Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3045596 A1 2018/06/07

(21) 3 045 596

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2017/12/01
(87) Date publication PCT/PCT Publication Date: 2018/06/07
(85) Entrée phase nationale/National Entry: 2019/05/30
(86) N° demande PCT/PCT Application No.: JP 2017/043219
(87) N° publication PCT/PCT Publication No.: 2018/101454
(30) Priorité/Priority: 2016/12/02 (JP2016-234808)

(51) Cl.Int./Int.Cl. *C12N 15/09* (2006.01), *C12N 1/21* (2006.01), *C12N 9/42* (2006.01), *C12P 21/02* (2006.01)

(71) Demandeur/Applicant:
DAIICHI SANKYO COMPANY, LIMITED, JP

(72) Inventeurs/Inventors:
ITO, HANAKO, JP;
ONO, YASUNORI, JP;
NAKAMURA, KENSUKE, JP

(74) Agent: MARKS & CLERK

(54) Titre : NOUVELLE ENDO-β-N-ACETYLGLUCOSAMINIDASE
(54) Title: NOVEL ENDO-β-N-ACETYLGLUCOSAMINIDASE

(57) Abrégé/Abstract:
Provided are: a novel endo-β-N-acetylglucosaminidase that is isolated from a fungus belonging to the genus Rhizomucor and capable of maintaining the activity under high temperature conditions; various mutant enzymes thereof; a gene encoding the enzyme; a recombinant plasmid; a transformant transformed by the recombinant plasmid, etc.

Abstract

The present invention provides a novel endo-β-N-acetylglucosaminidase that is isolated from a fungus belonging to the genus *Rhizomucor* and is active under high-temperature conditions; various mutant enzymes thereof; genes encoding the enzymes; a recombinant plasmid; a transformant transformed with the plasmid; and the like.

Description

Title of Invention: NOVEL ENDO-β-N-ACETYLGLUCOSAMINIDASE

Technical Field

[0001]

The present invention relates to an endo-β-N-acetylglucosaminidase that is active under high-temperature conditions, a gene encoding the enzyme, a recombinant plasmid, a transformant transformed with the plasmid, and the like.

Background Art

[0002]

Glycoproteins are widely found in tissues of animals and plants, cell membranes and the walls of eukaryotic microorganisms, and the like. Recently, it has been revealed that sugar chains in glycoproteins have important roles in mechanisms such as cell differentiation, carcinogenesis, and intercellular recognition. To elucidate these mechanisms, studies on the correlation between the structure and function of sugar chains have been pursued. In drug discovery studies, attempts such as remodeling of sugar chains, in which sugar chains in glycoproteins including antibodies are substituted with a uniform structure of sugar chain, and glycosylation of peptides or small molecules has been

pursued. Such drug discovery studies often use sugar chains cleaved from naturally-occurring glycoproteins/glycopeptides with a uniform sugar chain using an enzyme such as endo-β-N-acetylglucosaminidase.

[0003]

Representative sugar chains in animal glycoproteins include N-linked sugar chains attached to asparagine side chains. N-linked sugar chains are classified into high-mannose, hybrid, and complex types depending on their structure but have a chitobiose structure having two GlcNAc residues linked to the reducing end as a common structure. Endo-β-N-acetylglucosaminidase is an enzyme having both an activity that hydrolyzes the glycosidic linkage in chitobiose structures and transglycosylation activity that transfers the cleaved sugar chain onto an acceptor having a specific structure. Endo-β-N-acetylglucosaminidases, which have been isolated from various biological species, have respective different substrate specificities. Different endo-β-N-acetylglucosaminidases have been used for different purposes. Among these, endo-β-N-acetylglucosaminidases that use complex-type sugar chains as substrates have been reported to include those as described below.

[0004]

Endo-M, which is an enzyme derived from *Mucor hiemalis*, has a substrate specificity that is an activity of 4.4% on a complex-type biantennary sugar chain

(agalacto biantennary PAsugar) when the activity on high-mannose-type Man8GlcNAc2 was set to 100% (Non Patent Literature 1: Fujita et al., (2004) Arch Biochem Biophy. 432: p41-49). It is reported in the same literature that Endo-M was expressed as an insoluble aggregate in all of the inductions of expression at 37°C in *Escherichia coli* and the induction temperature being changed to 20°C resulted in low enzymatic activity in a soluble fraction and thus the expression of Endo-M was attempted in yeast. Thus, it is believed that it is difficult to express Endo-M suitably in *E. coli*. Endo-M is also known to inactivate at 40°C or more.

[0005]

It has been reported that Endo-Om, which is an enzyme derived from the yeast *Ogataea minuta* (Patent Literature 1: WO2013/051608 or US2014-0313246), uses complex-type sugar chains as substrates and has an optimum temperature of 50°C in a hydrolytic reaction. The expression of Endo-Om in organisms except yeast is unknown.

[0006]

Endo-F2 and Endo-F3 are enzymes derived from *Elizabethkingia miricola* (Non Patent Literature 2: Tarentino AL et al., (1993) J Biol Chem. 268: p9702-9708). Endo-F2 hydrolyzes high-mannose-type and biantennary complex-type sugar chains and has no hydrolytic activity on hybrid-type sugar chains. On the other hand, Endo-F3

hydrolyzes biantennary or triantennary complex-type sugar chains and has no hydrolytic activity on high-mannose-type and hybrid-type sugar chains.

[0007]

Endo-S, which is an enzyme derived from *Streptococcus pyogenes*, hydrolyzes only biantennary complex-type sugar chains and has no hydrolytic activity on high-mannose-type and hybrid-type sugar chains (Non Patent Literature 3: Goodfellow JJ et al., (2012) J Am Chem Sci. 134: p8030-8033).

[0008]

Endo-CE, which is an enzyme derived from *Caenorhabditis elegans*, hydrolyzes high-mannose-type and biantennary complex-type sugar chains and it is unknown whether Endo-CE can cleave hybrid-type sugar chains (Non Patent Literature 4: Kato T et al., (2002) Glycobiology 12: p581-587). Endo-CE is reported to have an optimum temperature of 20°C in a hydrolytic reaction.

[0009]

In contrast to Endo-M and Endo-Om known so far, Endo-CC is an enzyme that reportedly can be expressed in *E. coli*. Non Patent Literature 5 (Y. Eshima et al., (2015) PLoS One. 21; 10(7): e0132859) describes that Endo-CC was expressed in *E. coli* in an amount of 0.1 mg/250 mL of culture (= 0.4 mg/L of culture). Endo-CC is reported to have an optimum temperature of 35°C in a hydrolytic reaction.

[0010]

In the light of improvement of reaction efficiency and prevention of contamination, enzymes that are active at a high temperature of around 50°C are desired when cleaved sugar chains are used as raw materials of pharmaceuticals. It is also said that when a plurality of recombinant materials are used as raw materials of pharmaceuticals, the same host species should be used to reduce safety concerns in pharmaceuticals. It is desirable to have an ability to produce in an amount greater than a usual amount in *E. coli* because *E. coli* is often selected as a host for producing such raw materials. However, an enzyme that is active at a high temperature and has the ability to be produced in a high yield in *E. coli* remains undiscovered among known endo-β-N-acetylglucosaminidases that use complex-type sugar chains as substrates.

[0011]

Rhizomucor pusillus (*R. pusillus*), which is a species of thermophilic fungi, has an optimal growth temperature of 35-45°C and is known to be a fungus that produces rennet for cheese production. The genomic sequence of *R. pusillus* including sequence information from *R. pusillus* strain CBS 183.67 is published in the database provided by The Genozymes Project (Concordia University). This database includes a gene having 41.56% homology to Endo-M and the amino acid sequence (SEQ ID

NO: 7) encoded by the gene but does not include annotation about the activity of the amino acid sequence. There is no report so far of actually obtaining endo-β-N-acetylglucosaminidase derived from *R. pusillus*. The amino acid sequence from positions 191-198 of the amino acid sequence of SEQ ID NO: 7 (hereinafter referred to as the "known putative sequence") assumed to be endo-β-N-acetylglucosaminidase in the database is Leu-Ala-Asn-Thr-Tyr-Tyr-Ile-Arg (LANTYYIR).

Citation List

[0012]

Patent Literature

Patent Literature 1: WO2013/051608 or US2014-0313246

Non Patent Literature

[0013]

Non Patent Literature 1: Fujita et al., (2004) Arch Biochem Biophy. 432: p41-49

Non Patent Literature 2: Tarentino AL et al., (1993) J Biol Chem. 268: p9702-9708

Non Patent Literature 3: Goodfellow JJ et al., (2012) J Am Chem Sci.134: p8030-8033

Non Patent Literature 4: Kato T et al., (2002) Glycobiology 12: p581-587

Non Patent Literature 5: Y. Eshima et al., (2015) PLoS One. 21; 10(7): e0132859

Summary of Invention

Technical Problem

[0014]

The present invention aims to provide a novel endo-β-N-acetylglucosaminidase that can be produced by *E. coli* and has hydrolytic activity on complex-type sugar chains under high-temperature conditions.

Solution to Problem

[0015]

The present inventors have conducted intensive studies directed towards achieving the aforementioned object. As a result, the inventors have found that culture supernatants from a plurality of strains belonging to *Rhizomucor pusillus* have good hydrolytic activity on complex sugar chains under high-temperature conditions and that endo-β-N-acetylglucosaminidase cloned from the strains exhibits good expression efficiency in an *E. coli* production system and the produced enzyme has the aforementioned hydrolytic activity. The present inventors further pursued the study, thereby completing the present invention.

[0016]

The present invention provides the following aspects of the invention.

[0017]

(1) A polypeptide having the following properties (A) and (B):

(A) the polypeptide comprises an amino acid sequence having 75% or more identity and 95% or more similarity to the amino acid sequence shown in SEQ ID NO: 1 and is an amino acid sequence different from the amino acid sequence of SEQ ID NO: 7; and

(B) the polypeptide exhibits, at any temperature from 45 to 60°C, 40% or more of the maximal activity value of hydrolytic activity and/or transglycosylation activity on complex sugar chains.

[0018]

(2) The polypeptide according to (1), having the following properties (A) and (B):

(A) the polypeptide comprises an amino acid sequence having 75% or more identity and 95% or more similarity to the amino acid sequence shown in SEQ ID NO: 1 and an amino acid sequence of Leu-Ala-Lys-Leu-Leu (LAKLL) at positions corresponding to the amino acid sequence from amino acid positions 191 to 195 of SEQ ID NO: 1;

(B) the polypeptide exhibits, at any temperature from 45 to 60°C, 40% or more of the maximal activity value of hydrolytic activity on complex sugar chains.

[0019]

(3) The polypeptide according to (1), wherein the hydrolytic activity on complex sugar chains at 50°C is 60% or more of the maximal activity value.

[0020]

(4) The polypeptide according to any of (1) to (3), wherein, in addition to having the properties (A) and (B), (C) the polypeptide is produced in an amount of 10 mg/L of culture or more in recombinant expression in *E. coli*.

[0021]

(5) The polypeptide according to any of (1) to (4), wherein the polypeptide has 85% or more sequence identity to the region from amino acid positions 54 to 341 of the amino acid sequence of SEQ ID NO: 1.

[0022]

(6) The polypeptide according to any of (1) to (5), wherein at least one of amino acids D276, V223, W225, Y247, and W248 remains unchanged and preferably at least D276 remains unchanged.

[0023]

(7) The polypeptide according to any of (1) to (3), wherein the polypeptide is a polypeptide consisting of the amino acid sequence of SEQ ID NO: 1 or a polypeptide consisting of an amino acid sequence having at least one mutation selected from the group consisting of A128T, D333G, I434L, V460A, I527V, K569T, F610S, and H626R in the amino acid sequence of SEQ ID NO: 1.

[0024]

(8) The polypeptide according to (1), wherein the polypeptide consists of an amino acid sequence of any of SEQ ID NOS: 1 to 6.

[0025]

(9) The polypeptide according to any of (1) to (6), wherein the polypeptide has an amino acid sequence satisfying the property (A) and has at least one of the mutations contained in SEQ ID NO: 23.

[0026]

(10) The polypeptide according to (9), wherein the polypeptide has a mutation in at least one amino acid selected from N172, D176, Y214, S216, L245, N246, T275, L306, F307, and A310 and has increased transglycosylation activity.

[0027]

(11) The polypeptide according to (7), wherein the polypeptide has at least one mutation selected from the group of:

a mutation in which N172 is substituted with Gln, Asp, Gly, Ala, Phe, Cys, His, Ile, Ser, Thr, Val, or Met; a mutation in which D176 is substituted with Arg; a mutation in which Y214 is substituted with Phe; a mutation in which S216 is substituted with Val; a mutation in which L245 is substituted with Ser; a mutation in which N246 is substituted with Asp; a mutation in which T275 is substituted with Ile; a mutation in which F283 is substituted with Ser; a mutation in which L306 is substituted with Ile; a mutation in which F307 is substituted with Tyr; a

mutation in which A310 is substituted with Asp; and a mutation in which E314 is substituted with Gln.

[0028]

(12) The polypeptide according to any of (9) to (11), wherein the polypeptide has a mutation in which W278 is substituted with Phe or Tyr. The polypeptide according to (11), wherein, more specifically, the polypeptide has mutation N172Q, N172D, W278F, N172Q/W278F, N172D/W278F, or Y214/L306I/L307Y in the amino acid sequence of any of SEQ ID NOS: 1 to 6.

[0029]

(13) A polynucleotide encoding the polypeptide according to any of (1) to (12).

[0030]

(14) The polynucleotide according to (13), having a nucleotide sequence of any of the nucleotide sequences from nucleotide positions 1 to 2088 of SEQ ID NOS: 8 to 17 and the nucleotide sequences from nucleotide positions 1 to 2091 of SEQ ID NOS: 18 to 19.

[0031]

(15) An expression plasmid comprising the polynucleotide according to (13) or (14).

[0032]

(16) A host cell transformed with the plasmid according to (15).

[0033]

(17) The host cell according to (16), wherein the host cell is *E. coli* transformed with a plasmid comprising a polynucleotide comprising the nucleotide sequence from nucleotide positions 1 to 2088 of SEQ ID NO: 9, 11, 13, 15, or 17 or the nucleotide sequence from nucleotide positions 1 to 2091 of SEQ ID NO: 19.

[0034]

(18) A method of producing the polypeptide according to any of (1) to (12), comprising culturing the host cells according to (16) or (17) and collecting the polypeptide according to any of (1) to (12) from the resulting culture.

[0035]

(19) A reagent comprising the polypeptide according to any of (1) to (12).

Advantageous Effects of Invention

[0036]

The novel endo-β-N-acetylglucosaminidases of the present invention have good hydrolytic activity on complex-type sugar chains under high-temperature reaction conditions of 50°C or more so that they can safely and efficiently provide sugar chains under high-temperature conditions, which can prevent bacterial growth, in the manufacture of pharmaceuticals using these enzymes and can lead to high reaction efficiency. The enzymes of the present invention can be produced in a heterologous

expression system using *E. coli* in high yield. In cases where biological raw materials (such as a bioactive peptide/protein and other enzymes) other than the present enzymes are produced by *E. coli* in the manufacture of pharmaceuticals, use of the same host species in production of the present enzymes can facilitate assessment of the effect of the host species, from which the biological raw materials originate, on the safety of the final product.

Brief Description of Drawings

[0037]

[Figure 1] Figure 1 is a schematic representation showing a hydrolytic reaction in which Endo-Rp uses SGP as a substrate.

[Figure 2] Figure 2 represents charts showing the results of LC-MS analysis of SGP solution before enzyme treatment (upper panel) and crude solution after enzyme treatment (lower panel) under analysis condition A. The large peak detected at 3 to 4 minutes represents SGP and the peaks detected around 4.5 to 5 minutes represent SG(10).

[Figure 3] Figure 3 is a graph showing the temperature dependence of hydrolytic activities of crude enzymes derived from various *R. pusillus* strains (NBRC 9740 (open triangle), NBRC 9741 (filled triangle), NBRC 9742 (open square), and NBRC 9743 (filled square)) and Endo-M (filled circle) on SGP. X axis represents the reaction

temperature (°C) and Y axis represents the hydrolysis rate.

[Figure 4] Figure 4 shows the alignment of amino acid sequences of Endo-Rp and various homologs thereof, Endo-Rm and the known putative sequence.

[Figure 5] Figure 5 is a graph showing the time-dependent change of hydrolytic activities of Endo-Rp (open circle), Endo-M (filled circle), Endo-S (filled square), and Endo-Om (filled triangle) on SGP. X axis represents the duration after the start of reaction and Y axis represents the hydrolysis rate.

[Figure 6] Figure 6 is a graph showing the temperature dependence of hydrolytic activity of Endo-Rp on SGP. X axis represents the reaction temperature (°C) and Y axis represents the hydrolysis rate.

[Figure 7] Figure 7 is a graph showing the pH dependence of hydrolytic activity of Endo-Rp on SGP. X axis represents the pH value of the reaction solutions and Y axis represents the hydrolysis rate.

[Figure 8] Figure 8 represents charts showing the results of LC-MS analysis of the reaction solution of Endo-Rp in the presence of SGP alone (upper panel) or SGP + acceptor ((GlcNAc-)Asn) (lower panel). The large peak detected near 1 minute represents SGP, and the peaks detected at around 4 to 5 minutes represent SG(10), and the peak detected at around 3 minutes represents (SG-)Asn.

