleavages of Lys-C digestion.

3. Results and discussion

3.1. Principle of the novel two-step protease digestion and glycopeptide capture approach

Glycopeptide capture, especially when it gets high selectivity (*e.g.* hydrazide chemistry), usually obtains relatively low sequence coverage, and lowers consequently the confidence of protein assignment. Thus it is necessary to improve the shortcoming while making use of its effectiveness on complexity reduction of sample and high specificity for glycopeptide separation. In our approach as shown in Fig. 1, before glycopeptide oxidization and coupling, Lys-C (cutting at K) was used as the first protease to digest proteins into relatively larger peptides. The larger peptides, containing oxidized sugar groups, were captured by hydrazide resin. After glycopeptide coupling and washing steps, trypsin (cutting at R here) was used to cut off small peptides (named as LT-peptides, in the dashed circle) from Lys-C produced glycopeptides captured by the solid phase. The LT-peptide supernatant was collected and analyzed by LC–MS/MS, the way as the later PNGase F released glycopeptides (DG-peptides, in the dashed rectangle) did.

According to the existing frequencies of K (5.73%) and R (5.63%) residues from Swiss-Prot human protein database, statistically, a Lys-C-derived peptide would contain an R residue and could be digested by trypsin to generate another smaller peptides. Thus, when Lys-C was chosen as the first protease to generate relatively large glycopeptides for capture, subsequent in situ tryptic digestion would generate an appropriate amount of LT-peptides. Meanwhile, owing to the high specificity (95.5% in our experiment, Appendix B, Sheet B.1) and protease-enduring of hydrazide chemistry method, we could reduced the non-specific adsorption to a low level, so the newly introduced LT-peptides could be produced in a similar complexity as DG-peptides, separately collected and reliably used. Combining sequences of both LT- and DG-peptides, the overall protein sequence coverage could be improved. Additionally, N-glycosites were given further proofs because LT-peptides could indicate the existences of adjacent glycans coupling to the resin.

The two-step protease digestion approach was tested by standard glycoprotein (Table 1) and applied to analysis of unknown glycoproteins in cell lysate. By incorporating LT-peptides identification result, 109 of 143 identified N-glycoproteins obtained notably sequence coverage increase (an average increase of 79.4%; Appendix A, Table A.1) and 189 of the identified N-glycosites were further confirmed, after careful interpretation of LT-peptides related information.

3.2. Validation of the approach by model glycoprotein

Asialofetuin (ASF), a representative and well characterized glycoprotein with three N-glycosites and four O-glycosites, was used to test the approach, to see whether LT-peptides would appear as we designed. The LT- and DG-peptide fractions were analyzed by MALDI MS and searched against database. Results were listed in Table 1. Both LT- and DG-peptides detected were in our expectation majorly and played their roles.

Routine N-glycosite identification of N_{99}, N_{156}, and N_{176} were achieved by four DG-peptides. Three LT-peptides were also detected: two of them matched with DG-peptides from parent peptides generated by Lys-C digestion and provided further confident glycosylation information for corresponding N-glycosites; another LT-peptide turned to be adjacent to four recorded O-glycosites, as annotated in Swiss-Prot database. Because PNGase F does not work on O-glycans, it is reasonable that no DG-peptide proof could be found for O-glycosylation. Hydrazide chemistry approach can be applied to glycans if only *cis*-diol exists, thus the existence of LT-peptide might be useful clue for O-glycosylation.

In summary, after incorporating LT-peptide result, a sequence coverage increase from 12% to 23.4% for ASF identification was achieved, N-glycosite identification benefited from double confirmation, and LT-peptides even dropped a hint on O-glycosylation.

3.3. Validation of the approach by cell lysate

To assess the applicability of this novel approach, we used the total lysate of 293 T cells as a testing sample with 1D

