Differentiate Minor Difference of Protein Structure in Biosimilar and Reference Products Using High-Resolution Orbitrap LC-MS/MS

Zhiqi Hao[1], David M. Horn,[1] Shiaw-Lin Wu,[2] Fan Zhang[2] and Patrick Bennett[1]

[1]Thermo Fisher Scientific, San Jose, CA [2]Barnett Institute, Northeastern University, Boston, MA

Overview

Purpose: To analyze difference of protein structure in biosimilar and reference products using a Thermo Scientific™ bench-top Orbitrap™ LC-MS/MS.

Methods: A top-ten data-dependent high-energy collision dissociation (HCD) method was performed to analyze the samples using a bench-top Orbitrap™ mass spectrometer. Data was analyzed using a new software that is under development.

Results: An LC-MS/MS workflow was developed for differentiating minor difference of protein structure in biosimilar and reference products using a benchtop Orbitrap LC-MS/MS and a new software that is under development. This workflow provides qualitative and quantitative biosimilar to reference product comparison.

Introduction

Biosimilar products are required by regulatory authorities to have appropriate and comparable quality, safety and efficacy with a reference biologic product. Mass spectrometry can offer in-depth characterization to explore the similarity and difference between a candidate biosimilar and a reference biologic. In this study, we developed a robust approach for a comparability study of biosimilar and reference products. Any minor difference in sequence modification and glycosylation can be well characterized and compared by using combination of high resolution Orbitrap LC-MS/MS with a powerful software to systematically interpret the results.

In this study, tissue plasminogen activator (TPA) and a generic variant of TPA (TNK) are well characterized by the robust approach. In addition, two TNK forms (G-TNK as a reference product and I-TNK as a biosimilar form) are also compared to explore the similarity and difference.

Methods

Samples

Three samples, TPA, I-TNK, G-TNK, were digested using trypsin after reduction and alkylation. Tenectelplase (TNK) is a recombinant TPA with the following minor sequence changes:

T103->N (Becomes N-glycosylation site)
N117->Q (Removes N-glycosylation site)
KHRR (296-299) -> AAAA

Liquid chromatography

Peptides were separated using with an a Thermo Scientific™ EASY-Spray™ column setup containing a 50-cm C_{18} column (2 µm particle size) and a Thermo Scientific™ EASY-nLC (U-HPLC).
LC solvents are 0.1% formic acid in H_2O (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B). Flow rate was 250 µL/min. A 60 min gradient was used to elute peptides from the column.

Mass spectrometry

A top-ten data-dependent high-energy collision dissociation (HCD) method was performed to analyze the samples. The following MS and MS/MS settings were used: MS scan range 100-2000 *m/z*. FT-MS was acquired at 70,000 resolution at *m/z* 200 with AGC target of $1x10^6$. MS2 was acquired at 17,500 resolution at *m/z* 200 with AGC target of $2x10^5$. The spray voltage was 1.8kV. Capillary temperature was 275 °C. S-lens level was set at 55.

Data analysis

Data was analyzed using a new software that is under development. This software provides automated analyses of liquid chromatography/tandem mass spectrometry (LC-MS/MS) data for large-scale identification and quantification of known and unknown modifications. Peptide identification is achieved by comparing the experimental fragmentation spectrum to the predicted spectrum of each native or modified peptide. Peak areas of related peptide ions under their selected-ion chromatograms (SIC) are used for relative quantification of modified peptides.

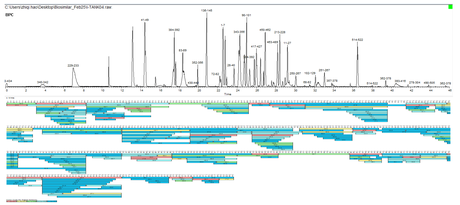
Results

Nine LC-MS/MS data files, three repeat runs for each of the sample: TPA, I-TANK and G-TANK, were analyzed and the results were compared.

1. Peptide identification and protein sequence coverage

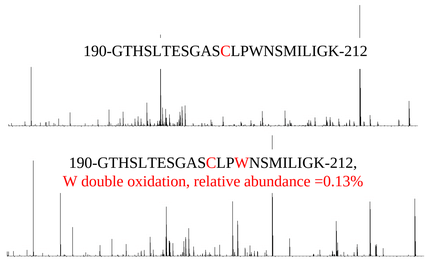
The top ten data-dependant acquisition using the Thermo Scientific™ Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer produced high quality, high resolution and accurate mass MS/MS spectra which yielded high successful rate in identification. For each raw file, 40% -50% of the MS/MS spectra resulted in high confidence peptide identification (data not shown). 100% protein sequence coverage was achieved for each of the nine data files. Figure 1 shows an example of the peptide map and sequence coverage view for one of the data file.

