play a significant role in diabetes treatment and also have preventive and therapeutic effects on cardiovascular diseases [9] and neurological disorders [10, 11]. Additionally, GLP-1 can bind to GLP-1 receptors in organs such as the kidneys and skin, influencing tissue metabolism and related diseases [12].

The GLP-1 fusion protein disclosed in US Patent US8658174 comprises a GLP-1 peptide fused with an IgG/Fc domain and can be used for the treatment of diabetes.

Studies have shown that GLP1-Fc fusion protein undergoes post-translational modifications. Hou et al. reported in 2019 that the addition of nicotinamide and cysteine during cell culture process helps reduce the hydroxylation level of a GLP analog when fused with IgG4/Fc protein (dulaglutide) [13].

The use of genetic engineering and recombinant protein technology to produce therapeutic fusion proteins aims to express the properties of natural peptides. The process involves cellular engineering steps including transcription, translation, and post-translational modifications. The manufacturing process directly impacts the physicochemical characteristics, conformation, in vivo half-life, biological activity, and production yield of the drug. Therefore, there is still a need for improved GLP-1 receptor agonists and their fusion proteins that are suitable for large-scale production, possess higher yield and activity, and exhibit longer half-life, among other desirable features, for clinical treatment.

Summary

In one aspect, provided herein is a fusion protein comprising a GLP-1 polypeptide and an immunoglobulin Fc domain, wherein the GLP-1 polypeptide is covalently linked to the immunoglobulin Fc domain, and the GLP-1 polypeptide is selected from human GLP-1 (7-37), human GLP-1 (7-36), and DPP-IV resistant human GLP-1, and the GLP-1 polypeptide contains one or more amino acid substitutions selected from the group consisting of A8G, G22E, and R36G relative to native human GLP-1; and wherein the immunoglobulin Fc domain comprises or is an IgG2-Fc domain, and the IgG2-Fc domain contains one or more amino acid substitutions selected from the group consisting of C222S, A330S, and P331S.

In another aspect, provided herein is a dimer comprising two identical peptide chains connected by a disulfide bond, wherein each peptide chain comprises the fusion protein as described herein.

In another aspect, provided herein is a nucleic acid molecule comprising a nucleotide sequence encoding the fusion protein as described herein.

In another aspect, provided herein is a vector comprising the nucleic acid molecule as described herein.

In another aspect, provided herein is a cell comprising a nucleic acid molecule as described herein, or a vector as described herein.

In another aspect, provided herein is a composition comprising the fusion protein as described herein, or the dimer as described herein, or the nucleic acid molecule as described herein, or the vector as described herein, or the cell as described herein.

In another aspect, provided herein is a method for constructing a cell as described herein, comprising:

a) introducing a nucleic acid molecule encoding the fusion protein into a vector to construct an expression vector;

b) introducing the expression vector into a cell that recombinantly expresses or naturally expresses lysine hydroxylase to obtain a recombinant cell; preferably, the lysine hydroxylase level or activity expressed by the recombinant cell is higher than the level or activity of lysine hydroxylase expressed by COS-7 cells; more preferably, the cell is a CHO cell, particularly CHO-K1 cells; more preferably, the vector is a pKN012 vector.

In another aspect, provided herein is a method for constructing a recombinant cell, comprising:

a) inserting the nucleotide sequence as set forth in SEQ ID NO: 26 into the NcoI and HindIII sites of a pKN012 vector to generate a pKN012-GLP1-IgG2/Fc expression vector;

b) introducing the pKN012-GLP1-IgG2/Fc expression vector into a CHO-K1 cell to obtain the recombinant cell.

In another aspect, provided herein is a method for producing a fusion protein, comprising the step of using the cell as described herein or the cell obtained by the construction method as described herein to produce the fusion protein.

In yet another aspect, provided herein is a method for detecting the quality of a fusion protein, wherein the fusion protein comprises a GLP-1 polypeptide and an immunoglobulin IgG2-Fc domain, the method comprises the following step: detecting the hydroxylation level of the fusion protein at position K34 relative to native human GLP-1.

In yet another aspect, provided herein is use of the fusion protein as described herein, or the dimer as described herein, or the nucleic acid molecule as described herein, or the vector as described herein, or the cell as described herein, or the composition as described herein, in the manufacture of a medicament for the treatment or prevention of diseases.

In yet another aspect, provided herein is use of the fusion protein as described herein, or the dimer as described herein, or the polynucleotide as described herein, or the vector as described herein, or the cell as described herein, or the composition as described herein, and additional therapeutic agents, in the preparation of medicaments for treating or preventing diseases.

In yet another aspect, provided herein is use of a fusion protein comprising the amino acid sequence shown in SEQ ID NO: 7 and metformin in the preparation of a medicament for the treatment of diabetes.

In yet another aspect, provided herein is use of a fusion protein comprising the amino acid sequence shown in SEQ ID NO: 7 and gamma-aminobutyric acid in the preparation of a medicament for the treatment of neurodegenerative diseases.

