# Molecular Cloning and Expression of Endo-β-N-Acetylglucosaminidase D, Which Acts on the Core Structure of Complex Type Asparagine-Linked Oligosaccharides<sup>1</sup>

Hisako Muramatsu, Hiroshi Tachikui, Hiroaki Ushida, Xiao-jun Song, Yunhai Qiu, Satoshi Yamamoto, and Takashi Muramatsu<sup>3</sup>

Department of Biochemistry, Nagoya University School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550

Received March 2, 2001; accepted March 23, 2001

Endo-β-N-acetylglucosaminidase D (Endo D) produced by Streptococcus pneumoniae cleaves the di-N-acetylchitobiose structure in asparagine-linked oligosaccharides. The enzyme generally acts on complex type oligosaccharides after removal of external sugars by neuraminidase,  $\beta$ -galactosidase, and  $\beta$ -N-acetylglucosaminidase. We cloned the gene encoding the enzyme and expressed it as a periplasm enzyme in Escherichia coli. The first 37 amino acids in the predicted sequence are removed in the mature enzyme, yielding a protein with a molecular mass of 178 kDa. The substrate specificity of the recombinant enzyme is indistinguishable from the enzyme produced by S. pneumoniae. Endo-β-N-acetylglucosaminidase A (Endo A) from Arthrobacter protophormiae, the molecular mass of which is 72 kDa, had 32% sequence identity to Endo D, starting from the N-terminal sides of both enzymes, although Endo A hydrolyzes high-mannose-type oligosaccharides and does not hydrolyze complex type ones. Endo D is not related to endo- $\beta$ -N-acetylglucosaminidases H, F<sub>1</sub>, F<sub>2</sub>, or F<sub>3</sub>, which share common structural motifs. Therefore, there are two distinct groups of endo-β-N-acetylglucosaminidases acting on asparagine-linked oligosaccharides. The C-terminal region of Endo D shows homology to  $\beta$ -galactosidase and  $\beta$ -N-acetylglucosaminidase from S. pneumoniae and has an LPXTG motif typical of surface-associated proteins of Gram-positive bacteria. It is possible that Endo D is located on the surface of the bacterium and, together with other glycosidases, is involved in virulence.

Key words: as paragine-linked oligosaccharides, endo- $\beta$ -N-acetylglucosaminidase, endogly cosidase, *Streptococcus pneumoniae*, virulence.

Asparagine-linked oligosaccharides are classified into complex type, which have external sialyl-galactosyl-N-acetylglucosamine chains, high mannose type, which have exposed larger clusters of mannosyl residues, and hybrid type (1). The di-N-acetylchitobiose structure in asparagine-linked oligosaccharides is cleaved by endo- $\beta$ -N-acetylglucosaminidases [EC 3.2.1.96], a group of enzymes with different specificity for other portions of the substrates (2, 3). Endo- $\beta$ -N-acetylglucosaminidase D (Endo D), which is produced by *Streptococcus pneumoniae*, has a strict specificity for the mannosyl core structure, and generally acts on complex type oligosaccharides after the removal of external sugars by neuraminidase,  $\beta$ -galactosidase and  $\beta$ -N-acetylglucosaminidase (4-6). Endo D was the first endoglycosidase found to act on glycoprotein-bound oligosaccharides

© 2001 by The Japanese Biochemical Society.

(7), and played important roles in elucidating the processing pathway for the biosynthesis of oligosaccharides (2, 8), and clarifying the function of carbohydrates in IgG (9). This enzyme is still of general interest because of its ability to remove various complex type oligosaccharide structures from intact glycoproteins in the absence of detergents in collaboration with neuraminidase,  $\beta$ -galactosidase and  $\beta$ -*N*-acetylglucosaminidase (10). Although *N*-glycanase is effective for removing most asparagine-linked oligosaccharides, it often requires substrate denaturation for efficient action (11, 12). Here, we describe the molecular cloning of Endo D and compare the predicted structure with those of other glycosidases.

## MATERIALS AND METHODS

Materials—DEAE-Sephadex A-25, concanavalin A (Con A)—Sepharose and HiTrap<sup>TM</sup> were purchased from Amersham Pharmacia Biotech AB (Uppsala, Sweden). PVDF membrane, Centriplus, and Centricon were purchased from Millipore (Bedford, MA, USA). Polyethylene glycol 20,000 was from Wako (Tokyo). Complex type glycopeptides from bovine thyroglobulin were prepared as described (13); then those that bound to Con A–Sepharose and were eluted with 0.1 M methyl mannoside were pooled and used. Exoglycosidase digestion of the glycopeptides was performed as de-

<sup>&</sup>lt;sup>1</sup> This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan. The nucleotide sequence for endo- $\beta$ -N-acetylglucosaminidase D reported in this paper has been submitted to DDBJ with accession number AB055806.

 $<sup>^2</sup>$  To whom correspondence should be addressed. Tel: +81-52-744-2059, Fax: +81-52-744-2065, E-mail: tmurama@med.nagoya-u.ac.jp Abbreviations: Con A, concanavalin A; Endo, endo- $\beta$ -N-acetylglucosaminidase.