Figure 1. Peptide map (top) and sequence coverage (bottom) of I-TNK



The Q Exactive mass spectrometer instrument has very high throughput and sensitivity. More than 5 orders magnitude of abundances of identified peptides was routinely achieved (data not shown), which ensures confident identification of low abundance modifications, non specific cleavage versions as well as sequence variants. Figure 2 shows the high quality MS/MS spectra of a peptide (top) and its double oxidized (on W) version (bottom) which is of 0.1% in abundance.

Figure 2 MS/MS spectra of a peptide (top) and its double oxidized version (bottom) which is of 0.13% in abundance.



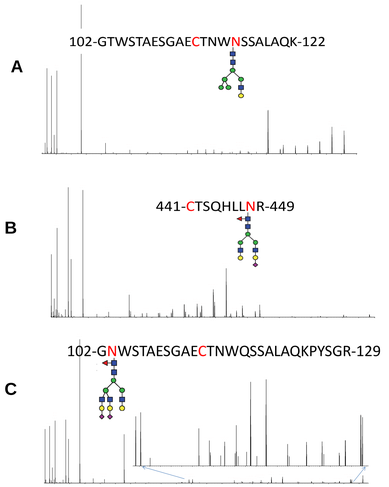
2. Glycosylation of TPA, I-TNK and G-TNK

A total of four glycosylation sites were identified, among which three of them are over 99% glycosylated. They are N 448 in all of the three samples, N103 in I-TNK and G-TNK, and N117 in TPA. The forth glycosylation site, N184, was identified only in I-TNK and only 19% of this site is glycosylated (Table 1). I-TNK has an additional glycosylation site (N184) compared to G-TNK even though these two proteins share the same amino acid sequence, suggesting difference in manufacturing procedure. Examples of MS/MS spectra of three identified glycopeptides are shown in Figure 3.

Table 1. Identified glycosylation sites, percentage of glycosylation and the number of glycoforms identified with high confidence

Site of glycosylation	Sample	# glycoforms	% glycosylation
N 103	I-TNK	18	>99
N 103	G-TNK	11	>99
N117	TPA	14	>99
N 184	I-TNK	12	19
N 448	TPA	44	>99
N 448	I-TNK	36	>99
N 448	G-TNK	47	>99

Figure 3. Examples of HCD spectra of identified glycopeptides. A: glycosylation on N117. B: glycosylation on N448. C: glycosylation on N103.



The type of glycosylation forms and their relative abundance in the three samples were compared and the following were observed :

1. Glycosylation forms on N448 and their relative abundance are consistent among all the three samples (Table 2A). Most of glycans on this site contain sialic acid.
2. Glycoforms on N103 are similar between I-TNK and G-TNK, while the relative abundance profile is quite different. Although the most abundant form, A2S1G1F, is the same in the two samples, the second and the third most abundant forms are different. For the top five most abundant forms, only two of them were shared in the two samples (Table 2B).
3. Glycans on N117 are of the type of high mannose, which is completely different from the glycans identified on other sites (Table 2C).
4. Glycosylation on N184 was only identified in I-TNK (Table 2D) and all of the glycans contain sialic acid.

Table 2. Comparison of glycoforms in the three samples. Only those with relative abundance higher than 1% in at least one of the samples are included. Abbreviations for glycan structure (1): Antenna A, core fucose (Fuc) F, mannose (Man) M, galactose (Gal) G, *N-acetyl neuraminic acid (NANA) S, N-glycolyl neuraminc acid (NGNA) Sg*

A

N448 Glycoform	TPA	I-TNK	G-TNK
N448+A2G2F	6.41%	5.40%	3.23%
N448+A2S1G0	5.18%	2.57%	ND
N448+A2S1G0F	0.52%	0.21%	1.79%
N448+A2S1G1F	23.11%	16.86%	14.43%
N448+A2S2F	37.96%	35.34%	37.59%
N448+A3G3F	0.59%	1.29%	0.80%
N448+A2Sg1S1F	1.32%	0.70%	0.56%
N448+A3S1G2F	1.59%	2.48%	0.91%
N448+A3S2G0	1.43%	0.86%	0.57%
N448+A3S2G1F	5.19%	7.00%	5.04%
N448+A4S2G2F	0.98%	ND	2.20%
N448+A4S1G3F	0.39%	1.16%	0.56%
N448+A3S3F	9.33%	11.61%	16.50%
N448+A4S3G1F	1.17%	6.55%	2.62%
N448+A4S4F	1.67%	7.20%	6.51%

