Structure Characterization and Differentiation of Biosimilar and Reference Products Using Unique Combination of Complementary Fragmentation Mechanisms

Zhiqi Hao,[1] Chen Li,[2] Shiaw-Lin Wu,[2,3] David M. Horn[1] and Jonathan Josephs[1]
[1]Thermo Fisher Scientific, San Jose, CA; [2]BioAnalytix Inc, Cambridge, MA;
[3]Barnett Institute, Northeastern University Boston, MA

Thermo SCIENTIFIC

Overview

Purpose: To analyze difference of protein structure in biosimilar and reference products using Orbitrap LC-MS/MS

Methods: A unique data-dependent instrument method that utilizes two different fragmentation mechanisms was applied for peptide sequence and PTM identification and quantification using a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer. While generating HCD MS/MS spectra on peptides in a data-dependent experiment, the method identifies glycopeptides on-the-fly using the diagnostic ions from glycan fragmentation. A subsequent ETD fragmentation is then triggered on the same peptide to produce information of amino acid sequence and site of glycosylation. The new Thermo Scientific™ PepFinder™ 1.0 software for peptide imaging, was used for data analysis.

Results: A LC-MS/MS workflow was developed for differentiating minor difference of protein structure in biosimilar and reference products using an Orbitrap Fusion LC-MS/MS and PepFinder 1.0 software. This new approach offers efficient, confident and comprehensive analysis, not only for biosimilar comparability study but also for lot-to-lot comparison of a same compound.

Introduction

Biosimilars are subsequent versions of innovator biopharmaceutical products created after the expiration of the patent on the innovator product. The approval of a biosimilar product by a regulatory agency requires thorough characterization that demonstrates comparability with a reference product in quality, safety and efficacy. High resolution mass spectrometry provides accurate characterization of various protein properties including primary structure, type and location of post-translational modifications (PTMs), and low abundant sequence variants or impurities. In this study, we developed a robust approach for comparability study of biosimilar and reference product. Minor differences in products including glycosylation were systematically compared using high resolution LC/MS/MS with complementary fragmentation methods and a new peptide mapping software package.

Methods

Samples

An original drug, a recombinant variant and its biosimilar product, TPA, I-TNK and 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 a Thermo Scientific™ EasySpray™ source setup containing 50-cm C_{18} column (2 µ particle size) and a high-pressure easy nanoLC (U-HPLC). The LC solvents were 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 70 min gradient was used to elute peptides from the column.

Mass spectrometry

Samples were analyzed using an Fusion mass spectrometer with a Thermo Scientific™ EASY-ETD™ ion source. An instrument method designed for glycopeptide analysis was used for this study. This method primarily acquires HCD MS/MS spectra on peptides in a data-dependent top-ten experiment. However, if diagnostic sugar oxonium ions from glycan fragmentation are detected in the HCD MS/MS spectrum, a subsequent ETD fragmentation is then triggered on the same precursor peptide to produce amino acid backbone sequence information to identify the site of glycosylation. Therefore, for each glycopeptide, this HCD product-dependent ETD method (HCDpdETD) generates a pair of HCD and ETD spectra, producing information for the peptide sequence and the site of glycosylation as well as confirming glycan structure. Orbitrap MS spectra were acquired at 120,000 resolution (at *m/z* 200) with an AGC target of $4x10^5$. MS/MS spectra were acquired at 30,000 resolution (at *m/z* 200) with an AGC target of $5x10^4$. Capillary temperature was set to 275 °C and the S-lens level was set at 60. The priority for precursor selection for data-dependent MS/MS was for the highest charge state followed by the lowest *m/z*. HCD collision energy was 30 and ETD activation time was charge dependent based on the standard calibration.

Data analysis

Data was analyzed using PepFinder 1.0 software. This software provides automated analyses of liquid chromatography/tandem mass spectrometry 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. A mass tolerance of 5 ppm was used to ensure accurate identification.

Results

1. Peptide identification and protein sequence coverage

The data was analyzed and the results were compared. Peptide mapping results indicated 100% sequence coverage for all of the data files. The relative abundance of each modified peptide forms was calculated and compared between files. A five order magnitude dynamic range for identified peptide abundance was achieved, which allowed identification of modified peptides with less than 0.01% in abundance of the unmodified versions (data not shown). Figure 1 shows an example of the sequence coverage view for one of the data files.