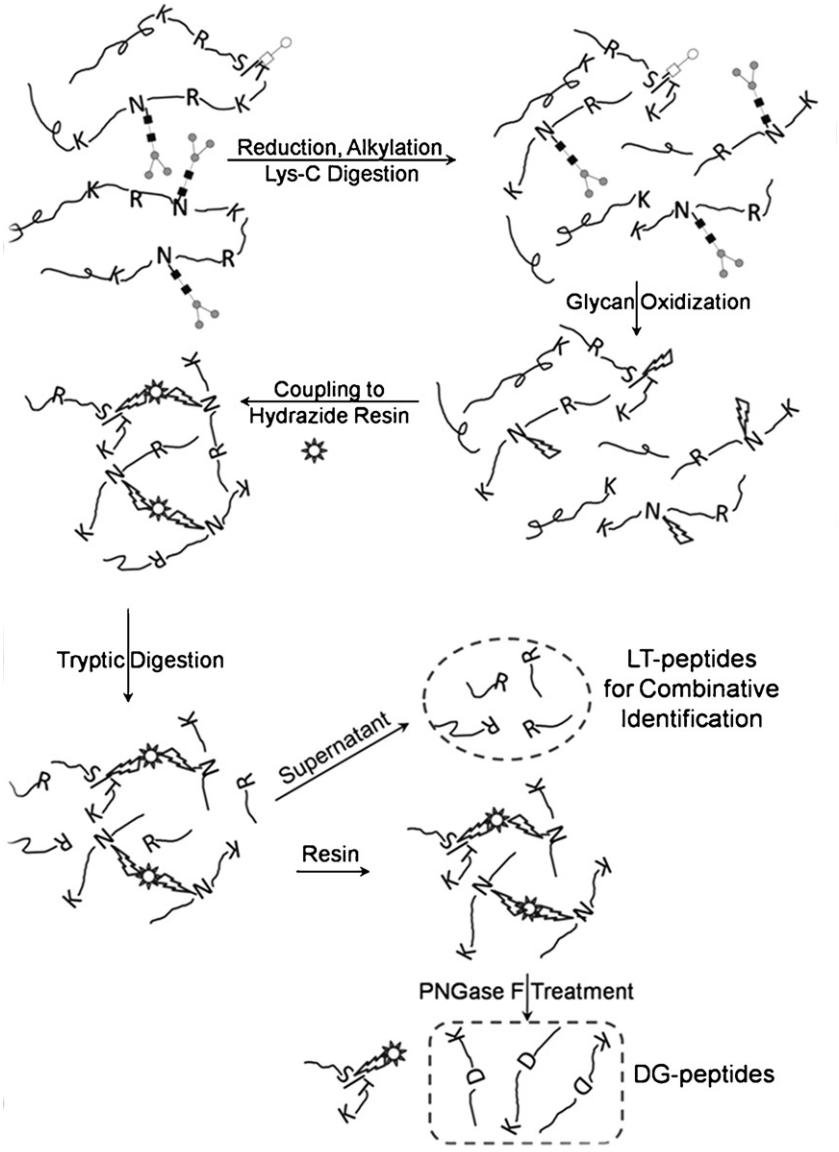

Fig. 1. Workflow of the two-step protease digestion and glycopeptide capture approach.

LC/ESI-MS/MS analysis carried out on LTQ-Orbitrap XL system. The system can provide accurate mass and high resolution similar to those achievable with FTICR instrumentation [37]. Finally, 2386 spectrum identification from DG-fraction and 1234 spectrum identification from LT-fraction resulted in 143 N-glycoproteins identification (Appendix B, Sheet B.1 and B.2).

On one hand, as expected, DG-fraction presented a consistently high specificity: 95.5% identifications contained N modification on NXS/T (X ≠ P) sequon (assigned to 359 N-glycosites). On the other hand, we evaluated whether LT-peptides would specifically present, too. After careful data interpretation, 932 of them (over 75% of all LT-peptides detected in 143 glycoproteins) were found

Table 1
List of ASF peptides identified by MALDI mass spectrometry analysis applying the two-step approach.

Detected Peptides	Position	Observed Mr	Expected Mr	Amino acid sequence
DG-Peptides (Sequence Coverage 12.0%)	91–103	1442.8864	1441.8791	VLDPTPLAN_{99}#CSVR
	94–103	1115.6938	1114.6865	PTPLAN_{99}#CSVR
	145–159	1741.9969	1740.9896	LCPDCPLLAPLN_{156}#DSR
	173–187	1666.9646	1665.9573	AESN_{176}#GSYLQLVEISR
LT-Peptides (Sequence Coverage 11.4%)	104–120	1978.1071	1977.0998	QQTQHAVEGDCDIHVLK (N_{99})
	196–211	1667.9131	1666.9058	SVSVEFAVAATDCIAK (N_{156}, N_{176})
	238–245	816.4941	815.4868	ALGGEDVR
Other	29–45	1774.0007	1772.9934	EPACDDPDTEQAALAAV

#N-glycosites detected by recognizing 0.98 Da mass shifts in NXS/T sequons; ($N_{num.}$) after the LT-peptides sequences represents N-glycosites from the same Lys-C generated glycopeptides as the LT-peptides. N_{156} and N_{176} were two N-glycosites from the same Lys-C produced glycopeptide. "Other" peptide seemed to be an unspecific adsorption, detected in both LT- and DG-peptide fraction.