0

ATATAAABCGCTTTCTTGAAAACAACAAAATCAATCTTTTAGGAGGAGAAA

ATGAAGAATCCATTTTTTGAAAGACGTTGCGTTACAGTAATTCGTAGAAGTATCAGTAGGAGCCTGCTCGCTGATGATTGGTGCTGTTTATTGTGGTCCAGCCTTGGCTGAAGAAACT 120 θCAGTTCCTBAAAATAGCBGABCTAATACAGAGCTTGTTTCAGBAGAGAGTGABCATTCGACCAATGAAGCTGATAAGCAGAATGAAGBOBGAACATACTAGAGAAAAGCTABAAAGG A Υ Ρ Ε Ν S Θ Α Ν Τ Ε L V S Θ Ε S Ε Η S Τ Ν Ε Α D K Ο Ν Ε Θ Ε Η Τ R Ε Ν K L Ε K 240 80 GCAGAAGGAGTAGCAICAECATCTGAAACTGCAGAAGCAGCTAGCGAGCTGAAGCAGGTGAGGTGGAGGTGGAGGTAGCACACCATCTGCAGAAGCAAACCTAAGTCTGAC A E G V A T A S E T A E A A S A A K P E E K A G E V V A E T P S A E A K P K S D 360 120 MEGANACABAAGCAAAGCAAAGCAACTAACCAAGEGGATBABTCTAAACCAGCAGCAGAAGCTAATAAGACTGAAAAGACTGAACAGCAGATGTCCAGCAGATGTCCCCTAAAAATACAGAAAAACCA K E T E A K P E A T N Q G D E S K P A A E A N K T E K <u>E Y Q P D Y P K</u> N T E K T 480 160 TTANACCAAAGGAAATCAAATTTAATTCTTGGGAAGAATTGTTAAAATGGGAACCAGGTGCTCGTGJGAAGATGATGCTATTAACCGCGGGATCTGTTGTCCTCGCTTCACGTCGGAAGGT L K P K E I K <u>F N S W E E L L K <u>H</u> E P G A R</u> E D D A I N R <u>G S Y Y L A S R</u> R T G 600 200 CATTIAGICAATGAAAAGCTAGCAAGGAAGGAAAAGTICAAGCCTTATCAAACACCAATTCTAAAGCAATGCCTATGCTICTGTGGGGAGAAGAGTICAAGGCCTATGCTTTGAC H L V N E K A S K E A K <u>V Q A L S N T H S K</u> A K D H A S V G G E E F K A Y A F D 720 240 840 280 960 320 ATCANCCMGANACALCTIGGAGAATTTGGTTAAACCTCTTGGAGAAAAGATGGGGCCAGTTTATGCTCTATAGCAAGGAATATGCTGCTAAGGAAACCATCCAATCAAGTATTCTTGGTAT LN O E T T G D L V K P L G E K W R <u>Q F W L Y S K</u> E Y A A K V N H P I K Y S W Y 1080 360 GATGCCATGACCTATAACTATGGACGTTACCATCAAATGGTTIGGGAGAATACAACTACCAATTCATGCAACCAGAAGGAGAATAAAGTTCCGGGCAGATAACTTCTTTGCTAACTTAAC D A M T Y N Y G R <u>Y H Q D G L G E Y N</u> Y Q F N Q P E G D K V P A D N F F A N F N 1200 TBBBATANGBCTANAATBATTACACTATTBCAACTBCCAACTBBATTBGCBTAATCCCTATBATBTATTTBCAGGTTTGGAATBCCAACABBBTGGTTCCTACAABACAAABBTAB W D K A K N D Y T I A T A N W I G R N P Y D V F A G L E L O O G G S Y K T K V K 1320 TOGRATGACENTIAL ACCONTINUES CONTINUES TO THE ACCONTINUES AND A CONTINUES AND 1440 1560 520 AATACAGGTCATGGTAAAAATGGTTCSTAGATGGTAAGGTTTCTAAGGATTCTGAGTGGAATTATCGTTCAGTATCAGGTGTTCTTCCAACATGGCGCTGGTGGCAGACTTCAACAGGG N T G H G K K W F Y D G K V S K D S E W N Y R <u>S V S G V L P T W R W W O T S T G</u> 1680 GAMMACTICGTGCAGAATATGATTTTACAGATGCCTATAATGGCGGAMATACCCTTAAATGCCGGTGATGTGAGCCGGTAAGACAGATCAGGATGTGAGACCTTTATTCTACTAAGTA E\_K\_L\_R\_A\_E\_Y\_D\_F\_T\_D\_A\_Y\_N\_G\_G\_N\_S\_L\_K\_F\_S\_G\_D\_V\_A\_G\_K\_T\_D\_0\_D\_V\_R\_L\_Y\_S\_T\_K\_L 1800 600 1920 640 CTAACCCTTTCTGACAACTGAAGAAATGAAGAATTTGATCTCAGCTCACTAGCAGGTAAAACCATCTATGCAGTCAAACTATTTTTTGAGCATGAAGGTGCTGTAAAAGATTATCAGTTC LTLSDNWTNEEFDLSSLAGKIIIYAACCATCTAGCAGGTAAAACCATCTATGCAGTCAAACTATTTTTGAGCATGAAGGTGCTGTAAAAGATTATCAGTTC 2040 680 ANCETCEGRACULTIANCEATECTOSEACAUTCACCANGAGECACULTEGECEACAAGECITTECTETAGTAGTAGTACTECTETAUAATECCCAAGUGEGEGUAGEAGTTETGECMETTETAUA $N \ L \ G \ Q, L, T$  | S D N H Q E P Q S P T S F S V V K Q S L K <u>N A Q E A E A V V Q</u> F K 2160 720 BOCALCAAGGATBCABATTTCTATGAAGTTTATGAAAAAGATBGABAAAGCAGCTGGAAATTACTAACTBGCTCATCTTCTACAACTATTTATCTACCAAGTAAOCGCTCABCAABTBCT <u>G N K D A D F Y E Y</u> Y E K <u>D G D S W K L L T G S S S T I I</u> Y L P K V S R S A S A 2280 760 