Figure 1. 100% sequence coverage of I-TNK

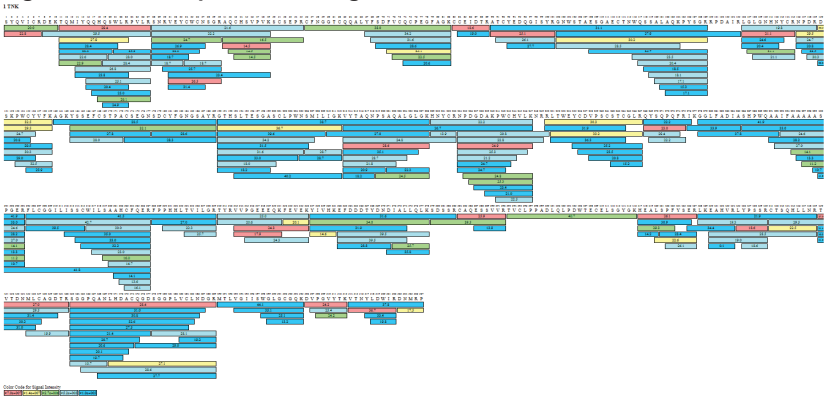
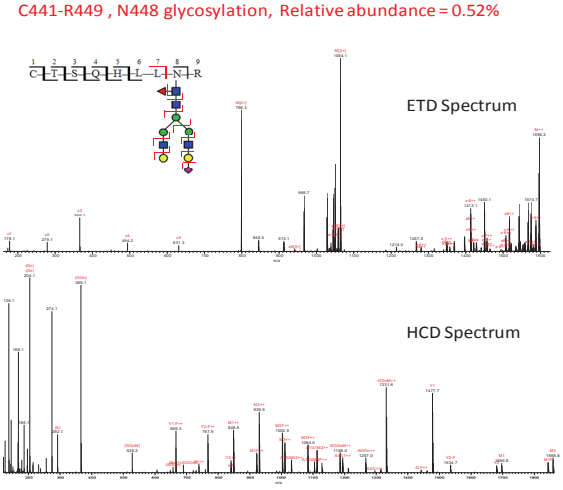


Table 1. Identified glycosylation sites, percentage of glycosylation and the number of glycoforms identified with high confidence

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 2. Characterization of glycopeptides using HCDpdETD. G-TNK peptide C441-R449 with glycosylation on N448. Top left is fragment ion coverage showing peptide backbone fragmentation from ETD (black, with glycan preserved) and fragmentation of peptide and glycan from HCD (red).



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

A total of four glycosylation sites were identified, three of which are over 99% glycosylated. N448 was glycosylated in all three samples, while N103 was detected in I-TNK and G-TNK and N117 only 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) even though it shares the same amino acid sequence as G-TNK, suggesting a different manufacturing process. Examples of two identified glycopeptides are shown in Figure 2 and Figure 3.

Figure 3. Characterization of glycopeptides using HCDpdETD. I-TNK peptide G102-R129 with glycosylation on N103. The fragment ion coverage at the top of this figure shows peptide backbone fragmentation from ETD (black, with glycan preserved) and fragmentation of peptide and glycan from HCD (red).

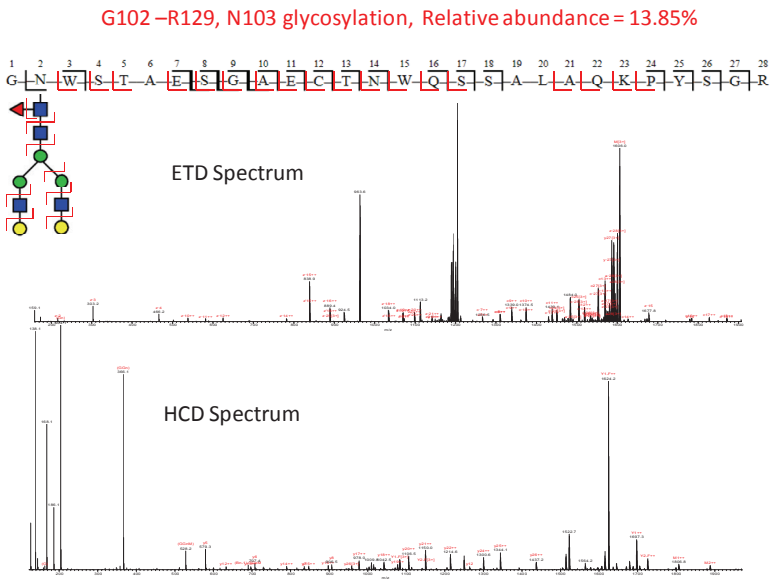


Table 2. Comparison of N448 glycoforms in the three samples. Only those with relative abundance higher than 1% in at least one of the samples are included. The five major glycoforms are highlighted in bold. Abbreviations for glycan structure: Antenna A, core fucose (Fuc) F, mannose (Man) M, galactose (Gal) G, *N-acetyl neuraminic acid (NANA) S, N-glycolyl neuraminc acid (NGNA) Sg*

N448 Glycoform	TPA	I-TNK	G-TNK
N448+A2G2F	**6.41%**	**5.40%**	**3.23%**
N448+A2S1G0	5.18%	2.57%	<1%
N448+A2S1G0F	<1%	<1%	1.79%
N448+A2S1G1F	**23.11%**	**16.86%**	**14.43%**
N448+A2S2F	**37.96%**	**35.34%**	**37.59%**
N448+A3G3F	<1%	1.29%	<1%
N448+A2Sg1S1F	1.32%	<1%	<1%
N448+A3S1G2F	1.59%	2.48%	<1%
N448+A3S2G0	1.43%	<1%	<1%
N448+A3S2G1F	**5.19%**	**7.00%**	**5.04%**
N448+A4S2G2F	<1%	<1%	2.20%
N448+A4S1G3F	<1%	1.16%	<1%
N448+A3S3F	**9.33%**	**11.61%**	**16.50%**
N448+A4S3G1F	1.17%	6.55%	2.62%
N448+A4S4F	1.67%	7.20%	6.51%

The type and relative abundance of gllycoforms were compared across the three samples and the following was observed:

1. The relative abundance and identity of the various glycoforms on N448 were consistent among all three samples (Table 2). Most of glycans on this site contain sialic acid.
2. The identity of the glycoforms on N103 are similar between I-TNK and G-TNK, but the relative abundance profiles are markedly different. Although the most abundant form, A2S1G1F, is the same in the two samples, the second and the third most abundant forms are not. For the top five most abundant forms, only two of them were common in the two samples (data not shown).
3. The glycoforms on N117 are primarily high mannose, which is very different from the glycans identified on any of the other sites (data not shown).
4. Glycosylation on N184 was only detected for the I-TNK sample (data not shown), with all of the glycans containing sialic acid.

3. Other identified and quantified modifications

Besides glycosylation, other covalent modifications that were indentified in these three samples included cysteine alkylation, deamidation, overalkyation, Cys+DTT, oxidation, formylation, and glycation. Figure 4 shows confident identification and localization of a low abundant double oxidation on W406. The relative abundance of the oxidized form is less than 0.1%.

A total of 12 N-deamidation sites were indentified with high confidence in the three samples. Deamidation on N140 was only identified in I-TNK and G-TNK, but not in TPA. Other sites of N-deamidation were consistent across all three samples (Table 3). Figure 5 shows examples of a peptide that were identified in 3 different forms: native and deamidated on two different Asp residues, respectively.

Figure 4. Identification of low abundant double oxidized peptide T393-K416 and localization of double oxidation to W406.

