Sequence of egV and Properties of EgV, a Ruminococcus albus Endoglucanase Containing a Dockerin Domain

Hiroki Ohara,† Jyunko Noguchi,† Shuichi Karita,† Tetsuya Kimura,† Kazuo Sakka,‡ and Kunio Omiya†

Faculty of Bioresources† and Center for Molecular Biology and Genetics,‡ Mie University, Tsu 514-8507, Japan

Received July 9, 1999; Accepted October 8, 1999

The Ruminococcus albus F-40 egV gene, encoding endoglucanase V (EGV), consists of an open reading frame of 1,833 nucleotides and encodes 611 amino acids with a deduced molecular weight of 67,103. The deduced EGV is a modular enzyme composed of a catalytic domain of family 5 of glycosyl hydrolases, a domain of unknown function, and a dockerin domain responsible for cellulosome assembly, suggesting that R. albus F-40 produces a cellulosome, and EGV is a component of the cellulosome. A truncated form of EGV with an apparent molecular weight of 42,000 was purified from a recombinant Escherichia coli and characterized since EGV suffered from partial proteolysis by E. coli protease(s). The truncated EGV was active toward carboxymethyl cellulose, xylan, lichenan, and acid-swollen cellulose. The pH and temperature optima of the enzyme were 7.0 and 40°C, respectively. By Western blot analysis using the antiserum raised against the truncated enzyme, EGV was detected in the whole cells but not in the culture supernatant of R. albus F-40, suggesting that EGV was located on the cell surface.

Key words: Ruminococcus albus; endoglucanase; dockerin domain; cellulosome

Microbial cellulolytic systems are generally composed of a number of cellulases, i.e., endoglucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91), which cleave β-1,4-glycosidic linkages and act cooperatively in converting cellulose into cellobiose. Recent molecular studies have established that many cellulases are modular enzymes typically composed of a catalytic domain connected to a cellulose-binding domain and/or other functional domain(s) via linker sequences. On the basis of amino acid sequence similarity, catalytic domains of all of the glycosyl hydrolases have been grouped into well-defined families; endoglucanases and cellobiohydrolases are now divided into 12 distinct families, i.e., families 5, 6, 7, 8, 9, 12, 44, 45, 48, 51, 61, and 74. The process of cellulose degradation seems to be different for different species, especially between aerobic and anaerobic microorganisms. Aerobic microorganisms such as Trichoderma reesei secrete a combination of endoglucanases and cellobiohydrolases, which attack their substrates individually but synergistically. By contrast, anaerobic cellulolytic microorganisms such as Clostridium cellulolyticum, Clostridium cellulovorans, Clostridium josui, and Clostridium thermocellum produce extracellular multienzyme complexes having high activity against crystalline cellulose; the complex is known as a cellulosome. A common feature of the clostridial cellulosomes is that they consist of a large number of catalytic components arranged around noncatalytic scaffolding proteins. The assembly of the cellulosomes are caused by the specific interaction between one of the reiterated cohesin domains of a scaffolding protein and a dockerin domain of each catalytic component. A dockerin domain is a docking domain that interacts with a cohesin domain of the scaffolding protein, and it consists of a pair of well-conserved 22-residue repeats and are highly conserved in catalytic components of the clostridial cellulosomes.

Ruminococcus albus is an anaerobic cellulolytic rumen bacterium. Previous studies suggested that the cellulase activities of R. albus are largely associated with enzyme aggregates of high molecular mass (>1.5 × 10⁷) and that protuberant structures are present on the cell surface. These observations were similar to those for the cellulosome of C. thermocellum. Thereafter, however, no genetic and biochemical evidence has accumulated supporting the occurrence of cellulosomes in R. albus. Although five endoglucanase genes and a xylanase gene were cloned and sequenced from some strains of R. albus, i.e., eg1 and eg2 from strain F-40, celA and

† To whom correspondence should be addressed; Phone: +81-59-231-9075; Fax: +81-59-231-9048; E-mail: karita@gene.recs.mie-u.ac.jp
Abbreviations: ASC, acid-swollen cellulose; CMC, carboxymethyl cellulose; PAGE, polyacrylamide gel electrophoresis
celB\textsuperscript{19} from strain SY3, and xynA (accession number: U43089) from strain 7, none of the encoded proteins contained a sequence segment similar to the clostridial dockerin domains. Recently, we have found that the egV gene from \textit{R. albus} F-40 encodes an endoglucanase containing a dockerin domain.\textsuperscript{20} This has been the first genetic evidence suggesting that \textit{R. albus} F-40 produce cellulases, and the translated product has not been characterized yet.