Table 2
Assorting of all 1234 identification of LT-peptides.

LT-peptides	Missed-cleavage (MC) Num.	Num. of identification	Percentage	
Within two MCs of identified N-glycosites	0	800	75.5%	91.5%
	1 and 2	132		
Within two MCs of undetected NXS/T sequon (X ≠ P)	0	181	16.0%	
	1 and 2	17		
Other	/	104	8.4%	
Sum	/	1234	100%	

just locating within two missed-cleavages of DG-peptide-revealed N-glycosites (Table 2). Despite that over three quarters of LT-peptides turned to be as our design, it seemed that LT-fraction was not as specific as DG-fraction, because the latter achieved a specificity of 95.5%. However, after attention was paid to the location of those "unspecific" LT-peptides, 198 of them were found appearing neighboring certain undetected NXS/T sequon. Actually, for either 932 direct or 198 indirect proved LT-peptides, the percentage of zero missed-cleavage was both dominant (86%, 91%) and similar with other routine identification, *e.g.* DG-fraction (89%), (Appendix A, Table A.2). Thus we could re-evaluate the specificity of our LT-peptide data as 91.5% (Table 2), which was comparable with DG-peptides and consistent with the enriching method.

Fig. 2 is an example of DG- and LT-peptide combination of the approach. Glycoprotein Q07954 was identified from ten traditional DG-peptides (underlined, bearing ten N-glycosites) and seven LT-peptides (bolded and bolded italic). Subsequent interpretation of LT-peptides showed four (bolded italic) located in the same Lys-C-produced peptides as corresponding DG-peptides, so that four N-glycosites got further proofs. Another three LT-peptides (bolded) just located within zero missed-cleavage of several undetected NXS/T sequon, all of which were annotated as "potential" N-glycosites in Swiss-Prot database. It is reasonable some LT-peptides detected without corresponding DG-peptides, because there may be N-glycosites insensitive to PNGase F or DG-peptides unsuitable for MS detection, which need further validation. Nevertheless, to focus on the one we could be most confident about,

sp|Q07954|LRP1_HUMAN

...K**NEPVDRPPVLLIANSQNILATYLSGAQVSTITPTSTR**...N_{239}ET...K...

...K...VNRFN_{446}STEYQVVTR...K...

...K...N_{1050}CT...**LDGLCIPLR**...K...

...KWTGHN_{1511}VTVVQR***TNTQPFDLQVYHPSR***...K...

...KLYWISSGN_{1763}HTINR...K...

...K***ADGSGSVVLR***N_{1825}STTLVMHMK...

...KDN_{2127}ATDSVPLR***TGIGVQLK***...

...K...N_{2472}DT...N_{2502}CS...**ILQDDLTCR**...N_{2521}SS...K...

...KLNLDGSN_{3048}YTLLK...

...KLTSCATN_{3788}ASICGDEAR...K...

...K...FGTCSQLCN_{3839}NTK...

...K...GVTHLN_{3953}ISGLK...

...K***LSVIGSIR***LN_{4075}GTDPIVAADSK...

Fig. 2. A glycoprotein (Q07954) identified by both DG-peptides (underlined) and LT-peptides (bolded and bolded italic).
Theoretical Lys-C generated peptides, which were related to LT- and DG-peptides, were ranked in sequence. For conciseness, suspension points were used to represent for unshown amino acids; K residues were always shown.

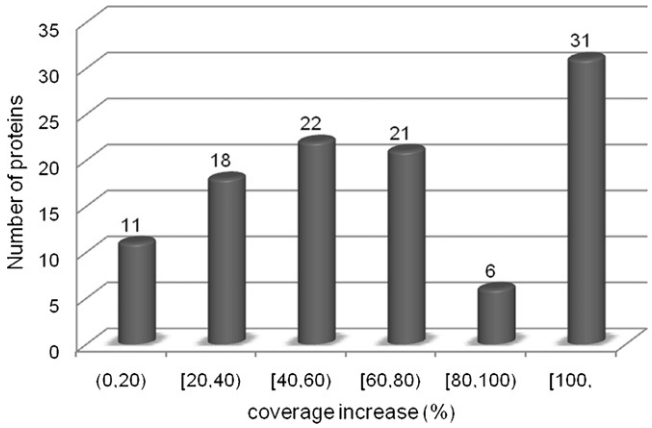

Fig. 3. Distribution of proteins realizing different extent of sequence coverage increase.
Horizontal axis represents sequence coverage increase (%) calculated as LT-peptide-coverage/DG-peptide-coverage × 100%. Vertical axis represents the numbers of proteins in different extent of increase intervals.

following analysis we presently showed was generated from 932 LT-peptides with direct DG-peptide proofs.

