SHIMADZU
Excellence in Science

Application News

Protein sequencer
MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight) Mass Spectrometer

Amino Acid Sequence Analysis of Semaglutide

Tomoko Kuriki, Kumiko Yamaguchi

User Benefits

- Possible to easily determine the amino acid sequences of proteins not listed in genome databases
- Effective for obtaining amino acid sequence information for peptides containing non-natural amino acids, as well as amino acid sequence information that includes the C-terminus
- Possible to obtain structural information on modified amino acids when combined with MALDI-TOF MS

■ Introduction

Middle-molecule drugs are attracting attention as a new therapeutic modality that combines characteristics of both small-molecule drugs and antibody drugs. Peptide drugs, one such category, offer high target specificity and are expected to reduce the risk of side effects. In particular, the market for these drugs is expanding in the fields of diabetes and obesity. Semaglutide, a representative peptide drug, is a middle-molecule drug that improves the stability and prolonged duration of action of human GLP-1. While peptide drugs are primarily manufactured using solid-phase peptide synthesis (SPPS), manufacturing methods utilizing cell expression systems are also being explored for certain targets.

When evaluating the quality of manufactured products, verification of the amino acid sequence is crucial from the perspectives of safety and efficacy. One such method for amino acid sequence analysis is the Edman degradation method, which allows for the confirmation of the amino acid sequence by sequentially cleaving and identifying amino acids starting from the N-terminus. In this report, we present an example of amino acid sequence analysis of semaglutide using the Edman degradation method and MALDI-MS analysis.

■ Semaglutide

Semaglutide is a peptide analog of human glucagon-like peptide-1 (GLP-1) consisting of 31 amino acid residues, and it shares 94% amino acid sequence homology with human GLP-1[1]. It is used as a treatment for type 2 diabetes and obesity. The amino acid sequence of semaglutide is shown in Fig. 1. Semaglutide is a modified analog of human GLP-1 designed to improve degradation resistance and prolong its duration of action. Specifically, the following three structural modifications have been made:

- Replacement of the amino acid at position 2 on the N-terminus with 2-aminoisobutyric acid (Aib)
- Fatty acid modification of the Lys side chain at residue 20
- Replacement of the amino acid at residue 28 with Arg

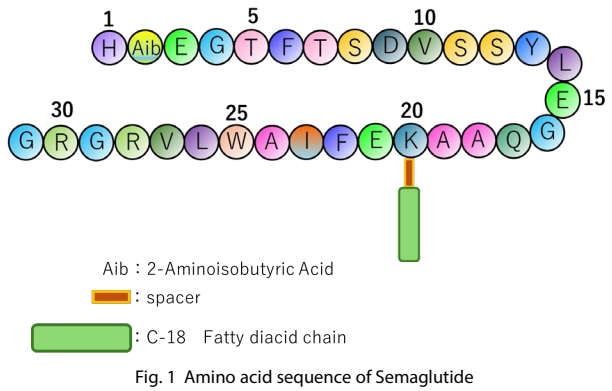

Fig. 1 Amino acid sequence of Semaglutide

■ Chromatogram of PTH-amino Acids (Phenylthiohydantoin Derivatives of Amino Acids)

In a protein sequencer using Edman degradation, amino acids are sequentially cleaved from the N-terminus of proteins and peptides; after derivatization into stable PTH-amino acids (phenylthiohydantoin derivatives of amino acids), they are separated and identified by HPLC. Table 1 shows the analytical conditions for a 25 pmol PTH-amino acids mixture standard using an isocratic system, and Fig. 2 (pink) shows the corresponding chromatogram. The amino acids that make up semaglutide include an amino acid (2-aminoisobutyric acid) that is not present in the PTH-amino acid mixture standard. The chromatogram of PTH-2-aminoisobutyric acid (PTH-Aib) is shown in Fig. 2 (blue). PTH-Aib elutes close to PTH-Arg, but can be distinguished due to its different retention time.

Table 1 Analytical conditions of Isocratic system

Column	: Wakopak Wakosil PTH-II (S-PSQ) (250 mm x 4.6 mm I.D.)
Mobile phase	: PTH-amino Acids Mobile Phase
Flow Rate	: 1.0 mL/min
Time program	: T. Flow 1.0 mL/min(0-21.25 min) - 0.3 mL/min(21.5 -45.25 min) - 1.0 mL/min(45.5 - 45.51 min)
Column temp.	: 40 °C
Detection	: UV 269 nm (SPD-M30A) High Sensitivity Flow Cell