In this paper, we describe the nucleotide sequence of the egV gene encoding a modular endoglucanase containing a dockerin domain. We also deal with characterization of the enzyme purified from a recombinant \textit{Escherichia coli} and identification of EGV in the \textit{R. albus} cells, probably on the cell surface.

\section*{Materials and Methods}

\textit{Bacterial strains and plasmids.} \textit{R. albus} F-40, used for isolation of the chromosomal DNA, was described previously.\textsuperscript{21} The \textit{E. coli} strains used were JM109 \texttt{(recA1 endA1 gyrA96 thi hsrl7 supE44 relA1 (lac-proAB) M15}, lacI\textsuperscript{Q} [traD36 proAB lacI\textsuperscript{Q} ZDM15]), XL1-Blue \texttt{(recA1 endA1 gyrA96 thi-I hsrl7 supE44 relA1 lac F\textsuperscript{−} proAB lacI\textsuperscript{Q} ZDM15 Tn10 (Tet\textsuperscript{Q})]}, and DH5\textsubscript{a} \texttt{(deoR endA1 gyrA96 hsdrl7 recA1 relA1 supE44 thi-1 Δ(lacZYA-argF) U169 φ80lacZM15 F\textsuperscript{−} λ\textsuperscript{−})}. Phage \texttt{λZapII} (Stratagene), which provides a pBluescript derivative containing the inserted fragment by \textit{in vivo} excision, was used for cloning of the egV gene from \textit{R. albus} F-40. Plasmids pUC19 (Toyobo) and pBluescript II SK(−) (Stratagene) were used for cloning and sequencing of egV.

\textit{Construction of the gene library and screening of endoglucanase-producing recombinants.} The \textit{R. albus} chromosomal DNA was isolated as described previously.\textsuperscript{22} The genomic DNA was digested with EcoRI and fragments of 5 to 7 kb were isolated by agarose gel electrophoresis followed by recovery with the GeneClean kit (Bio101, La Jolla). These fragments were ligated into the EcoRI site of phage \texttt{λZapII} with T4 DNA ligase. \textit{E. coli} XL1-Blue was infected with the recombinant phages after \textit{in vitro} packaging and endoglucanase-productivity was examined by the carboxymethylcellulose (CMC)-Congo red technique described previously.\textsuperscript{23} \texttt{λ.R.acelI} (Fig. 1) is one of the endoglucanase-productive clones. Plasmid pNK-1 (Fig. 1) was directly derived from \texttt{λ.R.acelI} according to the \textit{in vivo} excision protocol (Stratagene).

\textit{Cloning of a DNA fragment encoding the N-terminal end of EGV.} The \textit{R. albus} chromosomal DNA was completely digested with EcoRV and DNA fragments of about 2.3 kb were isolated by agarose gel electrophoresis followed by the GeneClean protocol. Libraries of these fragments were constructed by ligation into the dephosphorylated \texttt{PvuII} site of pUC19. \textit{E. coli} XL1-Blue was transformed with the ligation mixture and screened for the existence of the DNA fragment encoding EGV by endoglucanase productivity. The resulting plasmid was referred to as pEV-1 (Fig. 1).

\textit{Cloning of a DNA fragment encoding the C-terminal end of EGV by PCR.} Targeted gene walking PCR\textsuperscript{30} involved digestion of \textit{R. albus} genomic DNA with EcoRI, ligation of the genomic fragments to pUC19, and PCR from the chimeric plasmid library with an internal primer specific for egV (5'\textsuperscript{-}TAT-GACGGCAATGAAACCAAAGCGTG-3') and M13 reverse primer (5’-GAACGGATAACCAT- TTCACACAGG-3’). The PCR product was digested with \texttt{PstI} and EcoRI, recovered from an agarose gel as described above, and then ligated to pBluescript II SK(−) which had been linearized with the same enzymes to yield pEI1.2 (Fig. 1). The ligation mixture was used to transform \textit{E. coli} JM109.