By applying this approach, a notable sequence coverage increase was obtained for glycoprotein identification from cell lysate, without sacrificing the high specificity. Based on previous studies, glycopeptides are usually present in relatively low abundance (2–5%) in tryptic peptide mixtures of glycoproteins [38]. In methodological researches, very high enriching specificity for glycopeptides is normally a desirable and exact goal. Nevertheless, high specificity might result in quite low sequence coverage for glycoprotein identification. After incorporating the newly introduced LT-peptides, 109 glycoproteins obtained improved sequence coverage with increase up to 350%; about one-third proteins doubled their coverage or even more; a 79.4% increase in average was got. Different extent of sequence coverage increase was summarized in Fig. 3 with details listed in Table A.1. Particularly, a part of the glycoproteins were only identified by single DG-peptide hit (*e.g.* 31 protein IDs marked by asterisk in Table A.1). In combination of their LT-peptides, an average increase of 111% for sequence coverage was obtained. Moreover, LT-peptides played another role besides coverage increase. Their appearances further confirmed the existence of former glycosylation which happened on the same Lys-C-generated glycopeptides with them. Finally, 189 DG-peptide-identified N-glycosites in 109 glycoproteins were further confirmed because of their neighboring LT-peptides (Appendix B, Sheet B.3). Thus assignment confidence was enhanced.

4. Conclusions

Proteins can be identified by individual signature proteolytic peptides with MS and identification from multiple peptides improves the confidence of protein assignment [39], so that it is ideal to use multiple peptides to identify a protein. We novelly introduced two-step protease digestion into a popular

glycopeptide capture approach. Through controllably releasing non-glycopeptides and de-glycopeptides, high specificity was maintained, while low glycoprotein sequence coverage was improved and more confirmed glycosite identification was achieved. In future works, other suitable combination of proteases as releasing reagents may be considered if necessary. We believe this approach could be useful when applied to various high throughput glycoproteomic researches.

Competing interests

No competing interests are declared.

Acknowledgments

The work was supported by National Science and Technology Key Project of China (2007CB914100, 2009CB825607, and 2010CB912704), National Natural Science Foundation of China (20875016, 31070732, 21025519 and 20975024), MOE of China (20080246011), Shanghai Projects (Shuguang, Eastern Scholar and B109).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.talanta.2011.03.029.

