Endoglycoceramidase (EGCase) is a hydrolase that hydrolyzes the linkage between the oligosaccharide and ceramide of various glycosphingolipids. This paper describes the molecular cloning and expression of EGCase II, one of the isoforms of EGCases. The gene encoding EGCase II was obtained by screening of a genomic DNA library from Rhodococcus sp. strain M-777 constructed in pUC19 with oligonucleotide probes deduced from a partial amino acid sequence of the enzyme protein. Recombinant Escherichia coli cells in which the EGCase II gene was expressed produced 14 units of the enzyme per liter of culture medium but did not produce sphingomyelinase. Recombinant EGCase II was a functioning enzyme with substrate specificity identical to that of the wild-type enzyme. Sequence analysis showed the presence of an open reading frame of 1470 base pairs encoding 490 amino acids. The N-terminal region of the deduced amino acid sequence had the general pattern of signal peptides of secreted prokaryotic proteins. Interestingly, the consensus sequence in the active site region of the endo-1,4-β-glucanase family A was found in the amino acid sequence of EGCase II.

Glycosphingolipids are amphipathic compounds consisting of oligosaccharides and ceramide moieties and are cell surface components of all vertebrates. Some glycosphingolipids are tumor-associated antigens and receptors of bacterial toxins and hormones and modulate cell growth and differentiation (1). Some of these biological functions of glycosphingolipids have been elucidated in experiments in which glycosphingolipids are linked to the GenBank EBI Data Bank with accession number(s) U39554.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank/EBI Data Bank with accession number(s) U39554.

† To whom correspondence should be addressed. Tel.: 81-775-43-7217; Fax: 81-775-43-2494.

‡ Laboratory of Marine Biochemistry, Faculty of Agriculture, Kyushu University, Hakozaki 8-10-1, Higashi-ku, Fukuoka 812-81, Japan

§ From the Biotechnology Research Laboratories, Takara Shuzo Co., Ltd., Seta 3-4-1, Otsu, Shiga 520-21 and the Laboratory of Marine Biochemistry, Faculty of Agriculture, Kyushu University, Hakozaki 8-10-1, Higashi-ku, Fukuoka 812-81, Japan

# The abbreviations used are: EGCase, endoglycoceramidase; kbp, kilobase pairs. Please see Table I for definitions of glycosphingolipids.
the order of elution. Amino acid sequences of the isolated peptides were identified by automated Edman degradation (sequence model 477A, Applied Biosystems).

**Molecular Cloning**—Genomic DNA (5 μg) was digested with MluI, and the digest was fractionated by 0.7% agarose gel electrophoresis by the standard method (22). DNA was transferred from agarose gels to nylon membranes (Hybond N+, Amersham Corp.) as described by the manufacturer. On the basis of the amino acid sequence from residues 1 to 15 of peptide L1 (see Table I) obtained by lysylendopeptidase digestion, and in the expectation that the residue at position -1 of peptide L1 must be lysine, a mixed oligonucleotide probe, AATGTCGGCCTGAGCTCCTGCTCTGCTGCGCCCGCACTTCCAGGGATAGCGCAAAGGGCGCCCATGCGCCTGC- GC, was designed and designated probe 1. T4 polynucleotide kinase from the MEGALABEL kit was used to label the probe with [γ-32P]ATP. Hybridization was done in 5 X SSC (1 X SSC = 0.15 M NaCl and 0.05 M sodium citrate) containing 0.5% SDS, 5 X Denhardt’s solution (1 X Denhardt’s solution = 0.02% (w/v) each of Ficoll 400, bovine serum albumin, and polyvinylpyrrolidone-40), and 0.1 mg/ml denatured salmon sperm DNA at 65 °C for 16 h. After hybridization, the membranes were washed for 30 min with 0.2 X SSC containing 0.1% SDS at 65 °C and used to expose an imaging plate, which was examined later on an imaging analyzer (LAS2000, Fuji Photo Film Co., Tokyo). Hybridization with probe 1 to Southern blots of the MluI digest showed that only the 4.4-kbp fragment contained the EGCase II gene. For cloning of this gene, a digest was prepared with 20 μg of genomic DNA. Restriction fragments of genomic DNA of Rhodococcus sp. strain M-777 were fractionated by preparative 0.7% agarose gel electrophoresis. Fragments (4.4 kbp long) were extracted from the gel by adsorption to glass beads from the EASYTRAP Version 2 kit. A phosphorylated MluI linker, pGACGCGTC, was inserted into the HindII site of pUC19, and the resulting plasmid was designated pUC19M. The MluI fragments from the fractionated DNA digest were ligated to the MluI site of pUC19M. The recombinant plasmids obtained were used to transform E. coli JM109, used in the preparation of a gene library enriched with the EGCase II gene hybridized with probe 1. Colony hybridization was done by the standard procedure (22), and hybridized clones were detected with probe 1 under the same conditions as in Southern blotting. One clone was selected and the plasmid in the clone was designated pEGCM36.

A DNA probe was prepared by digestion of pEGCM36 with HindII. The probe, which was the 3'-end of the EGCase II gene in pEGCM36, was 210 base pairs long. This probe 2 was random prime labeled with [α-32P]dCTP using the BEST labeling kit. In the same way as when pEGCM36 was prepared, a gene library enriched with the 3'-end of the EGCase II gene was constructed from genomic DNA digested with BamHI with probe 2 used for hybridization and detection. BamHI fragments (2.7 kbp long) were ligated to the BamHI site of pUC19 and used to transform E. coli JM109, used in the preparation of an enriched gene library. The library was screened by colony hybridization with probe 2.

**DNA Sequencing and Sequence Analysis**—The nucleotides were sequenced by the dideoxy chain termination method with SeqCTest DNA polymerase and a DNA sequencer (Applied Biosystems, model 373A). Computer analysis including comparison of DNA sequences was done with DNAsIS and GENEBRIGHT software (Hitachi Software Engineering, Tokyo). Frame analysis was done as described elsewhere (24).

**Construction of Expression Plasmid with EGCase II Gene**—The vector pTV118N was treated with SnaI, Klenow fragment, and SpI. An insert that included the 5'-end of the EGCase II gene was prepared by digestion of clone pEGCM36 with AccIII, blunting digestion with MluI, and gel purification. An insert that included the 3'-end of the EGCase II gene was prepared by digestion of clone pEGC20 with MluI and SpI and then gel purification. The inserts and vector were ligated and used to transform E. coli JM109. The recombinant plasmid was purified and designated pTGE3.

**Expression and Purification of Recombinant EGCase II**—E. coli JM109 cells transformed with pTGE3 were grown at 37 °C in Luria-Bertani medium containing 100 μg/ml ampicillin upon reaching the optical density (absorbance at 600 nm) of about 0.5. Then isopropylthio-β-D-galactopyranoside was added to the final concentration of 1 mM to cause transcription, and culture was continued at 37 °C for 4 h more. Cells were harvested by centrifugation, suspended in extraction buffer (10 mM Tris-HCl, pH 8.0, and 0.5 mM 4-(2-aminoethoxy)benzenesulfonyl fluoride hydrochloride), and sonicated.Crude extracts were further purified by ion-exchange chromatography on Q-Sepharose.

**Enzyme Assay**—EGCase activity was assayed with purified asialo-lysosome from the substrate in the presence of Triton X-100 as described before (8). One unit of enzyme was defined as the amount needed to catalyze the hydrolysis of 1 μmol of substrate/min. The substrate specificity was examined as described elsewhere (8).

**RESULTS AND DISCUSSION**

**Molecular Cloning of EGCase II Gene**—The N-terminal amino acid of the EGCase II purified from Rhodococcus sp. strain M-777 was blocked, so we could not identify the N-terminal sequence. Seven peptides, L1 to L7, were purified from a lysylendopeptidase digest and sequenced (Table I). Southern blotting with probe 1 showed a 4.4-kbp band in the hybridization pattern of the genomic DNA digested with MluI. A clone, pEGCM36, containing a 4.4-kbp insert was isolated from the gene library enriched with the EGCase II gene hybridized with probe 1. Its nucleotides were sequenced, and it was found to contain a putative open reading frame. The amino acid sequences L1, L2, L3, L4, L6, and L7 were found in the deduced amino acid sequence of the open reading frame, but sequence L5 and a stop codon were not found so we screened the genomic library for the missing sequences. A probe 2 was prepared from pEGCM36 for use in this screening. Southern blotting with probe 2 of the BamHI digest gave a 2.7-kbp band. A clone, pEGCB20, containing a 2.7-kbp insert was isolated from the gene library enriched with the EGCase II gene and sequenced. The deduced amino acid sequence contained sequence L5 and a stop codon in the same frame. A partial sequence of the 3'-end of the EGCase II gene in pEGCM36 was found in pEGCB20. pEGCM36 and pEGCB20 overlapped (Fig. 1).

**DNA and Amino Acid Sequence Analysis**—We cloned and sequenced 2012 nucleotides of the two contiguous clones, pEGCM36 and pEGCB20, including the coding regions of the EGCase II gene and found an open reading frame at positions 1–1860. The initiation codon was not identified because the N-terminal amino acid of native EGCase II was blocked. In an attempt to tentatively identify the initiation codon and open reading frame, we analyzed the nucleotide sequence by frame

---

**TABLE I**

Sequences of the peptides derived from lysylendopeptidase digestion of EGCase II

| Peptide | Sequence |
|---------|----------|
| L1 | SAPDGPMQTFEDALAREYADMTNF |
| L2 | IDOPRAQQGQRIAYFPYILPDDIG |
| L3 | AWRVADRFDADVXLYMNFEPFGGGQLQ |
| L4 | VNLSDMHDQVSGAIFPENSGN2GAG1GNG |
| L5 | PYPRAVGTPEITSSTXKDRQLQ |
| L6 | HPELVEYAK |
| L7 | DDDGRSL1IRGFTASSAK |

---

**FIG. 1. Restriction map of two overlapping clones including the EGCase II gene.** pEGCM36, obtained by screening of the genomic library with probe 1, and pEGCB20, obtained by screening of the library with probe 2, overlapped as shown here. The closed box corresponds to the putative signal sequence. Coding regions of EGCase II are indicated by shaded boxes. The open boxes contained several AccIII sites (not shown here) and SpI sites in addition to the one shown here. Coordinates are given as the nucleotide numbers (base pairs (bp)) for the sequenced region.
analysis plotting, which shows codon position-specific differences in the GC content. In organisms with a high GC content, bases of coding regions have a high GC content at the third codon position, a low content at the second position, and an intermediate content at the first position (24). Frame analysis plotting of the 2012 nucleotides that included the EGCase II gene suggested that there was a coding region between the nucleotide positions of approximately 350 and 1850 (Fig. 2). A hydrophobic motif was found in the deduced amino acid sequence (Fig. 3). This sequence motif, a putative secretion signal peptide (25), had a positively charged N terminus followed by a hydrophobic core and a string of polar residues. The finding of a signal peptide sequence was in agreement with EGCase II being secreted into the culture medium. The signal sequence was coded at nucleotides 391–480, starting with GTG. A possible Shine-Dalgarno ribosome binding sequence started 4 bases upstream from the GTG. These results were in complete agreement with nucleotides 391–393 (GTG) being initiation codons. The DNA sequence and deduced amino acid sequence of the open reading frame of EGCase II are shown in Fig. 4. The open reading frame was 1470 base pairs long with 490 codons. All of the peptide sequences shown in Table I were in the deduced amino acid sequence.

Expression of EGCase II—The expression plasmid pTEG3 was constructed by insertion of a fragment of the coding sequence at nucleotide positions 482–1997, without a putative secretion signal sequence, into plasmid pHV118N between the SalI and SphI sites and in frame with the initiation codon of the plasmid. In pTEG3, transcription of recombinant genes is controlled by the promoter plac and can be induced by isopropyl-β-D-galactopyranoside. E. coli JM109 cells transformed with pTEG3 were cultured in a medium containing 1 mM isopropyl-β-D-galactopyranoside and separated from the medium by centrifugation. The EGCase II activity of the cell lysate was assayed with asialo-GM1 as the substrate. Recombinant E. coli cells produced 14 units of EGCase II activity per liter of culture medium, but Rhodococcus sp. strain M-777 produced 3 units of a mixed enzyme activity of EGCase II and EGCase I (8). Extracts from the negative control strain con-
Molecular Cloning of Endoglycoceramidase Gene

TABLE II
Specificity of endoglycoceramidase

| Substrate                      | Structure                     | Hydrolysis by Recombinant enzyme | Hydrolysis by Native enzyme |
|-------------------------------|-------------------------------|----------------------------------|-----------------------------|
|                               |                               | Condition I                      | Condition II                 | Condition I | Condition II |
|                               |                               | %                                | %                           | %           | %           |
| Ganglio series                |                               |                                  |                             |             |             |
| O
| 3Gb1                         | NeuAcα2-3Galβ1-3GalNAcβ1-4    | 100                              | 100                         | 100         | 100         |
| O
| 3Gd1                         | NeuAcα2-3Galβ1-3GalNAcβ1-4    | 100                              | 100                         | 100         | 100         |
| O
| 3Gm1                         | Galβ1-3GalNAcβ1-4             | 80.5                             | 100                         | 85.1        | 98.1        |
| O
| 3Mg2                         | GalNAcβ1-4                   | 72.6                             | 100                         | 74.5        | 100         |
| O
| 3M3                          | NeuAcα2-3Galβ1-4              | 100                              | 100                         | 100         | 100         |
| Asialo-Galβ3                  | Galβ1-3GalNAcβ1-4-Galβ1-Cer   | 98.3                             | 100                         | 100         | 100         |
| Globo series                  |                               |                                  |                             |             |             |
| Forssman antigen              | GaINacβ1-3Galα1-4Galβ1-Cer    | 0                                | 19.1                        | 0           | 20.3        |
| Gb4-ceramide                  | GaINacβ1-3Galα1-4Galβ1-Cer    | 0                                | 10.3                        | 0           | 12.4        |
| Lacto series                  |                               |                                  |                             |             |             |
| Sialosylparagloboside         | NeuAcα2-3Galβ1-4Galβ1-Cer     | 93.4                             | 100                         | 100         | 100         |
| Lactosylceramide              | GaINacβ1-4Galβ1-Cer           | 50.2                             | 100                         | 53.2        | 100         |

* In condition I, 0.3 milliunit of the enzyme was incubated with the appropriate substrate at 37 °C for 2 h.
* In condition II, 1.5 milliunits of the enzyme was incubated with the appropriate substrate at 37 °C for 16 h.

With various glycosphingolipids, we found that activator II increased the activities of both recombinant and native EGCase II. However, only native EGCase II hydrolyzed cell surface glycosphingolipids in the presence of activator II (15). That is, recombinant EGCase II did not hydrolyze the cell surface glycosphingolipids under the conditions we used (data not shown); the reason is not known. The N-terminal amino acid of native EGCase II was blocked (15). We speculate that the N-terminal structure of native EGCase II could be needed for the hydrolysis of cell surface glycosphingolipids.

In the putative amino acid sequence of EGCase II, those sequences of the N terminus of EGCase I and the sequence of activator II that have been isolated from a Rhodococcus sp. (15) were not found. These results suggested that the EGCase II gene is independent of the EGCase I and activator II genes.

Comparison with other nucleotide sequences published in computer data bases did not show similarity. However, a search of protein sequence data bases showed that an amino acid sequence of EGCase II was similar to that of the active site region of bacterial endo-1,4-β-glucanases (Fig. 5). Cellulase and endo-1,4-β-glucanases can be grouped into families of related enzymes on the basis of similarities of the amino acid sequence of their putative catalytic domains (26, 27). All cellulase and endo-1,4-β-glucanase sequences listed in Fig. 5 belong to family A. The sequence Asn-Glu-Pro, conserved in almost all

With various glycosphingolipids, we found that activator II increased the activities of both recombinant and native EGCase II. However, only native EGCase II hydrolyzed cell surface glycosphingolipids in the presence of activator II (15). That is, recombinant EGCase II did not hydrolyze the cell surface glycosphingolipids under the conditions we used (data not shown); the reason is not known. The N-terminal amino acid of native EGCase II was blocked (15). We speculate that the N-terminal structure of native EGCase II could be needed for the hydrolysis of cell surface glycosphingolipids.

In the putative amino acid sequence of EGCase II, those sequences of the N terminus of EGCase I and the sequence ofactivator II that have been isolated from a Rhodococcus sp. (15) were not found. These results suggested that the EGCase II gene is independent of the EGCase I and activator II genes.

Comparison with other nucleotide sequences published in computer data bases did not show similarity. However, a search of protein sequence data bases showed that an amino acid sequence of EGCase II was similar to that of the active site region of bacterial endo-1,4-β-glucanases (Fig. 5). Cellulase and endo-1,4-β-glucanases can be grouped into families of related enzymes on the basis of similarities of the amino acid sequence of their putative catalytic domains (26, 27). All cellulase and endo-1,4-β-glucanase sequences listed in Fig. 5 belong to family A. The sequence Asn-Glu-Pro, conserved in almost all
cellulase and endo-1,4-β-glucanase sequences of family A, was found in EGCase II at amino acid positions 232–234. EGCase II hydrolyzes d-glucose-β,1,4-ceramide glycosidic linkages but does not hydrolyze d-glucanose-β,1,4-ceramide (8). The Aan-Glu-Pro sequence of EGCase II may recognize β-d-glucosyl residues and may have a catalytic role similar to that of endo-1,4-β-glucanase. Hydrolysis by glycosyl hydrolases often involves general acid catalysis, usually promoted by aspartic acid, glutamic acid, or both. Results of site-directed mutagenesis of the general acid catalysis, usually promoted by aspartic acid, glutamic acid, and Erwinia chrysanthemi in family A again suggest that the conserved glutamic acid residue in Aan-Glu-Pro is part of the active site (28, 29). Active site residues are usually highly conserved during evolution. EGCase II and endo-1,4-β-glucanases may have evolved from the same ancestral protein, although they have different substrate specificities.

Acknowledgment—We thank Dr. T. Yamagata for encouragement throughout this work.

REFERENCES
1. Hakomori, S. (1981) Annu. Rev. Biochem. 50, 733–764
2. Ito, M., and Yamagata, T. (1986) J. Biol. Chem. 261, 14278–14282
3. Ashida, H., Yamamoto, K., Kumagai, H., and Tochikura, T. (1992) Eur. J. Biochem. 205, 729–735
4. Li, Y.-T., Ishikawa, Y., and Li, S.-C. (1987) Biochem. Biophys. Res. Commun. 149, 167–172
5. Zhou, L., Xue, G.-P., Orpin, C. G., Black, G. W., Gilbert, H. J., and Hazlewood, G. P. (1994) J. Biol. Chem. 271, 15791–15799
6. Muramoto, K., Kasahara, M., Kobayashi, K., Ito, M., Yamagata, T., and Kuroda, Y. (1994) Biochem. Biophys. Res. Commun. 202, 398–402
7. Hansson, G. C., Li, Y.-T., and Karlsson, H. (1989) Biochemistry 28, 6672–6678
8. Higashi, H., Ito, M., Zhang, G., and Yamagata, T. (1990) Glycobiology 5, 343–350
9. Py, B., Bortoli-German, I., Haiech, J., Chippaux, M., and Barras, F. (1990) Glycobiology 2, 271–274
10. Ito, M., and Yamagata, T. (1989) J. Biol. Chem. 264, 9510–9519
11. Ito, M., Ikegami, Y., and Yamagata, T. (1991) J. Biol. Chem. 266, 7919–7926
12. Shimamura, M., Hayase, T., Ito, M., Rasilo, M.-L., and Yamagata, T. (1988) J. Biol. Chem. 263, 12124–12128
13. Ohara, K., Sano, M., Kondo, A., and Kato, I. (1991) J. Chromatogr. 586, 35–41
14. Cunningham, C., McPherson, C. A., Martin, J., Harris, W. J., and Flint, H. J. (1991) Mol. Gen. Genet. 223, 217–223
15. Inoue, M., Ikegami, Y., and Yamagata, T. (1991) J. Biol. Chem. 266, 7919–7926
16. Ito, M., Ikegami, Y., Tai, T., and Yamagata, T. (1995) Eur. J. Biochem. 218, 637–643
17. Ito, M., Ikegami, Y., and Yamagata, T. (1993) Eur. J. Biochem. 218, 645–649
18. Ito, M., Ito, M., Zhang, G., and Yamagata, T. (1995) Glycobiology 5, 343–350
19. Basu, S. S., Dastgheib-Hosseini, S., Hoover, G., Li, Z., and Basu, S. (1994) EMBO J. 4, 3901–3906
20. Ito, M., and Komori, H. (1996) J. Biol. Chem. 271, 12655–12660
21. Baird, S. D., Hefford, M. A., Johnson, D. A., Sung, W. L., Yaguchi, M., and Seligy, V. L. (1990) Biochem. Biophys. Res. Commun. 175, 1910–1918
22. Ito, M., Ikegami, Y., and Yamagata, T. (1991) J. Biol. Chem. 266, 9253–9258
23. Ito, M., and Komori, H. (1996) J. Biol. Chem. 271, 12655–12660
24. Bibb, M. J., Findlay, P. R., and Johnson, M. W. (1984) Gene 270–274
25. Be’guin, P. (1990) Anal. Biochem. 198, 1–5
26. Ito, M., and Komori, H. (1996) J. Biol. Chem. 271, 12655–12660
27. Py, B., Bortoli-German, I., Haiech, J., Chippaux, M., and Barras, F. (1991) Eur. J. Biochem. 202, 398–402
28. Ashida, H., Yamamoto, K., Kumagai, H., and Tochikura, T. (1992) Eur. J. Biochem. 205, 729–735
29. Ito, M., Ikegami, Y., and Yamagata, T. (1991) J. Biol. Chem. 266, 9510–9519
30. Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132
31. Zhou, L., Xue, G.-P., Orpin, C. G., Black, G. W., Gilbert, H. J., and Hazlewood, G. P. (1994) J. Biol. Chem. 269, 223–228
32. Inoue, M., Ikegami, Y., and Yamagata, T. (1991) J. Biol. Chem. 266, 9510–9519
33. Meinke, A., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr., and Warren, B. A. J. (1993) J. Bacteriol. 175, 1910–1918
34. Ito, M., and Komori, H. (1996) J. Biol. Chem. 271, 12655–12660
35. Enkein, A., Gilkes, N. R., Kibbun, D. G., Miller, R. C., Jr., and Warren, B. A. J. (1993) J. Bacteriol. 175, 1910–1918
36. Hall, J., Hazlewood, G. P., Barker, P. J., and Gilbert, H. J. (1988) Gene (Amst.) 69, 29–38
37. Higashi, H. (1990) Trends Glycobiol. 2, 399–402
38. Ito, M. (1990) Trends Glycosci. Glycotechnol. 3, 390–402