In yet another aspect, provided herein is a pharmaceutical combination comprising the fusion protein as described herein, or the dimer as described herein, or the polynucleotide as described herein, or the vector as described herein, or the cell as described herein, or the composition as described herein, and an additional therapeutic agent.

In yet another aspect, provided herein is the fusion protein as described herein, or the dimer as described herein, or the polynucleotide as described herein, or the vector as described herein, or the cell as described herein, or the composition as described herein, and the fusion protein as described herein, or the dimer as described herein, or the polynucleotide as described herein, or the vector as described herein, or the cell as described herein, or the composition as described herein, and additional therapeutic agents, for use in therapy.

The content of the invention

The purpose of the present invention is to provide an improved GLP-1 receptor agonist and its fusion protein with higher activity and yield. The fusion protein can be obtained through various approaches or methods, such as amino acid substitutions at specific positions in the protein sequence, or modifications like hydroxylation and oxidation. For example, hydroxylation at position K34 of the GLP-1 peptide and/or reduction of oxidation of the GLP-1 peptide can be performed. Furthermore, specific amino acid substitutions or modifications at certain positions of the GLP-1 fusion protein in the present invention significantly extend the half-life of the improved fusion protein, while exhibiting excellent preventive and therapeutic effects in human diseases and animal disease models.

Therefore, one of the advantages of the present invention is to provide an improved GLP-1 fusion protein with increased yield, activity, and/or extended half-life. Additionally, the improved

d) Transfecting the pKN012-GLP1-IgG2/Fc expression vector into CHO-K1 cells to obtain recombinant cells.

In another aspect, the present invention provides a method for producing the fusion protein, comprising the step of obtaining the fusion protein using the cells described in this document or the cells prepared using the construction method.

In another aspect, the present invention provides a method for assessing the quality of the fusion protein, wherein the fusion protein comprises GLP-1 peptide and immunoglobulin IgG2-Fc domain. The method comprises the following steps: detecting the level of hydroxylation of the fusion protein at position K34 relative to native human GLP-1.

In another aspect, the present invention provides a use of the fusion protein, dimer, nucleotide, vector, cell, or composition as described herein, in the preparation of pharmaceuticals for the treatment or prevention of diseases.

The other features and advantages of the present invention will become apparent from the following detailed description in conjunction with the accompanying embodiments. The detailed description and specific examples provided herein are given for illustrative purposes while embodying the preferred embodiments of the present invention. Various changes and modifications within the spirit and scope of the present invention will be readily apparent to those skilled in the art.

BRIEF DESCRIPTION OF THE DRAWINGS:

Figure 1 is a schematic diagram of the pKN012-GLP1-IgG2/Fc vector.

Figure 2 shows the mass spectrum of the enzymatic cleavage peptide segment 27-34: EFIAWLVK (SEQ ID NO: 44) (+16Da) modified with GLP-1 fusion protein YN-011 after Lys-C digestion. A: Primary mass spectrum of EFIAWLVK (SEQ ID NO: 44); B: Primary mass spectrum of EFIAWLVK (SEQ ID NO: 44) (+16 Da) modification; C: Secondary mass spectrum of EFIAWLVK(SEQ ID NO: 44) ; D: Secondary mass spectrum of EFIAWLVK (SEQ ID NO: 44) (+16 Da) modification.

Figure 3A shows the UV detection data for the modified peptide 27-34 (aa21-28) with EFIAWLVK (SEQ ID NO: 44) (+16 Da) modification, the unmodified peptide 54-79 (aa48-73) without the +16 Da modification, and the unmodified peptide 224-240 (aa218-234) after Lys-C digestion. Figure 3B is an enlarged view of the inset in Figure 3A.

NAME	Amino Acid Sequence	SEQ ID NO:
IgG2/Fc domain (human IgG2 heavy chain Fc region residues 219-445)	ERKCCVECPPCPAPPVAGPSVFLFPPKPKDTLMI SRTPEVTCVVVDVSHEDPEVQFNWYVDGVEV HNAKTKPREEQFNSTFRVVSVLTVVHQDWLN GKEYKCKVSNKGLPAPIEKTISKTKGQPREPQV YTLPPSREEMTKNQVSLTCLVKGFYPSDIAVE WESNGQPENNYKTTPPMLDSDGSFFLYSKLTV DKSRWQQGNVFSCSVMHEALHNHYTQKSLSL SPG	5
IgG2/Fc domain mutation (human IgG2 heavy chain Fc region residues 219-445, including C222S, A330S, and P331S site mutations)	ERKSCVECPPCPAPPVAGPSVFLFPPKPKDTLMI SRTPEVTCVVVDVSHEDPEVQFNWYVDGVEV HNAKTKPREEQFNSTFRVVSVLTVVHQDWLN GKEYKCKVSNKGLPSSIEKTISKTKGQPREPQV YTLPPSREEMTKNQVSLTCLVKGFYPSDIAVE WESNGQPENNYKTTPPMLDSDGSFFLYSKLTV DKSRWQQGNVFSCSVMHEALHNHYTQKSLSL SPG	6

The corresponding positions of the A8G, G22E, R36G substitutions, and K34 hydroxylation in the GLP-1 peptide mentioned in this application, relative to SEQ ID NO: 3 and SEQ ID NO: 7, are shown in the following table. Additionally, the corresponding positions of the C222S, A330S, P331S substitutions, and M253 oxidation on the IgG2-Fc domain mentioned in this application, relative to QRG33935.1 and SEQ ID NO: 7, are also shown in the following table.