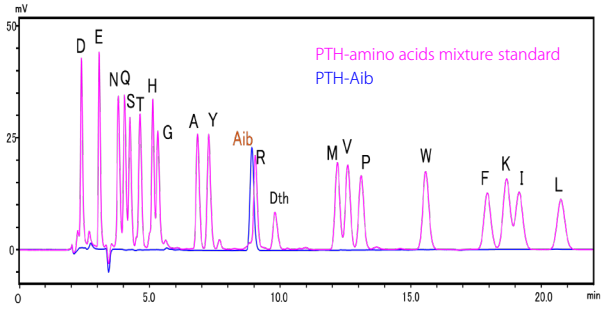

Fig. 2 Analysis of PTH-amino acids mixture standard (25 pmol) (Isocratic system)

Table 2 shows the analytical conditions for a 10 pmol PTH-amino acids mixture standard using a gradient system, and Fig. 3 (pink) shows the chromatogram of the PTH-amino acids mixture standard. As with the isocratic system, PTH-Aib (blue) can be easily separated and identified using the gradient system as well.

Table 2 Analytical Conditions of Gradient System

Column	:	Wakopak Wakosil PTH-GR (S-PSQ) (250 mm x 2.0 mm I.D.)
Mobile phase A	:	PTH-amino Acids Mobile Phase A (for Gradient Elution)
Mobile phase B	:	PTH-amino Acids Mobile Phase B (for Gradient Elution)
Flow Rate	:	0.3 mL/min
Time program.	:	B conc. 0%(0 min)-0% 0%(4 min) - 100%(17 - 30 min) -0%(30.01 - 45 min)
Column temp.	:	35 ˚C
Detection	:	UV 269 nm (SPD-M30A) High Sensitivity Flow Cell

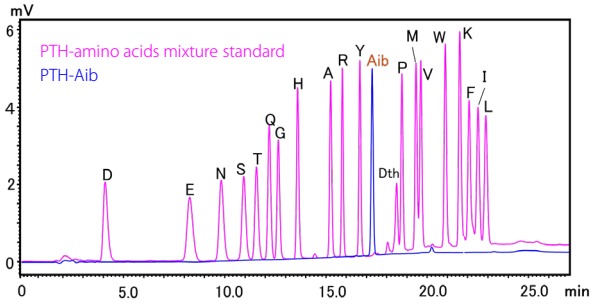

Fig. 3 Analysis of PTH-amino acids mixture standard (10 pmol) (Gradient system)

■ Sample Preparation

Semaglutide (Merck) was used as a sample. Semaglutide was dissolved in a 5% (vol/vol) acetic acid aqueous solution to prepare a sample solution with a concentration of 100 pmol/µL.

■ Amino Acid Sequence Analysis of Semaglutide (Isocratic System)

A 5 µL sample solution (500 pmol) was applied to a polybrene-treated glass fiber disk; after drying, amino acid sequence analysis was performed using PPSQ™-51A/53A. The resulting chromatogram is shown in Fig. 4.

In the second cycle, identification by automatic estimation was difficult because the elution position of PTH-Aib was close to that of PTH-Arg. However, by verifying the retention times, we were able to confirm it as PTH-Aib through manual analysis (Fig. 4A).

Fig. 4B shows the chromatograms for cycles 19 through 22. In cycle 20, no PTH-amino acids, which had been specifically increasing, were detected. Possible reasons for this include the following:

・As shown in Fig. 1, the side chain of the 20th amino acid (Lys) is modified by a fatty acid, so Edman degradation does not proceed beyond this cycle.

・The phenylthiohydantoin derivative of the 20th modified Lys is highly hydrophobic and interacts strongly with the analytical column, preventing it from eluting from the column.

・The side-chain modification was degraded during Edman degradation.

On the other hand, the results in Fig. 4B show that the amino acid sequence from the 21st cycle onward has been determined, indicating that the Edman degradation itself proceeds without being affected by this modified Lys.

Although the PTH-Glu peak showed a specific increase, the yield of PTH-Glu obtained was lower compared to the preceding and following cycles. However, in cycles where PTH-Glu was detected, a peak (14.3 min) corresponding to a PTH-Glu byproduct resulting from Edman degradation was also detected (blue arrow in Fig. 4B). Since the peak of the PTH-Glu byproduct had increased significantly in cycle 21, this cycle was identified as PTH-Glu.

Fig. 4C shows the chromatogram near the C-terminus. Although semaglutide is a 31-residue peptide, no PTH-amino acids, which typically show a specific increase starting from the 31st cycle, were detected in the chromatogram, and therefore the amino acid could not be identified (Fig. 4C). When analyzing the amino acid sequence of a peptide using a protein sequencer, fragments or amino acids from the C-terminus may elute from the sample carrier during the washing step of the Edman degradation process. Depending on the peptide sequence, it may be difficult to determine the sequence of the C-terminus.