[Figure 9] Figure 9 is a graph showing the time-dependent change of transglycosylation activities of Endo-Rp (open circle) and Endo-Rp N172Q (filled square). X axis represents the duration (hour) after the start of reaction and Y axis represents the transglycosylation rate (%).

Description of Embodiments

[0038]

The present invention will now be described in detail.

[0039]

The present invention provides an endo-β-N-acetylglucosaminidase which is a polypeptide having properties (A) and (B):

(A) the polypeptide comprises an amino acid sequence having 75% or more identity and 95% or more similarity to the amino acid sequence shown in SEQ ID NO: 1 (Endo-Rp amino acid sequence) and is an amino acid sequence different from the amino acid sequence of SEQ ID NO: 7; and

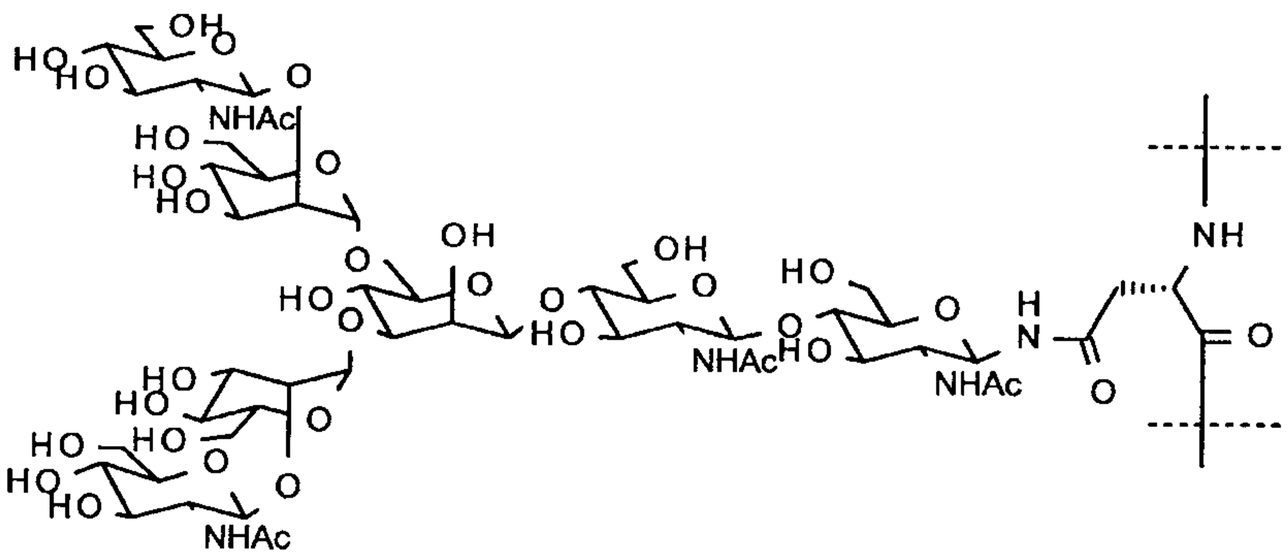
(B) the polypeptide exhibits 40% or more of the maximal activity value of hydrolytic activity and/or transglycosylation activity on complex-type sugar chains at a temperature ranging from 45 to 60°C.

[0040]

In the present invention, "complex-type sugar chain" means a sugar chain that is among human N-linked sugar chains and has a basic structure consisting of formulae (I) or (II) as described below. The complex-type sugar chain has a structure in which each of two branched chains (1-3 chain and 1-6 chain), branched from mannose (β mannose) near the reducing end, has GlcNAc. The structure will vary depending on the presence or absence of galactose and sialic acid at the non-reducing end as well as its valence isomerism and position isomerism. As long as the complex-type sugar chain has this basic structure, it may also have another branched structure or a structure that has been chemically modified at some of the hydroxy groups of a carbohydrate or a carbonyl group of sialic acid at the non-reducing end. It is reported that, as an example of such a chemically modified complex-type sugar chain, SGP modified by oximation after oxidative cleavage of diol in the sialic acid of SGP can be used as an Endo-M substrate (Org. Biomol. Chem, 2016, 14, 9501-9518).

[0041]

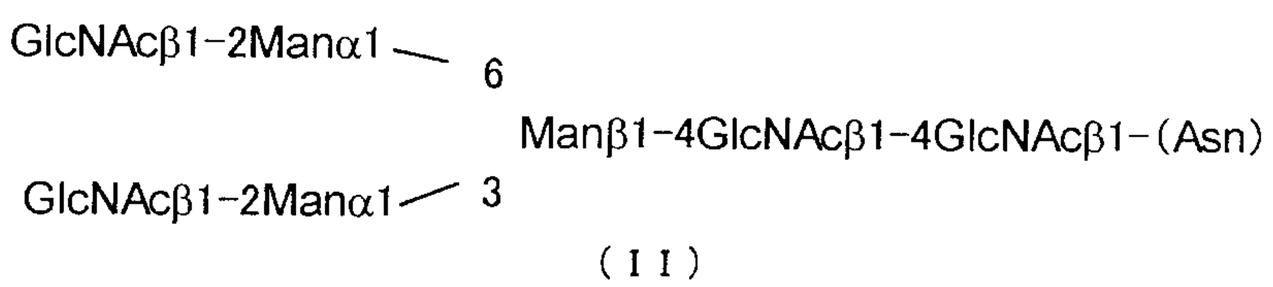
[Formula (I)]



(I)

[0042]

[Formula (II)]



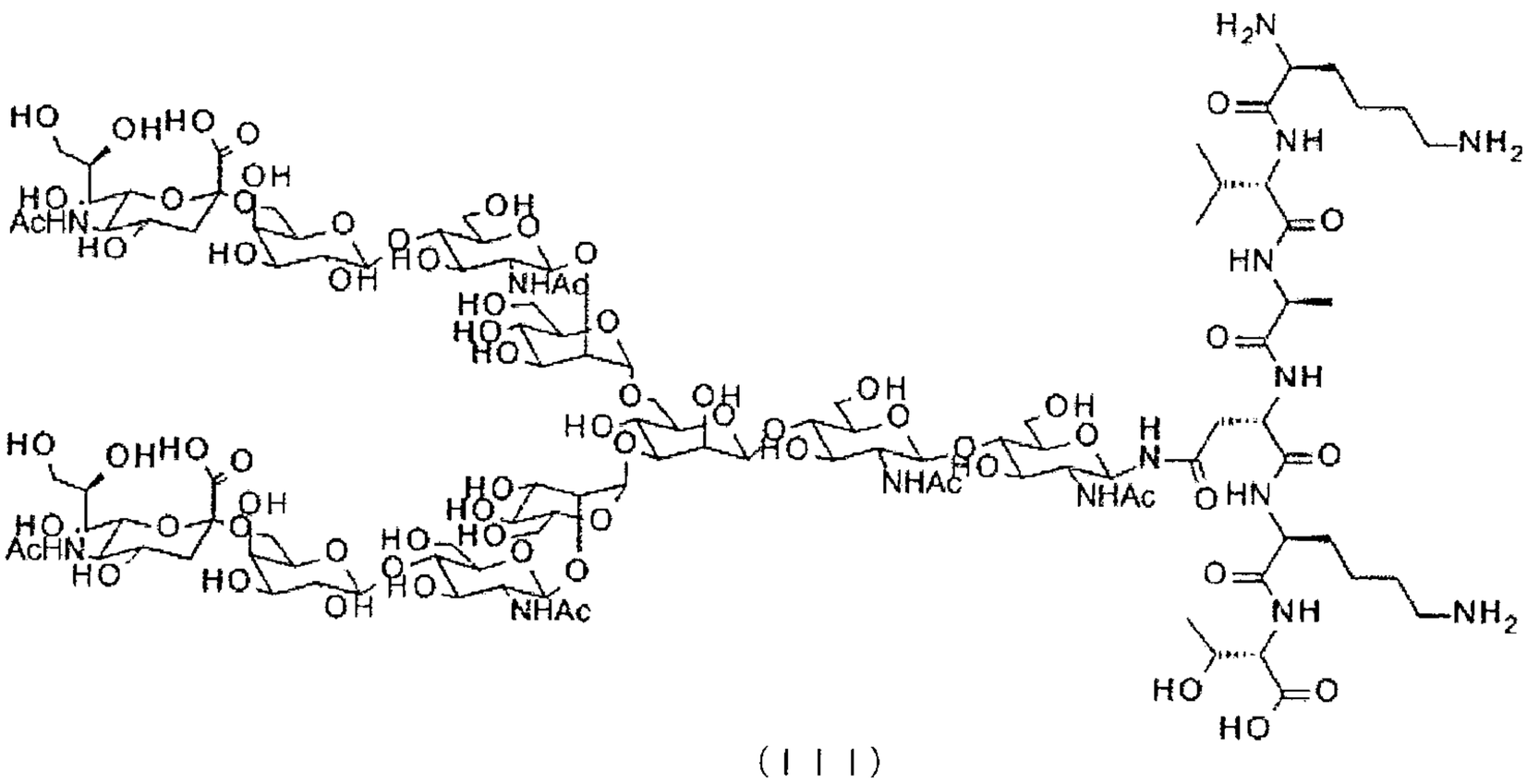
(I I)

[0043]

Complex-type sugar chains can typically include a sialylglycan (SG) included in sialylglycopeptide (SGP: formula (III) and (IV) below) extracted from chicken egg yolk. SGP can be purified from avian egg yolk but purified SGP is commercially available and can be purchased, for example, from Tokyo Chemical Industry Co., Ltd. or the like.

[0044]

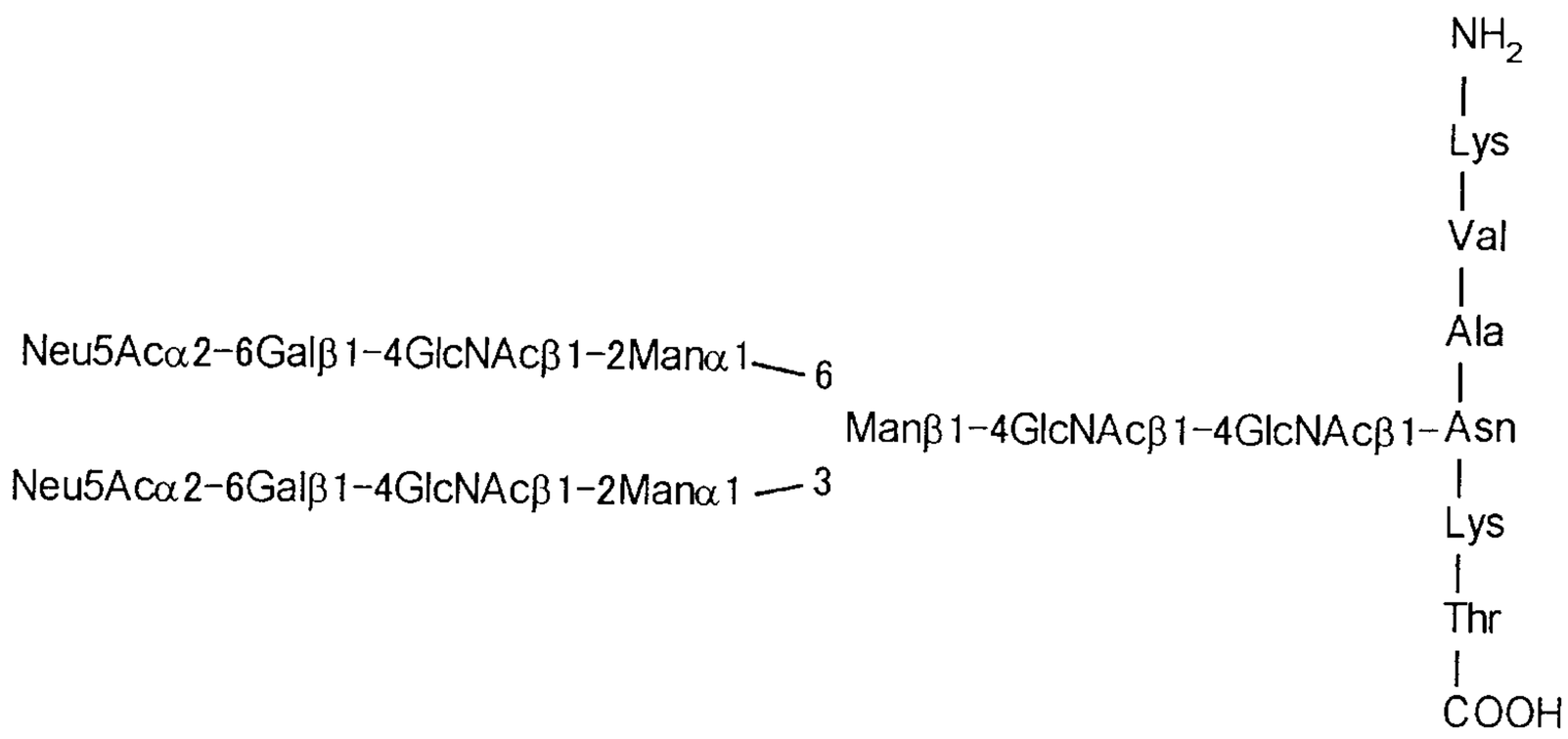
[Formula (III)]



(III)

[0045]

[Formula (IV)]



(IV)

[0046]

In the present invention, "endo-β-N-acetylglucosaminidase" is an enzyme that recognizes a sugar chain as its substrate and has both hydrolytic activity and transglycosylation activity. The amino acid sequence of naturally-occurring endo-β-N-

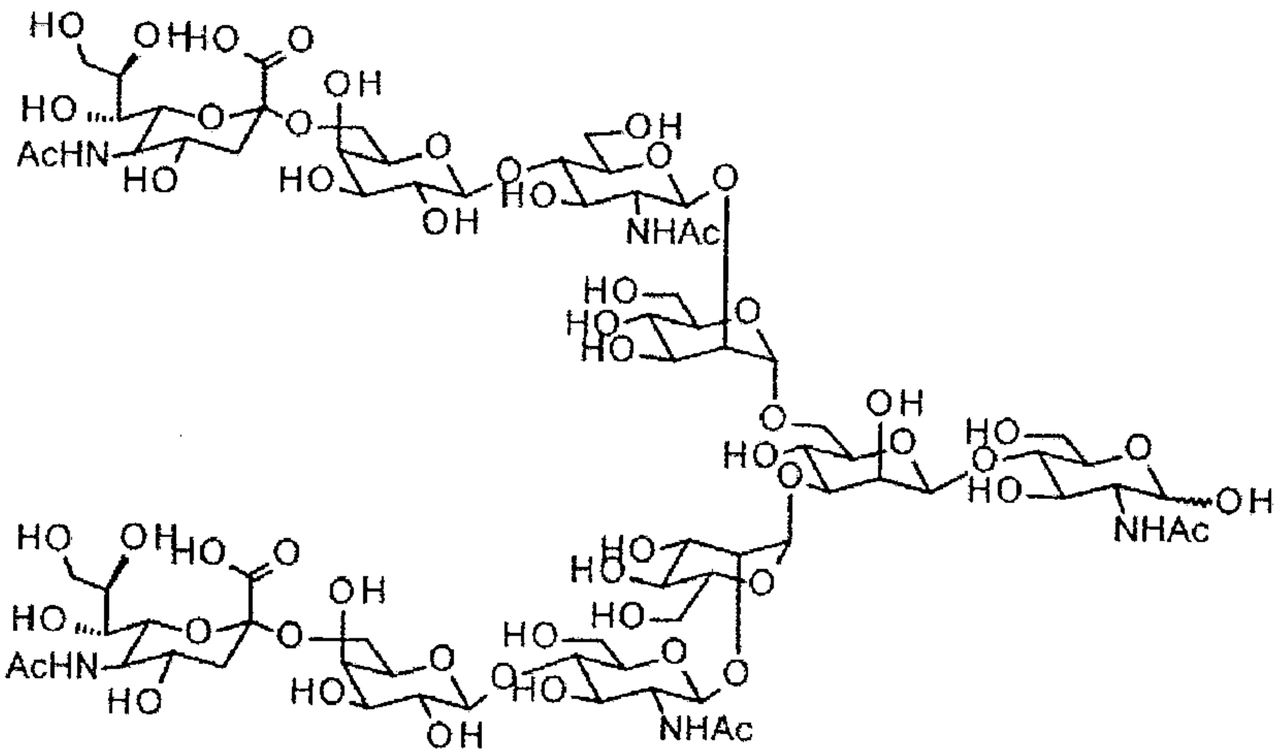
acetylglucosaminidase, which has both of these activities, can be modified to generate a mutant adjusted to enhance or reduce either one of the activities and an enzyme that has only hydrolytic activity or transglycosylation activity by allowing either one of the activities to disappear. The present invention includes not only the naturally-occurring enzymes but also such mutant enzymes.