\textit{DNA sequencing.} Nucleotide sequencing was done on an Applied Biosystems model 373A or Licor model 400L automated DNA sequencer, with appropriate dye primers and a series of subclones. The nucleotide sequence data was analyzed with GENETYX computer software (Software Development Co. Ltd.). Sequence similarity searches in GenBank were done with a BLAST program.

\textit{Construction of the full-length egV gene.} The
region encoding the C-terminal end of EGV was amplified by PCR using two specific primers, 5'-AACTGCAGGAAATAGGGATGCATCAGTC-3' and 5'-AACTGACAACTGGTTGTCGGTG-AATAG-3'. The PCR product was digested with PstI and ligated into the PstI site of pEV-1, yielding pEGV. This plasmid encoded the full-length EGV (Fig. 1).

**Purification of the recombinant enzyme.** E. coli DH5α harboring pEGV was cultured in 2 liters of LB medium containing a mpicillin (50 μg/ml) at 37°C for 12 h. Cells were harvested from the culture, suspended in 200 ml of 20 mM potassium phosphate buffer (pH 7.0), and disrupted by ultrasonication. Cell debris was removed by centrifugation. The cell-free extracts were put onto a CM-Toyopearl 650M column (1.0 by 20 cm; Tosoh) equilibrated with 20 mM potassium phosphate buffer (pH 7.0) and eluted with a linear gradient of 0 to 1.0 M sodium chloride in the same buffer. Active fractions were combined and put on a HiLoad Superdex 75pg column (16/60) (Amersham Pharmacia Biotech) chromatoctaphy, and the enzyme was eluted with 20 mM potassium phosphate buffer (pH 7.0). Active fractions were combined and used as the purified enzyme.

**Enzyme assays.** Enzyme activity of EGV toward CMC (low viscosity, Sigma) was measured by 5-min incubation at 37°C in 20 mM potassium phosphate buffer (pH 7.0). Reducing sugars released from the substrate were measured with the dinitrophenylhydrazine reagent as described previously. When oat-gluten xylan (Fluka Ag), lichenan (Sigma), laminarin (Nacalai Tesque Co.), Avicel (Merck), and acid-swollen cellulose (ASC) were used as the substrates, reducing sugars released from the substrates were measured with the 3,5-dinitrosalicylic acid reagent as described by Miller. ASC was prepared in this laboratory from Avicel. One unit of activity is defined as the amount of enzyme that produced 1 μmol of reducing sugars as a glucose or xylose standard per min. Protein concentrations were measured by the method of Lowry et al. with bovine serum albumin (Sigma) as a standard.

**Thin-layer chromatography (TLC).** TLC of the hydrolysis products was done on a silica gel 60-plastic sheet (Merck) developed with a solvent of n-butanol-acetic acid-water (2:1:1), and the oligosaccharides were made visible by spraying the plate with a diphenylamine-aniline-phosphate reagent.

**SDS-PAGE and Western blot analyses.** SDS-PAGE was done in 8% acrylamide gel as described by Laemmli. Antiserum against the truncated EGV purified from the recombinant E. coli was raised in a mouse. Western blot analysis was done with the antiserum and peroxidase conjugated to goat anti-mouse immunoglobulins as described previously. R. albus F-40 was cultivated at 37°C for 2 days in 5 ml of the cellulose medium specified for *Ruminococcus* species including 1.0% ball-milled-cellulose as carbon source, under anaerobic conditions. Cells were harvested by centrifugation at 10,000 × g for 10 min. Proteins from the culture supernatant were concentrated 5-fold by ultrafiltration through a Ultrafree-CL (nominal molecular weight cutoff of 10,000) (Millipore). The pelleted cells were suspended in SDS-PAGE sample buffer and heated at 100°C for 3 min. The proteins in the culture supernatant and the whole-cell proteins were analyzed by SDS-PAGE and Western blot techniques.

**N-terminal amino acid sequencing.** EGV purified from a recombinant E. coli was fractionated by SDS-PAGE and transferred onto a PVDF sequencing membrane (Millipore) by electroblotting. The blotted protein was cut from the membrane and its N-terminal amino acid sequence was analyzed on an Applied Biosystems model 476A protein sequencer.

**Nucleotide sequence accession number.** The nucleotide sequence of egV reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence data bases with the accession number AB028320.

**Results**

**Cloning of the egV gene**

Phage λR.acell was isolated from the *R. albus* genomic DNA library constructed by ligating the DNA fragments in the size range of 5 to 7 kb into λZapII. However, when pNK-1 was excised as a derivative of pBluescript from λR.acell by *in vivo* excision, this plasmid was found to contain a 1.6-kb fragment, suggesting that the cloned fragment suffered from unexpected deletion during the step of *in vivo* excision. Therefore, we attempted to clone the DNA fragment containing egV again. Since Southern hybridization analysis suggested that a 2.3-kb *Eco*RV fragment carried *egV*, we cloned this fragment into pUC19 to yield pEV-1 (Fig. 1). Although pEV-1 conferred endoglucanase activity on E. coli, DNA sequencing of this gene showed that the open reading frame of *egV* extended beyond the 3' end of the DNA insert. Gene walking PCR was used to isolate a DNA fragment downstream of pEV-1 and a 1.2-kb PCR product was cloned in pBluescript II SK(-), yielding pEI1.2 (Fig. 1), and sequenced. To construct the full-length *egV* gene, the region encoding the C-terminal end of EGV was amplified by PCR and a 0.4-kb PCR product was
cloned in pEV-1, yielding pEGV (Fig. 1). Nucleotide sequence of the egV gene Figure 2 shows the complete nucleotide sequence of the egV structural gene along with its flanking regions. The ORF of egV consists of 1,833 nucleotides encoding a protein of 611 amino acids with a predicted molecular weight of 67,103. The assigned ATG initiation codon at nucleotide position 619 is preceded by a spacing of 6 bp by a potential ribosome-binding sequence (AGGAG) which was similar to the consensus Shine-Dalgarno sequence. The sequences TTGAAA and TATAAT, with a 16-bp spacing, showing strong similarity to the −35 and −10 E. coli promoter consensus sequences, TTGACA and TATAAT, respectively, were identified upstream of the coding region.

Amino acid sequence of EGV Comparison of the amino acid sequence of EGV with those registered in protein data bases such as SWISS PROT and PIR clearly showed that the mature EGV consists of 3 distinct domains, i.e., a family 5 catalytic domain of glycosyl hydrolases, a domain of unknown function, and a dockerin domain known to be involved in cellulose assembly in some cellulolytic clostridia. As shown in Fig. 3, the family 5 catalytic domain of EGV, extending from amino acid position 30 to 378, had extensive sequence similarity with the other catalytic domains in family 5 from various organisms, e.g., Enda of R. flavefaciens 17 (46% sequence identity), CelB of Orpinomyces sp. (46%), CelB of Neocallimastix patriciarum (45%), and CelA of Clostridium longisporum (45%). A dockerin domain was found at the C terminus of EGV, suggesting that EGV was a component of the R. albus cellulose since dockerin domains are known to play a role in cellulose assembly by docking the various catalytic subunits to a noncatalytic scaffolding protein. Recently, some dockerin domains have been found in the enzymes of R. flavefaciens and Ruminococcus sp. Amino acid sequences of dockerin domains from R. albus F-40 and other Ruminococcus species are aligned in Fig. 4. The region between the family 5 catalytic domain and the dockerin domain did not show any sequence similarity to amino acid sequences in the protein databases.

Purification of EGV from E. coli The translated product of egV was purified 36.5-fold from the cell-free extract of E. coli DH5α (pEGV), with a recovery of 6% by CM-Toyopearl 650M and HiLoad Superdex 75pg column chromatographies. The final preparation gave a single band in SDS-PAGE, and the molecular weight of the enzyme was estimated to be around 42,000 (Fig. 5A, lane 3). The N-terminal amino acid sequence of this protein was identified as Lys-Asp-Val-Ser-Ser-Met-Thr-Ala-Val-Glu-Ile-Ala-Lys-Asp-Met-Gly-Leu.

---

Fig. 2. Nucleotide and Deduced Amino Acid Sequences of egV.
Fig. 3. Alignment of Family 5 Catalytic Domains of _R. albus_ (R.a) EGV, _R. flavofaciens_ (R.f) EndA, _Orpinomyces_ sp. (O.sp) CelB, _Neocaldimastricta_ patriciarum (N.p) CelB, and _Clostridium longiporum_ (C.l) CelA.

Glu residues identified as a proton donor is marked with an asterisk, and a nucleophile is marked with a sharp. Amino acids which are conserved in at least three of the five sequences are shaded. Dashes indicate gaps left to improve alignment. Numbers refer to amino acid residues at the start of the respective lines; all sequences are numbered from Met-1 of the peptide.

Fig. 4. Alignment of Dockerin Domains of EGV and EGV1 of _R. albus_ (R.a); EndA, XynB, and XynD of _R. flavofaciens_ (R.f); XynI of _Ruminococcus_ sp. (R.sp); CelA of _C. thermocellum_ (C.t); and CelC of _C. cellulolyticum_ (C.c).

Sequences of _R. albus, R. flavofaciens, Ruminococcus_ sp., _C. thermocellum_, and _C. cellulolyticum_ proteins are from references 12, 20, and 33. Asterisks indicate amino acid residues involved in calcium binding. A repeated pair in _C. thermocellum_ CelA or _C. cellulolyticum_ CelC are double underlined. Amino acids that are conserved or have similar chemical properties (I, M, V, K, R, S, and T) in at least five of eight sequences are shaded. Dashes indicate gaps left to improve alignment. Numbers refer to amino acid residues at the start of the respective lines; all sequences are numbered from Met-1 of the peptide.

Gly-Trp-Asn, which was found in the deduced amino acid sequence of EGV at amino acid positions 30 to 49 (Fig. 2), suggesting that the N-terminal sequence of 29 amino acids mediates secretion of the protein to the periplasmic space as a signal peptide. However, the molecular weight estimated by SDS-PAGE was apparently lower than that of the mature EGV excluding the signal peptide deduced from the nucleotide sequence (64,080), and it was likely that the recombinant enzyme obtained here arose from a parental protein by partial proteolysis at its C-terminal region.

_Detection of EGV proteins in recombinant E. coli strain and R. albus F-40_
To analyze molecular species produced from the egV gene in E. coli, we analyzed the proteins in the cell-free extract of E. coli DH5α (pEGV) by Western blotting using the antiserum raised against the purified enzyme. Some proteins with strong intensity were detected on the Western blot (Fig. 5B, lane 4). Among them, one of the major proteins with an apparent molecular weight of 42,000 appeared to corresponding to the purified enzyme. However, a 64-kDa protein correspond to the full-length EGV was narrowly detected as a faint band. These results suggested that EGV suffered severely from partial proteolysis by protease(s) of E. coli, resulting in the formation of many active proteins including the 42-kDa protein.

The anti-EGV antiserum reacted with a protein with an apparent molecular weight of 64,000 in the whole cells (Fig. 5B, lane 1). The size of this immunoreactive protein was in good agreement with that calculated from the deduced amino acid sequence of EGV, suggesting that the egV gene was expressed in R. albus. In addition, some proteins in the molecular weight range of 40,000 to 110,000 were also detected in the whole cells. A similar result was obtained for the proteins in the culture supernatant although the 50-kDa and 64-kDa proteins were not detected (Fig. 5B, lane 2).

**Characterization of EGV**

The truncated EGV protein (42-kDa) was used for general characterization of the enzymatic properties, since use of protease inhibitors did not prevent EGV from proteolysis by E. coli protease(s). The purified enzyme showed high activity toward CMC (197 U/mg) and lichenan (26.9 U/mg), and low activity toward oat-spelt xylan (3.1 U/mg) and ASC (1.0 U/mg), but no activity against Avicel or laminarin. The enzyme activity was completely inhibited by CuCl₂ and MnCl₂, and was partly inhibited by FeCl₃, CdCl₂, CoCl₂, ZnCl₂, and EDTA at a concentration of 1 mM. The optimum pH for activity was found to be 7.0 when the enzyme activity was assayed in Britton-Robinson’s universal buffer solution (40 mM phosphoric acid, 40 mM boric acid, 40 mM acetic acid; pH adjusted with NaOH) at various pHs. The enzyme was stable in the range of pH 5 to 9 when incubated in the same buffer solution at various pHs at 37°C for 30 min. The enzyme was optimally active at 40°C and stable under heat treatment at 40°C for 10 min. The action of EGV was qualitatively analyzed on cellooligosaccharides (Fig. 6). This enzyme hydrolyzed cellolactose, cellopentaose, and cellohexaose to yield cellobiose and celletriose as major products, and glucose as a minor product. By contrast, it was less active toward celletriose and not active at all toward cellobiose.

**Discussion**

*R. albus* EGV is a modular endoglucanase composed of a family 5 catalytic domain of glycosyl hydrolases, a domain of unknown function, and a dockerin domain. Among the glycosyl hydrolase families, family 5 is the family containing the greatest number of cellulases, accounting for around 40% of all the cellulase sequences reported so far. Extensive biochemical studies including site-directed mutagenesis along with tertiary structure analysis identified a strictly conserved glutamate residue as the catalytic nucleophile and a second glutamate residue as the proton donor, and showed that the family 5 enzymes form closely related eight-stranded α/β barrel structures. Residues Glu-307, identified as the nucleophile, and Glu-170, identified as the proton donor, in *Clostridium cellulolyticum* CelA are conserved as Glu-304 and Glu-181, respectively, in EGV (Fig. 3). Although some endoglucanases classified in family 5 have been reported from *R. albus*, i.e., EGI, EGV, CelA, and CelB, none of them contained a dockerin domain. Therefore, this is the first paper describing the complete amino acid sequence and some enzymatic properties of an endoglucanase containing a dockerin domain from *R. albus*.

A dockerin domain is a docking domain that mediates the binding of a catalytic component to a scaffolding protein, resulting in the incorporation of the catalytic component into the cellulosome. Therefore, a dockerin domain is essential in the cellulosome assembly. Some early studies suggested the occurrence of the cellulosome in *R. albus*, i.e., Wood _et al._ presumed that *R. albus* cellulase existed as an aggregate of low-molecular-weight cellulse components although the concept of the cellosome had not been established then; electron
microscopic examinations showed that the cells had specialized cell surface structures. Nevertheless, it was not necessarily believed that *R. albus* produces the cellulose as its cellulolytic system since extensive biochemical studies have not been done on the *R. albus* cellulose and early cloning experiments failed to show the presence of a dockerin domain in enzymes from *R. albus*. The identification of the dockerin domain in EGV strongly suggest that *R. albus* F-40 produces the cellulose and EGV is a component of the cellulose. In Western blot analysis (Fig. 5B), the antiserum raised against the truncated EGV identified a immunoreactive protein, the size of which corresponded to that calculated from the deduced amino acid sequence of EGV, and some minor proteins in the whole-cell proteins rather than in the proteins of culture supernatant of *R. albus* F-40. These observations suggest that some endoglucanases classified in family 5 are mainly located in the cells, most likely on the cell surface, although free cellulases and free cellules are also expected to be present in the culture supernatant (unpublished result). This is consistent with the early observation that the cellulase aggregates were released from the cell surface by washing the cells of *R. albus*.

Clostridial dockerin domains consist of a pair of well-conserved 22-residue repeats and are highly conserved in catalytic components of the cellulases. In clostridial dockerin domain sequences, five residues are almost completely conserved as D or N in position 1 to 5 for calcium-binding (Fig. 4). In the ruminococcal dockerin domains, amino acid residues in positions 2, 3, and 4 of the latter moieties of the repeats are less conservative, but those in the other positions are highly conserved. Pagès *et al.* showed that the cohesin-dockerin interaction in the *C. thermocellum* and *C. cellulolyticum* cellulases is a species-specific phenomenon, and they predicted that four amino acid residues, which comprise a repeated pair (AL or AI for *C. cellulolyticum*, ST or SS for *C. thermocellum*) located between positions 4 and 5, are critical to binding specificity as a recognition code. Amino acid residues at the corresponding positions of the *R. albus* F-40 dockerin domains are different from those conserved in the clostridial dockerin domains, suggesting that the dockerin domain of EGV specifically bind to a cohesin domain of an expected scaffolding protein of *R. albus* F-40. Recent findings about the presence of dockerin domains in the enzymes of *R. flavefaciens* and *Ruminococcus* sp. suggest that *Ruminococcus* species produce the cellulases in common.