B

N103 Glycoform	I-TNK	G-TNK
N103+A2G0F	ND	1.61%
N103+A2G1F	0.27%	4.49%
N103+A2G1M4F	ND	27.99%
N103+A2S1G0F	ND	1.72%
N103+A2G2	2.36%	ND
N103+A2G2F	14.89%	ND
N103+A2S1G1	5.82%	1.91%
N103+A2S1G1F	41.74%	51.76%
N103+A2S2	3.15%	ND
N103+A2S2F	26.09%	9.94%
N103+A3S1G2F	2.19%	1.89%
N103+A3S2G1F	2.16%	1.08%

C

N117 Glycoform	TPA
N117+A1G1M5	3.57%
N117+A1S1M4	2.63%
N117+A1S1M5	6.74%
N117+M5	52.41%
N117+M6	28.46%
N117+M7	6.00%

D

N 184 Glycoform	I-TNK
N184+A2S1G1F	3.22%
N184+A2S2F	4.74%
N184+A3S2G1F	2.01%
N184+A3S3F	2.99%
N184+A4S3G1F	1.50%
N184+A4S4F	1.67%

3. Other Covalent modifications identified and quantified

Besides glycosylation, other covalent modifications that were indentified in these three samples included cysteine alkylation, deamidation, overalkyation, Cys+DTT, oxidation, formylation, glycation etc. Also identified are low abundance semi-tryptic and non tryptic peptides (data not shown).

Figure 4 shows an example of a peptide that were identified in 3 different forms: native and deamidated on two different Asp residues, respectively. A total of 12 deamidation sites were indentified with high confidence in the three samples. Deamidation on N140 was only identified in I-TNK and G-TNK, not in TPA. Other sites and relative abundance of N-deamidation were consistent across all three samples (Table 3).

Figure 4. Identification and localization of two deamidation sites, N-140 and N-142 , on peptide 136-LGLGNHNYCR-145. Base peak chromatogram (A) and high resolution HCD spectra (B) of this peptide in native form or with deamidation either on N-140 or on N-142.

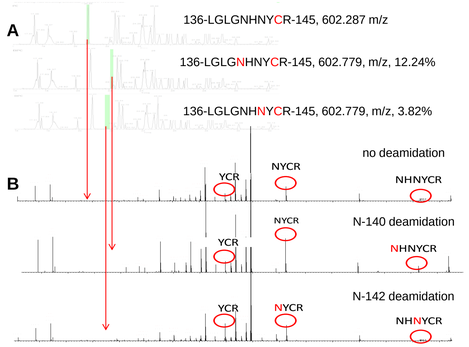


Table 3. Identified deamidation sites and relative abundance of deamidation

Location of N-deamidation	TPA	I-TNK	G-TNK
N140	ND	12.24%	10.21%
N142	3.68%	3.82%	2.70%
N205	2.08%	1.61%	0.15%
N218	0.63%	0.11%	0.31%
N234	0.15%	ND	ND
N37	29.83%	22.83%	19.64%
N370	8.24%	13.56%	0.50%
N454	3.62%	2.71%	2.27%
N469	3.71%	2.05%	1.24%
N486	11.20%	10.80%	7.64%
N516	3.68%	2.87%	2.20%
N524	1.32%	0.51%	1.80%

Conclusion

A LC-MS/MS workflow was developed for differentiating minor differences of protein structure in biosimilar and reference products. This workflow provides qualitative and quantitative biosimilar to reference product comparison.

1. 100% sequence coverage was obtained for all the nine data files analyzed. A five order magnitude dynamic range for identified peptide abundance was achieved.
2. The identified covalent modifications, both expected and un-expected, include cysteine alkylation, deamidation, overalkyation, Cys+DTT, oxidation, formylation, glycation, etc. Relative abundance of the modified forms was calculated and comparison between files was generated.
3. The site and type of glycosylation were identified and relative abundance of glycoforms was calculated. Comparison of glycosylation sites, type and relative abundance of glycoforms indicates the differences on glycosylation among the three samples.

References

1. Zhongqi Zhang and Bhavana Shah. Prediction of Collision-Induced Dissociation Spectra of Common N-Glycopeptides for Glycoform Identification Anal. Chem. 2010, 82, 10194–10202