[0047]

The hydrolytic activity of the enzymes of the present invention is an activity that specifically hydrolyzes a β1,4 glycosidic linkage in core chitobiose consisting of two consecutive GlcNAc units at the reducing end of the complex-type sugar chain described above (as used herein, "hydrolytic activity" means this activity unless stated otherwise). For example, as shown in Figure 1, in a hydrolytic reaction in which endo-β-N-acetylglucosaminidase uses SGP as a substrate, SG(10) (structures of formula (V) and (VI) below) consisting of a structure without one GlcNAc at the reducing end of SG is produced.

[0048]

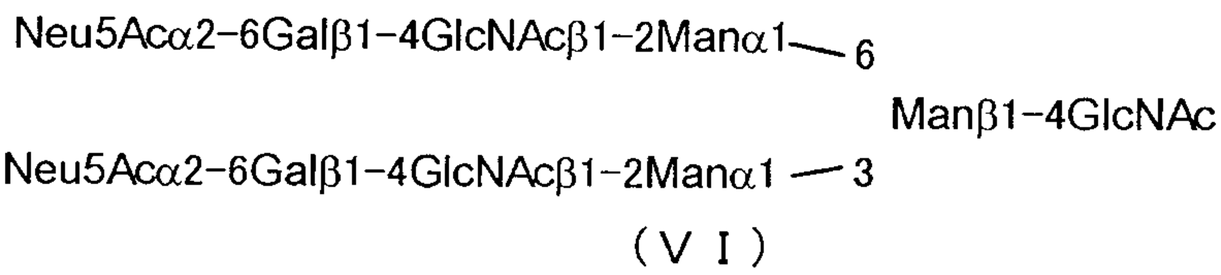
[Formula (V)]



(V)

[0049]

[Formula 6]



(VI)

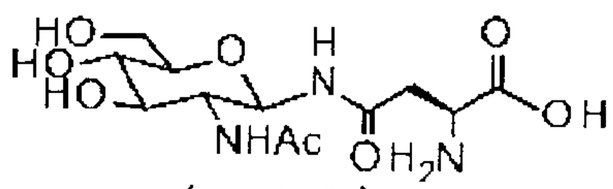
[0050]

The transglycosylation activity of the present enzymes is an activity (hereinafter referred to as "transglycosylation activity") that forms β1,4 glycosidic linkage by binding the reducing end of the sugar chain derived from a sugar chain donor, wherein the reducing end has GlcNAc, to a molecule having only GlcNAc as a carbohydrate unit or a molecule comprising a sugar chain having GlcNAc at the non-reducing end (hereinafter referred to as "acceptor molecule"). For example, in a transglycosylation reaction that uses (GlcNAc-)Asn having

the structure represented by formula (VII) below as an acceptor molecule and SGP as a donor molecule, SG(10) derived from SGP is transferred to GlcNAc unit of the acceptor molecule to produce (SG-)Asn represented by formula (VIII) below. Similarly, in a transglycosylation reaction that uses GlcNAc-AcA having the structure represented by formula (IX) below as an acceptor molecule and SGP as a donor molecule, SG-A represented by formula (X) below is produced.

[0051]

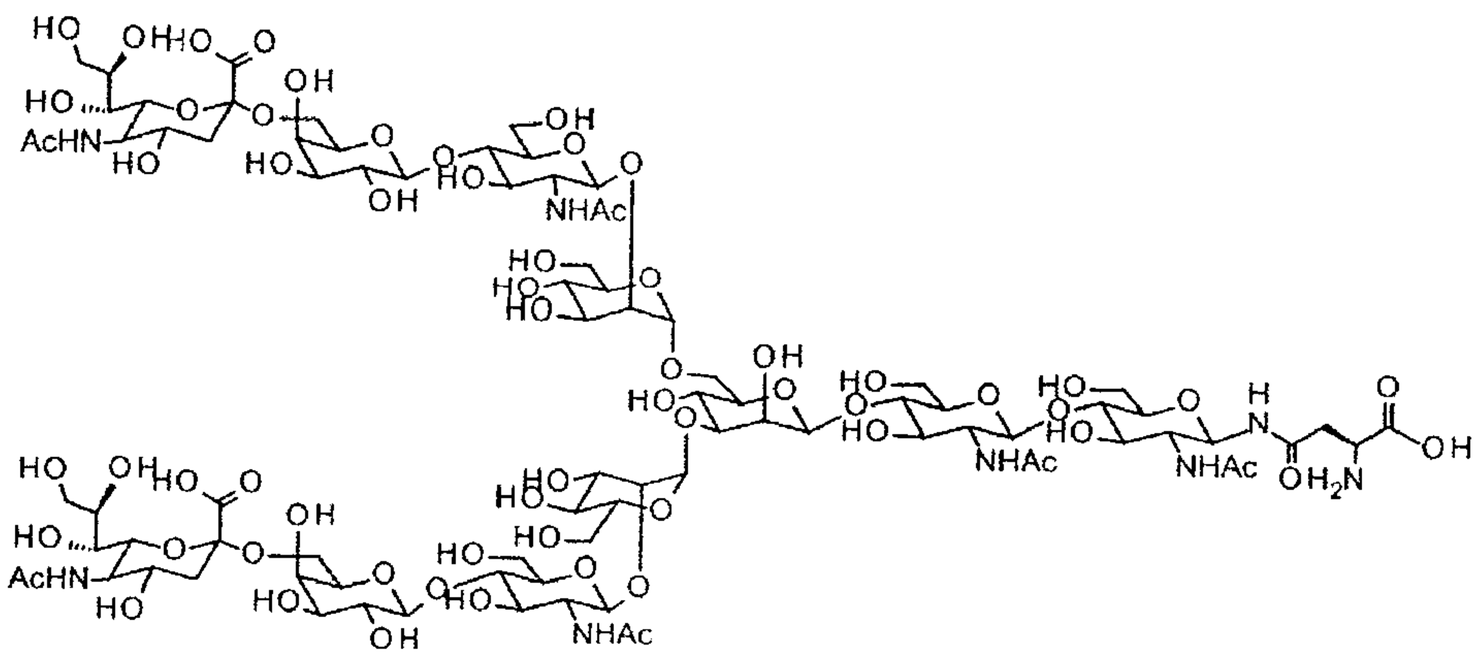
[Formula (VII)]



(VII)

[0052]

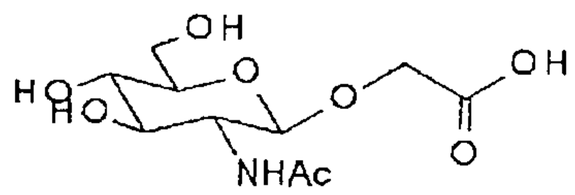
[Formula (VIII)]



(VIII)

[0053]

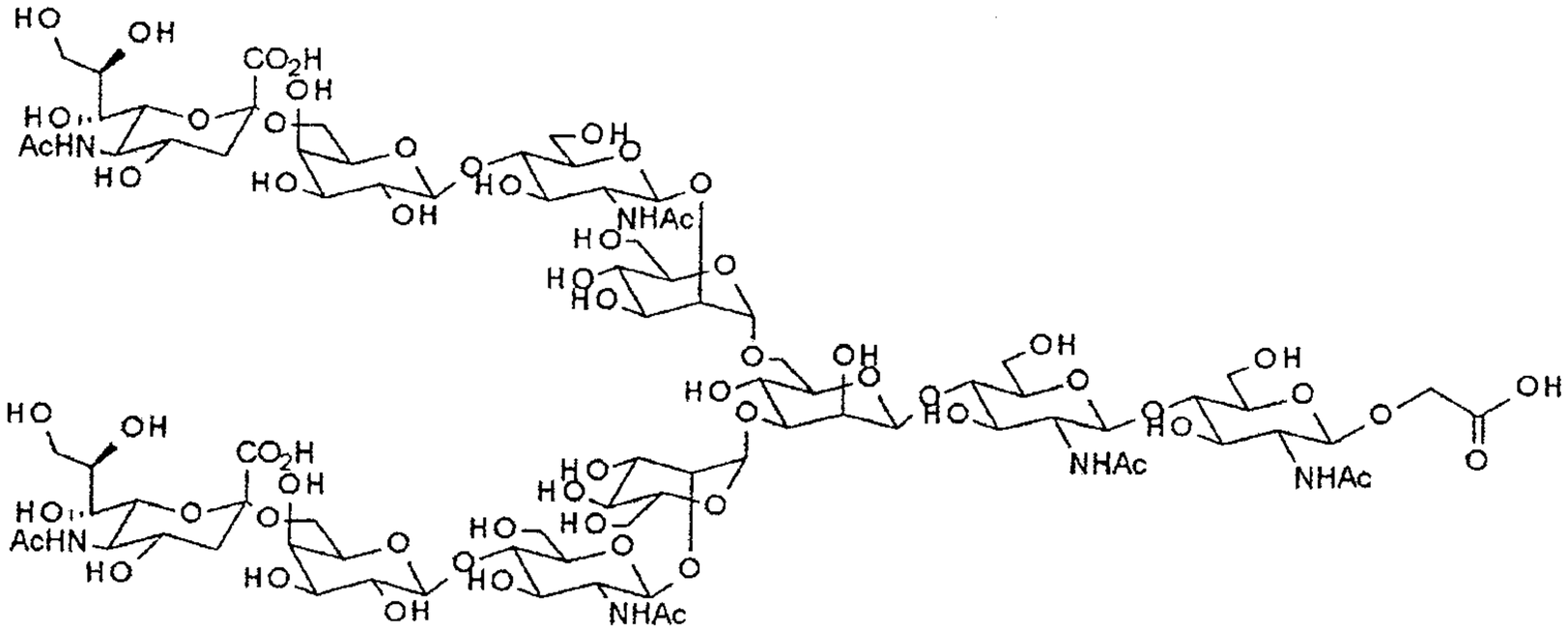
[Formula (IX)]



(IX)

[0054]

[Formula (X)]



(X)

[0055]

Enzymes and amino acid sequences

The enzymes of the present invention are not limited to enzymes that were obtained in the Examples and have specific sequences as long as they have properties as described above. The enzymes of the present invention may be isolated from naturally occurring sources or artificially produced or modified based on the sequence information of the enzymes of the present invention. For isolation from naturally occurring sources, biological species that are used as isolation sources preferably include, but are not particularly limited to, fungi, more preferably thermophilic fungi, still more preferably fungi belonging to the genus *Rhizomucor*, yet still more

preferably fungi belonging to *Rhizomucor pusillus* or *Rhizomucor miehei*.

[0056]

In the present invention, endo-β-N-acetylglucosaminidases having the present properties are cloned from a plurality of strains belonging to *Rhizomucor pusillus* (*R. pusillus*) which are thermophilic fungi. The enzyme derived from *R. pusillus* strain NBRC 9742 is named as Endo-Rp (amino acid sequence: SEQ ID NO: 1, nucleic acid sequence derived from the strain: SEQ ID NO: 8); the enzyme derived from strain NBRC 9740 is named as Endo-Rp2 (amino acid sequence: SEQ ID NO: 2, nucleic acid sequence derived from the strain: SEQ ID NO: 10); the enzyme derived from strain NBRC 9741 is named as Endo-Rp3 (amino acid sequence: SEQ ID NO: 3, nucleic acid sequence derived from the strain: SEQ ID NO: 12); and the enzyme derived from strain NBRC 9743 is named as Endo-Rp4 (amino acid sequence: SEQ ID NO: 4, nucleic acid sequence derived from the strain: SEQ ID NO: 14). All of the strains of *R. pusillus* described above are available from NBRC. The place of origin of each of the strains is Japan.

[0057]

The amino acid sequence of SEQ ID NO: 7 (its putative nucleic acid sequence being: SEQ ID NO: 20) is published as the putative amino acid sequence of endo-β-N-acetylglucosaminidase based on the sequence information

of the *R. pusillus* strain with its published genome. Optimization of the nucleic acid sequence for an *E. coli* expression system to express it in *E. coli* resulted in little expression and very low enzymatic activity of the enzyme (Example 3). In the amino acid sequence of SEQ ID NO: 7, the amino acid corresponding to Lys at position 193 of SEQ ID NO: 1 is substituted with amino acids Asn-Thr-Tyr-Tyr-Ile-Arg (see Figure 4). An enzyme consisting of the amino acid sequence (SEQ ID NO: 5) in which the Asn-Thr-Tyr-Tyr-Ile-Arg from positions 193 to 198 in SEQ ID NO: 7 has been substituted with Lys as in SEQ ID NO: 1 showed adequate expression and hydrolytic activity in *E. coli* and the enzyme having this amino acid sequence is named as Endo-Rp5. In conclusion, the sequence corresponding to positions 191 to 195 of SEQ ID NO: 1 is identified as a region very important for the properties of the enzymes of the present invention.

[0058]

Based on the sequence information of Endo-Rp identified in the present invention, a related sequence was also identified from the published genomic information (*R. miehei* strain CAU432) of *Rhizomucor miehei* (*R. miehei*), a related species of *R. pusillus*. Optimization of the nucleic acid sequence for *E. coli* expression to express the protein in question in *E. coli* resulted in production of the enzyme having the enzymatic activity of the present invention to some degree. This

enzyme was named as Endo-Rm (amino acid sequence: SEQ ID NO: 6, nucleic acid sequence derived from the fungus: SEQ ID NO: 18).

[0059]

Amino acid sequence identity between Endo-Rp and its homologs, Endo-Rp2, Endo-Rp3, Endo-Rp4, and Endo-Rp5 is 99% or more. Endo-Rp2 is a homolog having the mutations A128T and K569T in SEQ ID NO: 1. Endo-Rp3 is a homolog having the mutations D333G, I434L, I527V, K569T, F610S, and H626R in SEQ ID NO: 1. Endo-Rp4 is a homolog having the mutations V460A and K569T in SEQ ID NO: 1. Endo-Rp5 is a homolog having the mutation K569T in SEQ ID NO: 1. The amino acid sequence of Endo-Rm has 96% similarity and 77% identity to the amino acid sequence of SEQ ID NO: 1. The region from positions 54 to 340 in this sequence has a high level of identity (identity: 89%, similarity: 97%)) and the region from positions 190 to 195, His-Leu-Ala-Lys-Leu-Leu (HLAKLL) is completely identical, which satisfies property (A) of the enzymes of the present invention.

[0060]

When the enzyme of the present invention is an enzyme derived from fungi, the protein may be isolated from culture supernatants of fungi or fungi homogenates or expressed using a heterologous expression system such as *E. coli* or yeast. It is known that many known endo-β-N-acetylglucosaminidases are produced in a low yield

using a heterologous expression system in *E. coli*, whereas the amino acid sequence of the enzyme of the present invention has the property of showing excellent production efficiency in an *E. coli* expression system. Nucleic acid sequences of fungi-derived proteins are generally optimized for *E. coli* expression if the proteins are to be expressed in *E. coli*. Even when such optimization was performed, for example, Endo-M N175Q (single amino acid mutant) was produced in *E. coli* in a yield of only 3.7 mg/L of culture. It is also reported that Endo-CC was produced in a yield of 0.4 mg/L of culture. In contrast to this, the enzyme of the present invention is expressed in *E. coli* in an amount of 10 mg/L of culture or more (preferably 12 mg/L of culture or more, and more preferably 15 mg/L of culture or more) and shows a good production efficiency in *E. coli* as compared with conventional enzymes. As shown in Example 4, from the results of production efficiencies of the known putative sequence (SEQ ID NO: 7) and Endo-Rp5 (SEQ ID NO: 5) in *E. coli*, this property seems to be a property imparted by a certain amino acid sequence (particularly, a region comprising amino acid positions 191 to 195 of SEQ ID NO: 1, among others, the sequence from positions 192 to 194, and more particularly, Lys at position 193).

[0061]

The amino acid sequences of the enzymes of the present invention are each an amino acid sequence that

comprises an amino acid sequence that has 75% or more identity and 95% or more similarity to the full-length amino acid sequence of SEQ ID NO: 1 and is different from SEQ ID NO: 7 (preferably the amino acid corresponding to amino acid position 193 of SEQ ID NO: 1 is Lys; preferably the amino acid sequence corresponding to the amino acid sequence from amino acid positions 192 to 194 is Ala-Lys-Leu (AKL); more preferably the amino acid sequence corresponding to the amino acid sequence from amino acid positions 191 to 195 is Leu-Ala-Lys-Leu-Leu (LAKLL)).

[0062]

The identity of the amino acid sequence refers to a numeric value quantified from the match rate of amino acids, wherein if an amino acid exactly matches with the amino acid present at the corresponding position in the full-length sequence, these amino acids are considered to be the same amino acid. On the other hand, the similarity of amino acid sequence refers to a numeric value quantified from the relationship between two amino acid sequences, wherein if an amino acid has a property similar to that of the amino acid present at the corresponding position, these amino acids are considered to have similarity. Sequence identity and similarity in the present invention are calculated by the sequence analysis software GENETYX-SV/RC (manufactured by GENETYX CORPORATION). This algorithm is commonly used in the art.