CAGEGTACAACTCAAGAACTGAAGGTTETASCAGTCGBTAAAAATGGAGTTCGTCAGAAGCTGCAACCACAACCTTTGATTGGGGTATBACTGTAAAAGATACCAGGCTACCAAAACT 0 g t t 0 e l k v v a v g k n g v r <u>s e a a t t t f d w g w t v</u> k d t s l p k p 2400 800 CTAGCTBAAAATATCBTTCCABGTGCAACAGTTATTGATABTACTTTCCCTAAGACTGAABGTGCABAAGBTATTGAABGTATGTABGTACGATAACTAGCTTGTCABATAAATBG L A E N I V P B A T V I D S T F P K <u>T E G B E G I E G N</u> L N G T I T S L S D K W 2520 840 TCTTCAGCTCAGTTGAGTGGTABTGTGGATATTCGTTTGACCAGGCCACGTACCGTTGCTAGGGTCATGGATCATGCAGGGGCTGGTGGTGAGGGTGAGGCTGGTGATGACC S S A O L S G S V D I R L T K P R T V V R W V N D H A G A G G E S V N D G L M N 2640 880 ACCANAGACTITGACCTITATTATAAAGATGCAGATGGTGAGTGGBAAGCTAGCAAGGCAGCTGGCAACAAGGCAGCAGGTGACAGATATCACTCTTGATAAACCAATGACTGCTGAT T K D F D L Y Y K D A D G E W K L A K E V R G N K A H V T D I T L D K P I T A Q 2760 920 ĸ GACTOGGGGTTURAATGTTUSTCACTTCUGAAATGGAACTCCATGGAAGGGTATTCGTATCCTATAACTGGAAATUTAGAAAGGTTGATACTGGAGGTGTCAATATTCCGATGGCCAAG D W R L N Y Y T S D N G T P W K A I R I Y N W K W Y E K L D T E S Y N I P M A K 2880 GCTOCAGCCCGTTCTCTAGOCATAACAAGGTACAAGTTGGCTTIGCAGATGTACAGGCTGGAGCAACTATTACCGTTTATGATAATCCAAATTCTCAAACTCCGCTCGCAACCTGGAG A A A R S L G N N K V Q V G F A D V Q A G A T I T V Y D N P N S Q T P L A T L K 3000 1000 ASCEANGTTGEAGEGAGACCTASCAAGTGCACCATTGGATTTGACAAATCAATCTGGTCTTCTTTATTATCGTACCCAGTAGCCAAGGAAATTABTAATGTCCTASCAGTTTCGTT S E V 9 G D L A S A P L D L T N 0 S G L L Y Y R T 0 L P G K <u>E I S N V L A V S V</u> 3120 3240 1090 3360 1120 AATTIGICAGTGACTGGCCAAGACGAAGCAAGTCCGAAAACTATTITIGGGAATTGAAGTAAGTCAGAAACCGAAAAACTAGCTAGTGGTGATAGCTTAGACTTGTCTGAA N L S V T V T G Q D E A S P K T I L G I E V S Q K P K <u>K D Y L V G D S L D</u> L S E 3480 1160 3600 1200 3720 1240 TTTBCTTCACCABALETALABALECTTTCTTCAAABCAATTBALGCBBCCCGAACAABTGTTGALABCCATBALATTABCLCCCAAGATCAABTCAATBALCCBACTTAATABCAAT FASPEVKEAFLKALIEAAAEQVLKDHEISTQ 3840 1280 3960 1320 GOCCTAGCTOCCCTTCTTBAGAUUACAAGGTCTTGGTTGGUUAGTAGATTTGAGTCCABAAGAGCTTGCAACAGCGAULCAGAGTCTAAUAGATCTGGTTGCTTTATTGUUAGAAGAC A L A P L L E K N K V L V E K V D L S P E E L A T A K Q S <u>L K D L V A L L</u> K E D 4080 1360 ANGCCAGCAGTCTTTTCTBATAGTAAAACAGGTGTTGAAGTACACTTCTCAAATAAMGAGAAGACTGTCATCAABGGTTTGAAAGTAGAGGGTGTTCAAGCAAGTAGTAGAAGAAGAA K P A V F S D S K T G V E V H F S N K E K T V I K <u>G L K V E R V Q A S A E E K K</u> 4200 1400 TACTTIGCTGGAGAAGATGCTCATGTCTTTGAAATAGAAGGTTIGGATGAAAAAGGTCAAGATGTTGATCTCTCTTACGCTTCTATTGTGAAAATCCCCAATGAAAAGATAGAAAGTT Y F A G E D A H Y F E I E G L D E K G O D Y D L S Y A S I V K I P I E K D K K V 4320 MANNETATTTTCTTACCTANGCCANARAGCAGTABAATIGCCTTTGAACHCOGATAGTCATGTTATCTTACCGCACACCTTACTCACTATATGCCTTIGTTATGATTT 4440 1480 GCTRANAMACCACAMOCTGCTAMACCAGCACCACAMAGCACAMAGTCCTTOCAMAACCTACTTATCAACAGGGCTTCTRATCAACAAGGCTCCTAAATTGGAAGTTCAACAAGGGTT A E K P Q P A K P A P Q N K V L P K P T Y Q P A S D Q Q K A P K L E V Q E E K V 4560 1520 GCCTTTCATCATCATGAGAGATGATAATBCTGAGATGCTAGTTGGGGAGAACAACGAGTCATCATACAGGGAGAGAAGATGGACTGTTAAGACATGTCTTGAAGAACGGTCAGGGT A F H R Q E H E N A E M L V G E O R V I I Q G R D G L L R H V F E V D E N G O R 4680 1560 4800 1600 4920 1640 AAAAGAGAAGAAGAATAAAGATTAAATATCGAAAAATCTTGTGAAAATCTTICCG K R E D K D  $\star$ 4970 1646