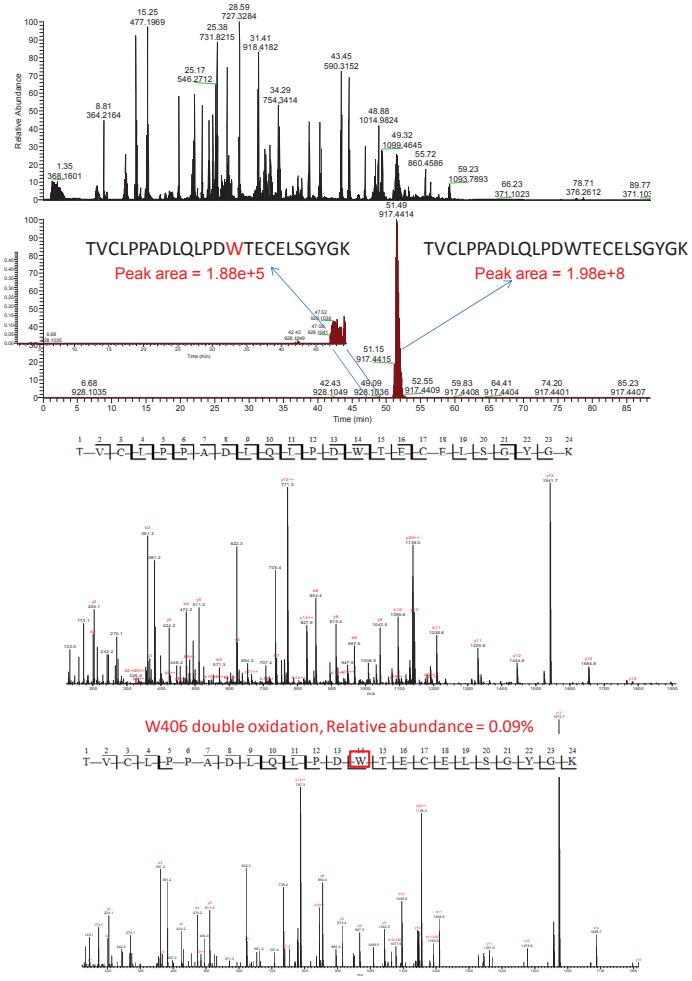


Figure 5. Identification and localization of two deamidation sites, N140 and N142 , on peptide L136-R145. High resolution HCD spectrum of this peptide in native form (top), with deamidation either on N142 (middle) or on N140 (bottom).

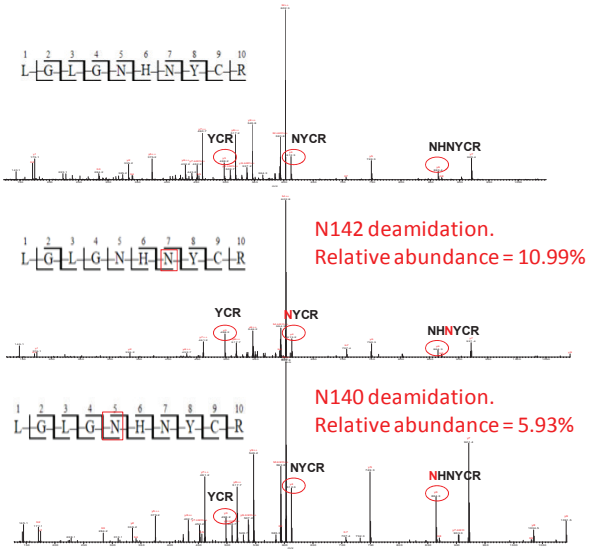


Table 3. Identified N-deamidation sites and relative abundance of deamidation

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%	<0.5%
N218	0.63%	<0.5%	<0.5%
N234	<0.5%	<0.5%	<0.5%
N37	29.83%	22.83%	19.64%
N370	8.24%	13.56%	<0.5%
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%	<0.5%	1.80%

Conclusion

A LC-MS/MS workflow was developed to differentiate minor differences in protein structure for biosimilar and reference products using an Orbitrap Fusion instrument and new peptide mapping software, PepFinder 1.0. This workflow provides qualitative and quantitative biosimilar to reference product comparison.

1. 100% sequence coverage was obtained for all the nine data files analyzed .
2. The identified covalent modifications, both expected and un-expected, include cysteine alkylation, deamidation, overalkyation, Cys+DTT, oxidation, formylation, glycation and glycosylation. The relative abundance of the modified forms was calculated and compared between datasets. Confident identification and precise localization of low abundant PTMs was achieved.
3. Glycosylated peptides were characterized using the unique HCDpdETD method which generates information of peptide sequence, site of glycosylation as well as glycan structure. Comparison of glycosylation sites as well as the type and relative abundance of glycoforms indicated that there are significant differences in glycosylation between the three samples.

www.thermoscientific.com

Thermo Fisher Scientific, San Jose, CA USA is ISO 9001:2008 Certified.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China 800 810 5118 (free call domestic)
400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 9 3291 0200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 45 453 9100
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC

A Thermo Fisher Scientific Brand

PN-64145-EN-0614S