[0063]

The active domain of Endo-Rp is expected to be a region from amino acid positions 1 to 374 of SEQ ID NO: 1 based on the alignment with Endo-A that has been analyzed for its crystal structure (Zhenlian Ling et al, Journal of Molecular Biology (2009), Vol. 389, No.1, Pages 1-9). In fact, the amino acid sequences of the enzymes of the present invention (Endo-Rp1 to 5 and Endo-Rm) have a sequence identity of 75% or more to SEQ ID NO: 1 for the full-length alignment. The sequence from amino acid positions 54 to 341 of SEQ ID NO: 1 in the region expected to be the active domain has a very high identity of 89%. The full-length amino acid sequence has an identity of preferably 80% or more, more preferably 90% or more, still more preferably 95% or more, and most preferably 99% or more. The region from amino acid positions 54 to 341 of SEQ ID NO: 1 has an identity of preferably 85% or more, more preferably 90% or more, further preferably 95% or more, even more preferably 98% or more, and most preferably is completely identical to the amino acid sequence in this region.

[0064]

In the specification, the notation of amino acids included in a molecule follows the practice in the art. A mutation site is represented by using one-letter notation of the wild-type amino acid (or nucleic acid) and the position of the mutation (for example, Asn at

position 172 is referred to as "N172"). A mutation is also represented by using one-letter notation of the wild-type amino acid (or nucleic acid), the position of the mutation, and one-letter notation of the amino acid (or nucleic acid) after mutation (for example, the mutation in which Asn at position 172 is substituted with Gln is referred to as "N172Q"). A mutant having a particular mutation is represented by using the molecule name and mutation (for example, the mutant having Asn at position 172 of Endo-Rp substituted with Gln is referred to as "Endo-Rp N172Q"). A mutant having a plurality of mutations is represented by the expression delimited by "/" between the plurality of mutations (for example, the Endo-Rp N172Q mutant having an additional mutation in which Trp at position 278 is substituted with Phe is referred to as "Endo-Rp N172Q/W278F").

[0065]

The enzymes of the present invention may include an amino acid mutation (substitution), deletion, insertion, and/or addition as long as they have a certain level or more of identity or similarity to SEQ ID NO: 1, provided that, at least in SEQ ID NO: 1, the sequence corresponding to positions 193 to 198 of SEQ ID NO: 7 is a sequence different from Asn-Thr-Tyr-Tyr-Ile-Arg and preferably the site corresponding to the amino acid at position 193 (preferably positions 192 to 194, more preferably positions 191 to 195) of SEQ ID NO: 1 has an

amino acid sequence completely identical to SEQ ID NO: 1. The results of Example 8 showed that both hydrolytic activity and transglycosylation activity almost disappeared when the D276 amino acid in SEQ ID NO: 1 was mutated and that each of the mutants E126A, V223R, W225H, Y247F, and W248N greatly reduced both of the activities. This indicates that these amino acids in the enzyme of the present invention are preferably amino acids similar or identical to those in SEQ ID NO: 1. Thus, the region in the enzyme of the present invention having the amino acid sequence completely identical to SEQ ID NO: 1 is preferably the region from amino acid positions 118 to 332, more preferably the region from positions 54 to 341, and further preferably the region from positions 1 to 374.

[0066]

The enzymes of the present invention may include substitution, deletion, insertion, and/or addition of some (preferably 10 or fewer, more preferably 7 or fewer, and still more preferably 5, 4, 3, 2, or 1) amino acids per site in some sites (preferably 5 sites or fewer, more preferably 3, 2, or 1 site) in any region in the amino acid sequences of SEQ ID NOS: 1 to 6, except completely identical regions described above, as long as they have the sequence identity/similarity described above. Such amino acid deletion may provide polypeptides having the properties of the enzymes of the present invention even when, for example, several amino acids are deleted from

the N- and/or C-terminus of any of the amino acid sequences of SEQ ID NOS: 1 to 6. Particularly, the region ranging from position 375 to the C-terminus of SEQ ID NO: 1 is not an active domain and therefore accepts many amino acid modifications (substitution, deletion, insertion, and/or addition). Polypeptides with an amino acid addition include those that have amino acids or peptides that are known not to affect the aforementioned activities added to the N- and/or C-terminus of any of the amino acid sequences of SEQ ID NOS: 1 to 6. Such peptides to be added include a tag peptide (such as a His tag and a GST tag) to be added for protein purification.

[0067]

For the position of an amino acid mutation in the enzymes of the present invention, it has been confirmed that an enzyme having a mutation in at least one position represented by Xaa in SEQ ID NO: 23 retains hydrolytic activity and/or transglycosylation activity. The present enzymes are expected to have the active domain at amino acid positions 1 to 374 in SEQ ID NO: 1. Mutants that keep original activities or are engineered to have the desired activities can be designed by making mutations in the active domain based on the findings of structure-activity relationships in known endo-enzymes and the disclosure of the present invention.

[0068]

The results from various homologs indicate that the enzymes of the present invention accept amino acid substitutions at A128, D333, I434, V460, I527, K569, F610, and H626 in SEQ ID NO: 1. Also, as shown in Example 8, many mutants with substitution of amino acids in the active domain were generated. Many of these mutants were confirmed to reduce apparent activities as compared with wild-type Endo-Rp but to have a certain level of hydrolytic activity and/or transglycosylation activity sufficient to be available for transglycosylation (the specific activity value of any one of the activities is at least 0.5% or more of that of wild-type Endo-Rp). It was also confirmed that certain amino acid substitutions are acceptable in this region. Specifically, the amino acid mutations described below are acceptable.

[0069]

Although N172 seems to be an active residue which contacts with the substrate, substitution with any amino acid except Trp, Arg, and Tyr allows the activities to be maintained and therefore substitution with any amino acid except amino acids with bulky side chains is widely acceptable. It has been confirmed that substitution with an amino acid having a highly polar side chain such as Gln and Asp; and Gly, Ala, Phe, Cys, His, Ile, Ser, Thr, Val, Met, or the like can increase the transglycosylation activity ratio (transglycosylation activity value /hydrolytic activity value).

[0070]

The substitution of D176 with Arg allowed the activities to be maintained and therefore the substitution with many amino acids may be acceptable. The substitution of D176 with Arg is also useful for generating mutants with an increased transglycosylation activity ratio.

[0071]

The substitution of Y214 with an amino acid having a small side chain like Ala greatly reduces both of the activities and therefore Y214 will accept substitution with amino acids having a relatively large side chain. The substitution of Y214 with Phe is useful for generating mutants with an increased transglycosylation activity ratio.

[0072]

The substitution of S216 with an amino acid having a small side chain (preferably Ala or Val) allows transglycosylation activity to be maintained and hydrolytic activity to be reduced. This mutation is useful for generating mutants with an increased transglycosylation activity ratio.

[0073]

The substitution of L245 with Ser maintained the activities and therefore the substitution of L245 with many amino acids may be acceptable. The substitution of

L245 with Ser is also useful for generating mutants with an increased transglycosylation activity ratio.

[0074]

Although N246 seems to resist acceptance of major change, the substitution of N246 with Asp greatly reduced hydrolytic activity while maintaining transglycosylation activity. Such a mutation is useful for generating mutants with an increased transglycosylation activity ratio.

[0075]

The substitution of T275 with Ile maintained the activities and therefore the substitution of T275 with many amino acids may be acceptable. The substitution of T275 with Ile is also useful for generating mutants with an increased transglycosylation activity ratio.

[0076]

It is believed that the hydrophobic interaction between W278 and the substrate sugar chain contributes to the activities. W278 may accept amino acids with a highly hydrophobic side chain, such as Tyr, Phe, Ala, Leu, and Ile.

[0077]

The substitution of F283 with Ser allowed the activities to be maintained and therefore F283 may accept substitution with many amino acids. The substitution of F283 with Ser is also useful for generating mutants with an increased transglycosylation activity ratio.

[0078]

The substitution of L306 with Ile causes less significant changes in the activities but tends to maintain transglycosylation activity and reduce hydrolytic activity and therefore the substitution may be useful for generating mutants with an increased transglycosylation activity ratio. The predicted conformation of the substitution product indicates that substitution with an amino acid with a bulky side chain or charged side chain may be unfavorable.

[0079]

F307 may accept various amino acids such as His and Tyr. The substitution of F307 with His or Tyr greatly reduces hydrolytic activity rather than transglycosylation activity and is useful for generating mutants with an increased transglycosylation activity ratio.

[0080]

The substitution of A310 with an amino acid having a side chain with a polar residue or a charged residue (preferably Asp, Glu, Lys, Ser, or the like) greatly reduces hydrolytic activity rather than transglycosylation activity. These mutations are useful for generating mutants with an increased transglycosylation activity ratio.

[0081]

The substitution of E314 with Gln causes no large variation of the activities and may be acceptable.

[0082]

As regards the amino acid substitution/mutation, the enzymes of the present invention include an amino acid sequence that preferably satisfies the property (A) of the present invention and has one or more amino acid substitutions selected from the mutations shown in SEQ ID NO: 23 (A128, N172, D176, Y214, S216, L245, N246, T275, W278, F283, L306, F307, A310, E314, D333, I434, V460, I527, K569, F610, and H626 in SEQ ID NO: 1); more preferably an amino acid sequence that has at least one mutation selected from A128T (Rp2), D333G (Rp3), I434L (Rp3), V460A (Rp4), I527V (Rp3), K569T (Rp2-4), F610S (Rp3) and H626R (Rp3) in SEQ ID NO: 1; or an amino acid sequence that further has mutation at at least one amino acid of N172, D176, Y214, S216, L245, N246, T275, W278, F283, L306, F307, A310, and E314 in the amino acid sequence, wherein a plurality of mutations may simultaneously occur and all of the mutations may occur.

[0083]

Endo-β-N-acetylglucosaminidases have both hydrolytic activity and transglycosylation activity. The enzymes having high hydrolytic activity also hydrolyze, as substrates, the sugar chain that has been transferred to the acceptor molecule due to its transglycosylation activity. This may prevent such enzymes from

appropriately producing the desired transglycosylated molecule. For this reason, such mutated enzymes for transglycosylation are also important in the synthesis of glycosylated compounds.

[0084]

As mentioned above, it has been confirmed that some mutations at N172, D176, Y214, S216, L245, N246, T275, F283, L306, F307, A310, and E314 in SEQ ID NO: 1 increase the transglycosylation activity of the enzymes of the present invention or greatly reduce the hydrolytic activity relative to the transglycosylation activity. Thus, these mutations are useful for designing a mutated enzyme for transglycosylation with an increased transglycosylation activity ratio (transglycosylation activity/hydrolytic activity). The present invention also provides such a mutated enzyme for transglycosylation. Specific mutations in the mutated enzyme for transglycosylation include a mutation in which N172 is substituted with Gln or Asp; a mutation in which D176 is substituted with Arg; a mutation in which Y214 is substituted with Phe; a mutation in which S216 is substituted with Val; a mutation in which L245 is substituted with Ser; a mutation in which T275 is substituted with Ile; a mutation in which W278 is substituted with Phe or Tyr; a mutation in which F283 is substituted with Ser; a mutation in which L306 is substituted with Ile; a mutation in which F307 is

substituted with Tyr; a mutation in which A310 is substituted with Asp; and a mutation in which E314 is substituted with Gln in SEQ ID NO: 1. The mutations introduced into the mutated enzymes for transglycosylation of the present invention may be at least one of these mutations. The mutations may be a single mutation or multiple mutations comprising some of these mutations. Preferable mutations are N172Q, N172D, W278F, N172Q/W278F, N172D/W278F, or Y214/L306I/F307Y.

[0085]

Hydrolytic activity and transglycosylation activity

The enzymes of the present invention have hydrolytic activity and/or transglycosylation activity on complex-type sugar chains. The activities represent a property exhibiting, at any temperature from 45 to 60°C, 40% or more of the maximal activity value.

[0086]

The enzymes of the present invention have an activity of hydrolyzing various complex-type sugar chains as substrates. The enzymes are identified by assessing the activity of hydrolyzing SGP as a substrate to provide SG (10) according to the method described below.

[0087]

In a hydrolytic reaction, a reaction solution (total volume: 100 μL) containing a 200 mM potassium phosphate buffer (pH 6.25), 69 mM SGP, and 0.02 μM enzyme (for blank, an equal volume of a buffer instead of the enzyme)

(all concentrations listed represent the final concentration in the reaction solution) is prepared and incubated at a predetermined temperature for 18 hours. The resulting reaction solution and blank solution were analyzed by LC-MS to quantify SGP and SG (10) and calculate the hydrolysis rate and specific activity according to the formulae described below.

[0088]

LC-MS analysis condition A

MS apparatus: 6130 Quadrupole LC-MS (Agilent Technologies, Inc.)

Ionization: ESI

Mode: Positive

HPLC: 1260 Infinity LC (Agilent Technologies, Inc.)

Column: Inertsil ODS-3 3 μm ϕ3.0 x 50 mm (GL Sciences Inc.)

Column temperature: 40°C

Mobile phase A: H_2O + 0.1% HCOOH

Mobile phase B: Acetonitrile + 0.1% HCOOH

Gradient (mobile phase B%): 0% (0 min), 10% (5 min), 30% (7 min)

Flow rate: 0.6 mL/min

[0089]

Hydrolysis Rate

Hydrolysis rate is calculated according to the following formula:

Hydrolysis rate (%) = the concentration of SG (10) after reaction (M) / the concentration of SGP in blank (M) x 100

[0090]

Specific Activity

Specific activity is calculated according to the following formula:

Specific activity (μmol/min/μg) = the amount of SG (10) produced (μmol) / the duration of reaction (min) / the amount of enzyme (μg)

[0091]

The enzymes of the present invention have the activity of recognizing various sugar chain donors and acceptor molecules and transferring sugar chains. The enzymes are identified by assessing the activity of transferring the sugar chain of SGP used as a substrate to provide SG-A according to the method described below.

[0092]

In a transfer reaction, a reaction solution (total volume: 30 μL) containing a 1.6 M potassium phosphate buffer (pH 6.25), 69 mM SGP, 690 mM GlcNAc-AcA, and 1.0 μM enzyme (for blank, an equal volume of a buffer instead of the enzyme) (all concentrations listed represent the final concentration in the reaction solution) is prepared and incubated at a predetermined temperature. The reaction solutions after 1 hour, 2 hours, 4 hours, 8 hours, 24 hours, 48 hours, and 96 hours and the blank

solution were analyzed by LC-MS to quantify SGP, SG(10), and SG-A and calculate the hydrolysis rate and specific activity according to the formulae described below.

[0093]

LC-MS analysis condition B

MS apparatus: 6130 Quadrupole LC-MS (Agilent Technologies, Inc.)

Ionization: ESI

Mode: Positive

HPLC: 1260 Infinity LC (Agilent Technologies, Inc.)

Column: Inertsil ODS-3 2 μm ϕ2.1 x 50 mm (GL Sciences Inc.)

Column temperature: 40°C

Mobile phase A: H2O + 0.1% HCOOH

Mobile phase B: Acetonitrile + 0.1% HCOOH

Gradient (mobile phase B%): 0.8% (0 min), 8% (0.1 min), 8% (5 min)

Flow rate: 0.7 mL/min

[0094]

Transfer rate

Transfer rate is calculated according to the following formula:

Transfer rate (%) = the concentration of SG-A after reaction (M) / the concentration of SGP in blank (M) x 100

[0095]

Specific Activity

Specific activity is calculated according to the following formula:

Specific activity (μmol/min/mg) = the amount of SG-A produced (μmol) / the duration of reaction (min) / the amount of enzyme (mg)

[0096]

When hydrolytic activity and/or transglycosylation activity are determined according to the method described above at temperatures appropriately varying from, for example, 10°C to 70°C (for example, in 2, 3, 5, or 10°C increments) and the maximum of the activity values calculated from all of the temperature conditions is set to 100%, the enzymes of the present invention exhibit, at any temperature condition from 45°C to 60°C (preferably at 50°C), a relative activity value of 40% or more, preferably 50% or more, more preferably 60% or more, still more preferably 80% or more and most preferably the maximal activity value in the temperature range.

[0097]

Specific activity is generally determined under the optimum temperature condition for the enzyme. The enzymes of the present invention exhibit a high specific activity. For example, Endo-Rp exhibits a specific activity of 0.21 μmol/min/μg when it uses 69 mM SGP as a substrate at 50°C. This specific activity is an excellent activity that is 30-times more than the specific activity of Endo-M at 37°C (0.0071 μmol/min/μg).

[0098]

The enzymes of the present invention are not limited to the enzymes and mutants specifically described in the Examples herein. The enzymes of the present invention include various polypeptides as long as they satisfy properties (A) and (B). The enzymes of the present invention may be those that exhibit at least one of their hydrolytic activity and transglycosylation activity at a certain level or higher, which level is preferably 0.5% or more, more preferably 1% or more, still more preferably 5% or more, even more preferably 10% or more, and most preferably 30% or more of the activity of the enzyme consisting of the amino acid sequence of SEQ ID NO: 1 under the temperature condition as described above.

[0099]

Gene

The present invention further provides genes having nucleic acid sequences encoding the enzymes of the present invention as described above.

[0100]

The genes can be cloned from nature (for example, thermophilic fungi, preferably fungi belonging to the genus *Rhizomucor*, and more preferably *R. pusillus* or *R. miehei*) as a gene encoding the present enzyme based on the gene information from *R. pusillus* or *R. miehei*. The genes can also be generated as a recombinant gene based

on the amino acid sequence of the enzymes according to known genetic engineering techniques.