Fig. 1. DNA sequence of the Endo D gene and its predicted protein sequence. Solid lines indicate peptides identified in purified Endo D after in-gel trypsin digestion. The double solid line indicates the N-terminal sequence of recombinant Endo D. The 3'- and 5'-untranslated regions were also cloned by PCR and a portion of each respective sequence is shown.

scribed previously (4). Ovalbumin glycopeptides  $(Man)_6$ -(GlcNAc)<sub>2</sub>Asn and  $(Man)_6$ (GlcNAc)<sub>2</sub>Asn (6) were gifts from Seikagaku Kogyo (Tokyo).

Purification of Endo D—Endo D was purified as described previously (10). The major active fractions were pooled, concentrated by dialysis against polyethylene glycol 20,000, dialyzed further against 10 mM Tris-HCl buffer, pH 7.0, containing 0.15 M NaCl, and concentrated by Centriplus and Centricon (Millipore).

SDS Polyacrylamide Gel Electrophoresis and Determination of Partial Amino Acid Sequences-Proteins were separated by SDS-PAGE in 7% gels (14). Protein bands lightly stained by Coomassie Brilliant Blue were excised and digested in the gel with trypsin essentially as described by Rosenfeld et al. (15). The resulting peptides were separated on a reverse phase column (Monitor C18, 1 × 150 mm) with a linear gradient of acetonitrile, 0.1% trifluoroacetic acid at a flow rate of 50 µl/min using a MAGIC2002 (Michrom Bio-Resources, Auburn, USA). Amino acid sequence analysis of the resulting peptides was carried out using a 494A protein sequencer (Applied Biosystems, Foster City, USA). For the N-terminal amino acid sequence, purified recombinant enzyme was subjected to SDS-PAGE, and transferred to a PVDF membrane. The 180 kDa band was cut out, and its amino acid sequence was analyzed with a 494A protein sequencer.

Molecular Cloning of the Endo D Gene—A gene in S. pneumoniae (sp234 gene) was found to show homology to endo- $\beta$ -*N*-acetylglucosaminidase. To amplify the sequence, PCR was performed after denaturation at 94°C for 5 min with 30 cycles of 94°C for 40 s, 55°C or 60°C for 40 s, and 72°C for 2 min using genomic DNA from S. pneumoniae as a template. The gene was amplified in 5 fragments with the following primers: figures in the parentheses after Endo D indicate nucleotide numbers of the Endo D gene in Fig. 1; introduced restriction enzyme sites are also shown. Fragment I, 5'-ATCTTTGGTACCCTCGAGATGAAGAATCCAT-TITTTGA-3' (KpnI-XhoI-Endo D1-20) and 5'-GATGGT-TTACCTTAGCAGCA-3' (Endo D1060-1041); Fragment II, 5'-AAAGATTTGCTGAAGCTTTG-3' (Endo D860-879) and 5'-TCACCTGGATCCTGGCCAGTCGGGTCTCCTTG-3' (BamHI-Endo D1472-1453); Fragment III, 5'-CTTGGTG-AATTCGCCCCAGATACCATTACAAG-3'(EcoRI-Endo D1375-1394) and 5'-ATATCCGGATCCCCACTCAACTGA-GCTGAAGA-3' (BamHI-Endo D2540-2521); Fragment IV, 5'-CCCTAAGACTGAAGGTGGAG-3' (Endo D2448-2467) and 5'-TATCAGGAATTCCTGCTGGCTTGTCTTCTTC-3' (EcoRI-Endo D4090-4070); Fragment V, 5'-GTAGATGAAT-TCCCAGAAGAGCTTACAACAGC-3' (EcoRI-Endo D4018-4037) and 5'-CACAAGGCGGCCGCTTCGAAAATCTTTAT-CTTCTCTTTTC-3' (BstBI-NotI-Endo D4939-4920). Final-



Fig. 2. Cloning of the Endo D gene. Fragments I to V were produced as described in "MATERIALS AND METHODS," and digested with the appropriate restriction enzymes. The products (I to V) were ligated utilizing newly formed restriction sites shown in the figure.

Production of Recombinant Endo D-The coding sequence of the enzyme was ligated into the expression vector pBAD/gIII (Invitrogen, Carlsbad, CA, USA), producing a protein with a myc epitope and (His)<sub>e</sub> tag at the C-terminus that can be secreted into the periplasm. The recombinant gene was transformed into Escherichia coli TOP10, which were then cultured at 37°C in 1 liter of LB medium containing 50  $\mu$ g/ml of ampicillin to an OD<sub>600</sub> of 0.5. Protein expression was induced by adding 0.002% arabinose to the culture medium. After 4 h, the cells were collected by centrifugation at 3,000 ×g for 10 min at 4°C, and suspended in 20 mM Tris-HCl buffer, pH 8.0, containing 2.5 mM EDTA and 20% sucrose, and incubated on ice for 10 min. After centrifugation for 1 min at 4°C, the cell pellets were resuspended in 20 mM Tris-HCl buffer, pH 8.0, containing 2.5 mM EDTA, incubated on ice for 10 min, and centrifuged for 10 min at 4°C. The supernatant was dialyzed against 20 mM phosphate buffer, pH 7.5, containing 0.15 M NaCl, and applied to a HiTrap chelating column. The recombinant enzyme was eluted with 100 mM imidazole according to the manufacturer's instructions. The eluted enzyme was further purified by gel filtration on a Sephadex G-200 column as described previously (4).

Determination of Enzyme Activities—The glycopeptides were N-acetylated with [<sup>14</sup>C]acetic anhydride (7.20 mCi/ mmol, NEN Life Science Products, Boston, MA, USA) as described previously (4). Enzyme assay and calculation of the enzyme unit were performed as described (4), except that the substrate and the product were separated by Con A-Sepharose affinity chromatography (10).