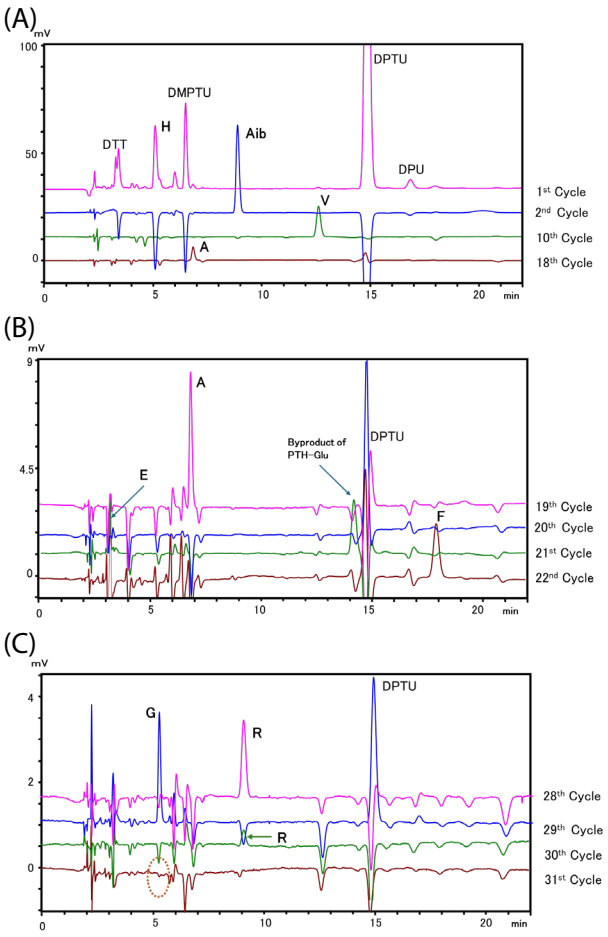

Fig. 4 Chromatograms of the amino acid sequence analysis of Semaglutide (Isocratic system)
A: Chromatograms for cycles 1, 2, 10, and 18
B: Chromatograms for cycles 19, 20, 21, and 22
C: Chromatograms for cycles 28, 29, 30, and 31

■ Amino Acid Sequence Analysis of Semaglutide (Gradient System)

Similar to the analysis performed using the isocratic system, we conducted amino acid sequence analysis of semaglutide using the PPSQ-51A/53A gradient system. The chromatogram is shown in Fig. 5. In the gradient system, PTH-Aib in the second cycle was detected as a peak that increased specifically around 17.4 minutes and could be easily identified (Fig. 5A). Fig. 5B shows the chromatograms for cycles 19 through 22. In cycle 20, no PTH-amino acids showing a specific increase, as observed in the isocratic system, were detected.

Furthermore, for the Glu in the 21st cycle, as with the isocratic system, not only was the peak for PTH-Glu clearly detected, but a peak (at 21.4 min) corresponding to a byproduct of PTH-Glu resulting from Edman degradation was also clearly detected (blue arrow in Fig. 5B). Based on this, the 21st cycle was identified as PTH-Glu.

Fig. 5C shows the chromatograms near the C-terminus. In cycle 31, we confirmed a specific increase in PTH-Gly. In the subsequent chromatograms, no PTH-amino acids showing a specific increase were detected. Based on these results, we determined that the C-terminal amino acid is Gly. The gradient system was shown to detect PTH-amino acids with higher sensitivity.