[0101]

Nucleic acid sequences encoding the enzymes of the present invention encode a polypeptide that comprises an amino acid sequence that has 75% or more identity and 95% or more similarity to the amino acid sequence shown in SEQ ID NO: 1 and is an amino acid sequence different from the amino acid sequence of SEQ ID NO: 7. Examples of the nucleic acid sequences may include, for example, nucleic acid sequences naturally occurring in fungi as shown in SEQ ID NO: 8, 10, 12, 14, 16, or 18 (including the termination codon). Nucleic acids having nucleic acid sequences optimized for expression in *E. coli* can be generated by utilizing for example, GeneArt Strings DNA Fragments (manufactured by Thermo Fisher Scientific, Inc.) based on the amino acid sequences of the enzymes. The nucleic acid sequences of SEQ ID NOS: 9, 11, 13, 15, 17, and 19 are nucleic acid sequences designed based on the amino acid sequences of SEQ ID NO: 1 (Endo-Rp), 2 (Endo-Rp2), 3 (Endo-Rp3), 4 (Endo-Rp4), 5 (Endo-Rp5), and 6 (Endo-Rm), respectively. These nucleic acid sequences are optimized for expression in *E. coli* of the enzyme with a His tag (His x 6) added to the C-terminus. By referring to these nucleic acid sequences, a nucleic acid sequence efficiently expressing a polypeptide having various modified amino acids, such as for example the

amino acid sequence of SEQ ID NO: 23, in *E. coli* can be designed.

[0102]

Specific examples of the nucleic acid sequences encoding the enzymes of the present invention can include nucleic acid sequences having 80% or more, preferably 95% or more, more preferably 95%, even more preferably 98% or more identity to the nucleic acid sequence shown in any of nucleotide positions 1 to 2088 of SEQ ID NOs: 8 to 17 and nucleotide positions 1 to 2091 of SEQ ID NOs: 18 to 19 (preferably nucleotide positions 1 to 2088 of SEQ ID NOs: 8 to 17 encoding Endo-Rp or homologs thereof) and nucleic acids encoding Leu-Ala-Lys-Leu-Leu in a region from positions 570 to 585 in each nucleic acid sequence.

[0103]

Production and purification of enzyme

The present invention further provides a gene construct including a plasmid and an expression vector comprising a recombinant gene encoding an enzyme of the present invention; a host cell transformed with the gene construct; and a method of producing the enzyme of the present invention comprising collecting the enzyme of the present invention from the culture of the host cells. The gene encoding the enzyme of the present invention is introduced into a suitable plasmid/expression vector depending on the host cell type (including animal cells, plant cells, *E. coli*, and yeast; any cells commonly used

in protein production or other cells can be appropriately selected) to be transformed with the plasmid/expression vector. The transformed cells are cultured under suitable conditions. The enzyme of the present invention can be collected from the culture.

[0104]

The gene constructs for transformation can be generated according to known techniques using a vector or plasmid that is generally known in the art of genetic engineering and is selected depending on cell type in which the vector or plasmid is expressed. For example, vectors that can be used for expression in *E. coli* include, but are not limited to, pET vector, pCold vector, pFLAG vector, and the like.

[0105]

Examples of the *E. coli* can include BL21 (DE3) and Origami (DE3).

[0106]

Examples of the culture conditions of *E. coli* for producing the enzyme of the present invention include a method in which transformed bacterial culture is inoculated into 25 mL of TB medium (50 μg/mL Kanamycin) in a 100 mL flask and cultured with shaking at 37°C overnight (160 rpm, O/N). Another method which can also be used, but which is not limiting, comprises inoculating 20 mL of the preculture medium into 1 L of TB medium (50 μg/mL Kanamycin, 0.01% antifoam 204, 2 mM $MgSO_4$) in a 2.5

L baffled flask; culturing with shaking at 37°C (200 rpm, 2 hours); lowering the incubator temperature to 16°C and culturing for 3 hours, and thereafter adding IPTG at a final concentration of 0.2 mM; and further culturing it for 24 hours. Media that can be used include common media such as LB medium and M9 medium, in addition to TB medium.

[0107]

The enzyme of the present invention can be collected by utilizing the physical properties of the enzyme appropriately in combination with any common purification technique. For convenient collection, the gene construct is previously designed to express the enzyme linked to a tag peptide such as a His-tag or a GST-tag so as to collect the enzyme by utilizing the affinity of the tag peptide. The tag peptide may be removed after purification but the enzyme linked to the tag peptide may be directly used for reactions including hydrolysis if the tag peptide has no effect on the enzymatic activity. The enzymes of the present invention include an enzyme having an amino acid sequence linked to such a tag peptide. Specific examples of the enzyme include amino acid sequences having a His-tag (His x 6) linked to the C-terminus of the amino acid sequences of SEQ ID NOS: 1 to 6.

[0108]

The enzymes of the present invention have excellent production efficiency in *E. coli* due to the properties of the amino acid sequences. The enzymes of the present invention exhibit an expression efficiency of 2 mg or more per 1 L of culture, preferably 4 mg or more, more preferably 8 mg or more, further preferably 12 mg or more, and even more preferably 15 mg or more per 1 L of culture when they are produced in *E. coli* under the culture conditions described above.

Examples

[0109]

Hereinafter, the present invention will be specifically described with reference to the following Examples. The description in the Examples is an example of one embodiment of the present invention. These Examples are not intended to limit the present invention.

[0110]

Protein concentration as described herein was quantified using a Microvolume Spectrophotometer NanoDrop 1000 (manufactured by Thermo Fisher Scientific, Inc.) or NanoDrop 2000 (manufactured by Thermo Fisher Scientific, Inc.).

[0111]

In the Examples, when the hydrolytic activity of the enzyme endo-β-N-acetylglucosaminidase was determined, SGP and SG (10) were detected in the reaction solutions using

the above-mentioned LC-MS analysis condition A. Hydrolysis rate and specific activity were calculated according to the above-mentioned formulae for calculation.

[0112]

Example 1 Discovery of endo-β-N-acetylglucosaminidase derived from *R. pusillus*

R. pusillus strains NBRC 9740, NBRC 9741, NBRC 9742, and NBRC 9743 were inoculated on a GSYFe slant (5% Glucose, 2% Soytone, 1% Yeast extract, 0.05% $FeSO_4 \cdot 7H_2O$, 1.5% agar) and cultured at 40°C for 5 days. Fungal cells were collected and disrupted in 2 mL of a 100 mM potassium phosphate buffer (pH 6.25) using a glass homogenizer followed by filter sterilization to give a crude enzyme solution.

[0113]

To 50 μL of the crude enzyme solution, 3 mg of SGP was added (to a final concentration of 21 mM) and incubated at 50°C. The reaction solution was analyzed by LC-MS and disappearance of SGP and production of SG (10) were observed (Figure 2). Thus, we considered that *R. pusillus* NBRC 9742 has an endo-β-N-acetylglucosaminidase.

[0114]

Furthermore, hydrolytic reactions were performed at varying reaction temperatures. In each sample, relative activity was calculated such that the activity at the temperature at which the reaction progressed most efficiently was set as 100%. The results are shown in

Figure 3. It can be confirmed that the optimum temperature of the crude enzyme derived from *R. pusillus* is 55°C or more although Endo-M (manufactured by Tokyo Chemical Industry Co., Ltd.) used as a comparative control was inactivated at 45°C.

[0115]

Example 2 Cloning of endo-β-N-acetylglucosaminidase gene derived from *R. pusillus*

The method described below was used to clone an endo-β-N-acetylglucosaminidase gene from *R. pusillus* strain NBRC 9742 and extract the related sequence based on the published genomic sequence.

[0116]

(1) Gene cloning from microorganisms

First, *R. pusillus* strain NBRC 9742 was inoculated on a GSYFe slant and cultured at 40°C for 5 days. Fungal cells were collected together with a metalcone into a 2 mL screw-capped microtube and frozen at -80°C. The fungal cells were repeatedly disrupted using a Multi-beads Shocker (Yasui Kikai Corporation) at 2,000 rpm through 5 cycles of on-time 30 seconds and off-time 30 seconds while cooling the frozen sample at 4°C. mRNA solution was obtained from this sample using a NucleoSpin RNA Plant (MACHEREY-NAGEL GmbH & Co. KG) and an Oligotex-dT30<Super> mRNA Purification Kit (Takara Bio Inc.).

[0117]

The resulting mRNA was used as a template with primer 1 (SEQ ID NO: 22) and 3'-Full RACE Core Set (Takara Bio Inc.) to amplify the gene encoding the endo-β-N-acetylglucosaminidase and the amplified product was cloned into a pUC19 vector. This gene consists of 2091 bases including the termination codon (SEQ ID NO: 8) and encodes a protein with a molecular weight of 78,874 consisting of 696 amino acid residues (SEQ ID NO: 1). This protein was named as Endo-Rp.

[0118]

Endo-Rp homologs Endo-Rp2 (the amino acid sequence of SEQ ID NO: 2), Endo-Rp3 (the amino acid sequence of SEQ ID NO: 3), and Endo-Rp4 (the amino acid sequence of SEQ ID NO: 4) were also cloned from *R. pusillus* strains NBRC 9740, 9741, and 9743 in a similar manner. NBRC 9740 had only the Endo-Rp2 sequence while NBRC 9741 had the Endo-Rp2 and the Endo-Rp3 sequences and NBRC 9743 had the Endo-Rp3 and the Endo-Rp4 sequences. Each of the sequences had 99% identity to the amino acid sequence of Endo-Rp.

[0119]

(2) Comparison with published genomic sequence

The sequences of Endo-Rp, Endo-Rp2, Endo-Rp3, and Endo-Rp4 were compared with putative sequences published in the database provided by The Genozymes Project. The amino acid corresponding to Lys at position 193 in SEQ ID NO: 1 was substituted with Asn-Thr-Tyr-Tyr-Ile-Arg in SEQ

ID NO: 7, leading to the different length of the protein (Figure 4). Genomic DNA was extracted from *R. pusillus* strain NBRC 9742 and the full-length ORF sequence of the Endo-Rp gene was amplified by PCR and analyzed. The analysis results indicated that the putative sequence of the strain with the published genome has different predicted positions of the introns. Therefore, this could result in the difference in amino acid sequences as described above.

[0120]

The amino acid sequence of Endo-Rp was also used to perform a BLAST search against the database provided by NCBI. Results from the search showed that the genomic sequence of *Rhizomucor miehei* strain CAU432 which is a related species of *R. pusillus* has the sequence corresponding to Endo-Rp. This sequence information was used to identify the nucleic acid sequence expected to encode an endo-β-N-acetylglucosaminidase (SEQ ID NO: 18) and its amino acid sequence (SEQ ID NO: 6). The protein consisting of this amino acid sequence was named as Endo-Rm. Endo-Rm was 67% homologous to Endo-Rp.

[0121]

Example 3 Expression of Endo-Rp in *E. coli*

A nucleic acid sequence (SEQ ID NO: 15) optimized for heterologous expression in *E. coli* was designed from the fungus-derived endo-β-N-acetylglucosaminidase gene obtained in Example 2 using GeneArt Strings DNA Fragments

provided by Thermo Fisher Scientific, Inc. The sequence was used to generate a gene encoding a polypeptide with 6 x His-tag added to the C-terminus of Endo-Rp. This was cloned into a pET24b(+) vector, which was then introduced into *E. coli* BL21 (DE3) to be transformed. The bacterial culture after transformation was inoculated into 25 mL of TB medium (50 μg/mL Kanamycin) in a 100 mL flask and cultured with shaking at 37°C overnight (160 rpm, O/N). 20 mL of the preculture medium was inoculated into 1 L of TB medium (50 μg/mL Kanamycin, 0.01% antifoam 204, 2 mM $MgSO_4$) in a 2.5 L baffled flask and cultured with shaking at 37°C (200 rpm, 2 hours). After lowering the incubator temperature to 16°C and culturing for 3 hours, IPTG was added at a final concentration of 0.2 mM and additional culturing was performed for 24 hours.

[0122]

Harvested bacterial cells were suspended in 100 mL of a binding buffer (50 mM HEPES (pH 8.0), 0.5 M NaCl, 20 mM Imidazole, 5% Glycerol), sonicated, and centrifuged. The supernatant from centrifugation was purified with a Ni Sepharose 6 Fast Flow and a HiLoad 16/60 Superdex 200 pg column (GE Healthcare). The yield (calculated from A_{280} and extinction coefficient) was 16.9 mg/L of broth.

[0123]

Similar heterologous expression was also performed for Endo-Rp homologs and Endo-Rm. Also performed was the expression of Endo-Rp homologs including the enzyme

identified in Example 2 and the sequence (Endo-Rp5, SEQ ID NO: 5) that has a Lys residue substituted for Asn-Thr-Tyr-Tyr-Ile-Arg from positions 193 to 198 in SEQ ID NO: 7 as in other homologs. The Endo-Rm gene sequence optimized for *E. coli* (SEQ ID NO: 20) was obtained by utilizing GeneArt Strings DNA Fragments (Thermo Fisher Scientific, Inc.) based on the amino acid sequence of SEQ ID NO: 6. Other genes were obtained by introducing a mutation into a vector for *E. coli* expression of Endo-Rp comprising the base sequence of SEQ ID NO: 15 according to the standard protocol of the PrimeSTAR Mutagenesis Basal Kit (Takara Bio Inc.) (Endo-Rp2: SEQ ID NO: 16, Endo-Rp3: SEQ ID NO: 17, Endo-Rp4: SEQ ID NO: 18, Endo-Rp5: SEQ ID NO: 19).

[0124]

The expression was induced under the above-mentioned conditions except that the main culture was in 100 mL of TB medium in a 500 mL baffled flask. Harvested bacterial cells were suspended in 6 mL of a binding buffer, sonicated, and centrifuged. The supernatant from centrifugation was purified with a His GraviTrap column (GE Healthcare).

[0125]

After purification, the bands of expressed proteins except for SEQ ID NO: 7 were observed on SDS-PAGE. The yield of each protein was 2.067 mg (20.67 mg/L of culture) for Endo-Rp, 2.139 mg (21.39 mg/L of culture)

for Endo-Rp2, 2.765 mg (27.65 mg/L of culture) for Endo-Rp3, 2.301 mg (23.01 mg/L of culture) for Endo-Rp4, 2.187 mg (21.87 mg/L of culture) for Endo-Rp5, and 1.650 mg (16.50 mg/L of culture) for Endo-Rm.

[0126]

Example 4 Hydrolytic activity of Endo-Rp on complex-type sugar chains

The hydrolytic activity of enzymes obtained in Example 3 was determined using the above-mentioned method. Endo-M (manufactured by Tokyo Chemical Industry Co., Ltd.), EndoS, and Endo-Om were used as a comparative control for this determination.

[0127]

Reaction solutions (total volume: 100 μL) containing a 200 mM potassium phosphate buffer (pH 6.25), 69 mM SGP, and 0.02 μM enzyme (all concentrations listed represent the final concentration in the reaction solutions) were prepared. The reaction solution containing Endo-Rp or Endo-Om was incubated at 50°C while the reaction solution containing Endo-M or Endo-S was incubated at 37°C. The reaction solutions were sampled after 1 hour, 2 hours, 6 hours, 12 hours, 18 hours, and 24 hours and analyzed by LC-MS under the above-mentioned analysis condition A. The time-dependent change of the reaction yields is shown in Figure 5. The specific activities were 0.21 μmol/min/μg for Endo-Rp, 0.0071 μmol/min/μg for Endo-M, 0.0022 μmol/min/μg for Endo-S, and 0.060 μmol/min/μg for

Endo-Om. The specific activity of Endo-Rp was 30-times higher than that of Endo-M, 100-times higher than that of Endo-S, and 4-times higher than that of Endo-Om. This demonstrates that Endo-Rp has very high hydrolytic activity on SGP as compared with known enzymes.

[0128]

Endo-Rp homologs expressed and purified in Example 3 were also determined for their activities. The reaction solutions (total volume: 30 μL) containing a 200 mM potassium phosphate buffer (pH 6.25), 69 mM SGP, and 0.2 μM enzyme (all concentrations listed represent the final concentration in the reaction solutions) were prepared and incubated at 50°C for 1 hour. The hydrolysis rate after 1 hour was 87% for Endo-Rp, 91% for Endo-Rp2, 84% for Endo-Rp3, 79% for Endo-Rp4, 94% for Endo-Rp5, and 23% for Endo-Rm. This demonstrates that Endo-Rp2, Endo-Rp3, Endo-Rp4, and Endo-Rp5 have hydrolytic activity somewhat similar to that of Endo-Rp on SGP.

[0129]

Example 5 Determination of optimal reaction condition for Endo-Rp

Optimal reaction temperature and pH for Endo-Rp were determined.

[0130]

For the determination of reaction temperature, the reaction solutions (total volume: 100 μL) containing a 200 mM potassium phosphate buffer (pH 6.25), 69 mM SGP,

and 0.02 μM enzyme (all concentrations listed represent the final concentration in the reaction solutions) were prepared and incubated at a temperature of 45, 50, 52, 55, 57, 60, and 63°C for 18 hours to calculate the hydrolysis rates. The results are shown in Figure 6.

For the determination of reaction pH, the reaction solutions (volume: 100 μL) at pH 4.5, 5.0, 5.5, 5.8, 6.0, 6.2, 6.5, 7.0, and 7.5 containing a 200 mM sodium acetate buffer or potassium phosphate buffer, 69 mM SGP, and 0.02 μM enzyme (all concentrations listed represent the final concentration in the reaction solutions) were prepared and incubated at 50°C for 23 hours to calculate the hydrolysis rates. The results are shown in Figure 7.

[0131]

The results of the determination showed that the optimal reaction temperature for Endo-Rp is around 55°C and the optimal reaction pH is around 5.8.

[0132]

Example 6 Transglycosylation activity of Endo-Rp

Whether Endo-Rp has transglycosylation activity was determined as follows.

[0133]

The reaction solutions (total volume: 30 μL) containing a 1.6 M potassium phosphate buffer (pH 6.25), 69 mM SGP, 690 mM (GlcNAc-)Asn (manufactured by WATANABE CHEMICAL INDUSTRIES, LTD.) and 0.1 μM enzyme (all concentrations listed represent the final concentration

in the reaction solutions) were prepared and incubated at 50°C for 4 hours.

[0134]

The reaction solutions were analyzed by LC-MS under the conditions as described below.

MS apparatus: 6130 Quadrupole LC-MS (Agilent Technologies, Inc.)

Ionization: ESI

Mode: Positive

HPLC: 1260 Infinity LC (Agilent Technologies, Inc.)

Column: Inertsil ODS-3 2 μm ϕ2.1 x 50 mm (GL Sciences Inc.)

Column temperature: 40°C

Mobile phase A: H_2O + 0.1% HCOOH

Mobile phase B: Acetonitrile + 0.1% HCOOH

Gradient (mobile phase B%): 0.8% (0 min), 2% (5 min), 2% (6 min)

Flow rate: 0.7 mL/min

[0135]

The results are shown in Figure 8. The peaks of SGP, which is a donor, and SG (10), which is a hydrolysate, were only detected under a condition without an acceptor. The peaks of SG(10) and (SG-)Asn, which is a transglycosylation reaction product, were detected under the conditions with (GlcNAc-)Asn added as an acceptor. These results demonstrate that Endo-Rp has transglycosylation activity.

[0136]

Example 7 Generation of Endo-Rp N172Q

To obtain an enzyme with suppressed hydrolytic activity on sugar chains while retaining transglycosylation activity of Endo-Rp, asparagine (N) at position 172, which is the active center, was substituted with glutamine (Q) to generate an N172Q variant. The N172Q variant was prepared by substituting AAC from positions 514 to 516 with CAA in the base sequence encoding Endo-Rp shown in SEQ ID NO: 9 using the method as described in Example 3.

[0137]

Transglycosylation activity was assessed as follows. The reaction solution (total volume: 30 µL) containing a 1.6 M potassium phosphate buffer (pH 6.25), 69 mM SGP, 690 mM (GlcNAc-)Asn, and 1.0 µM enzyme (all concentrations listed represent the final concentration in the reaction solution) was prepared and incubated at 50°C. The reaction solution was sampled after 20 minutes, 40 minutes, 1 hour, 2 hours, 4 hours, 8 hours, 20 hours, 24 hours, and 48 hours of reaction and analyzed by LC-MS under the conditions as described in Example 6.

[0138]

The time-dependent change of the reaction yields is shown in Figure 9. Endo-Rp WT rapidly hydrolyzed a transglycosylation reaction product (SG-)Asn whereas the

N172Q variant caused less hydrolysis, which demonstrated that the N172Q variant produces (SG-)Asn in a high yield.

[0139]

Example 8 Modification of Endo-Rp

To obtain an enzyme with increased transglycosylation activity in comparison with that of the Endo-Rp N172Q variant, modification was attempted. Various mutants listed in Table 1 were designed based on the structures of Endo-A (PDB ID: 3FHQ) and Endo-D (PBD ID: 2W92). Hydrolytic activity and transglycosylation activity of these mutants on SGP were determined.

[0140]

Transglycosylation activity was assessed as follows. The reaction solution (total volume: 30 μL) containing a 1.6 M potassium phosphate buffer (pH 6.25), 69 mM SGP, 690 mM GlcNAc-AcA, and 1.0 μM enzyme (all concentrations listed represent the final concentration in the reaction solution) was prepared and incubated at 50°C. The reaction solution was sampled after 1 hour, 2 hours, 4 hours, 8 hours, 24 hours, 48 hours, and 96 hours of reaction and analyzed by LC-MS under the above-mentioned LC-MS analysis condition B.

[0141]

Hydrolytic activity was assessed as follows. The reaction solution (total volume: 20 μL) containing a 1.6 M potassium phosphate buffer (pH 6.25), 69 mM SG-A, and 0.2 μM enzyme (all concentrations listed represent the

final concentration in the reaction solution) was prepared and incubated at 50°C. The reaction solution was sampled after 1 hour, 2 hours, 4 hours, 8 hours, 24 hours, 48 hours, and 72 hours of reaction and analyzed by LC-MS under the above-mentioned LC-MS analysis condition B.

[0142]

Table 1 shows the relative specific activities (hydrolysis and transglycosylation) of each variant when specific activities for transfer and hydrolysis of Endo-Rp were set to 100. The results showed that variants with mutations N172A, N172C, N172D, N172E, N172G, N172H, N172I, N172M, N172S, N172T, N172V, D176R, Y214F, S216V, L245S, N246D, T275I, F283S, L306I, F307Y, F307H, A310D, and E314Q had suppressed hydrolytic activity with maintained transglycosylation activity and an increased transglycosylation activity ratio (transglycosylation activity/hydrolytic activity; the ratio of transfer/hydrolysis in the table). Variants with mutations W278F and W278Y had also both increased transglycosylation activity and hydrolytic activity but decreased the transglycosylation activity ratio.

[0143]

[Table 1-1]

Mutant	Relative transglycosylation activity %	Relative hydrolysis activity %	Ratio of transfer/ hydrolysis
WT	100	100	1.00
N172Q	25	6.7	3.71
N172H	16	4.4	3.59
N172A	9.5	2.2	4.27
N172C	6.2	3.1	2.01
N172D	48	27	1.81
N172E	9.2	5.3	1.74
N172F	ND	1.4	-
N172G	10	2.4	4.12
N172I	13	3.3	3.86
N172K	ND	1.5	-
N172L	ND	1.7	-
N172M	6.7	3.2	2.13
N172P	0.67	0.94	0.71
N172R	ND	0.34	-
N172S	6.7	5.8	1.16
N172T	11	6.2	1.75
N172V	12	6.6	1.83
N172W	ND	0.15	-
N172Y	ND	0.36	-
N172Q/E126A	0.15	0.36	0.42
N172Q/Y214F	5.5	3.1	1.77
N172Q/Y214A	0.16	0.48	0.33
N172Q/V223R	0.020	0.30	0.05
N172Q/W225H	0.92	0.54	1.69
N172Q/W248N	0.55	1.2	0.45
N172Q/N246L	ND	0.32	-

[0144]

[Table 1-2]

N172Q/N246A	ND	ND	-
N172Q/Y247F	2.0	1.0	1.98
N172Q/W278E	0.080	0.18	0.44
N172Q/W278R	ND	0.23	-
N172Q/W278F	28	8.1	3.46
N172Q/W278Y	29	9.5	3.02
N172Q/F307Y	4.2	1.5	2.86
N172Q/A310D	0.60	0.60	0.99
W278F	93	155	0.60
W278F/D176R	104	129	0.81
W278F/Y214F	189	98	1.92
W278F/S216V	65	13	5.17
W278F/L245S	120	74	1.63
W278F/N246D	15	0.60	24.71
W278F/N246V	ND	ND	-
W278F/T275I	156	129	1.21
W278F/D276N	2.2	0.46	4.83
W278F/R280E	ND	0.080	-
W278F/F283S	119	50	2.38
W278F/L306I	113	141	0.80
W278F/F307Y	207	62	3.32
W278F/F307H	22	9.5	2.31
W278F/A310D	93	23	3.98
W278F/E314Q	95	141	0.67
W278F/N172D	34	33	1.01
W278F/N172D/D176R	20	19	1.06
W278F/N172D/Y214F	15	6.6	2.34
W278F/N172D/S216V	1.3	ND	-
W278F/N172D/L245S	19	14	1.33
W278F/N172D/N246D	0.17	ND	-
W278F/N172D/N246V	ND	ND	-
W278F/N172D/T275I	22	19	1.19
W278F/N172D/D276N	0.11	ND	-
W278F/N172D/R280E	ND	ND	-
W278F/N172D/F283S	19	5.8	3.36
W278F/N172D/L306I	33	27	1.20
W278F/N172D/F307Y	14	1.5	9.43
W278F/N172D/F307H	2.3	0.16	14.05
W278F/N172D/A310D	8.2	0.87	9.41

[0145]

[Table 1-3]

W278F/N172D/E314Q	18	22	0.79
N172Q/Y214F/F307Y	0.25	0.13	1.85
Y214F/F307Y	20	6.6	3.03
Y214F/L306I/F307Y	34	8.3	4.12

ND: Not detected.

Claims

[Claim 1]

A polypeptide having the following properties (A) and (B):

(A) the polypeptide comprises an amino acid sequence having 75% or more identity and 95% or more similarity to the amino acid sequence shown in SEQ ID NO: 1 and is an amino acid sequence different from the amino acid sequence of SEQ ID NO: 7; and

(B) the polypeptide exhibits, at any temperature from 45 to 60°C, 40% or more of the maximal activity value of hydrolytic activity and/or transglycosylation activity on complex sugar chains.

[Claim 2]

The polypeptide according to claim 1, having the following properties (A) and (B):

(A) the polypeptide comprises an amino acid sequence having 75% or more identity and 95% or more similarity to the amino acid sequence shown in SEQ ID NO: 1 and an amino acid sequence of Leu-Ala-Lys-Leu-Leu (LAKLL) at positions corresponding to the amino acid sequence from amino acid positions 191 to 195 of SEQ ID NO: 1;

(B) the polypeptide exhibits, at any temperature from 45 to 60°C, 40% or more of the maximal activity value of hydrolytic activity on complex sugar chains.

[Claim 3]

The polypeptide according to claim 1, wherein the hydrolytic activity on complex sugar chains at 50°C is 60% or more of the maximal activity value.

[Claim 4]

The polypeptide according to any of claims 1 to 3, wherein, in addition to having the properties (A) and (B), (C) the polypeptide is produced in an amount of 10 mg/L of culture or more in recombinant expression in *E. coli*.

[Claim 5]

The polypeptide according to any of claims 1 to 4, wherein the polypeptide has 85% or more sequence identity to the region from amino acid positions 54 to 341 of the amino acid sequence of SEQ ID NO: 1.

[Claim 6]

The polypeptide according to any of claims 1 to 5, wherein at least one of amino acids D276, V223, W225, Y247, and W248 remains unchanged and preferably at least D276 remains unchanged.

[Claim 7]

The polypeptide according to any of claims 1 to 3, wherein the polypeptide is a polypeptide consisting of

the amino acid sequence of SEQ ID NO: 1 or a polypeptide consisting of an amino acid sequence having at least one mutation selected from the group consisting of A128T, D333G, I434L, V460A, I527V, K569T, F610S, and H626R in the amino acid sequence of SEQ ID NO: 1.

[Claim 8]

The polypeptide according to claim 1, wherein the polypeptide consists of an amino acid sequence of any of SEQ ID NOS: 1 to 6.

[Claim 9]

The polypeptide according to any of claims 1 to 6, wherein the polypeptide has an amino acid sequence satisfying the property (A) and has at least one of the mutations contained in SEQ ID NO: 23.

[Claim 10]

The polypeptide according to claim 9, wherein the polypeptide has a mutation in at least one amino acid selected from N172, D176, Y214, S216, L245, N246, T275, L306, F307, and A310 and has increased transglycosylation activity.

[Claim 11]

The polypeptide according to claim 7, wherein the polypeptide has at least one mutation selected from the group of:
a mutation in which N172 is substituted with Gln, Asp, Gly, Ala, Phe, Cys, His, Ile, Ser, Thr, Val, or Met; a mutation in which D176 is substituted with Arg; a mutation in which Y214 is substituted with Phe; a mutation in which S216 is substituted with Val; a mutation in which L245 is substituted with Ser; a mutation in which N246 is substituted with Asp; a mutation in which T275 is substituted with Ile; a mutation in which F283 is substituted with Ser; a mutation in which L306 is substituted with Ile; a mutation in which F307 is substituted with Tyr; a mutation in which A310 is substituted with Asp; and a mutation in which E314 is substituted with Gln.

[Claim 12]

The polypeptide according to any of claims 9 to 11, wherein the polypeptide has a mutation in which W278 is substituted with Phe or Tyr.

[Claim 13]

A polynucleotide encoding the polypeptide according to any of claims 1 to 12.

[Claim 14]

The polynucleotide according to claim 13, having a nucleotide sequence of any of the nucleotide sequences from nucleotide positions 1 to 2088 of SEQ ID NOS: 8 to 17 and the nucleotide sequences from nucleotide positions 1 to 2091 of SEQ ID NOS: 18 to 19.

[Claim 15]

An expression plasmid comprising the polynucleotide according to claim 13 or 14.

[Claim 16]

A host cell transformed with the plasmid according to claim 15.

[Claim 17]

The host cell according to claim 16, wherein the host cell is *E. coli* transformed with a plasmid comprising a polynucleotide comprising the nucleotide sequence from nucleotide positions 1 to 2088 of SEQ ID NO: 9, 11, 13, 15, or 17 or the nucleotide sequence from nucleotide positions 1 to 2091 of SEQ ID NO: 19.

[Claim 18]

A method of producing the polypeptide according to any of claims 1 to 12, comprising culturing the host cells according to claim 16 or 17 and collecting the

polypeptide according to any of claims 1 to 12 from the resulting culture.

[Claim 19]

A reagent comprising the polypeptide according to any of claims 1 to 12.

Figure 1

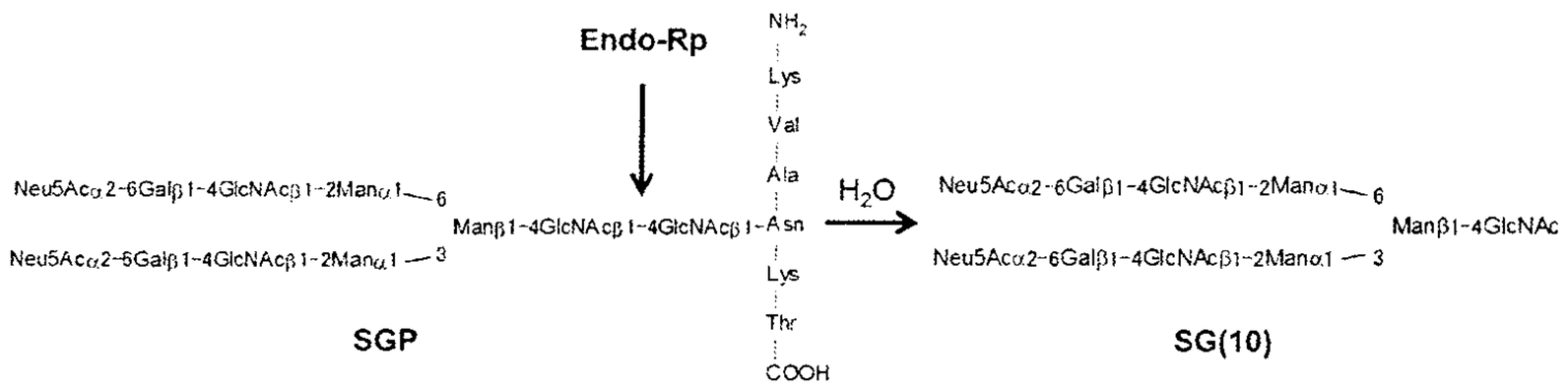


Figure 2

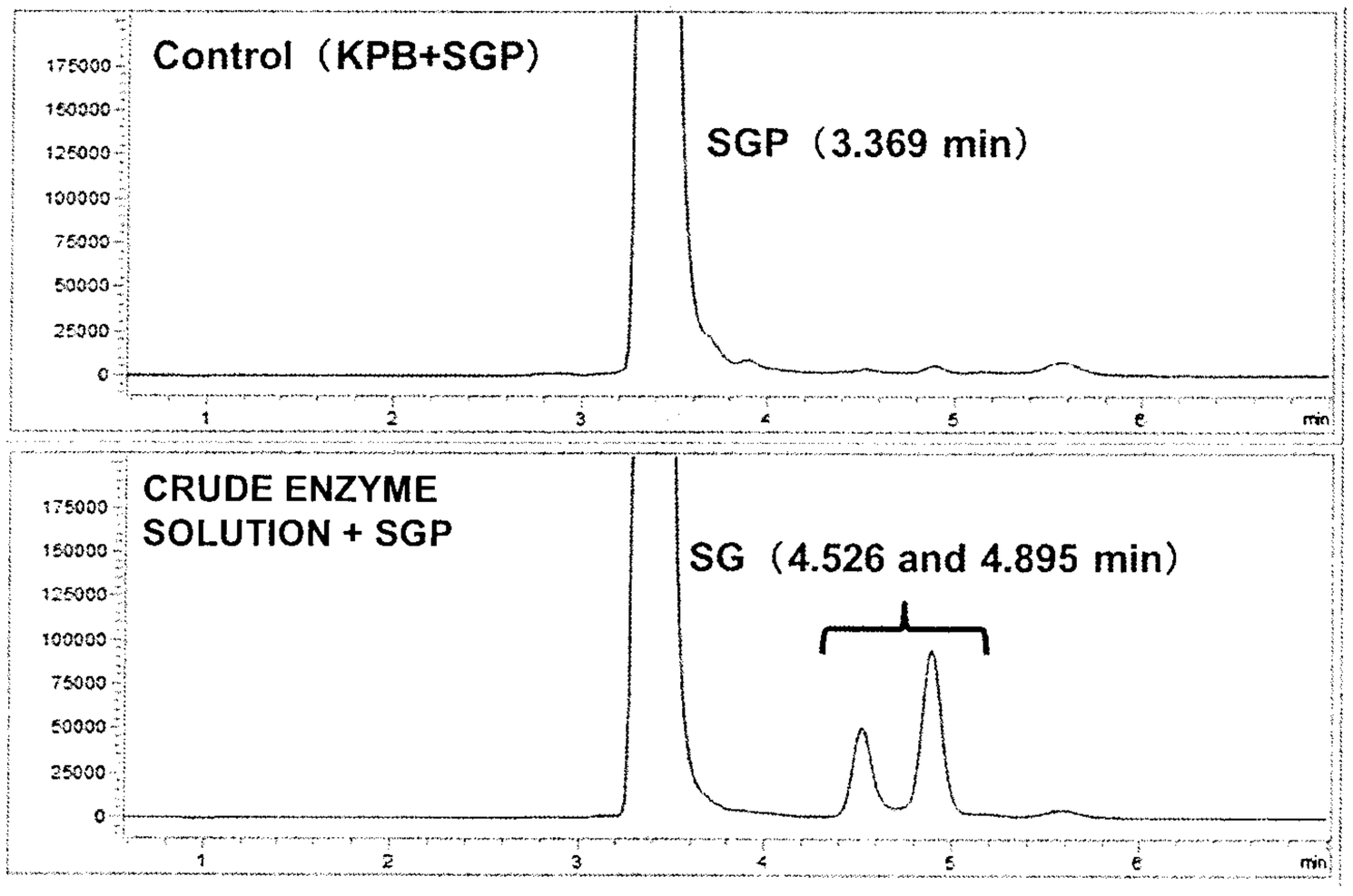


Figure 3

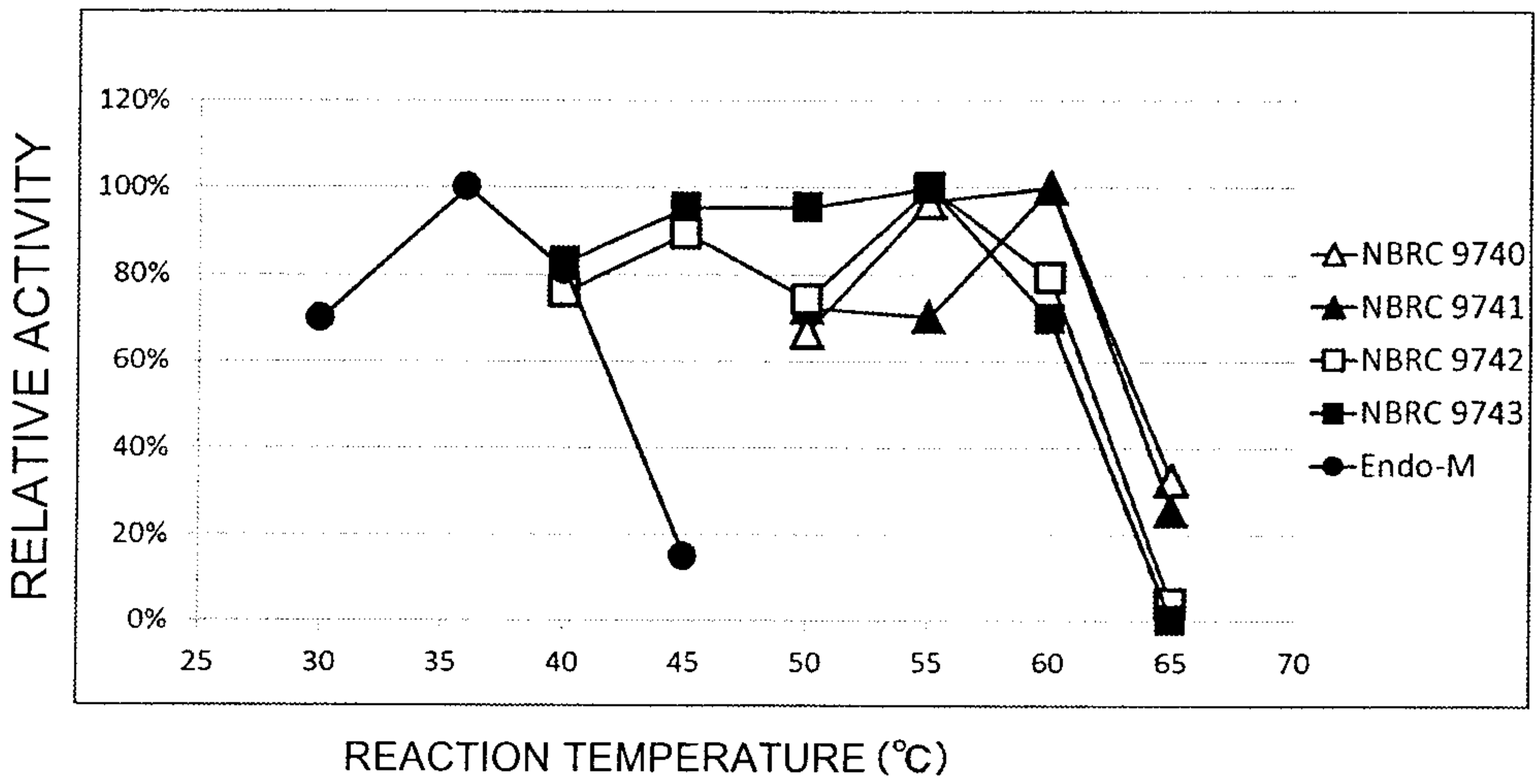

RELATIVE ACTIVITY
120%
100%
80%
60%
40%
20%
0%
25
30
35
40
45
50
55
60
65
70
REACTION TEMPERATURE (℃)
NBRC 9740
NBRC 9741
NBRC 9742
NBRC 9743
Endo-M

Figure 4

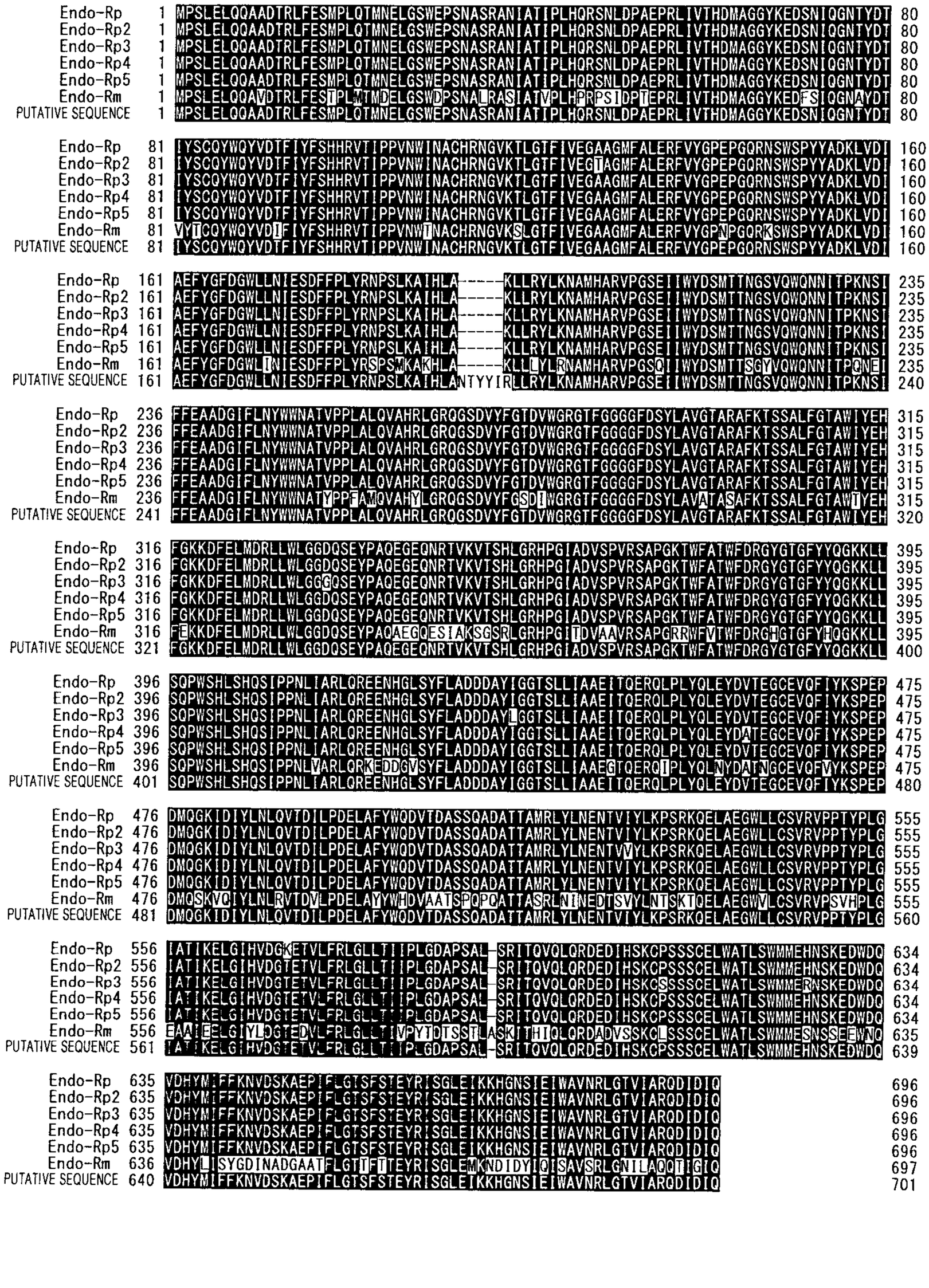

Endo-Rp 1 MPSLELQQAADTRLFESMPLQTMNELGSWEPSNASRANIATIPLHQRSNLDPAEPRLIVTHDMAGGYKEDSNIQGNTYDT 80
Endo-Rp2 1 MPSLELQQAADTRLFESMPLQTMNELGSWEPSNASRANIATIPLHQRSNLDPAEPRLIVTHDMAGGYKEDSNIQGNTYDT 80
Endo-Rp3 1 MPSLELQQAADTRLFESMPLQTMNELGSWEPSNASRANIATIPLHQRSNLDPAEPRLIVTHDMAGGYKEDSNIQGNTYDT 80
Endo-Rp4 1 MPSLELQQAADTRLFESMPLQTMNELGSWEPSNASRANIATIPLHQRSNLDPAEPRLIVTHDMAGGYKEDSNIQGNTYDT 80
Endo-Rp5 1 MPSLELQQAADTRLFESMPLQTMNELGSWEPSNASRANIATIPLHQRSNLDPAEPRLIVTHDMAGGYKEDSNIQGNTYDT 80
Endo-Rm 1 MPSLELQQAVDTRLFESTPLMTMDELGSWDPSNALRASIATVPLHPRPSIDPTEPRLIVTHDMAGGYKEDFSIQGNAYDT 80
PUTATIVE SEQUENCE 1 MPSLELQQAADTRLFESMPLQTMNELGSWEPSNASRANIATIPLHQRSNLDPAEPRLIVTHDMAGGYKEDSNIQGNTYDT 80
Endo-Rp 81 IYSCQYWQYVDTFIYFSHHRVTIPPVNWINACHRNGVKTLGTFIVEGAAGMFALERFVYGPEPGQRNSWSPYYADKLVDI 160
Endo-Rp2 81 IYSCQYWQYVDTFIYFSHHRVTIPPVNWINACHRNGVKTLGTFIVEGTAGMFALERFVYGPEPGQRNSWSPYYADKLVDI 160
Endo-Rp3 81 IYSCQYWQYVDTFIYFSHHRVTIPPVNWINACHRNGVKTLGTFIVEGAAGMFALERFVYGPEPGQRNSWSPYYADKLVDI 160
Endo-Rp4 81 IYSCQYWQYVDTFIYFSHHRVTIPPVNWINACHRNGVKTLGTFIVEGAAGMFALERFVYGPEPGQRNSWSPYYADKLVDI 160
Endo-Rp5 81 IYSCQYWQYVDTFIYFSHHRVTIPPVNWINACHRNGVKTLGTFIVEGAAGMFALERFVYGPEPGQRNSWSPYYADKLVDI 160
Endo-Rm 81 VYTCQYWQYVDIFIYFSHHRVTIPPVNWTNACHRNGVKSLGTFIVEGAAGMFALERFVYGPNPGQRKSWSPYYADKLVDI 160
PUTATIVE SEQUENCE 81 IYSCQYWQYVDTFIYFSHHRVTIPPVNWINACHRNGVKTLGTFIVEGAAGMFALERFVYGPEPGQRNSWSPYYADKLVDI 160
Endo-Rp 161 AEFYGFDGWLLNIESDFFPLYRNPSLKAIHLA-----KLLRYLKNAMHARVPGSEIIWYDSMTTNGSVQWQNNITPKNSI 235
Endo-Rp2 161 AEFYGFDGWLLNIESDFFPLYRNPSLKAIHLA-----KLLRYLKNAMHARVPGSEIIWYDSMTTNGSVQWQNNITPKNSI 235
Endo-Rp3 161 AEFYGFDGWLLNIESDFFPLYRNPSLKAIHLA-----KLLRYLKNAMHARVPGSEIIWYDSMTTNGSVQWQNNITPKNSI 235
Endo-Rp4 161 AEFYGFDGWLLNIESDFFPLYRNPSLKAIHLA-----KLLRYLKNAMHARVPGSEIIWYDSMTTNGSVQWQNNITPKNSI 235
Endo-Rp5 161 AEFYGFDGWLLNIESDFFPLYRNPSLKAIHLA-----KLLRYLKNAMHARVPGSEIIWYDSMTTNGSVQWQNNITPKNSI 235
Endo-Rm 161 AEFYGFDGWLINIESDFFPLYRSPSMKAKHLA-----KLLLYLRNAMHARVPGSQIIWYDSMTTSGYVQWQNNITPQNEI 235
PUTATIVE SEQUENCE 161 AEFYGFDGWLLNIESDFFPLYRNPSLKAIHLANTYYIRLLRYLKNAMHARVPGSEIIWYDSMTTNGSVQWQNNITPKNSI 240
Endo-Rp 236 FFEAADGIFLNYWWNATVPPLALQVAHRLGRQGSDVYFGTDVWGRGTFGGGGFDSYLAVGTARAFKTSSALFGTAWIYEH 315
Endo-Rp2 236 FFEAADGIFLNYWWNATVPPLALQVAHRLGRQGSDVYFGTDVWGRGTFGGGGFDSYLAVGTARAFKTSSALFGTAWIYEH 315
Endo-Rp3 236 FFEAADGIFLNYWWNATVPPLALQVAHRLGRQGSDVYFGTDVWGRGTFGGGGFDSYLAVGTARAFKTSSALFGTAWIYEH 315
Endo-Rp4 236 FFEAADGIFLNYWWNATVPPLALQVAHRLGRQGSDVYFGTDVWGRGTFGGGGFDSYLAVGTARAFKTSSALFGTAWIYEH 315
Endo-Rp5 236 FFEAADGIFLNYWWNATVPPLALQVAHRLGRQGSDVYFGTDVWGRGTFGGGGFDSYLAVGTARAFKTSSALFGTAWIYEH 315
Endo-Rm 236 FFEAADGIFLNYWWNATYPPFAMQVAHYLGRQGSDVYFGSDIWGRGTFGGGGFDSYLAVATASAFKTSSALFGTAWTYEH 315
PUTATIVE SEQUENCE 241 FFEAADGIFLNYWWNATVPPLALQVAHRLGRQGSDVYFGTDVWGRGTFGGGGFDSYLAVGTARAFKTSSALFGTAWIYEH 320
Endo-Rp 316 FGKKDFELMDRLLWLGGDQSEYPAQEGEQNRTVKVTSHLGRHPGIADVSPVRSAPGKTWFATWFDRGYGTGFYYQGKKLL 395
Endo-Rp2 316 FGKKDFELMDRLLWLGGDQSEYPAQEGEQNRTVKVTSHLGRHPGIADVSPVRSAPGKTWFATWFDRGYGTGFYYQGKKLL 395
Endo-Rp3 316 FGKKDFELMDRLLWLGGGQSEYPAQEGEQNRTVKVTSHLGRHPGIADVSPVRSAPGKTWFATWFDRGYGTGFYYQGKKLL 395
Endo-Rp4 316 FGKKDFELMDRLLWLGGDQSEYPAQEGEQNRTVKVTSHLGRHPGIADVSPVRSAPGKTWFATWFDRGYGTGFYYQGKKLL 395
Endo-Rp5 316 FGKKDFELMDRLLWLGGDQSEYPAQEGEQNRTVKVTSHLGRHPGIADVSPVRSAPGKTWFATWFDRGYGTGFYYQGKKLL 395
Endo-Rm 316 FEKKDFELMDRLLWLGGDQSEYPAQAEGQESIAKSGSRLGRHPGITDVAAVRSAPGRRWFVTWFDRGHGTGFYHQGKKLL 395
PUTATIVE SEQUENCE 321 FGKKDFELMDRLLWLGGDQSEYPAQEGEQNRTVKVTSHLGRHPGIADVSPVRSAPGKTWFATWFDRGYGTGFYYQGKKLL 400
Endo-Rp 396 SQPWSHLSHQSIPPNLIARLQREENHGLSYFLADDDAYIGGTSLLIAAEITQERQLPLYQLEYDVTEGCEVQFIYKSPEP 475
Endo-Rp2 396 SQPWSHLSHQSIPPNLIARLQREENHGLSYFLADDDAYIGGTSLLIAAEITQERQLPLYQLEYDVTEGCEVQFIYKSPEP 475
Endo-Rp3 396 SQPWSHLSHQSIPPNLIARLQREENHGLSYFLADDDAYLGGTSLLIAAEITQERQLPLYQLEYDVTEGCEVQFIYKSPEP 475
Endo-Rp4 396 SQPWSHLSHQSIPPNLIARLQREENHGLSYFLADDDAYIGGTSLLIAAEITQERQLPLYQLEYDATEGCEVQFIYKSPEP 475
Endo-Rp5 396 SQPWSHLSHQSIPPNLIARLQREENHGLSYFLADDDAYIGGTSLLIAAEITQERQLPLYQLEYDVTEGCEVQFIYKSPEP 475
Endo-Rm 396 SQPWSHLSHQSIPPNLVARLQRKEDDGVSYFLADDDAYIGGTSLLIAAEGTQERQIPLYQLNYDATNGCEVQFVYKSPEP 475
PUTATIVE SEQUENCE 401 SQPWSHLSHQSIPPNLIARLQREENHGLSYFLADDDAYIGGTSLLIAAEITQERQLPLYQLEYDVTEGCEVQFIYKSPEP 480
Endo-Rp 476 DMQGKIDIYLNLQVTDILPDELAFYWQDVTDASSQADATTAMRLYLNENTVIYLKPSRKQELAEGWLLCSVRVPPTYPLG 555
Endo-Rp2 476 DMQGKIDIYLNLQVTDILPDELAFYWQDVTDASSQADATTAMRLYLNENTVIYLKPSRKQELAEGWLLCSVRVPPTYPLG 555
Endo-Rp3 476 DMQGKIDIYLNLQVTDILPDELAFYWQDVTDASSQADATTAMRLYLNENTVVYLKPSRKQELAEGWLLCSVRVPPTYPLG 555
Endo-Rp4 476 DMQGKIDIYLNLQVTDILPDELAFYWQDVTDASSQADATTAMRLYLNENTVIYLKPSRKQELAEGWLLCSVRVPPTYPLG 555
Endo-Rp5 476 DMQGKIDIYLNLQVTDILPDELAFYWQDVTDASSQADATTAMRLYLNENTVIYLKPSRKQELAEGWLLCSVRVPPTYPLG 555
Endo-Rm 476 DMQSKVQIYLNLRVTDVLPDELAYYWHDVAATSPQPQATTASRLNINEDTSVYLNTSKTQELAEGWVLCSVRVPSVHPLG 555
PUTATIVE SEQUENCE 481 DMQGKIDIYLNLQVTDILPDELAFYWQDVTDASSQADATTAMRLYLNENTVIYLKPSRKQELAEGWLLCSVRVPPTYPLG 560
Endo-Rp 556 IATIKELGIHVDGKETVLFRLGLLTIIPLGDAPSAL-SRITQVQLQRDEDIHSKCPSSSCELWATLSWMMEHNSKEDWDQ 634
Endo-Rp2 556 IATIKELGIHVDGTETVLFRLGLLTIIPLGDAPSAL-SRITQVQLQRDEDIHSKCPSSSCELWATLSWMMEHNSKEDWDQ 634
Endo-Rp3 556 IATIKELGIHVDGTETVLFRLGLLTIIPLGDAPSAL-SRITQVQLQRDEDIHSKCSSSSCELWATLSWMMERNSKEDWDQ 634
Endo-Rp4 556 IATIKELGIHVDGTETVLFRLGLLTIIPLGDAPSAL-SRITQVQLQRDEDIHSKCPSSSCELWATLSWMMEHNSKEDWDQ 634
Endo-Rp5 556 IATIKELGIHVDGTETVLFRLGLLTIIPLGDAPSAL-SRITQVQLQRDEDIHSKCPSSSCELWATLSWMMEHNSKEDWDQ 634
Endo-Rm 556 EAAIEELGIYLDGTEDVLFRLGLLTIVPYTDTSSTLASKITHIQLQRDADVSSKCLSSSCELWATLSWMMESNSSEEWNQ 635
PUTATIVE SEQUENCE 561 IATIKELGIHVDGTETVLFRLGLLTIIPLGDAPSAL-SRITQVQLQRDEDIHSKCPSSSCELWATLSWMMEHNSKEDWDQ 639
Endo-Rp 635 VDHYMIFFKNVDSKAEPIFLGISFSTEYRISGLEIKKHGNSIEIWAVNRLGTVIARQDIDIQ 696
Endo-Rp2 635 VDHYMIFFKNVDSKAEPIFLGTSFSTEYRISGLEIKKHGNSIEIWAVNRLGTVIARQDIDIQ 696
Endo-Rp3 635 VDHYMIFFKNVDSKAEPIFLGTSFSTEYRISGLEIKKHGNSIEIWAVNRLGTVIARQDIDIQ 696
Endo-Rp4 635 VDHYMIFFKNVDSKAEPIFLGTSFSTEYRISGLEIKKHGNSIEIWAVNRLGTVIARQDIDIQ 696
Endo-Rp5 635 VDHYMIFFKNVDSKAEPIFLGTSFSTEYRISGLEIKKHGNSIEIWAVNRLGTVIARQDIDIQ 696
Endo-Rm 636 VDHYLISYGDINADGAATFLGTTFTTEYRISGLEMKNDIDYIQISAVSRLGNILAQQTIGIQ 697
PUTATIVE SEQUENCE 640 VDHYMIFFKNVDSKAEPIFLGTSFSTEYRISGLEIKKHGNSIEIWAVNRLGTVIARQDIDIQ 701

Figure 5

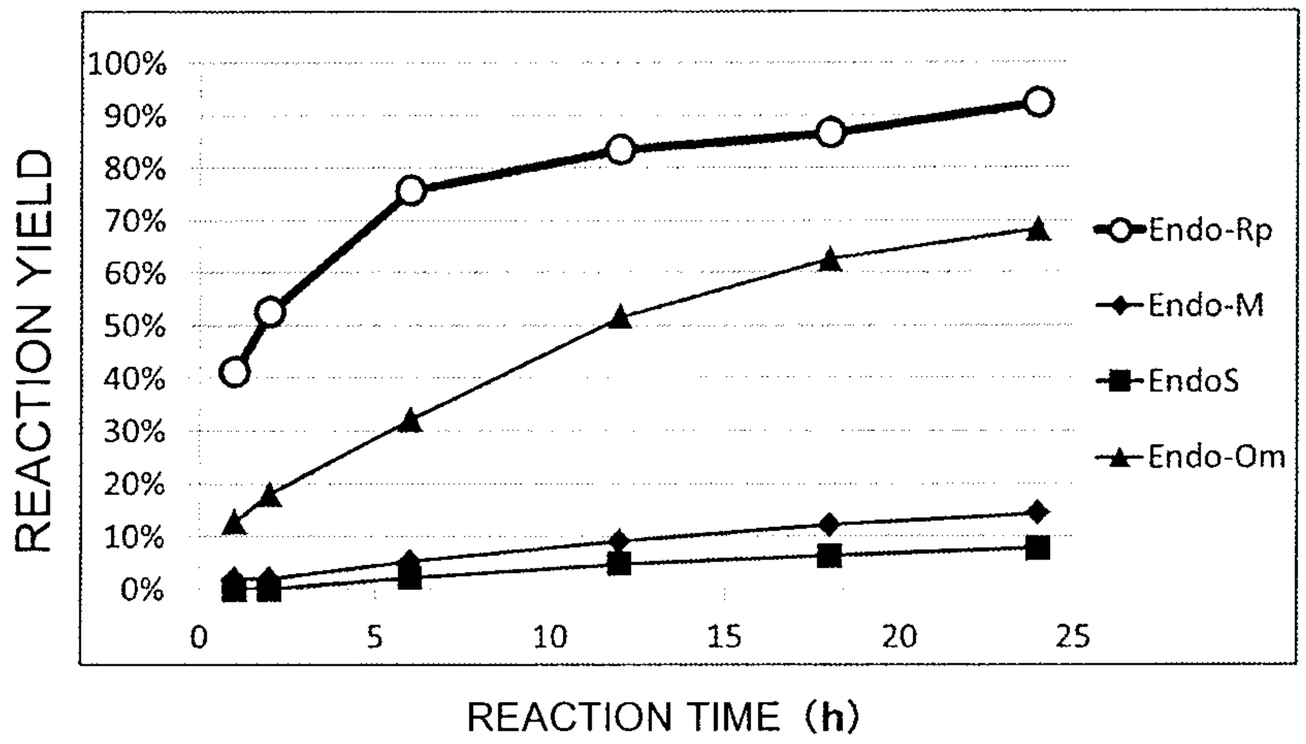


Figure 6

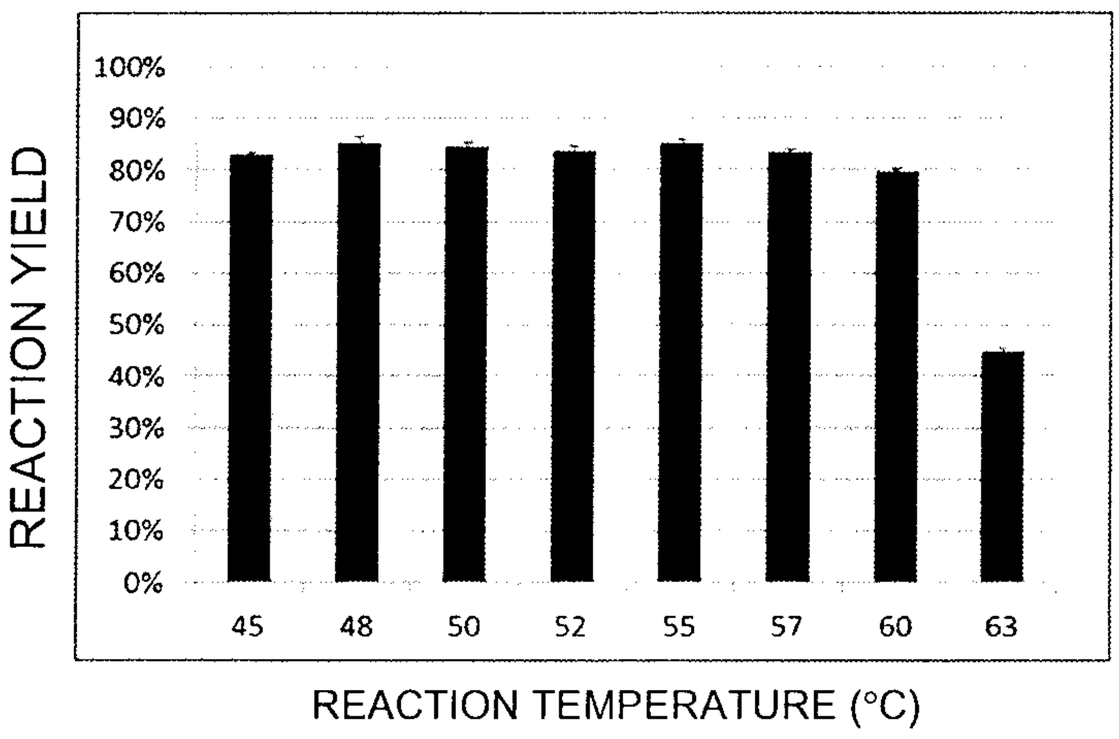


Figure 7

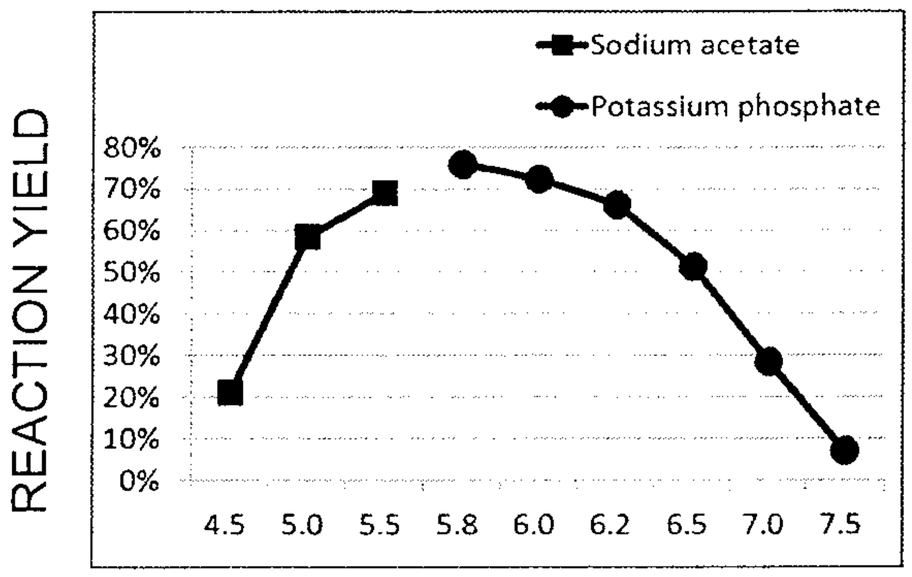


Figure 8

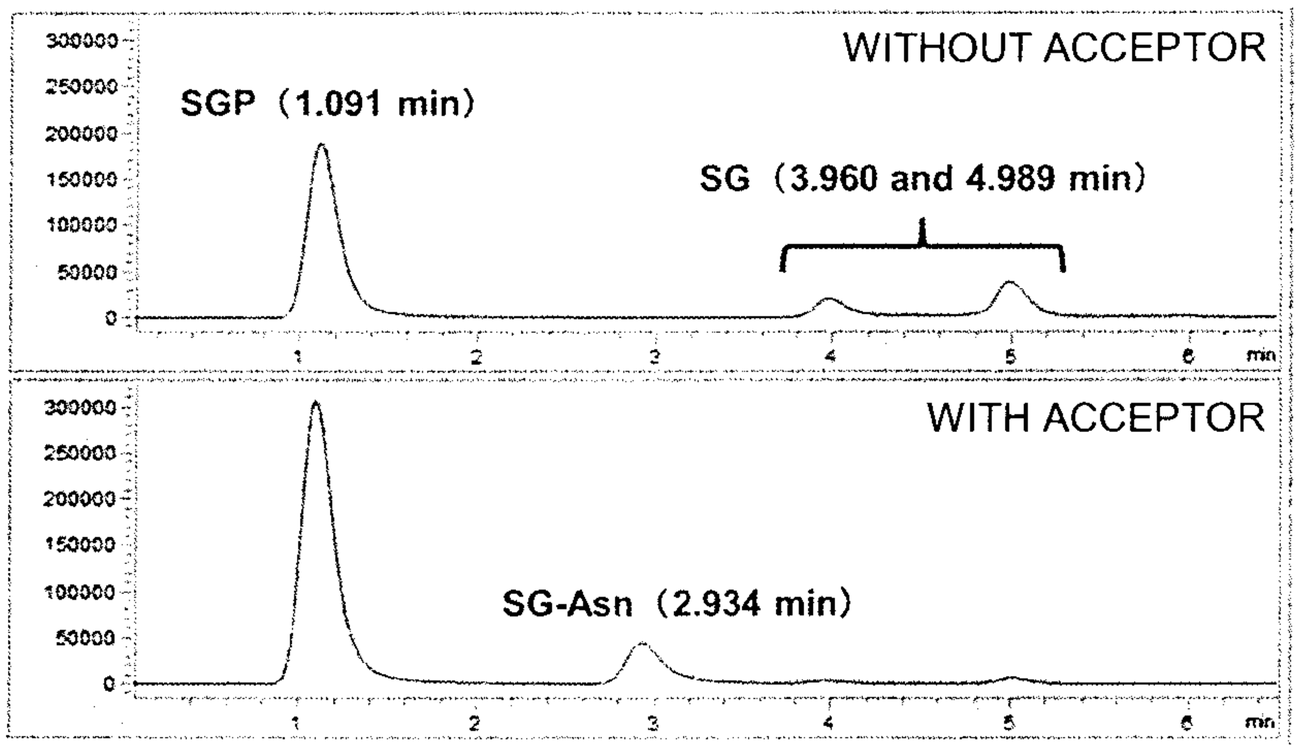


Figure 9