Other Methods—The amount of protein was determined using a BCA protein assay kit (Pierce, Rockford, IL, USA).

#### RESULTS

Molecular Cloning of Endo D—Two different methods were used to clone Endo D. First, the S. pneumoniae genome in the TIGR (The Institute for Genomic Research) database was searched for homology to known endo- $\beta$ -Nacetylglucosaminidases, *i.e.* those from Arthrobacter protophormiae (Endo A, Ref. 16), Streptomyces plicatus (Endo H, Ref. 17), and Flavobacterium meningosepticum (Endo F<sub>1</sub>, Ref. 18; Endo F<sub>2</sub>, Ref. 19; Endo F<sub>3</sub>, Ref. 19). Only Endo A



Fig. 3. **SDS-PAGE of Endo D.** The recombinant enzyme produced by *E. coli* (A) and the enzyme produced by *S. pneumoniae* (B) were analyzed in 7% gels. Protein bands were detected by staining with Coomassie Brilliant Blue. Arrowheads indicate the positions where standard proteins migrated; their molecular masses are shown in kDa.

showed a homologous 5,110 nucleotide sequence (sp234 gene) with 66% sequence identity at the nucleotide level (35% identity at the protein level). This result was surprising, since Endo A hydrolyzes high-mannose type oligosaccharides but not complex type oligosaccharides (16). The second approach was to determine the protein structure of Endo D, which was reported to have a molecular mass 150 kDa (10). In our hands, purified Endo D electrophoresed as three closely related bands with molecular masses around 180 kDa (Fig. 3). The difference can be explained by the improved electrophoretical techniques employed in the present study. The central band was subjected to in-gel trypsin digestion, and the sequences of the resulting peptides were determined after separation by HPLC. All the identified peptides had sequences identical or closely related to that of the predicted protein related to Endo A (Fig. 1). Thus, it is highly likely that the sp234 gene encodes Endo D. To test this possibility, we cloned the sp234 gene. The cloned DNA showed 98.5% identity to the previously reported DNA sequence of the gene; the minor sequence difference might be due to subspecies differences or to errors in the sequence reported previously. All peptide sequences found in the purified Endo D protein completely matched portions of the predicted Endo D protein (Fig. 1). The whole Endo A protein, with a molecular mass of about 72 kDa, showed 32% identity to a large segment of the finally determined Endo D protein (Fig. 4).

Expression and Substrate Specificity of Recombinant Endo D-To express the enzyme, the DNA was placed into the expression vector pBAD/gIII. The expressed protein with a histidine tag was purified using a HiTrap chelating column, and minor impurities were removed by gel filtration. Only a single band was detected in the molecular mass region around 180 kDa (Fig. 3). The three bands in the preparation from S. pneumoniae might have been produced by proteolysis during secretion, as discussed later. Nterminal sequence analysis of the 180 kDa recombinant protein revealed that the N-terminal 37 amino acids are removed from the mature protein, probably during secretion (Fig. 1). We inferred that the N-terminal peptide serves as a signal sequence since sequences around the putative cleavage site agree with those of a signal peptide (20). The calculated molecular mass of the processed mature protein

TABLE I. Substrate specificity of recombinant Endo D.

Substrates*	Relative activities
(Man) <sub>s</sub> (GlcNAc) <sub>2</sub> Asn	100
(Man) <sub>6</sub> (GlcNAc) <sub>2</sub> Asn	< 0.01
Complex type glycopeptides from thyroglobulin	< 0.01
Complex type glycopeptides from thyroglobulin	98.8
after digestion with neuraminidase,	
$\beta$ -galactosidase, and $\beta$ -N-acetylglucosaminidase	

\*Used after <sup>14</sup>C-N-acetylation.

B	37	ETELDRISPETOPBARFRRSSIPEREREVLYTVRD TOOTDRIKLVALSALRONTSGVPSDGONEFFANTFGFROVVBLUVVRIAGSABEGI I TPPSODVTDAAFRROVPLLBAVFFPPKVYGG	156
D.	172	EELLADEFGARERDA I NROSVALASRRTGHLVNEKRSBENKVOALSNTRSKAKDHASVGGEGRAVITADERVARI ISBNY IIIFGLVFTPBV FOADBAR	284
A.	38	EELLADEFDSDPDAPFINBSHVPLEPOBVANRVNANDDRDVHLVSLSALURHTSGVPSDGAPVITYENTBSTUHVTDLINVRIAGSABEGI I TVPSSDVTDAAFRROVP	157
B	157	OEEN VDKMLVRGEDGSFPAADKLLEVAETYGFDGNFINOETESGTPET-AKDEDEFLEYLGENKPEGMFIN-VRSBIISNGDIRNGNYLTDENAMFFGAADRKVADSMFINOENNN	271
D∶	285	Doerfaealwedasgevenamer y gybertfinoen vkplgedarfinoen vkplgedarfen. Yn vetak var waant wy grymdenge yn voenn fe	404
∧∶	158	Olen Leon Leon Edgeffaadkli evad y grognfinoen foa en de graaffedaal vylgedarfegan fin - wydsniftaan wan hutdryk wylg-ngstrvadsmfia	272
B:	272	SOERSKOKAASLGRSPYDLYAGIDVEANGT-NTYMMOOVFPEGEPHTSLGIYRPDTAFKSTDTINNEFYDREOLFIN MPSGMAKVD-EHAENGGMATYFPEKTYIN ELPFMTHFNTGS	389
D:	405	KNDYTIATANNI IBGMPYGYFMGLELOOGESYKDGGGMNDILDENGNLRLSLGEFAPHTITSLGRMGEBYHKNDDIFFTEM GEDRTGOKPEDKOM MINHLADRAFAWON	524
A.	273	ROSNELADALGRSPYDLYAGYDVEARHT-STPMOHEGLFPEGEKAHTSLGLYRPDTAFOSSETMEAFYEKELGEN GSTGMAETD-GOSNBYGMANNFPAKSTATSVPM	388
B.	390	KOFFAINGKTKSDVSWNNKSLODLEPTWRWNAESEETPETVOFDWEESFYBGSSEKLSBOLSKENATNVKEYKTNELIEKOTEIOLTYRTPVKKHGLKVGISFLODPDEFIFFDLKKO	507
D:	525	GKKWFVDMVVSK <mark>DSENHYRS</mark> VSG <mark>VEPTWRWNGTBIDEKLGOEYDFTDWYNGGNSLKFSODVN</mark> GKTDOD <b>HALYSWCLEVTBKWCD</b> RVANNG <mark>G-KI</mark> SKVYWAFS-TTPDYKMDDADAWKELT	642
A.	389	GAOFSAE <u>GK</u> TVSEG <u>EWNNRSLODVLPTWRW</u> I-OHGOD-LEATFSWEEAFE <u>DGSEL</u> DWHDSLAFGEHAOIELYDIELPISETISLTWTFKSE-HGNDLNVGFR-LDGEEOFRYVEG-EORE	503
B:	508	GDHEMYTEXVKÜKKHKÖNDÜMNISLYVESDORITÖYÖLYVGELKÜSNKHONOOMPPEYTEVKYKNHSFEDGIYADLSLEWTÄGENVRHYEVYRHLANGEREFLÖATPNNVYYLSRLKRVG	627
D:	643	LSDHATMEEFDLSSLAMKÖTNYVKLEFDHEGAVKOVOTNLMOLTILLOULTILLOULTILLOUPEST-TSFSN-N-KMAGLEAVVOFKONKDADFVEVYEKDGDSMOLLTB	758
A.	504	SINOMTDWTLPLDAFAGOTTTGLAFAAEGNETGLA-EEYIL <u>BOL</u> AVGADSEK-PA-APNYNY-HG-YDPDP-SGIOLVWEKOSNVHHYRVYKETKHG-KEETGLGESG	614
B	628	KEEATTLEVVPWS-ETFASTOPSGTOVTFERPPYPKPRA-DFAVSOTVAAPGESTO-FFWRSSEVTES-VENTFEG-GOPNISDELN-PMIVYEEEGTYSWK-LVAK-NSEGEDMD	734
D∙	759	Sagge 1999. Kvmavgkngvege matterprend tyddtslpkplaen i vpgatvedstypkteggeggiggen i slsdkwsbaagessed i rltkprtvmrwvmdhagagessendde	878
A'	615	Kondvrlhieals-Etfvp-Sdarn i diksgsf	645
B:	735	ĬR-ENF-VTITE-EAN-KUTNLALOKKATASGOC-ASS-EGPDKAVDGNVSDNSK-PCATBANOM-L-O-VDLGEV-YTIAKFVLKHAEAGGEPAAF-NTKA-YTTETSVNGOSNKPVVE	841
D:	879	NNTKOFDLYYKDADGENGLAKEVRGNKAHVTDITLDKPITAODURLNVVTBONGTPUKAIRIYNDKNYEKLDTESDNIPMAAAARSLGNNKVOVOFADVONGATUTYVONPNBOTPLAT	998
B:	842	VVNNTKAT-SEHSIGTRTVCKIVDRS <mark>T</mark> PTRWRSSGTND	878
D.	999	LKSEVGGDLASAPLDLTNOSGLLYV <b>RHD</b> LPGKEISNVLAVSVPKDDRRIKSVSLETGPKKTSYAEGEDLDLRGGVLRVQYEGGTEDELIRLTHAGVSVSGFDTHHKGEONLTLQYLGOPV	1118

Fig. 4. Comparison of the protein sequences of endo-β-N-acetylglucosaminidases. B, enzyme from *Bacillus halodurans*; D, Endo D; A, Endo A. Identical amino acids are boxed.

Endo D	YOFASDOORAPYCHEVIDELKVADHROEHEINAEMUVGEORVIIIORROBLIR-HVFFVDENGO	1559
β−Gal	-Kpvse-ktvaavoedlpkdeuvekdl-ayktvekrostlylge-tr-veol-gkvigk	2132
β−GlcNAc	-e-vKi-prodioxitelphttvirgvpllukgrovutkgevig-hrsspstsad	1202
Endo D	– RRLRS– TE–VII–VEAIPETVEIGI– KVKTV––––– PAVVATOE–2–––– DAO––– DAO––– NTA–V	1601
β−Gai	Eniftainposkerklrivvevpidrivlyg– Tkpvaoea– KRD–– Ovs– Eka–Dikpi	2186
β−GicNAc	Gkevktlvnsvadiavtgivevgi– Kvthygdenggaa iaeerskleidsopapstapa	1261
Endo D	KSEE-(AS-KOLPATGTADANEALIAGTASLGLASLALILRKKREDKD)	1646
β−Gai	DSKEASOTNKAO-IPSICSAASDAAVAAGLILLGLS-AGIVVTKGKKE-D	2233
β−GicNAc	EESKALPODPAPVVTERK <u>IPEIGTRØS</u> AGLVVAGLAS <mark>-TIAAVGLI-KKR-D</mark> -D	1311

Fig. 5. Comparison of the C-terminal sequences of Endo D,  $\beta$ -galactosidase ( $\beta$ -Gal), and  $\beta$ -N-acetylglucosaminidase ( $\beta$ -GlcNAc). All three enzymes were produced in *S. pneumoniae.* Identical amino acids are indicated by closed boxes. The LPXTG motif is underlined. was 178 kDa, in good agreement with the observed molecular mass of the native enzyme, which is 180 kDa.

The purified recombinant enzyme protein hydrolyzed  $(Man)_{\delta}(GlcNAc)_{2}Asn-[^{14}C]acetyl with specific activity of 12 units/mg protein min, a rate similar to that of the native enzyme (10). The enzyme did not hydrolyze <math>(Man)_{\theta}(GlcNAc)_{2}Asn-[^{14}C]acetyl (Table I).$  Furthermore, complex type glycopeptides from thyroglobulin were hydrolyzed only after treatment with neuraminidase,  $\beta$ -galactosidase and  $\beta$ -*N*-acetylglucosaminidase (Table I). This substrate specificity is identical to that of Endo D (4). No enzymatic activity was found in the culture medium or in untransfected *E. coli.* From these results, the cloned protein was concluded to be Endo D.

### DISCUSSION

We cloned Endo D, and the identity of the cloned enzyme was confirmed by both its enzyme specificity and protein structure. The GenBank/EMBL/DDBJ database contains two *S. pneumoniae* genes under the name of endo- $\beta$ -*N*-acetylglucosaminidase. One gene encodes a protein comprising 658 amino acids and hydrolyzes the bacterial cell wall (21), while the other encodes a  $\beta$ -*N*-acetylglucosaminidase of 1,311 amino acids (22). The latter enzyme is an exoglycosidase and is eluted in a position different from that of Endo D on DEAE-Sephadex chromatography (4). We concluded that a mistake was made during the assembly of the database.

A search of the database with the Endo D protein sequence revealed the strongest homology to endo-\u00c3-N-acetylglucosaminidase from Bacillus halodurans (GenBank accession number, AP001509). The molecular mass of the enzyme is about 99 kDa, and the whole protein shows 32% sequence identity to the corresponding portion of Endo D (Fig. 4). As mentioned above, Endo D and Endo A also share extensive homology (Fig. 4). Homology starts from the N-terminal sides of the three enzymes and they share many identical amino acids (Fig. 4). It is noteworthy that the molecular mass of Endo D is much larger than those of Endo A or the Bacillus enzyme. Endo D recognizes subtle differences in  $\alpha$ -mannosyl residues in the core structure of asparagine-linked oligosaccharides; Endo D requires an exposed C-2 hydroxyl group in the  $\alpha$ -mannosyl residue linked by 1-3 linkage to the inner  $\beta$ -mannosyl residue (5, 6, 23). The large molecular size of the enzyme might play a role in the precise recognition of the key  $\alpha$ -mannosyl residue. Since Endo D is the first cloned endo-B-N-acetylglucosaminidase with the above-mentioned specificity, we are currently analyzing the structural basis of its enzyme action.

Endo D shows no significant homology to Endo H, Endo F<sub>1</sub>, Endo F<sub>2</sub>, or Endo F<sub>3</sub>, although conserved amino acids have been found among the latter four endoglycosidases and chitinase (19). Endo H and Endo F<sub>1</sub> act on high-mannose type oligosaccharides and Endo F<sub>2</sub> and F<sub>3</sub> on certain complex type oligosaccharides (18, 19, 24). Therefore, there are at least two protein groups among endo- $\beta$ -N-acetylglu-cosaminidases acting on asparagine-linked oligosaccharides, one group including Endo H and Endo F, and the other consisting of Endo D, Endo A, and the Bacillus enzyme. Both groups include enzymes with different substrate specificities.

We also noticed that the 37 amino acid N-terminal sequence in the precursor protein of Endo D shows 59.4% sequence identity to the N-terminal region of neuraminidase from S. pneumoniae (25). Both enzymes are extracellular, and the high degree of sequence similarity is consistent with the view that the N-terminal region in the Endo D precursor serves as a signal sequence. About 150 amino acid residues at the C-terminal end of Endo D also share about 30% sequence identity with  $\beta$ -galactosidase (26) and  $\beta$ -N-acetylglucosaminidase (22) from S. pneumoniae (Fig. 5). This region may serve as an anchor to associate Endo D with the surface of the bacterial cell. An LPXTG motif, which is typical of cell surface-associated proteins in Grampositive bacteria, is present in the homologous region of the three enzymes (Fig. 5). This motif is also present in the corresponding position of neuraminidase from S. pneumoniae (25).  $\beta$ -Galactosidase from S. pneumoniae is proven to be a cell-surface associated enzyme and is considered to be a virulence factor (26). Neuraminidase,  $\beta$ -galactosidase,  $\beta$ -Nacetylglucosaminidase and Endo D are all secreted into the culture medium of S. pneumoniae (4, 27, 28). However, prolonged culture is required for secretion. These enzymes may be cell surface-associated enzymes and contribute to pathogenesis of the bacteria by removing complex type oligosaccharides through their concerted actions. As an example, IgG becomes inefficient in complement-mediated cell lysis after removal of the oligosaccharides by exoglycosidases and Endo D (9).

#### REFERENCES

- 1. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J., eds. (1999) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Muramatsu, T. (1992) Endo-β-N-acetylglucosaminidases in CRC Handbook of Endoglycosidases and Glycoamidases (Takahashi, N. and Muramatsu, T., eds.) pp. 13–40, CRC Press, Boca Raton, Ann Arbor, London
- Yamamoto, K. (1994) Microbial endoglycosidases for analysis of oligosaccharide chains in glycoproteins. J. Biochem. 116, 229– 235
- Koide, N. and Muramatsu, T. (1974) Endo-β-N-acetylglucosaminidase acting on carbohydrate moieties of glycoproteins: purification and properties of the enzyme from *Diplococcus* pneumoniae. J. Biol. Chem. 249, 4897-4904
- Muramatsu, T. (1978) Endo-β-N-acetylglucosaminidase D from Diplococcus pneumoniae. Methods Enzymol. 50, 555–559
- Tai, T., Yamashita, K., Ogata-Arakawa, M., Koide, N., Muramatsu, T., Iwashita, S., Inoue, Y., and Kobata, A. (1975) Structure of two ovalbumin glycopeptides in relation to endo-β-Nacetylglucosaminidase specificity. J. Biol. Chem. 250, 8569-8575
- Muramatsu, T. (1971) Demonstration of an endo-glycosidase acting on a glycoprotein. J. Biol. Chem. 246, 5535–5537
- Hunt, L.A., Etchison, J.R., and Summers, D.F. (1978) Oligosaccharide chains are trimmed during synthesis of the envelope glycoprotein of vesicular stomatitis virus. *Proc. Natl. Acid. Sci.* USA 75, 754-758
- Koide, N., Nose, M., and Muramatsu, T. (1977) Recognition of IgG by Fc receptor and complement: effects of glycosidase digestion. Biochem. Biophys. Res. Commun. 75, 838-843
- Muramatsu, T., Koide, N., and Maeyama, K. (1978) Further studies on endo-β-N-acetylglucosaminidase D. J. Biochem. 83, 363-370
- Muramatsu, T. (1992) Protein deglycosylation in CRC Handbook of Endoglycosidases and Glycoamidases (Takahashi, N. and Muramatsu, T., eds.) pp. 333–350, CRC Press, Boca Raton, Ann Arbor, London

- Tarentino, A.L., Gomez, C.M., and Plummer, T.H., Jr. (1985) Deglycosylation of asparagine-linked glycans by peptide: N-glycosidase F. Biochemistry 24, 4665–4671
- Spiro, R.G. (1966) Characterization of carbohydrate units of glycoproteins. Methods Enzymol. 8, 26-52
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680– 685
- Rosenfeld, J., Capderville, J., Guillemot, J.C., and Ferrara, P. (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis Anal. Biochem. 203, 173-179
- Takegawa, K., Yamabe, K., Fujita, K., Tabuchi, M., Mita, M., Izu, H., Watanabe, A., Asada, Y., Sano, M., Kondo, A., Kato, I., and Iwahara, S. (1997) Cloning, sequencing and expression of *Arthrobacter protophormiae* endo-β-N-acetylglucosaminidase in *Escherichia coli. Arch. Biochem. Biophys.* 338, 22-28
- Robbins, P.W., Trimble, R.B., Wirth, D.F., Hering, C., Maley, F., Maley, G.F., Das, R., Gibson, B.W., Royal, N., and Biemann, K. (1984) Primary structure of the Streptomyces enzyme endo-β-N-acetylglucosaminidase H. J. Biol. Chem. 259, 7577-7583
- Tarentino, A.L., Quinones, G., Schrader, W.P., Changchien, L.M., and Plummer, T.H. (1992) Multiple endoglycosidase (Endo)F activities expressed by *Flavobacterium meningosepti*cum. Endo F<sub>1</sub>: molecular cloning, primary sequence and structural relationship to Endo H. J. Biol. Chem. 267, 3868-3872
- Tarentino, A.L., Quinones, G., Changchien, L.M., and Plummer, T.H. (1993) Multiple endoglycosidase F activities expressed by *Flavobacterium meningosepticum* endoglycosidases F<sub>2</sub> and F<sub>3</sub>. Molecular cloning, primary sequence and enzyme expression. J. Biol. Chem. 268, 9702–9708
- 20. Von Heijne, G. (1984) Analysis of the distribution of charged

residues in the N-terminal region of signal sequence: implications for protein export in prokaryotic and eukaryotic cells. *EMBO J.* **3**, 2315–2318

- Garcia, P., Gonzalez, M.P., Garcia, E., Lopez, R., and Garcia, J.L. (1999) LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol. Microbiol.* 31, 1275-1277
- Clarke, V.A., Platt, N., and Butters, T.D. (1995) Cloning and expression of the β-N-acetylglucosaminidase gene from Streptococcus pneumoniae. Generation of truncated enzymes with modified aglycon specificity. J. Biol. Chem. 270, 8805–8814
- Mizuochi, T., Amano, J., and Kobata, A. (1984) New evidence of the substrate specificity of endo-β-N-acetylglucosaminidase D. J. Biochem. 95, 1209–1213
- Tarentino, A.L. and Maley, F. (1974) Purification and properties of an endo-β-N-acetylglucosaminidase from Streptomyces griseus. J. Biol. Chem. 249, 811-817
- Camara, M., Boulnois, G.J., Andrew, P.W., and Mitchell, T.J. (1994) A neuraminidase from Streptococcus pneumoniae has the features of a surface protein. Infect. Immun. 62, 3688-3695
- Zahner, D. and Hakenbeck, R. (2000) The Streptococcus pneumoniae β-galactosidase is a surface protein. J. Bacteriol. 182, 5919-5921
- 27. Hughes, R.C. and Jeanloz, R.W. (1963) The extracellular glycosidases of *Diplococcus pneumoniae*. Purification and properties of a neuraminidase and a  $\beta$ -galactosidase. Action on the  $\alpha$ 1-acid glycoprotein from human plasma. *Biochemistry* 3, 1535-1542
- 28. Hughes, R.C. and Jeanloz, R.W. (1963) The extracellular glycosidases of *Diplococcus pneumoniae*. Purification and properties of a  $\beta$ -N-acetylglucosaminidase. Action on a derivative of the  $\alpha$ 1-acid glycoprotein. *Biochemistry* **3**, 1543–1548