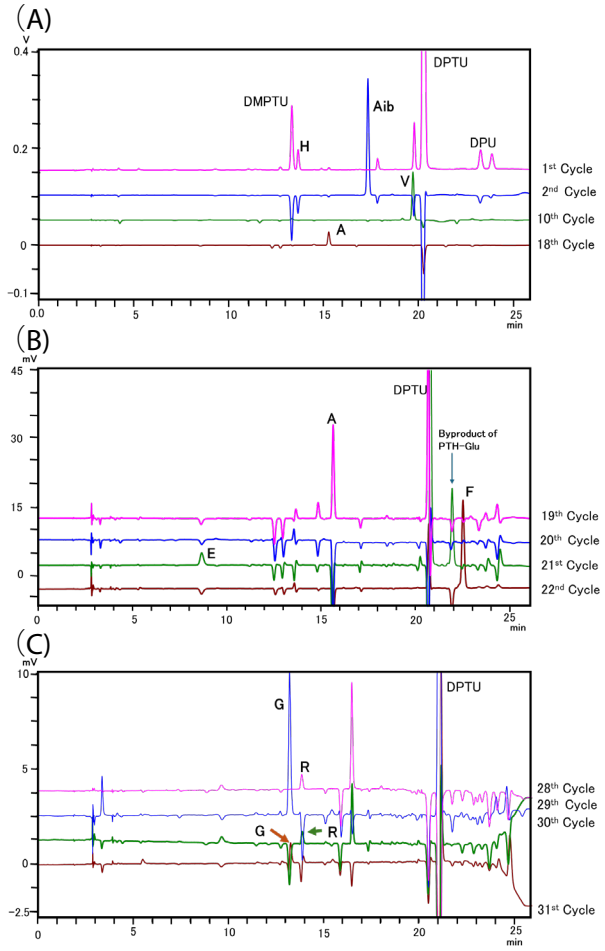

Fig. 5 Chromatograms of the amino acid sequence analysis of Semaglutide (Gradient system)
A: Chromatograms for cycles 1, 2, 10, and 18
B: Chromatograms for cycles 19, 20, 21, and 22
C: Chromatograms for cycles 28, 29, 30, and 31

■ Molecular Weight Measurement Using MALDI-TOF MS

To verify the molecular weight of semaglutide, we measured its molecular weight using a matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-TOF MS; MALDI-8030).

The sample solution was diluted with a 1:1 (vol/vol) mixture of acetonitrile and 0.1% trifluoroacetic acid aqueous solution to prepare a sample solution with a concentration of 10 pmol/μL.

α-Cyano-4-hydroxycinnamic acid (CHCA) was used as the MALDI matrix. The MALDI mass spectrum of semaglutide is shown in Fig. 6. The measurement results confirmed that the average molecular weight of semaglutide is $[M+H]^+ = m/z$ 4114.8.

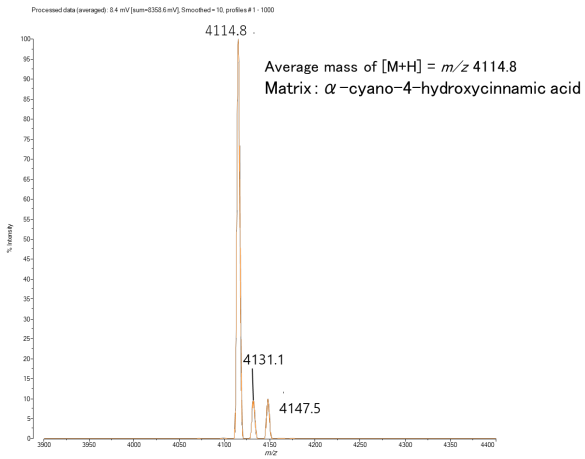

Fig. 6 MALDI MS spectrogram

Furthermore, based on the difference between the amino acid sequence data obtained using the PPSQ-53A gradient system and the total molecular weight obtained via MALDI, the average molecular weight of the modified amino acid at the 20th residue was estimated to be 844.2. By combining the molecular weight measurement results with the amino acid sequence data in this way, structural information regarding semaglutide can be obtained. Furthermore, molecular weight data is useful for quickly determining whether there is an incorrect amino acid composition, degradation, or the presence of modifications.

■ Conclusion

We found that the PPSQ-51A/53A system enables the confirmation of amino acid sequences in synthetic peptides containing non-natural amino acids. Furthermore, by combining this with molecular weight information obtained via MALDI-TOF MS, we were able to demonstrate that it is possible to obtain mass information regarding modified amino acids, which is difficult to confirm using sequence information alone. This method is expected to serve as a useful analytical approach in the research, development, and quality assessment of peptide drugs. However, we were unable to detect or identify the Lys residues in semaglutide whose side chains were modified with fatty acids.

In recent years, peptide drugs synthesized chemically have evolved into side-chain-modified peptides and cyclic peptides with more complex structures in order to enhance their activity, specificity, and stability in the body.

In future studies, it is believed that analyses combining multiple analytical methods will be essential for verifying the synthesis and analyzing the structures of such novel peptide drugs.

Reference

1) Lotte Bjerre Knudsen and Jesper Lau, "The Discovery and Development of Liraglutide and Semaglutide", Sec. Molecular and Structural endocrinology, 10 (155), 1-32 (2019)

01-01230-EN First Edition: Jul. 2026

SHIMADZU
Shimadzu Corporation
www.shimadzu.com/an/

› Please fill out the survey

Related Products Some products may be updated to newer models.

› PPSQ-51A/53A
Protein Sequencer

› MALDI-8030
Dual-Polarity Benchtop Linear MALDI-TOF Mass Spect···

Related Solutions

› Pharmaceutical and Biopharmaceutical

› Biopharmaceutical

› Life Science

› Proteomics

› Price Inquiry

› Product Inquiry

› Technical Service / Support Inquiry

› Other Inquiry