References

[1] D. Shental-Bechor, Y. Levy, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 8256–8261.
[2] R.A. Dwek, Chem. Rev. 96 (1996) 683–720.
[3] P.M. Rudd, R.A. Dwek, Biochem. Mol. Biol. 32 (1997) 1–100.
[4] M.A. Daniels, K.A. Hogquist, S.C. Jameson, Nat. Immunol. 3 (2002) 903–910.
[5] M. Dwek, H. Ross, A. Leathem, Proteomics 1 (2001) 756–763.
[6] N. Packer, Proteomics 6 (2006) 6121–6123.
[7] J. Zhao, W. Qiu, D.M. Simeone, D.M. Lubman, J. Proteome Res. 6 (2007) 1126–1138.
[8] J. Zhao, T.H. Patwa, W. Qiu, K. Shedden, R. Hinderer, D.E. Misek, M.A. Anderson, D.M. Simeone, D.M. Lubman, J. Proteome Res. 6 (2007) 1864–1874.
[9] Y.H. Qiu, T.H. Patwa, L. Xu, K. Shedden, D.E. Misek, M. Tuck, G. Jin, M.T. Ruffin, D.K. Turgeon, S. Synal, R. Bresalier, N. Marcon, D.E. Brenner, D.M. Lubman, J. Proteome Res. 7 (2008) 1693–1703.
[10] H. Zhang, X.J. Li, D.B. Martin, R. Aebersold, Nat. Biotechnol. 21 (2003) 660–666.
[11] T. Liu, W.J. Qian, M.A. Gritsenko, D.G. CampII, M.E. Monroe, R.J. Moore, R.D. Smith, J. Proteome Res. 4 (2005) 2070–2080.
[12] P. Ramachandran, P. Boontheung, Y.M. Xie, M. Sondej, D.T. Wong, J.A. Loo, J. Proteome Res. 5 (2006) 1493–1503.
[13] P. Hägglund, J. Bunkenborg, F. Elortza, O.N. Jensen, P. Roepstorff, J. Proteome Res. 3 (2004) 556–566.
[14] S. Sekiya, Y. Wada, K. Tanaka, Anal. Chem. 77 (2005) 4962–4968.
[15] Y. Zhang, E.P. Go, H. Desaire, Anal. Chem. 80 (2008) 3144–3158.
[16] K. Sparbier, T. Wenzel, M. Kostrzewa, J. Chromatogr. B 840 (2006) 29–36.
[17] K. Sparbier, A. Asperger, A. Resemann, I. Kessler, S. Koch, T. Wenzel, G. Stein, L. Vorwerg, D. Suckau, M. Kostrzewa, J. Biomol. Tech. 18 (2007) 252–258.
[18] Q.B. Zhang, N. Tang, J.W.C. Brock, H.M. Mottaz, J.M. Ames, J.W. Baynes, R.D. Smith, T.O. Metz, J. Proteome Res. 6 (2007) 2323–2330.
[19] Z. Dai, J. Fan, Y.K. Liu, J. Zhou, D.S. Bai, C.J. Tan, K. Guo, Y. Zhang, Y. Zhao, P.Y. Yang, Electrophoresis 28 (2007) 4382–4391.
[20] J. Zhao, D.M. Simeone, D. Heidt, M.A. Anderson, D.M. Lubman, J. Proteome Res. 5 (2006) 1792–1802.
[21] T. Nakagawa, N. Uozumi, M. Nakano, Y. Mizuno-Horikawa, N. Okuyama, T. Taguchi, J. Gu, A. Kondo, N. Taniguchi, E. Miyoshi, J. Biol. Chem. 281 (2006) 29797–29806.
[22] C.A. McDonald, J.Y. Yang, V. Marathe, T.Y. Yen, B.A. Macher, Mol. Cell. Proteomics 8 (2009) 287–301.
[23] J. Cao, C.P. Shen, H. Wang, H.L. Shen, Y.H. Chen, A.Y. Nie, G.Q. Yan, H.J. Lu, Y.K. Liu, P.Y. Yang, J. Proteome Res. 8 (2009) 662–672.
[24] B. Sun, J.A. Ranish, A.G. Utleg, J.T. White, X. Yan, B. Lin, L. Hood, Mol. Cell. Proteomics 6 (2007) 141–149.
[25] E. Bause, Biochem. J. 209 (1983) 331–336.
[26] P.M. Angel, J.M. Lim, L. Wells, C. Bergmann, R. Orlando, Rapid Commun. Mass Spectrom. 21 (2007) 674–682.
[27] M. Kvaratskhelia, P.K. Clark, S. Hess, D.C. Melder, M.J. Federspiel, S.H. Hughes, Virology 326 (2004) 171–181.
[28] R.J. Chalkley, A.L. Burlingame, Mol. Cell. Proteomics 2 (2003) 182–190.
[29] S. Itoh, N. Kawasaki, A. Harazono, N. Hashii, Y. Matsuishi, T. Kawanishi, T. Hayakawa, J. Chromatogr. A 1094 (2005) 105–117.
[30] R. Chen, X.N. Jiang, D.G. Sun, G.H. Han, F.J. Wang, M.L. Ye, L.M. Wang, H.F. Zou, J. Proteome Res. 8 (2009) 651–661.
[31] Y.H. Chen, M.Q. Liu, G.Q. Yan, H.J. Lu, P.Y. Yang, Mol. BioSyst. 6 (2010) 2417–2422.
[32] J. Bunkenborg, B.J. Pilch, A.V. Podtelejnikov, J.R. Wiśniewski, Proteomics 4 (2004) 454–465.
[33] M.R. Larsen, P. Højrup, P. Roepstorff, Mol. Cell. Proteomics 4 (2005) 107–119.
[34] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Anal. Chem. 74 (2002) 5383–5392.
[35] J.E. Elias, S.P. Gygi, Nat. Methods 4 (2007) 207–214.
[36] J. Malmström, M. Beck, A. Schmidt, V. Lange, E.W. Deutsch, R. Aebersold, Nature 460 (2009) 762–766.
[37] M. Scigelova, A. Makarov, Proteomics 6 (S2) (2006) 16–21.
[38] G. Alvarez-Manilla, J. Atwood III, Y. Guo, N.L. Warren, R. Orlando, M. Pierce, J. Proteome Res. 5 (2006) 701–708.
[39] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, Anal. Chem. 75 (2003) 4646–4658.