Modified Sites	Position in SEQ ID NO: 3	Position in SEQ ID NO: 6 of IgG2 (219-445) Relative to the sequence shown in QRG33935.1	Position in SEQ ID NO: 7
A8G	AA2		AA2

Mass spectrometry analysis was performed using Lys-C digestion for identification. The YN-011 sample was denatured, reduced, and alkylated, followed by Lys-C enzyme hydrolysis to generate peptides. Online LC-MS/MS analysis was conducted using a Thermo LTQ Velos Orbitrap instrument, and the data were processed and analyzed using Mascot software. It was observed that the peptide segment 21-28: EFIAWLVK (SEQ ID NO: 44) exhibited a molecular weight increase of 16 Daltons (Da) due to modification. The primary and secondary mass spectra of peptides without the +16 Da modification and peptides with the +16 Da modification in the YN-011 sample are shown in Figure 2. The mass measurement error of the primary spectra was within 10 ppm. By analyzing the mass-to-charge ratio of the secondary fragment ions, the site of the +16 Da modification was determined to be K28. K28 is also referred to as K34 since the active GLP-1 (e.g., amino acids 7-37 of GLP-1) lacks the first six residues. The peptide 21-28, also known as 27-34 peptide, can be found in natural human GLP-1 as well.

The modification level at the K34 site in YN-011 was determined using UV280 and EIC (Extracted Ion Chromatography) methods, as shown in Table 1 and Figures 3A and 3B. The calculation method for UV280 is as follows:

$$k = \frac{\frac{k\varepsilon_w}{k\varepsilon_w + \varepsilon_Y} A_1}{\frac{k\varepsilon_w}{k\varepsilon_w + \varepsilon_Y} A_1 + A_o} \qquad \begin{matrix} \varepsilon_w = 5500 \\ \varepsilon_w = 1490 \end{matrix}$$

Where:

K: The ratio of modified peptides with +16 Da to the total amount of peptides in the amino acids 21-28 segment (sum of unmodified peptides and +16 Da modified peptides).

A1: Absorbance value at 280 nm for the amino acids 21-28 peptide segment without the +16 Da modification.

A0: Absorbance value at 280 nm for the amino acids 21-28 peptide segment with the +16 Da modification.

Table 1: Detection of K34 modification levels in YN-011 using UV280 and EIC methods.

Sample Name	Sequence Position	Amino Acid Sequence (SEQ ID NO: 44)	Translated Modification (PTMs)	UV280 Ratio (%)	EIC* Ratio (%)
YN-011 Sample 1	21-28	EFIAWLVK	/	/	/
		EFIAWLV**K**	+16Da	29.6%	31.8%
YN-011 Sample 2	21-28	EFIAWLVK	/	/	/
		EFIAWLV**K**	+16Da	37.0%	31.4%
YN-011 Sample 3	21-28	EFIAWLVK	/	/	/
		EFIAWLV**K**	+16Da	30.0%	31.1%
Dulaglutide	21-28	EFIAWLVK	/	/	/
		EFIAWLV**K**	+16Da	8.2%	4.4%

**Note: UV280 and EIC methods were used to detect the K34 modification levels in YN-011.*

2.2 K34 Hydroxylation Level in YN-011

After conducting mass spectrometry analysis and confirming that the +16Da modification occurs at the 34th position of lysine (Lys), considering that +16Da corresponds to the molecular weight of an oxygen atom, it suggests the possibility of an oxidation reaction. In fact, there are two enzymes in cells that catalyze the modification of lysine side chains. One is called lysyl oxidase (Lysyl oxidase, see reference https://en.wikipedia.org/wiki/Lysyl_oxidase), and the other is lysyl hydroxylase (Lysyl hydroxylase, see reference https://en.wikipedia.org/wiki/Lysyl_hydroxylase). Lysyl oxidase can oxidize the amino group of the lysine side chain at the 6th carbon, forming an aldehyde group. However, this modification only results in a 1Da difference compared to the unmodified form, which does not match the observed +16Da modification. On the other hand, lysyl hydroxylase adds a hydroxyl group to the carbon (γ position) of the lysine side chain, forming a stable hydroxylated lysine (hydroxylysine). Hydroxylation of lysine increases the molecular weight by 16Da, as it adds a hydroxyl group and reduces a hydrogen atom on the side chain. Therefore, the observed +16Da modification is consistent with lysine hydroxylation. The hydroxylation site at position K34 is shown below: