

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
22 October 2009 (22.10.2009)(10) International Publication Number
WO 2009/129086 A2

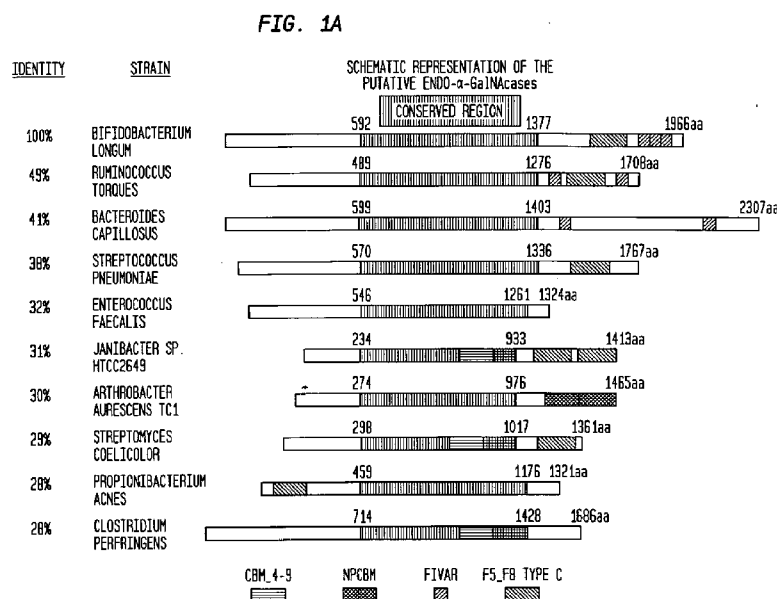
- (51) **International Patent Classification:**
C12N 9/24 (2006.01) *C12N 9/40* (2006.01)
- (21) **International Application Number:**
PCT/US2009/039735
- (22) **International Filing Date:**
7 April 2009 (07.04.2009)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/046,129 18 April 2008 (18.04.2008) US
- (71) **Applicant (for all designated States except US):** **NEW ENGLAND BIOLABS, INC.** [US/US]; 240 County Road, Ipswich, MA 01938 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** **KOUTSIOULIS, Dimitris** [GR/GR]; Aristeidou 4, GR-41222 Larissa (GR). **GUTHRIE, Ellen** [US/US]; 416 Lowell Street, Andover, MA 01810 (US).
- (74) **Agent:** **STRIMPEL, Harriet, M.**; New England Biolabs, Inc, 240 County Road, Ipswich, MA 01938 (US).
- (81) **Designated States (unless otherwise indicated, for every kind of national protection available):** AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States (unless otherwise indicated, for every kind of regional protection available):** ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** ENDOGLYCOSIDASES THAT CLEAVE O-LINKED GLYCANS

(57) **Abstract:** Methods and compositions have been described that relate to a newly identified polypeptide family wherein each member has *O*-glycosidase activity and specified sequence characteristics. This family of enzymes can be used for example for cleaving *O*-linked glycans and for synthesis of neoglycopeptides or neoglycoproteins.

ENDOGLYCOSIDASES THAT CLEAVE O-LINKED GLYCANS

BACKGROUND

5

Glycosylation is a common post-translational modification of proteins. Glycans are implicated in a wide range of biological events such as cell-cell interactions and recognition, inflammation, and autoimmune diseases (Ohtsubo and Marth *Cell* 126:855-867 (2006); Varki *Glycobiology* 3:97-130 (1993)). Detailed knowledge of glycan structures would facilitate structure-function correlations. This can be achieved by developing tools for highly sensitive analysis of glycan chains. For example, structural analysis of asparagine-linked carbohydrates (*N*-linked glycans) can be performed by releasing sugars from the protein backbone using enzymes such as PNGase F (Tarentino and Plummer *Methods Enzymol.* 230:44-57 (1994)). The *O*-linked glycans are most commonly attached to serine or threonine residues through the GalNAc residue at the reducing end. So far there is no enzymatic way of releasing of *O*-glycans intact. Consequently, this is achieved by chemical methods, typically by β -elimination with mild alkali (Takechi et al. *J Chromatogr A.* 680:209-215 (1994)) or mild hydrazinolysis (Royle et al. *Anal Biochem.* 304:70-90 (2002)). *O*-linked disaccharides (Core 1 type *O*-glycan) are among the most abundant core structures found in mucin glycoproteins, for example, the Thomsen-Freidenreich antigen (T antigen) immunodeterminant group and is used as a specific marker of carcinoma (Ohtsubo and Marth *Cell* 126:855-867 (2006); Varki *Glycobiology* 3:97-130 (1993)).

30

To date, endo- α -GalNAcases have been purified from *Clostridium perfringens* (Huang and Aminoff *J Biol Chem.* 247:6737–6742 (1972)); *Streptococcus pneumoniae* (Umemoto et al. *J Biol Chem.* 252:8609–8614 (1977), Glasgow et al. *J Biol Chem.* 252:8615–
5 8623 (1977) and Brooks and Savage *Glycoconj J.* 14:183–190 (1997)); *Alcaligenes sp.* (Fan et al. *Agric Biol Chem.* 54:233–234 (1990)); *Bacillus sp.* (Ashida et al. *Arch Biochem Biophys.* 373:394–400 (2000)); *Bifidobacterium longum* (Fujita et al. *J Biol Chem.* 280:37415–37422 (2005) and U.S. Publication No. 2006/0223140);
10 and *Streptomyces* (Ishii-Karakasa et al. *Biochem J.* 288:475–482 (1992) and Ishii-Karakasa et al. *Eur J Biochem.* 247:709–715 (1997)). All of the enzymes have a narrow substrate specificity, acting only on the α -linked disaccharide, Gal β 1,3GalNAc.

15

SUMMARY

In an embodiment of the invention, a polypeptide family is described for the first time in which each member is defined by: endo- α -N-acetylgalactosaminidase (endo- α -GalNAcase), also called
20 O-glycosidase activity; an amino acid sequence homology with EngBF measured by an expectation value of less than e^{-10} in a BLAST search; an amino acid sequence comprising SEQ ID NO:1; and an FDY amino acid sequence in a central conserved domain wherein the number of amino acids between the FDY and a C-
25 terminal end of the polypeptide member is no more than 200 amino acids. In particular, the O-glycosidase activity may result in cleavage of a Core 1 and a Core 3 glycan. Another feature of the polypeptide family may include that each polypeptide has a sequence containing less than 1400 amino acids. Another feature of

the polypeptide family may include an absence of a carbohydrate-binding domain at the C-terminal end.

In another embodiment of the invention, the polypeptide
5 member is a purified recombinant endo- α -glycosidase, the
glycosidase being capable of cleaving a glycosidic bond in
GlcNAc β 1,3GalNAc α IPNP.

In another embodiment of the invention, an isolated nucleic
10 acid is provided that encodes a polypeptide member having at least
35% amino acid sequence identity with SEQ ID NO:2 wherein the
polypeptide is capable of cleaving GlcNAc β 1,3GalNAc α IPNP.

In another embodiment of the invention, a method is provided
15 that includes the steps of cleaving an O-linked glycan consisting of
either Gal β 1,3GalNAc or GlcNAc β 1,3GalNAc linked α 1 to serine or
threonine from a glycoprotein or glycopeptide with a polypeptide
member characterized above, resulting in release of a disaccharide.

20 In other embodiments, a method is provided for synthesis of a
neoglycopeptide or neoglycoprotein using a polypeptide member in
the polypeptide family. A method is also provided for (a) performing
a BLAST search using a glycosidase in the family of glycosidases
described in claim 1 to identify homologs with an expectation value
25 of less than e^{-10} ; (b) determining the presence of SEQ ID NO:1 or a
sequence having at least 90% identity with SEQ ID NO:1; (c)
determining the presence of FDY and its location within 200 amino
acids from the C-terminal end such that if (a), (b) and (c) are
positive, verifying Core 1 and Core 3 glycosidase activity; and

(d) preparing by synthesis or cloning, a polypeptide member of a polypeptide family as described above.

BRIEF DESCRIPTION OF THE DRAWINGS

5

Figure 1A shows a schematic representation of the putative endo- α -GalNAcases. Alignment of the putative endo- α -GalNAcases that were revealed by BLAST search against the protein sequence of EngBF and the identity results are shown. The grey shaded boxes
 10 correspond to the center-located conserved region, while the black and patterned boxes indicate different types of sugar-binding domains (SBD also referred to as CBD). CBM_4_9: carbohydrate binding module (Pfam 02018), NPCBM: novel putative carbohydrate binding module (Pfam 08305), FIVAR: uncharacterized SBD (Pfam
 15 07554) and F5-F8 type C: member of the galactose-binding domain-like superfamily (Pfam 00754). (The Pfam database is a database offered by the Wellcome Trust (London, UK) and is accessible through <http://pfam.sanger.ac.uk>. See also Finn et al. *Nucleic Acid Research* 34:D247-D251 (2006)).

20

The protein accession numbers are: *B. longum* (AAX4493.1) (EngBF), *R. torques* (ZP_01966813) (EngRT),
B. capillosus (ZP_02035456), *S. pneumoniae* (YP_873926.1) (EngSP), *E. faecalis* (NP_815498.1) (EngEF), *Janibacter* sp.
 25 (ZP_00993766.1) (EngJSP), *A. aurescens* (YP_947239.1) (EngAA), *S. coelicolor* (NP_630440.1) (EngSC), *P. acnes* (YP_056270.1) (EngPA), *C. perfringens* (YP_695137.1) (EngCP).

Figures 1B-1 to 1B-15 show a ClustalW alignment of amino
 30 acid sequences of the proteins of Figure 1A: EngEF.pep (SEQ ID

NO:2), EngAA.pep (SEQ ID NO:4), EngBF.pep (SEQ ID NO:5), EngCP.pep (SEQ ID NO:6), EngJSP.pep (SEQ ID NO:7), EngPA.pep (SEQ ID NO:3), EngRT.pep (SEQ ID NO:8), EngSC.pep (SEQ ID NO:9), EngSP.pep (SEQ ID NO:10).

5

Figure 2 shows a purified endo- α -GalNAcases. SDS-PAGE analysis using 10-20% polyacrylamide gel, stained by Coomassie Brilliant Blue R-250. Lane M: molecular mass standards, lane 1: EngCP, lane 2: EngEF and lane 3: EngPA.

10

Figure 3 shows TLC analysis of the reaction products using *p*NP substrates and natural glycoproteins. Reaction mixtures were incubated with (lane +) and without (lane -) EngEF at 25°C for 16 h. Lane 1: Core 2 trisaccharide Gal β 1,3(GlcNAc β 1,6)GalNAc α 1*p*NP, lane 2: Core 3 disaccharide GlcNAc β 1,3GalNAc α 1*p*NP, lane 3: Gal β 1,3GlcNAc α 1*p*NP disaccharide, lane 4: GalNAc α 1*p*NP monosaccharide, lane 5: Core 1 disaccharide Gal β 1,3GalNAc α 1*p*NP, lane 6: Fetuin, lane 7: Fetuin + Neuraminidase, lane 8: Asialofetuin, lane 9: κ -casein glycopeptides, lane 10: κ -casein glycopeptides + Neuraminidase, lane 11: Glycophorin A, lane 12: Glycophorin A + Neuraminidase, lane 13: Mucin glycopeptides, lane 14: Mucin glycopeptides + Neuraminidase.

15

20

Figure 4 shows a TLC analysis of the transglycosylation reactions of five different endo- α -GalNAcases EngCP, EngEF, EngPA, EngSP and EngAL. Transglycosylation of disaccharide Gal β 1,3GalNAc α 1*p*NP (A, B) or GlcNAc β 1,3GalNAc α 1*p*NP (C) to various 1-alkanols by endo- α -GalNAcases. Lane 1: methanol, lane 2: ethanol, lane 3: 1-propanol, lane 4: 1-butanol, lane 5: 1-

25

pentanol, lane 6: 1-hexanol, lane 7: 1-heptanol, lane 8: 1-octanol, lane 9: 1-nonanol and lane S: Gal β 1,3GalNAc α 1pNP.

Figures 5-1 to 5-4 show a ClustalW sequence comparison of EngEF (SEQ ID NO:2) , EngPA (SEQ ID NO:3) and EngAA (SEQ ID NO:4), in which the conserved amino acid sequences that characterize the family of proteins of which the sequences are members, are highlighted. The grey box denotes the portion of the sequence that has been deleted in EngAA*.

10

DETAILED DESCRIPTION OF EMBODIMENTS

A family of novel *O*-linked glycosidases, which have broad enzyme specificity against Core 1 and Core 3 *O*-linked glycans, are described here. One or more of the following features can characterize these enzymes:

- (a) A sequence homology with EngBF as determined by a BLAST search in which the Expectation value is less than e^{-10} more specifically less than e^{-20} ;
 - (b) VDWQDAA (SEQ ID NO: 1);
 - (c) a C-terminal domain of no more than 200 amino acids measured from FDY in the central conserved region (see the grey shaded boxes in Figure 1); and
 - (d) a length of no more than 1400 amino acids in length.
- In addition to the above features, the family of glycosidases generally lacking any carbohydrate-binding domain (CBD).

A novel family of glycosidases have been cloned, purified and tested here. These enzymes possess both Core 3 and Core 1 *O*-glycan cleavage activity as demonstrated (see Tables I and II), for

30

example, using Gal β 1,3GalNAc α 1pNP and GlcNAc β 1,3GalNAc α 1pNP substrates as well as natural glycoproteins containing O-linked glycans that had been pre-treated with a neuraminidase.

O-glycosidase activity generally refers to endo- α -N-acetyl

- 5 galactosaminidase activity (endo- α -GalNAcase). Additional members of the family may be identified by means of a BLAST (Altschul et al. *Nucleic Acids Res.* 25:3389–3402 (1997)) search using the open reading frame for EngEF or EngPA as described in Example 1.

- 10 Examples of this family of O-glycosidases include EngEF, EngPA and a truncated EngAA (after removal of the two terminal CBDs). The amino acid sequences for these glycosidases are provided in Figure 5 and in Figure 1B, which also provides sequences for glycosidases with specificity for Core 1
- 15 O-linked glycans, demonstrated by cleavage with Gal β 1,3GalNAc α 1pNP. Figure 1A shows a cartoon describing the relative positions of the central conserved domains and the CBDs.

- The glycosidases within the newly characterized family share
- 20 a larger conserved region beyond required specific sequences as shown in Figure 1A. The central conserved domain is characterized by an overall % sequence identity of at least 4% greater than that observed for total protein (see Table III).

- 25 The glycosidic activity profile of the newly described O-glycosidases makes these enzymes a powerful tool for the release of O-glycan sugars from glycoproteins.

- In addition to glycosidic activity, the glycosidases were also
- 30 tested for transglycosylation activity. Transglycosylation can be

achieved using a glycosidase to attach a monosaccharide or disaccharide to a substrate. This can be achieved by driving the glycosidic reaction backwards in the presence of an excess of end product more particularly an excess of a disaccharide containing
5 Core 1 or Core 3 glycans. When incubated in the presence of Gal β 1,3GalNAc α 1pNP and several 1-alkanols, transglycosylation products could be detected up to 1-pentanol (Figures 4A and 4B) using the novel family of Endo- α -GalNAcases with a broad substrate specificity described here. A similar activity profile was observed
10 using GlcNAc β 1,3GalNAc α 1pNP as a donor and 1-alkanols as acceptors.

The references cited herein, as well as U.S. Provisional Application No. 61/046,129 filed April 18, 2008, are incorporated by
15 reference.

EXAMPLES

Abbreviations

20

Endo- α -GalNAcase: endo- α -*N*-acetylgalactosaminidase or *O*-glycosidase; pNP: *p*-nitrophenol; TLC: thin layer chromatography; CBD: carbohydrate-binding domain; SBD: sugar-binding domain; Gal β 1,3GalNAc: galactopyranosyl- β 1,3*N*-acetyl-D-galactosamine
25 pyranoside; GlcNAc β 1,3GalNAc: *N*-acetyl-D-glucosaminepyranosyl- β 1,3*N*-acetyl-D-galactosamine pyranoside; ORF: open reading frame; PCR: polymerase chain reaction.

Example 1: Endo- α -GalNAcases with a broad substrate specificity

30

A BLAST search (Altschul et al. *Nucleic Acids Res.* 25:3389–3402 (1997)) was performed using the EngBF protein sequence as a template. In this search, four potential glycosidases were selected from the BLAST results at a cut off expectation value of less than e^{-10} (and as low as $2e^{-96}$) and were cloned, purified and characterized. Two of these glycosidases were found to release Core 1 and Core 3 type O-glycans and a third was found to be a candidate for this activity if the sugar-binding domains were removed.

The structural and functional features of 4 endo- α -GalNAcases were compared (Figure 1A). These were (1) NP_815498.1 of *Enterococcus faecalis* (EngEF) (2) YP_056270.1 of *Propionibacterium acnes* (EngPA) that has an F5_F8 type C domain (member of the galactose-binding domain-like superfamily) close to the N-terminus, (3) YP_947239.1 of *Arthobacter aurescens* (EngAA) that has two NPCBM domains (novel putative carbohydrate-binding module) at the C-terminus and (4) YP_695137.1 of *Clostridium perfringens* (EngCP) that has a CBM4_9 domain.

Example 2: Cloning and expression of the *E. faecalis* endo- α -GalNAcase gene (*engEF*)

Based on the DNA sequence NP_815498.1 of *Enterococcus faecalis*, an oligonucleotide primer pair EFfor-NdeI (5' -CCCATATGAAACATGGAAAAATAAAACGATTTAGTAC-3' (SEQ ID NO:11), NdeI restriction site underlined) and EFrev-XhoI (5' -CCCTCGAGTTTTTTTGATTCCA CTGTGACCGTAAAG-3' (SEQ ID NO:12), XhoI restriction site underlined) was designed. The putative endo- α -GalNAcase gene was amplified by PCR using these primers and genomic DNA of *E. faecalis* ATCC 700802 as the

template. The PCR reaction mixture was incubated in an Applied Biosystems (Foster City, CA) thermal cycler (model 2720) for 30 cycles of 98°C for 10 sec, 56°C for 30 sec and 72°C for 2 min. The amplified product was isolated by agarose gel electrophoresis, gel
5 purified using QIAEX (Qiagen, Valencia, CA) and digested with NdeI and XhoI (New England Biolabs, Inc., Ipswich, MA). The resulting NdeI/XhoI fragment was inserted into an appropriate vector, in this case the pET-21a vector (Novagen, EMD Chemicals, Inc., Gibbstown, NY) previously digested with these enzymes. The
10 ligation mixture was used to transform competent bacterial cells resulting in plasmid pET21a-engEF. Plasmid pET21a-engEF was introduced into T7 Express lysY *E. coli* strain (New England Biolabs, Inc., Ipswich, MA). The transformed cells were grown overnight at 30°C with agitation in 20 ml LB medium (10 g/l tryptone, 5 g/l
15 yeast extract, 10 g/l NaCl, 1 g/l dextrose, 1 g/l MgCl₂) containing 100 µg/ml ampicillin. This culture was used to inoculate 1000 ml of fresh medium and antibiotics. The incubation temperature was shifted at 25°C and when the cultures reached an A₆₀₀ of 0.6-0.7, isopropyl thio-β-D-galactopyranoside at 0.3 mM was added.
20 Incubation continued for 12-14 h at 20°C. The cells were harvested by centrifugation (5.000 g for 20 min) and stored frozen at -70°C.

Example 3: Cloning and expression of the *C. perfringens* endo-α-GalNAcase gene (*engCP*)

25

Based on DNA sequence of YP_695137.1 of *C. perfringens*, an oligonucleotide primer pair CPfor-NdeI

(5' -CCCATATGGGTAGAAAATGCATGAATAAGAAGATTG-3'

(SEQ ID NO:13), NdeI restriction site underlined) and CPrev-XhoI

(5' - CCCTCGAGTCTAGCAGTTCTAACAGTTATTGATTCCTTAG -3' (SEQ ID NO:14), XhoI restriction site underlined) was designed. The putative endo- α -GalNAcase gene was amplified by PCR using these primers and genomic DNA of *C. perfringens* ATCC 13124 as the template and the amplification conditions described in Example 2. The amplified gene was cloned and expressed as described in Example 2.

Example 4: Cloning and expression of the P. acnes endo- α -GalNAcase gene (engPA)

Based on DNA sequence of YP_056270.1 of *P. acnes*, the putative endo- α -GalNAcase gene was chemically synthesized. The codons were optimized for gene expression in *E. coli* using DNAWorks software (Hoover and Lubkowski *Nucleic Acids Res*, 30, e43 (2002)). The optimized sequence was divided into six building blocks and synthesized (Hoover and Lubkowski *Nucleic Acids Res*, 30, e43 (2002)). After the sequence of each block was verified, the full sized gene was assembled using the USER method (Bitinaite et al. *Nucleic Acids Res*. 35:1992-2002 (2007)). Once assembled the synthesized *engPA* gene was amplified by PCR using the primer pair PA for (5' - GGAGACAUCCATATGAGTCGCACCC-3') (SEQ ID NO:15) and PArev (5' - GGGAAAGUTTAACGACCTTGACGTGAAAC-3') (SEQ ID NO:16) and inserted into the corresponding sites of pNEB206A using the USERTM Friendly Cloning Kit (New England Biolabs, Inc., Ipswich, MA). The resulting plasmid pNEB206A-*engPA* was transformed and expressed as previously described in Example 2.

Example 5: Cloning and expression of the truncated *A. aureescens* endo- α -GalNAcase gene (EngAA*)

Based on DNA sequence of YP_947239.1 of *A. aureescens*, an
5 oligonucleotide primer pair AAUR-noNPCBM-NdeI
(5' - CCCCCCCATATGCCCCGCTTGTCATCCC -3' (SEQ ID NO:17),
NdeI restriction site underlined) and AAUR-noNPCBM-HindIII
(5' - CCCCCCAAGCTTCAGCGTCCGCACGGTG -3' (SEQ ID NO:18),
HindIII restriction site underlined) was designed to remove the
10 signal sequence from the N-terminus of the protein and the two
NPCBM domains at the C-terminus. The truncated endo- α -
GalNAcase gene (EngAA*) was amplified by PCR using these
primers and genomic DNA of *A. aureescens* TC1 as the template. The
PCR reaction mixture was incubated in an Applied Biosystems
15 thermal cycler (model 2720) (Foster City, CA) for 30 cycles at 98°C
for 10 sec, and 72°C for 1 min 40 sec. The amplified product was
isolated by agarose gel electrophoresis, gel purified using QIAEX
(Qiagen, Valencia, CA) and digested with NdeI and HindIII (New
England Biolabs, Inc., Ipswich, MA). The resulting NdeI/HindIII
20 fragment was inserted into an appropriate vector, in this case the
pET-21a vector (Novagen, EMD Chemicals, Inc., Gibbstown, NY)
previously digested with these enzymes. The ligation mixture was
used to transform competent bacterial cells resulting in plasmid
pETAAUR_1465. The plasmid pETAAUR_1465 was introduced into
25 SHuffle™ T7 Express LysY Competent *E. coli* (New England Biolabs,
Inc., Ipswich, MA). The transformed cells were grown overnight at
30°C with agitation in 2 ml LB medium (10 g/l tryptone, 5 g/l yeast
extract, 10 g/l NaCl, 1 g/l dextrose, 1 g/l MgCl₂) containing 100
µg/ml ampicillin. This culture was used to inoculate 100 ml of fresh
30 medium and antibiotics. When the cultures reached an A₆₀₀ of 0.5-

0.6, isopropyl thio- β -D-galactopyranoside at 0.8 mM was added. Incubation continued for 4 h at 30°C. The cells were harvested by centrifugation (5.000 g for 20 min) and stored frozen at -20°C.

5 *Example 6: Method for purification of the expressed endo- α -GalNAcase EngEF*

All procedures were performed either on ice or at 4°C. Ten grams of frozen cells obtained above were thawed in 30 ml of 20
10 mM Tris-HCl pH 7.6, 200 mM NaCl, 1 mM dithiothreitol and were sonicated (6 x 20 sec with 30 sec intervals) using a Sonicator Ultrasonic processor model-375 (Misonix, Inc., Farmingdale, NY). Sample was subsequently centrifuged at 10,000 g for 15 min and the supernatant was collected for further purification.

15

The cell extract was diluted 3 fold with a buffer containing 20 mM Tris-HCl pH 7.6 (buffer A), applied onto a HisTrap™ Q HP column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) previously equilibrated in Buffer A and eluted with a linear gradient
20 of 0-1 M NaCl in Buffer A. Fractions containing the enzyme were eluted in ~0.2-0.75 M NaCl from the column. These fractions were applied to a HisTrap™ HP column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) previously equilibrated with Buffer B (20 mM Tris-HCl pH 7.6, 500 mM NaCl), followed by elution with a linear
25 gradient of 0-0.5 M imidazole in Buffer B. The enzyme was eluted at the concentration range of 0.06-0.27 M imidazole. These fractions were combined, dialyzed overnight against Buffer A and subsequently applied onto a Source 15Q column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) previously equilibrated in Buffer
30 A. Proteins were eluted with a linear gradient of 0.1-0.23 M NaCl in

Buffer A. Active EngEF were eluted in 0.14-0.18 M NaCl fractions, which were then combined and concentrated with a Amicon Centricon Concentrator 10 (Millipore, Billerica, MA) to about 5 ml. Concentrated samples were loaded onto a Superdex 75 column
5 previously equilibrated with Buffer C (20 mM Tris-HCl pH 7.6, 200 mM NaCl), and washed with Buffer C; and fractions with EngEF activity were pooled and concentrated. After the addition of glycerol at 50%, the purified enzyme preparation was stored at -20°C. Enzyme purity was judged by gradient polyacrylamide gel
10 electrophoresis under denaturing conditions (Laemmli *Nature* 227:680-685 (1970)). Protein concentration was determined using Bradford's dye binding assay (Bio-Rad Laboratories, Hercules, CA) (Bradford *Anal Biochem.* 72:248-254 (1976)) with bovine serum albumin as the protein standard.

15

Example 7: Method for purification of the expressed endo- α -GalNAcase EngCP

All procedures were performed either on ice or at 4°C. Ten
20 grams of frozen cells obtained above were thawed in 30 ml of 20 mM Tris-HCl pH 7.6, 200 mM NaCl, 1 mM dithiothreitol and were sonicated (6 x 20 sec with 30 sec intervals) using a Sonicator Ultrasonic processor model-375 (Misonix, Farmingdale, NY). Sample was subsequently centrifuged at 10,000 g for 15 min and the
25 supernatant was collected for further purification.

The cell extract was diluted 3 fold with Buffer A, applied onto a (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) 15Q column previously equilibrated in the same buffer and eluted with a linear
30 gradient of 0-1 M NaCl in Buffer A. Fractions in ~0.3-0.45 M NaCl

containing the enzyme were eluted from the column and applied onto a HisTrap™ HP column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) previously equilibrated with buffer B, followed by elution with a linear gradient of 0-0.5 M imidazole in buffer B. The enzyme was eluted in a concentration range of 0.12-0.19 M imidazole. These fractions were combined, dialyzed overnight against Buffer D (20 mM Sodium acetate pH 6.0) and subsequently applied onto a Source 15S column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) previously equilibrated in Buffer D. Proteins were eluted with a linear gradient of 0-1 M NaCl in Buffer D. Active EngCP (corresponding to 0.45-0.60 M NaCl) fractions were combined and concentrated with a Amicon Centricon Concentrator 10 (Millipore, Billerica, MA) to about 5 ml. Concentrated samples were loaded onto a Superdex 75 column previously equilibrated with Buffer E (20 mM Sodium acetate pH 6.0, 200 mM NaCl), and washed with Buffer E; and fractions with EngCP activity were pooled and concentrated. After the addition of glycerol at 50%, the purified enzyme preparation was stored at -20°C.

Example 8: Method for purification of the expressed endo- α -GalNAcase EngPA

All procedures were performed either on ice or at 4°C. Ten grams of frozen cells obtained above were thawed in 30 ml of 20 mM Tris-HCl pH 7.6, 200 mM NaCl, 1 mM dithiothreitol and were sonicated (6 x 20 sec with 30 sec intervals) using a Sonicator Ultrasonic processor model-375 (Misonix, Inc., Farmingdale, NY) the supernatant was collected for further purification.

The cell extract was diluted 3 fold with Buffer A, applied to a Source 15Q column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) previously equilibrated in the same buffer and collected in the flow through where most EngPA was detected. The fractions were
5 dialyzed against Buffer D. The dialyzed sample was applied onto a Source 15S column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ) previously equilibrated in Buffer D, followed by elution with a linear gradient of 0-1 M NaCl in Buffer D. Active EngPA (corresponded to 0.31-0.36 M NaCl) fractions were combined,
10 concentrated (5 ml) and loaded onto a Superdex 75 column previously equilibrated with Buffer E. Column was washed with Buffer E and fractions with EngPA activity were pooled and concentrated. After the addition of glycerol at 50%, the purified enzyme preparation was stored at -20°C.

15

*Example 9: Method for purification of the expressed truncated endo- α -GalNAcase EngAA**

All procedures were performed either on ice or at 4°C. 1 gram
20 of frozen cells obtained above were thawed in 10 ml of 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.2 mM PMSF and were sonicated (5 x 20 sec with 30 sec intervals) using a Sonicator Ultrasonic processor model-375 (Misonix, Inc., Farmingdale, NY). Sample was subsequently centrifuged at 20,000 g for 20 min and
25 the supernatant was discarded.

The pellet containing the inclusion bodies was washed two times by resuspending the pellet in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 1% Triton-X100 by vortexing and centrifuged at 20,000 g for
30 20 min. The resulting pellet was resuspended in 5 ml Buffer B

+PMSF (100 mM NaPO₄, 10 mM Tris-HCl, 8 M Urea, 0.2 mM PMSF, pH 8.0). This suspension was incubated slowly rotating at 25°C to solubilize the protein pellet for 1.5 hr. After incubation the protein suspension was diluted with 10 ml of Buffer B. The sample was
5 dialyzed overnight at 4°C against 2 liters of 100 mM NaH₂PO₄, 10 mM Tris-HCl pH 8.0. The next day any insoluble material was removed by centrifugation at 20,000 g for 20 min. The buffer of the solubilized protein (the supernatant) was exchanged and concentrated to 2 ml using a 5 KD Amicon filter (Millipore, Billerica,
10 MA) with two washes of 50 mM NaPO₄, 5 mM MgCl₂, pH7.5. After the addition of glycerol at 50%, the partially purified enzyme preparation was stored at -20°C.

Example 10: Characterization of endo- α -GalNAcase Activity

15

a) Substrate specificity on *p*NP substrates

The hydrolytic activity of the the purified enzymes EngCP, EngEF, EngPA and the commercially available EngSP and EngAL
20 were assayed using Gal β 1,3GalNAc α 1*p*NP (Core 1), GlcNAc β 1,3GalNAc α 1*p*NP (Core 3), Gal β 1,3(GlcNAc β 1,3)GalNAc α 1*p*NP (Core 2), Gal β 1,3GlcNAc α 1*p*NP and GalNAc α 1*p*NP as substrates and the released sugars were detected by colorimetric assay (Table I). Partially purified EngAA*
25 was assayed using Gal β 1,3GalNAc α 1*p*NP (Core 1), GlcNAc β 1,3GalNAc α 1*p*NP (Core 3), Gal β 1,3(GlcNAc β 1,3)GalNAc α 1*p*NP (Core 2) (Table I). The standard reaction mixture contained in a total volume of 100 μ l, 50 mM sodium phosphate buffer pH 7.5, 5 mM MgCl₂ and 0.25 mM
30 substrate. The released *p*-nitrophenol (*p*NP) was monitored by a 96-

well plate reader (SpectraMax M5, Molecular Devices, Inc., Sunnyvale, CA) at 405 nm at room temperature.

Gal β 1,3GalNAc α 1pNP was hydrolyzed at the most rapid rate by all enzymes tested. After a 16 h incubation, EngEF, EngPA and EngAA*

- 5 were the only enzymes capable of fully hydrolyzing the Core 3 disaccharide (GlcNAc β 1,3GalNAc α 1pNP). EngAL could partially hydrolyze it (27%) (Table I). EngAL could also partially release GalNAc while the rest of the enzymes released only traces of the monosaccharide (Table I). None of the enzymes tested could act on
- 10 Gal β 1,3GlcNAc α 1pNP and only EngAA* had significant activity that was detected when Core 2 trisaccharide was used as substrate (Gal β 1,3(GlcNAc β 1,6)GalNAc α 1pNP) (Table I).

b) Substrate specificity on pNP and glycoproteins using TLC

15

The substrate specificity was also determined for EngEF on TLC using various pNP glycosides and natural glycoproteins.

Gal β 1,3GalNAc α 1pNP, GlcNAc β 1,3GalNAc α 1pNP,

Gal β 1,3GlcNAc α 1pNP, GalNAc α 1pNP, calf κ -casein, human

- 20 glycophorin A, porcine mucin, calf fetuin, and calf asialofetuin were purchased from Sigma Alrich (St. Louis, MO).

Gal β 1,3(GlcNAc β 1,6)GalNAc α 1pNP was from Toronto Research Chemicals Inc. (North York, Ontario, Canada). For thin layer chromatography (TLC) analysis, a Silica Gel 60 plate (Merck,

- 25 Whitehouse Station, NJ) was developed in a solvent system of chloroform/methanol/water, 3/3/1 (v/v/v), and the sugars were visualized by spraying diphenylamine/aniline/phosphate reagent (Bailey and Bourne *J Chromatogr.* 4:206–213 (1960)).

Calf κ -casein, human glycoporphin A, porcine mucin, calf fetuin, and calf asialofetuin were incubated with EngEF. Based on the TLC analysis, sugars were only released when asialofetuin and mucin were used as substrates (Figure 3). Co-incubating the
5 substrates with neuraminidase increased the ability of the enzyme to release sugars. The released sugars migrated at the same height on the TLC plate as the Core 1 and 3 disaccharides (Figure 3).

c) Optimum pH

10

The pH dependence of enzyme activity was determined for EngEf, EngCP and EngPA in pH range 2.0–9.0 using the following buffers (50 mM): glycine-HCl (2.0–4.0), sodium acetate (3.5–6.0), sodium phosphate (5.5–8.0), and Tris-HCl (7.0–9.0). The pH
15 dependence can be similarly determined for EngAA*.

d) Steady-state enzyme kinetics

Glycosidases purified as described above were used for the
20 kinetics measurements for EngEF, EngCP and EngPA. Endo- α -GalNAcases from *Streptococcus pneumoniae* (EngSP) and *Alcaligenes sp.* (EngAL) were purchased from Roche (Basel, Switzerland) and Seikagaku Corporation (Tokyo, Japan), respectively. Steady-state enzyme kinetics were performed at 25°C.
25 The program HYPER v 1.01 was used to determine V_{max} and K_m values. The k_{cat} values were calculated from V_{max} using a molecular mass of 188000 Da for EngCP, 147000 Da for EngEF, 142000 Da for EngPA, 190000 Da for EngSP and 160000 Da for EngAL. Reported values were the average of three measurements. The standard
30 deviations did not exceed 5%.

The enzyme kinetics of EngCP, EngEF, EngPA, EngSP and EngAL was measured using Gal β 1,3GalNAc α 1pNP and GlcNAc β 1,3GalNAc α 1pNP (Table II). When Gal β 1,3GalNAc α 1pNP was used, EngEF exhibited the highest k_{cat} and EngPA had the lowest, about 25 times lower than EngEF activity. In the case of GlcNAc β 1,3GalNAc α 1pNP, the kinetic parameters could be measured only for EngPA and EngEF. EngPA was more active than EngEF and had 3 times higher k_{cat} . The enzyme kinetics can be similarly determined for EngAA*.

e) Transglycosylation activity

The transglycosylation reaction mixture contained in a total volume of 15 μ l, 50 mM sodium phosphate buffer pH 7.5, 5 mM MgCl₂, 0.8 mM GlcNAc β 1,3GalNAc α 1pNP or 1.6 mM Gal β 1,3GalNAc α 1pNP as substrates, 0.8 μ g of endo- α -GalNAcase, and various 1-alcanols as acceptors (13%, v/v) at room temperature for 16 h. Methanol, ethanol, 1-propanol, 1-butanol, 1-pentanol, 1-hexanol, 1-heptanol, 1-octanol and 1-nonanol were purchased from Sigma Aldrich (St. Louis, MO). The transglycosylation reaction mixtures were analyzed on a Silica Gel 60 TLC plate using chloroform/methanol/water 65/35/8 as the developing solvent and the sugars were visualized by spraying diphenylamine/aniline/phosphate reagent.

EngCP, EngEF, EngPA, EngSP and EngAL all exhibited similar transglycosylation activity. The transglycosylation products could be detected up to 1-pentanol (Figures 4A and 4B). While EngEF and EngPA which were capable of fully hydrolyzing GlcNAc β 1,3GalNAc α 1pNP compared to the quite low activity of the

rest of the endo- α -GalNAcases (Table I), no significant difference was observed in the transglycosylation products using EngEF, EngPA and EngSP with GlcNAc β 1,3GalNAc α 1pNP as donor and 1-alkanols as acceptors (Figure 4C). The transglycosylation activity
 5 can be similarly determined for EngAA*.

Table I. Substrate specificity using pNP substrates

	Eng CP	Eng EF	Eng PA	Eng SP	Eng AL	Eng AA*
% product released						
Gal β 1,3GalNAc α 1pNP Core 1	100	100	100	100	100	100
Gal β 1,3(GlcNAc β 1,6)GalNAc α 1pNP Core 2	2.5	2	0	0.6	0	24
GlcNAc β 1,3GalNAc α 1pNP Core 3	6	100	100	3	27	100
Gal β 1,3GlcNAc α 1pNP	0	0	0	0	0	NT
GalNAc α 1pNP	4.4	2.2	1.8	1.2	30	NT

10 Reaction mixtures were incubated with the different endo- α -GalNAcases at 25°C for 16 h. Product release was measured at 405 nm. NT= not tested.

Table II. Kinetic parameters of endo- α -GalNAcases using Gal β 1,3GalNAc α 1pNP (Core 1) and GlcNAc β 1,3GalNAc α 1pNP (Core 3) as substrates

Core 1	k_{cat} (1/sec)	K_m (μM)
EngCP	19.9	70.93
EngEF	51.17	47.85
EngPA	2.009	3.781
EngSP	10.51	40.37
EngAL	25.89	33.87
EngBF #	17.8	21.8
Core 3	k_{cat} (1/sec)	K_m (mM)
EngEF	9.434	20.03
EngPA	28.9	11.15

5

Data from Fujita et al. *J Biol Chem.* 280:37415-37422 (2005).

Table III. BESTFIT Analysis of endo- α -GalNAcases - % Identity and % Similarity

10

Proteins Compared	Total Protein		Conserved Region	
	% Identity	% Similarity	% Identity	% Similarity
EngEF to EngPA	35.2	43.8	39.2	47.6
EngAA to EngEF	38.7	47.0	44.1	52.4
EngPA to EngAA	44.9	51.8	50.2	56.8

In the claims

1. A polypeptide family in which each polypeptide member is characterized by:
 - 5 (i) O-glycosidase activity;
 - (ii) an amino acid sequence homology with EngBF measured by an expectation value of less than e^{-10} in a BLAST search;
 - (iii) an amino acid sequence comprising SEQ ID NO:1.; and
 - 10 (iv) an FDY amino acid sequence in a central conserved domain wherein the number of amino acids between the FDY and a C-terminal end of the polypeptide member is no more than 200 amino acids.
- 15 2. The polypeptide family according to claim 1, wherein each polypeptide member is capable of cleaving Core 1 and Core 3 O-glycans.
3. The polypeptide family according to claim 1, wherein each
20 polypeptide has a sequence containing less than 1400 amino acids.
4. The polypeptide family according to claim 1, wherein the polypeptide lacks a carbohydrate-binding domain at the C-terminal end.
25
5. The polypeptide family according to claim 1, wherein the polypeptide member is a purified recombinant endo- α -glycosidase, the glycosidase being capable of cleaving a glycosidic bond in GlcNAc β 1,3GalNAc α IpNP.

6. An isolated nucleic acid that encodes a polypeptide member according to claim 1, having at least 35% amino acid sequence identity with SEQ ID NO:2 wherein the polypeptide is capable of cleaving GlcNAc β 1,3GalNAc α IPNP.

5

7. A method, comprising:

cleaving an O-linked glycan consisting of either a Gal β 1,3 GalNAc or GlcNAc β 1,3GalNAc disaccharide linked α 1 to a serine or threonine from a glycoprotein or glycopeptide with the polypeptide member of claim 1.

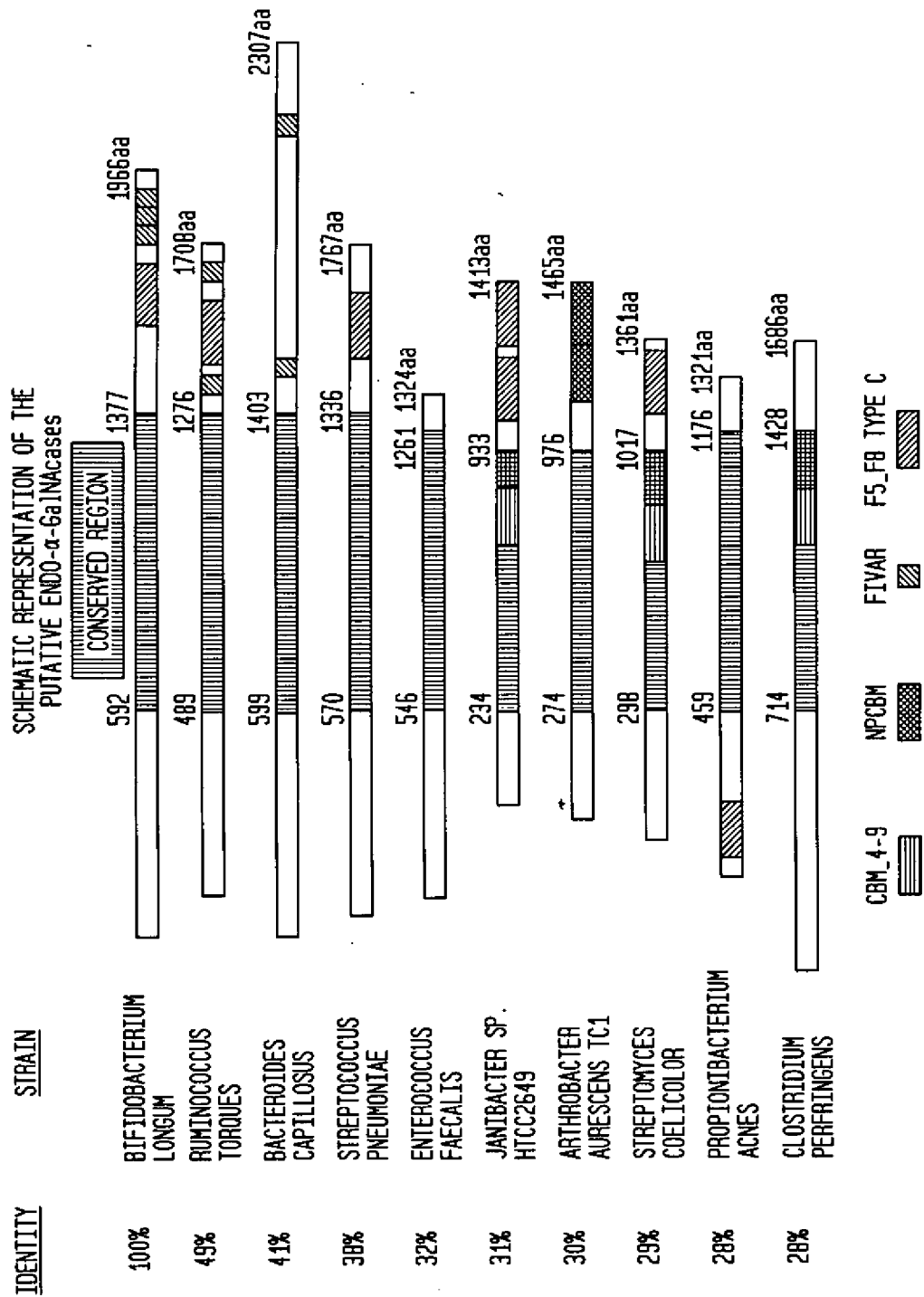
10

8. A method for synthesis of a neoglycopeptide or neoglycoprotein using a polypeptide member in the polypeptide family according to claim 1.

15

1/22

FIG. 1A



2/22

FIG. 1B-1

EngEF.pep	23	-----MKHGKIKRFSTLTMLASATILVP	10	20	30	40	50	60	70	80
EngAA.pep	0	-----								
EngBF.pep	40	-----MKKKYISAAALATLALTCGSGGGTAFVPLSDADLQTL								
EngCP.pep	80	-----MGRKCMKKIAAIIAAMVIGQLPISVLATPWNEAGDEINSESAILTNSDEEAEAYIQNYDRPEGITWTKLKSGSVEV								
EngJSP.pep	0	-----								
EngPA.pep	0	-----								
EngRT.pep	0	-----								
EngSC.pep	0	-----								
EngSP.pep	34	-----MKGLFEKRCYSIRKFSLSLGVASVIGAAPTGT								
EngEF.pep	78	-----LSTSAEETNSSTSTSSWEPTATEKLMQSDPFGKTYGWDVIG-----KTRRELAG	90	100	110	120	130	140	150	160
EngAA.pep	0	-----								
EngBF.pep	98	-----ASQIQINDTSDSATSETPSAQADAVEGWTIDSNIAQGEILEWANG-----WELKSTASN								
EngCP.pep	160	-----TDGFTSVTNGDTRIMEDQSPHIKNGELESKFTVGSQGLIFRATESHYGHNYNAGTGNVLEKSNWEDITGPRLNG								
EngJSP.pep	0	-----								
EngPA.pep	0	-----								
EngRT.pep	0	-----								
EngSC.pep	0	-----								
EngSP.pep	90	-----PVLADSVQSGSTANFLPADIALATATAKENDGDFEAPKVGEDQGSPE-----VTDGPKTEE								

	170	180	190	200	210	220	230	240	
EngEF.pep	ESLAISSD-----AAAGNNAVSLILDSFKLA-----								104
EngAA.pep	-----								0
EngBF.pep	GHAAMPSSNNWPVAVWGTDYDFSKAGSFHATIKSPQEGS-----								140
EngCP.pep	DWTVKATFVEKHLLTWNVSVDGEFTIYDKESDLIPLOAGKVGYRGCHNAKTTFPDYIKYSPMTIDRGPITYSINEVNVE-----								240
EngJSP.pep	-----								0
EngPA.pep	-----HSRTPRGRSIG-----								11
EngRT.pep	--MVMFN-----TFDPNKPGYFSFTLKSNNAHTG-----								28
EngSC.pep	-----								0
EngSP.pep	EILLALKE-----KPAEEKPKEDKPAAAKPETPKT-----								120
EngEF.pep	-DGEVETKYTAGSGRTGVIRGNTHDSNVFVGYNAN-----GKNLVESPNSHNDSISGPTINEDNYILLKRVYTG-----								175
EngAA.pep	-----								0
EngBF.pep	----AMRPGFYLGINDPSCGLFIGYDSGGHFNQTYTCGGSSGSWSYGARIAPSAHEEHDLRVSTNDANKVATLTVDGQKAF-----								216
EngCP.pep	TYPRVRKLPSSTVTNHENGASSIKDVSNIYIKESYSKPGTFKVEGTVEGTDVKALANVTYSSDLAYTETNFEETRGC-----								320
EngJSP.pep	-----								0
EngPA.pep	-----ALAVSAGTHLALLIAPTAPAHAETRYRIHQAA-----ITAVAADSAFAADPIISNTILDENPDTIHWTTHQNGK-----								78
EngRT.pep	-ISDSRVGVILGYNTDONGHYIGDNGGHFWQYKGG-NGDYIQOTRKPAFTKDQEVKVRIDMTADHKMTFFTLNGEWF-----								106
EngSC.pep	-----								0
EngSP.pep	----VTPENQIVANKEQGQGVTTIREKGVRYNQLSSYA-----QNDNAGRPALFEKKIGITVDANGHATVDLITFDKDSKK-----								190

EngEF . pep	-----KIGVAPKLPQVKVYTYNTGSEAN-----EAVRWNEIDPDAYKEPCTGEVDGTLENTLIKAKASIVWAKN-----	313
EngAA . pep	-----MPLRSSPG-----RLASLSLACVAVSSSLGILAIPTPAAAAPSTQPADIVSAAD-----	48
EngBF . pep	-----VRLDKVVKYGADGTFSFAD-----IEEGEHTLSIAKEGTVDSQVTVSGADIALDPTLTKTK-----	337
EngCP . pep	GDGKYGHPPGCTTPEQCKYTRIKLKVEDTNI TNWVDEGKIGQVAVSNLPDVRGKVGILGTFGFKHNVTLDLWVVEELGGIM	480
EngJP . pep	-----MPSVDTAAA-----	9
EngPA . pep	-----IRKUTFAATKATCVAKVKVSSW-----GGGSGDEVSSMAEFNAFTVDCSDPSPDPTPSEPPTPEVP-----	210
EngRT . pep	-----VTTGMLTAETDDGKYSLQ-----IGAGKHLELYTKAGYQATTSVTVTEGVEAKAVKLEKT-----	228
EngSC . pep	-----KRSVRVDTSGSSSPPTGRARV-----RRHGAVVAALGLTAGLSLAALPAGAAPPRPAAAAAPAGA-----	68
EngSP . pep	-----NLPDVTTLPAVNDHLKNEKKI-----LLKAGSYDDERTFVSVKTDNQEGVKTEDTPAEKTEGCEVDD-----	326

	570	580	590	600	610	620	630	640
EngGF. pep	NVAVPEIG---	LRPTTLTVSEGOELAMKLSIRE---	EGTKIHTISIPHQGLISVNSTDEGATPAG	444				
EngAA. pep	VLTLPDYG---	NTVIKARLSVKKN-VVSFNIFEKO---	SARHQVRTIQLPRLALTVVGSQPGCSQVST	175				
EngBF. pep	TLKVRDEAKKIDAVITVQIVTKAN-QLLHLVTVKIKNNLSEGIPI---	EGNGVEENAIQTLSPHQSLVSVRSSEKNAQPTG	484					
EngCP. pep	TLNPFEDIG---	MTITLKATVWKN-KLNKEVTDIQE---	GDVKLOTLNFPNHSIASVSSLNKGTALSV	616				
EngQSP. pep	ALTFADLP---	GVRIELTVSVEGN-VLTWRVTAKVTD---	SSSVKAVGTLDIQELDLVSVSSAQAGATTAP	135				
EngPA. pep	KLTFREDLP---	GVVELTADIKVSDG-VHTWSIPIHVD---	TPDRHVTVTVSVPGELTILASTVSTDPAQLSS	338				
EngNT. pep	ENTVVKKEKHIDAVIATBELTAKDN-TVSEFELTVKVENKLTGEGPCTALESNGVNGPIQITELIPHESLVSVNSIQANLILIG		381					
EngSC. pep	TSTTFEDLP---	GLTITSSITVTKETTVVFAVEKISG---	KAAPGVRTILAI PQSLSVSVDSAEFGANILAR	195				
EngSP. pep	IMKLRDDAHLINAEHTVRLQVDKN-QLLHFDVTVKIVN-HNQVTPRG---	QKIDDESKILLSSISFELGHALVSVSSHQCTGAKTGDG	469					

6/22

FIG. 1B-5

		650	660	670	680	690	700	710	720	
EngEF .pep	VVNTGTNANNGKNGDTIQDLTTTSQETK--KYNVGLTNTANYAASFYNVAGDG-----SVDGSDNHRJHKQTKEA	516								
EngAA .pep	ANLSVDR-----SVTGDFTPIASTPLDAAKSSAYALANTATILGAVESNALYDT-----SSGPGAKDRGRFWRQAVSD	246								
EngBF .pep	ARMSSNT-----QKPGDTNFAVTEDTNVTDS--DYTGFTSCAGLSAGLSNSSEHDGTYVAAPVRGSSQNTRYVYATTQQT	557								
EngCP .pep	LITGDMN-----NINEKFTDVAKAKPGVKG--KTYAFINDDKFAVTINNTYIE-----GGRVVLTTTEKDTLPD	678								
EngJSP .pep	TKIDTNS-----TRNADVIAPVKATDAVQTAPVGAAGIYNTAQLAASVETNGTYDK-----PTSATHSNDRGFWRHQVRAM	206								
EngPA .pep	ANIWVDR-----HKYGDLEQLATADVSQDT---SWVAMANDSTLAAGFEDNATQDG-----LVG-SAAATVARFVHSISQV	405								
EngRT .pep	AAWSTQT-----KVSQDEYVEVKANTPARER--DYHYAFVSNNEHSAGLSNSSEYEGRNAGACSSGSSQNTRYVNSVSEKK	454								
EngSC .pep	TKISTDS-----TTTADRFPVPTGDTAPDKPGVPTPYAFVGAQISAGIITNATEDS-----PQDNTWNTRLQSLRYDE	266								
EngSP .pep	ATMSNNT-----HVSQDDHIDVTNPKOLAK--GYHYGFVSTDKLAAGVNSNSQV-----SYGGSSNDWTRLTAYKETV	536								
		730	740	750	760	770	780	790	800	
EngEF .pep	ATG-FVTTLSSGANTYRPFDAF-----EDYTTGETPEVKVKFSDNSDNRVDWQDAALGF	571								
EngAA .pep	GAGGVNMGLASQNLTRAEGST-----TTEELPWTNRVALTSDANNDGGVDWQDAALAM	299								
EngBF .pep	G-DATSLGLASAPWYHRTVTDKSG-----AKYTYAETALPQMAVALAGDENEDCAVNWQDGAIAIY	617								
EngCP .pep	NTNYKVGISNGTWTYKELQDTTD-----QGSKLQGEKFWSEVILARDENEDGQVDWQDGAIQY	739								
EngJSP .pep	SDGSKRVGLAPGQWTVRGDGAP-----AQSELPAKVVTVTGADANADAADVWQDGAIAF	259								
EngPA .pep	G-GTKVGALEPATWVERGKGSATPPTDLSGNKAVCOLPGGATVKDGIQDPPDTPYVRVKLIVADANADGKVDWQDAAVAT	484								
EngRT .pep	D-GYVSMGLGSSANWYHRTVTDSHN-----RTWVLEETENPKMKVWITGCNCGDKNVWQDGAIVAF	514								
EngSC .pep	GEGRRAELISAGTYTTPPEGATDPR-----YDTYELPATVTVLADANDRGTVDWQDGAIAH	323								
EngSP .pep	G-NANYVGLHSSENQWEKAYKG-----JVFPETTKELPSAKVWITEDANADKNVDWQDGAIAIY	593								

7/22

FIG. 1B-6

		810	820	830	840	850	860	870	880	
EngEF . pep	650	RSIMNPFGEKVP	DELVMQRI	PPNFASQATNP	PVLTIDESKRIY	NLDGLQ	MILAKGY	NECHDS	AHPDYG	-AIGQ
EngAA . pep	379	RSIQVSPKGEQ	PDNVTI	THIPNFASQ	AHPPLRTLD	DDVKRIS	LA	NDGLQ	VNMLKGY	TSECHDS
EngBF . pep	696	RDIMNPFSEK	VP	ELVAMRI	AMNFGSQA	QNPPLT	LDNVK	VA	LT	DGLQSV
EngCP . pep	817	RKNKIPVGG	EIKNQSY	-IDFNIG	-YTQNP	PLRS	LT	TIKKLS	NY	TDG
EngJSP . pep	339	RSIAQVAKG	ADQVPR	VVQHIPP	NFSS	LA	THPPLRT	LD	DDVKRIS	QETD
EngPA . pep	564	RDVTHKPTGS	EDVANKVI	THIPN	IVSQA	THPPLRT	LD	DDVKRIS	LA	TDGLQ
EngRT . pep	593	RDIMNPFSEK	VP	ELVAYRI	AMNFGSQA	QNPPLT	LDNVK	VA	MT	DGLQSV
EngSC . pep	403	REHRRPLG	ADRV	PERV	QRI	PPNFASQ	A	THPFLAKY	LD	NTKRI
EngSP . pep	672	RSIMNPFGEK	VKDOI	TAYRI	AMNFGSQA	QNPPLT	LDG	IK	IN	LT
		890	900	910	920	930	940	950	960	
EngEF . pep	725	GEQALNQ	LIDEG	CHK	MAV	FGV	ELND	TESY	PEAKG	FNEEL
EngAA . pep	455	GLEDLNT	LV	KEG	EN	WAS	FGV	ELN	ATEI	Y
EngBF . pep	776	GADDMNT	HEG	SKY	G	AR	FGV	AV	N	ASEY
EngCP . pep	892	GKEDFNT	LI	EQ	G	K	Y	N	A	I
EngJSP . pep	414	GLDMNT	LI	EQ	G	K	W	G	A	D
EngPA . pep	640	GKMDLEK	L	T	ES	G	R	Q	W	N
EngRT . pep	673	GPEDM	K	T	L	L	E	K	G	A
EngSC . pep	478	GKMDL	R	L	R	T	G	A	G	I
EngSP . pep	751	GVEDF	K	T	L	E	K	A	K	Y

10/22

FIG. 1B-9

	1290	1300	1310	1320	1330	1340	1350	1360	
EngEF . pep	E-ANTPYVIAE----	PDSIEPMFTGCTPPKDPGFNEANTYKNNKVVFGDGE	-----	VKKDANGDYVFSSEKER	1037				
EngAA . pep	D-AKQPYILAPKNNKALPKKADFGCGTAFFNDPGFNGTD---	LSPNHPAGPYT-----	QVRDDKGRFPFAGGATP	754					
EngBF . pep	D-AETPYVYVK---	GEAKQIQVNNSEGHVVDAGFNGSGSNTLTDNFTVGGSG	-----	KAEVEG--DNWAMRLT	1141				
EngCP . pep	K-QDTPYIVTK--	ERVVEKRLEDWGYGSEIADPGFDSQT---	FDXNNKESTAENTDHIITERNESVQKRLGNDVLKISGHE	1208					
EngJSP . pep	K-SNTAYVLTP-N-SVPAQADPQFGCGGIVKDPGFNAST---	LGNSPQGTVT-----	QEHLSFGQVQAIGAGA	714					
EngPA . pep	TKAKTAYVLTP-TSKVPAAKTPNNGEGSHFANPGFYSGD---	TAGNARGNVS-----	VKHNDRGHFHLEFGKAAQ	946					
EngRT . pep	E-SEVPYVVCCK---	GEKENIKVTNSEGHIVDAGFNSGS---	LDNFTKAGEG-----	TAQIAKSOHSHPMKILD	1041				
EngSC . pep	E-KGQPYVLVP-NGGRAPHRDAHYGEFTGLSDPGFNGGD---	LDANWASGGAE-----	IYRACNGDNVVRLGEDA	782					
EngSP . pep	L-ANQPYVLTP---	SKQTHPEMSNSEGHIIYDGFNSGT---	LKHFTISGDAS-----	KAEIVKSQGANMLRLIQ	1096				
	1370	1380	1390	1400	1410	1420	1430	1440	
EngEF . pep	T-----	EIKQDINLPKPG-KYSLYINTEHTR-KATVTVKIGGKKYTR-----	TYNNSVAQWYIQADINHYS--R-	1098					
EngAA . pep	S-----	SISQDVQ-LDAG-TQSVSANIEIKQKTRPTTILSVLDIGKT-----	ESVTIDSNAENYVAGDE----	812					
EngBF . pep	G---	KVDVSQRLTDLKAGQKALYGVGDNRSTGDASVTVTSGEKVLATN-----	STGSLAKNYIKAYGENTN--SN	1208					
EngCP . pep	G--	ADAKISQISGLEEGVTYSVANWKNDRREVTLGNVGGEDFT-----	N--VITSGEKVQGEVYK-----	1269					
EngJSP . pep	G-----	SISQQLTGKAGSRYAASANWQVAPGGRPTTVSVKSGVD-----	YANAFDRSTVYNSEASNE-----	774					
EngPA . pep	S-----	QISQVLM-LPAG-DHSLFANVQIDPTKTRPVGLAVDGTGTPIDHKGCGGHAEVITSTAINATASDE-----	-----	1015					
EngRT . pep	G---	KVSMTQKLTDLKAGQYAVLVGVGDNRSDAKASVEIKSGDKVLGSN-----	YTTTRSIAKNYKAYTHNTN--SS	1108					
EngSC . pep	S-----	GLAQRVRLGTLPEERYTLGADVGLGGERRETTILVRGGKDS-----	EARTFDITPANREMSDE-----	842					
EngSP . pep	GNEKKSLSLQKLTGLKPHTKYAVTVGVGDNRSNKASITVNTGKEKVTY-----	YTHKSLALNYKAYAHNTRDNA	1167						

11/22

FIG. 1B-10

EngBF.pep	1171	1450	1460	1470	1480	1490	1500	1510	1520
EngAA.pep	880								
EngBF.pep	1285								
EngCP.pep	1341								
EngJSP.pep	843								
EngPA.pep	1086								
EngRT.pep	1184								
EngSC.pep	915								
EngSP.pep	1244								
EngBF.pep	1233								
EngAA.pep	948								
EngBF.pep	1349								
EngCP.pep	1400								
EngJSP.pep	905								
EngPA.pep	1148								
EngRT.pep	1248								
EngSC.pep	989								
EngSP.pep	1308								

12/22

FIG. 1B-11

	1610	1620	1630	1640	1650	1660	1670	1680	
	-----+-----								

FIG. 1B-13

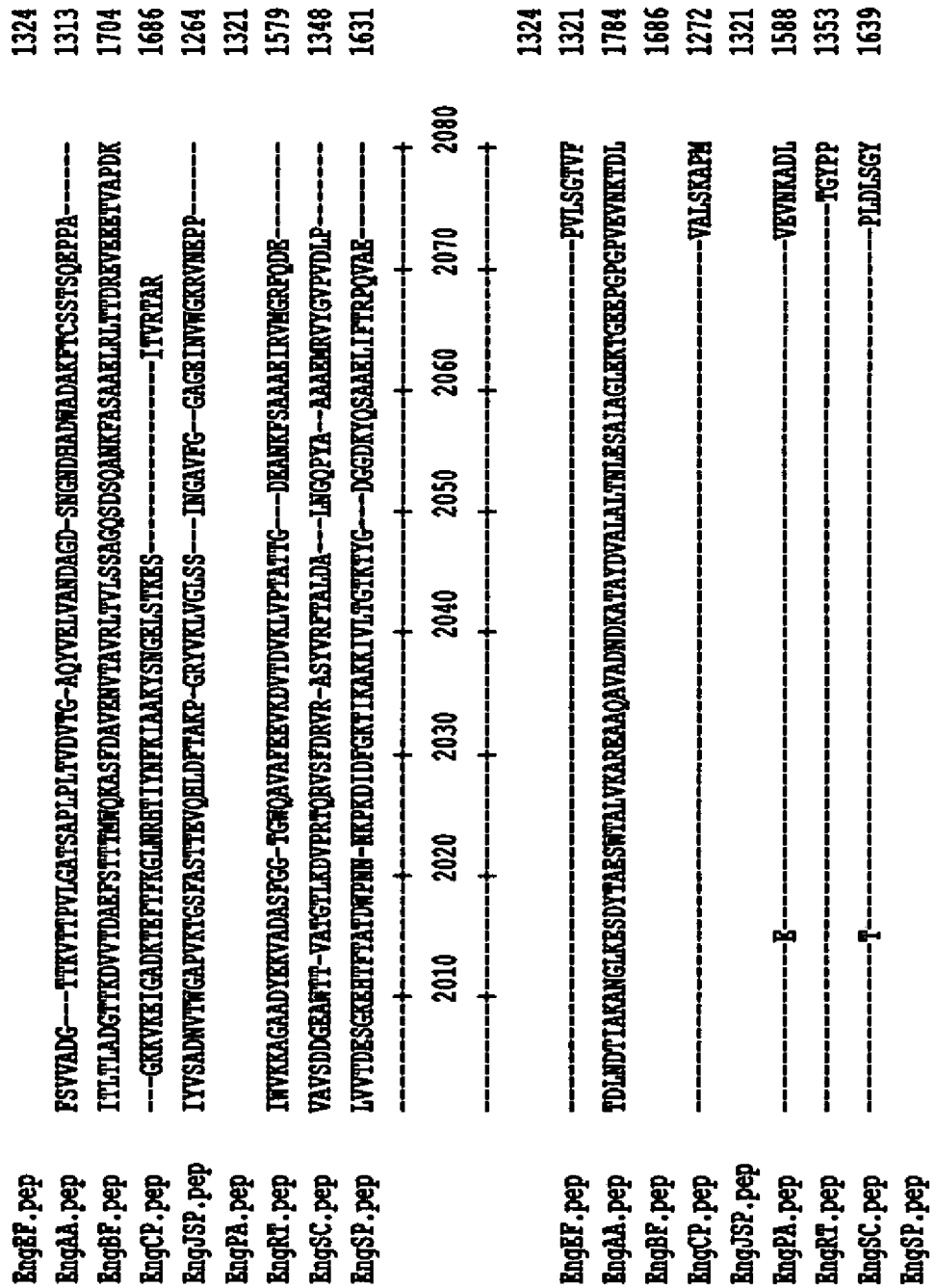


FIG. 1B-15

	-----+-----+-----			
	2250	2260		
	-----+-----+-----			
EngEF.pep	NGDDAGDDNGDWADAFSCA	[SEQ ID NO: 4]	1324	
EngAA.pep	SDVATIALAGLLIAGAGAAIAYRRREQI	[SEQ ID NO: 5]	1465	
EngBF.pep	N-SINGAIFAAIAELQFYAPATP	[SEQ ID NO: 7]	1966	
EngCP.pep	TSYFGAGAGIFGLFAAAAFERRRRQ	[SEQ ID NO: 8]	1686	
EngJSP.pep	ESQSDTALLIASVSLALSALFVVKTKD	[SEQ ID NO: 10]	1413	
EngPA.pep			1321	16/22
EngRT.pep			1708	
EngSC.pep			1361	
EngSP.pep			1767	

17/22

FIG. 2

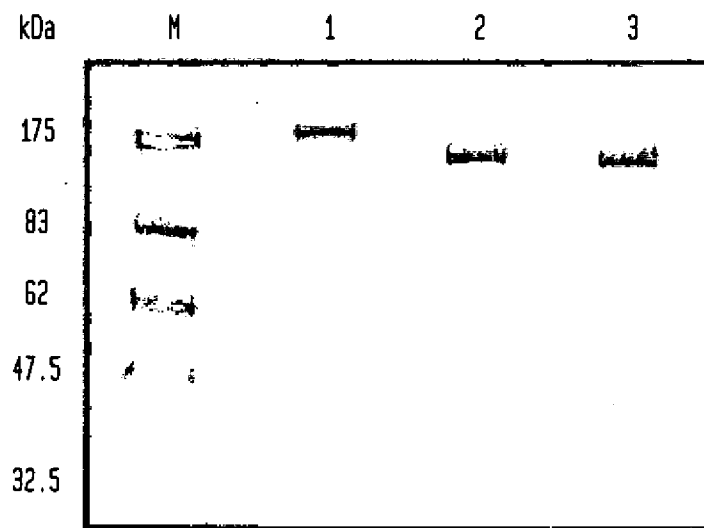
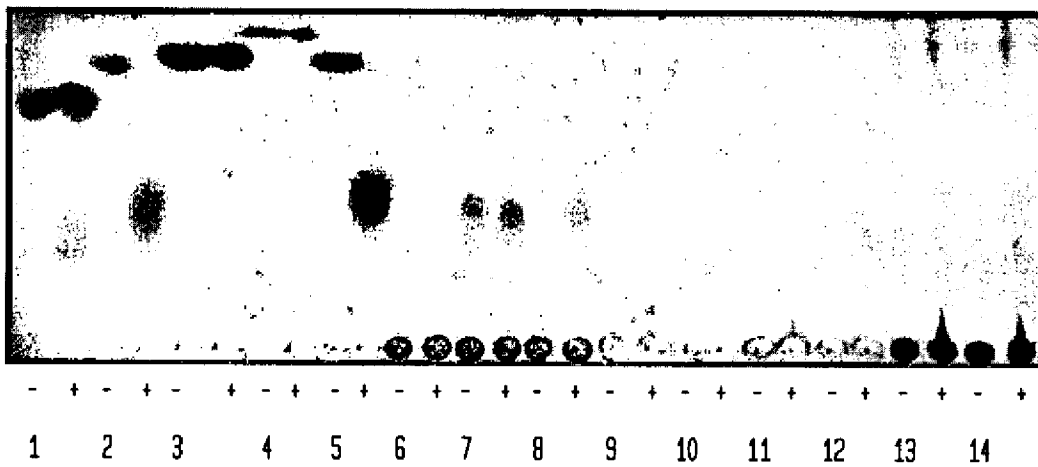


FIG. 3



18/22
FIG. 4A

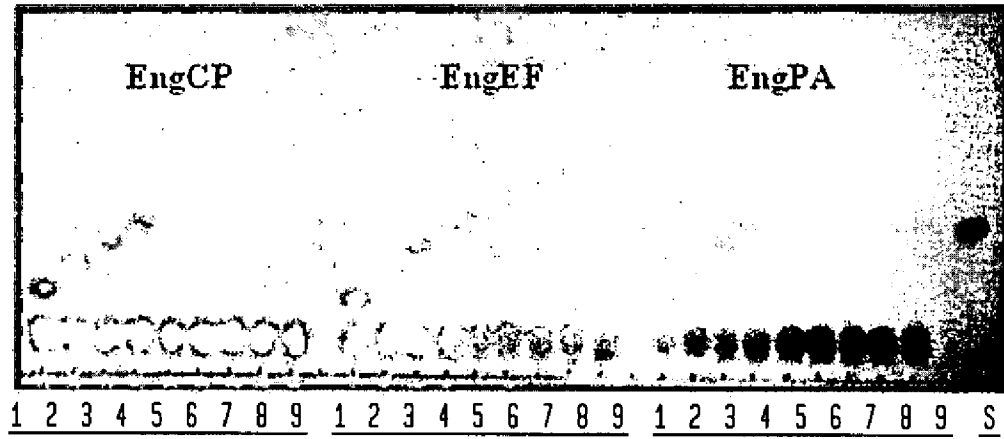


FIG. 4B

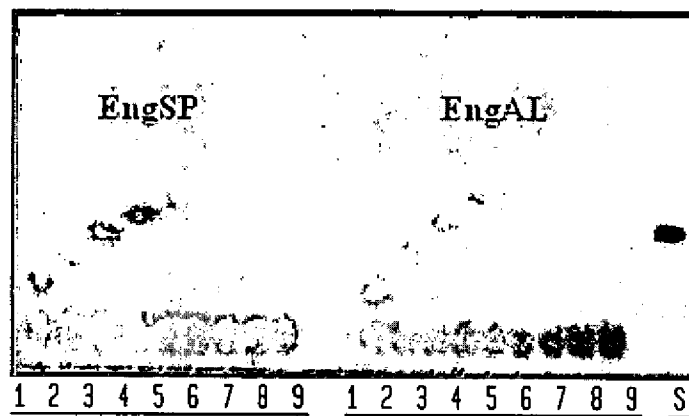


FIG. 4C

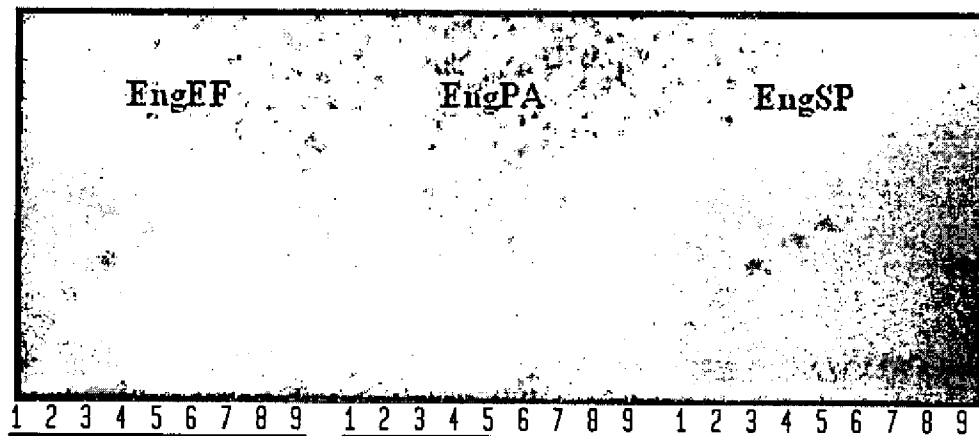


FIG. 5-1

	10	20	30	40	50	60	70	80	
EngEF.pep	MKHGKIKRFTLTLASATILVPLSTSAEETNSSTETSSSMVEPTATEEKLWQSDFPGGKTGEWQDVIGKTNRELAGE								80
EngPA.pep	MSRTPRGRSIGALAVSAGTMLALIAPTAPAHAEATRYRQINQAAITAVAADSATATDPISENTLDGNPDITWETTQWNGKDP								80
EngAA.pep	-----								0
	90	100	110	120	130	140	150	160	
EngEF.pep	LAI SRDAAAGNNVSLNLDSPKLADGEVETKFKYTAGSGRTGVIIRGNTKDSWVPVGYNANGKWLVESPNWSDSISGPT								160
EngPA.pep	LPEWIVFKLGDEAVNL-----GKVEITPRSSSNGSGRMHDYELY-----								119
EngAA.pep	-----MRLSSFG-----								8
	170	180	190	200	210	220	230	240	
EngEF.pep	LMEDTNYLLKVRVYVGEKITIWLNTTLIYEGEPVLANGDKIPTAGHVGVRLWYDKKIVNYDYFKNGPVDSIPEIVPEVTQ								240
EngPA.pep	-----TANTKTCKNAAFSSAKPVATGSYGA-----SDTSIRKITP								154
EngAA.pep	-----RLASLSLACVVASSSLG-----								25
	250	260	270	280	290	300	310	320	
EngEF.pep	IAPVKVFTKIGVAPKLPKQVKVTYNTGKEANEAVRWNEIDPDAYKEPGTFEVDGTLTENTNIKAKASIVVAKDNEAEKGDK								320
EngPA.pep	AATKATCVKVKVNSWGGDGSDEEVSSMAEFNAFTVDGSDPSDPDTP-----SEPPTPEVPKDAIS-----								215
EngAA.pep	-----LLAIPPAAP-----TQPADIVSAADTAT-----								51
	330	340	350	360	370	380	390	400	
EngEF.pep	ISSADLTAVVDPPFRIIRYEDPQSNQVIFNGQHEKIDQVMIDGKAYKATAEKQKSEANQAVYNVAVPEIG-LRFTTTLT								399
EngPA.pep	LSDGTVTVRARRDFPQVIDYTVGHAEMAGR--IGSPLTKVRINGADHVATVSAPTTTGSSASWKLTPFDLPVELTADIK								293
EngAA.pep	ITSGDLRVVDGTTFPQVLGYTDAASKARLDG-TTTLRLSTITLNGTEYTVSGTSAASGKDARDYVFLFLPDFGNTVIKARLS								130
	410	420	430	440	450	460	470	480	
EngEF.pep	VSEGQELAMKLSDIRE-EGTKIHTISIPNQGLISVNSTDEGATFAGVVMNTGTNANNGNKNGDTIQDLTTLT--SQEETK								476
EngPA.pep	VSDG-VMTWSIPHIVDTPDHRVNTVSVPGTLASVTSTDPAQALS-----SANIVVDRNKTGDLFQPLAT---ADVSDQT								364
EngAA.pep	VKKN-VVSFNITEIKDSEAHQVRTLQLPRLNLVTVGSTQPGSQVS-----TANLSVDRSVTGDFTPTITASTPLDAAKS								204

20/22

FIG. 5-2

	490	500	510	520	530	540	550	560								
EngEF.ppe	YMYGFLNTANYAASF	WTHAYGDGSDGSD	HNH----	IHKQTKEAATGF	VTTLS	SGANTYRPF	DAPEDYT-----		541							
EngPA.ppe	SWVAMANDSTLAAG	FEDNATQDGLVGS	AATVAR--	FVESISQVGGTK	VGAIEPATW	VERGKSATP	FPPTDSLGNK	AVCQL	442							
EngAA.ppe	SAYALANTATLGA	AVESNALYDTSS	GPGAKDRGRFW	RQAVSDGAGG	VNMGLASGQ	NLYRAEGSTT-----			269							
	570	580	590	600	610	620	630	640								
EngEF.ppe	-----	TGETPEVKVKFSK	SDNDN	VDWQDA	IGFRS	IMNPMGA	EKVP	ELVNQRI	PFNFASQATNP	FLVTLD	609					
EngPA.ppe	PGGATVKDGIG	PDPTPYVRVKI	VADANADG	VDWQDA	AVATRDV	TMKPTG	SGDVANK	VI	THIPFNIVS	QATHP	PLRTLD	522				
EngAA.ppe	-----	TEELPWTRVAIT	SDANNDG	VDWQDA	ATAMRSI	QVSPNK	GEQTPD	NVI	THIPFNIVS	QATHP	PLRTLD	337				
	650	660	670	680	690	700	710	720								
EngEF.ppe	ESKRIYNLTDGL	QMNLLKGYQNE	GDSAH	PDYGA	IGORP	GGEQALN	QLIDEGH	KLNAV	FGVHIND	TESYPEAK	GFNEE	688				
EngPA.ppe	DVKRISLATDGL	QQALLKGYQAE	GDSAH	PDYGGNV	SHRAGG	MKLEKL	TESGRQ	WNTPG	IEHNLV	ESYPEAN	HFQDN	602				
EngAA.ppe	DVKRISLATDGL	QVAMLKGYT	SEGDS	SANTDYG	NFN	TRAGGLE	DLNTLV	KEGKE	WNASPG	VHINATE	IYPEAKS	FSSED	417			
	730	740	750	760	770	780	790	800								
EngEF.ppe	LVDP-TRGND	WLDPSYFI	KQRPDTLS	GRRYERF	EKLQKAP---	NLDYI	YVDVW	GNQGES	GNASRQ	LSKEIN	SLGWFTT	764				
EngPA.ppe	ILVKPYQKAND	WMEQSYR	MDYAKDL	GSGQLF	ARLNQLR	KELGAKS	NLDWLY	PDN---	YPAGW	QNDRI	ANALNA	EGWRIG	679			
EngAA.ppe	LLRADKGLG	WNWLDQ	SYMNO	REDIN	SGKLAQ	RIKEL	REST--	NKNL	DFVY	VDVY---	YEPGW	LAERLQ	QELVKNG	FRVG	492	
	810	820	830	840	850	860	870	880								
EngEF.ppe	NEFFNALEYD	SVNNEWSA	EKDYGG	TTKGF	NSTIVR	FIRNHQ	KDTWII	SDNPL	GGAB	FEAYEG	WVGK	TNFTYR	QKTFA	844		
EngPA.ppe	SEWSSTYPR	YNQWSE	WANDEN	YG-TG	NKGYSS	RIIRFID	NSRRD	TWNP--	DPIL	GNSNV	VEYEG	WTSEN	DYNA	FLANVWQ	756	
EngAA.ppe	SEWADHLS	RNHTW	SEWAND	EKYGG	STNK	GINSQ	ILRFIN	NTQSD	VWNP--	DPKL	GVSH	IVEFEG	WTQND	FNAPSEN	VWT	570
	890	900	910	920	930	940	950	960								
EngEF.ppe	INVPTKFLQ	HYQITN	WETTTA	DGQIY	GTIKL	ANGAEK	VTVT---	QADAN	SPRSIT	LNETE	VLKGD	-AYLL	PWN---	VN	916	
EngPA.ppe	RNLPTKFLQ	RS	DIMSWQD	-----	GRIAF	ANGAV	ATSSK	SISG	HEIPT	ARTIT	FDGAT	VPKEG	GSYLL	PWS---	NG	824
EngAA.ppe	ANVPAKFLQ	HH	PITKTA	-----	ERIEL	ADG	VAVTGN	-----	TAEGR	NITV	GGTSV	LQGG	-TYLL	PWSSK	ENG	632

FIG. 5-3

	970	980	990	1000	1010	1020	1030	1040				
EngEF.pep	GQDKLYHWNPKGGTSTWSL	DKKMQGRTHLHL	YELTDQGR	IDKGAIATTNNQVT	IQAEANTPYVIAEPDS	----	IEPMTFG		992			
EngPA.pep	GSDRLYYWNPNGSATWKL	TNSWAAQKSVSLF	MLTDTGRVKVAEIP	VTNRSIRIPATKAKTAY	VLVLYPTSKVPAAKT	PNWG			904			
EngAA.pep	KVDKLYHNPPTGGASTWT	LTOEFAKSSSLEQ	FKLTDNGRUVKQVAD	VPVVGQVTVTADAKQ	PYILAPKNNKAELPK	KADFG			712			
	1050	1060	1070	1080	1090	1100	1110	1120				
EngEF.pep	TGTPFKDPGFNEANTL	KNNWVFRGDGEVKK	DANGDYVFSSEKERT	EIKQDINLPK--	PGKYSLYLWTTETED	RKATVTVK			1070			
EngPA.pep	EGSHFANPGFYSGDTAG	---WNARGNVSVKH	NDRGNPHLEPGKAQ	SQISQVLNLPAGDHS	LWAWVQIDPTKTRP	VGLAVD			981			
EngAA.pep	EGTAFNDPGFNGTDLSP	---WNPAGPVTQVR	DDKGRFAENGATPSS	ISQDVQLDAGTQSV	SAWIEIQGKTRPTL	SV			789			
	1130	1140	1150	1160	1170	1180	1190	1200				
EngEF.pep	IGGKK-----	YTRTVNNSVAQNYI	QADINHTSRKNPQ	YMNMRIDFEIPDMA	KKGSVTLAVDKGNS	SVTKFDDLR			1139			
EngPA.pep	GTGVTPIDHQKCGGHA	ESVITSTTAINATA	SDEYFG-----	TYHQRLRVAFHS--	DGRPVTVTLKALAG	NAIVSADDFR			1054			
EngAA.pep	IDGKT-----	ESVTIDSSNAENY	VAGDEKHG-----	TAFQRI RVLVDV	PRNNTKATVTVQA	ADGATVRVDDFR			853			
	1210	1220	1230	1240	1250	1260	1270	1280				
EngEF.pep	IVERQTDIMN-PDKQT	VIKQDFEDTQAVGL	YPVKGSAGGVEDP	RIHLSE	RNEPPTYQYGNW	GNLVS	VDVLEGNWSLKA	HKQ	1218			
EngPA.pep	VYDAAVPSDPHVTPAT	VLFPQNFEDVDTG	-YWPFTGSAGMEG	DARTQLSRRE	PPTYQKGNWGRAN	DSVLSGDWSLKM	HBE		1133			
EngAA.pep	AVKTR-----	VPTTGVLSEDFEN	VDQG-WGPFVKG	DAGGSTDPRT	HITERREFFTQ	KGWDANVIDE	VLDTGWSLIA	DE	927			
	1290	1300	1310	1320	1330	1340	1350	1360				
EngEF.pep	G-----	AGLMLQTI	PQNIKFE	PN-KKYTVQ	FDYQTDGENV	FTAGTINGEL	KNNNDPKPV	GELTSTAADGQTK	HYEAEII	1291		
EngPA.pep	R-----	NGIVLRTT	ASAPLTG	GGTRYRIS	FDYQADKP-G	YSFVTGHD	KVSGKSVKE	VITESHANGVAT	STTHFSTDIV	1206		
EngAA.pep	NRAPNGGPGMVYRTTE	ASVPQAG-HKYK	VSFDYQNSK	AGQYANWSGY	DSQAGPAVTGS	QAIEAKTSTTR	FEQILD	TGFC		1006		
	1370	1380	1390	1400	1410	1420	1430	1440				
EngEF.pep	GDASGN	TTFGIFTGADK	FIMDNFTV	VESKK	[SEQ ID NO: 2]				1324			
EngPA.pep	VKDQ--	PAWIGF	THEQGEGD	MSIDNLR	IEKLDPR	PISVTSTQAA	VFPD-----		1251			
EngAA.pep	GDYF--	VGLORT	GSSNGSD	FTLDN	FLVEDL	GASRAV	PACAQLS	AELOGDVVQOG	KADQFDYTT	FVSDEPAISGL	VALEL	1084

FIG. 5-4

	1450	1460	1470	1480	1490	1500	1510	1520	
EngEF.pep									1324
EngPA.pep	----ACKPTPEPIQPAQPSASAPTTSGSPQAPGTGN-----RPNRYALPRTGADGAG-----								1299
EngAA.pep	PEGWTATPSTPATAPTLPAAGTGLFTTWKITAPASADGDYPITAKASYTVSSSGIDPAGSRTISTTTTVRTLPKPPQATVP								1164
	1530	1540	1550	1560	1570	1580	1590	1600	
EngEF.pep									1324
EngPA.pep	-----LG-----FSSSEAASATAA								1313
EngAA.pep	ASDEHPWVSATNGWGPVEKDQSNGGTGAAGDGPLTLNGTVYARGLGARANGTVRYYLGGYCTAPTATVGIIDDAQPTRGSVK								1244
	1610	1620	1630	1640	1650	1660	1670	1680	
EngEF.pep									1324
EngPA.pep	VGVSRRGR [SEQ ID NO: 3]								1321
EngAA.pep	FSVVADGTTKVITPVLGATSAPLPLTVDVDTGAQYVELVANDAGDSNGNDHADWADAKFTCSSTSQEPAPVLSGTVFASD								1324
	1690	1700	1710	1720	1730	1740	1750	1760	
EngEF.pep									1324
EngPA.pep									1321
EngAA.pep	LPWIGSTNGWGAERDRANGEQNAGDGPALRLDGVVYSKGIGVHADSKISIAEAKCTAFTAVAGVDDAKLNKRLHGSVV								1404
	1770	1780	1790	1800	1810	1820			
EngEF.pep									1324
EngPA.pep									1321
EngAA.pep	FIVKGGGRELRLTPVLSADSAAFLNVDITGVQNVELIADKNAGDDAGDDWGDWADAKFSCA						[SEQ ID NO: 4]		1465