When we attempted to purify EGV from the recombinant *E. coli*, the main difficulty encountered was the cleavage of the protein during cultivation and purification. As a result, we obtained a truncated enzyme with an apparently molecular weight of about 42,000 in a purified form. Since the N-terminal amino acid sequence of the truncated EGV corresponds to that of the N-terminal end of the family 5 catalytic domain and the catalytic domain resides at N-terminal region of the peptide, it is apparent that proteolytic truncation occurs within the domain of unknown function. This region, which is enriched in hydroxyl amino acids (Thr and Ser), seems to be preferentially recognized by *E. coli* protease(s). The truncated EGV showed a wide substrate specificity, i.e., this enzyme was active toward xylan and lichenan in addition to cellulose substrates. This property is common to endoglucanases in family 5.

In conclusion, EGV is a modular endoglucanase containing a dockerin domain in addition to a family 5 catalytic domain and a domain of unknown function. Characterization of the truncated EGV showed that it has enzymatic properties common to other endoglucanases in family 5 reported so far, while the function of the C-terminal domain remains to be studied. EGV and some other proteins of family 5 were detected in the whole-cell fraction, suggesting that *R. albus* F-40 produces the cellulose on the cell surface. Isolation and biochemical characterization of the *R. albus* F-40 cellulose remains to be done to confirm this suggestion. Furthermore, a scaffolding protein, with which the dockerin domain of EGV specifically interacts, should be investigated to understand the mechanism of the cellulose assembly in *R. albus* F-40.

**Acknowledgment**

We are grateful to N. Kawai for his assistance in constructing the recombinant clones. This work was supported in part by a grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences, and a Grant-in-Aid for Scientific Research No. 09760243 from the Ministry of Education, Science, and Culture of Japan.

**References**

5) Henrissat, B. and Bairoch, A., Updating the sequence-based classification of glycosyl hydrolases.


7) Bélaich, J.-P., Tardif, C., Bélaich, A., and Gaudin, C., The cellulolytic system of Clostridium cel-


9) Kakiuchi, M., Isi, A., Suzuki, K., Fujiwara, T., Fujino, E., Kimura, T., Karita, S., Sakka, K., and

Ohmiya, K., Cloning and DNA sequencing of the genes encoding Clostridium josui scaffolding protein

CipA and cellulase CelD and identification of their gene products as major components of the cellulo-


10) Bayer, E. A., Shimon, L. J. W., Shoahm, Y., and Lamed, R., Cellulosomes—Structure and ultrastruc-


12) Pagès, S., Bélaich, A., Bélaich, J.-P., Morag, E., Lamed, R., Shoahm, Y., and Bayer, E. A., Speecies-

pecificity of the cohesin-dockerin interaction between Clostridium thermocellum and Clostridium cel-

ulolyticum: Prediction of specificity determinants of the dockerin domain. Proteins, 29, 517–

527 (1997).


14) Lamed, R., Naimark, J., Morgenstern, E., and Bayer, E. A., Specialized cell surface structures in cel-


16) Bayer, E. A. and Lamed, R., Ultrastructure of the cell surface cellulosome of Clostridium thermocellum


17) Ohmiya, K., Kajino, T., Kato, A., Shimizu, S., Structure of Ruminococcus albus endo-1,4-β-


19) Poole, D. M., Hazlewood, G. P., Laurie, J. I., Barker, P. J., and Gilbert, H. J., Nucleotide se-

quence of the Ruminococcus albus SY3 endo-


20) Karita, S., Sakka, K., and Ohmiya, K., Cellulos-

omes, cellulase complexes, of anaerobic microbes: Their structure models and functions. In “Rumen


21) Ohmiya, K., Shimizu, M., Taya, M., and Shimizu, S., Purification and properties of cellullobiose from


22) Asmundson, R. V. and Kelly, W. J., Isolation and characterization of plasmid DNA from Ruminococ-


23) Teather, R. M. and Wood, P. J., Use of Congo red-poly saccharide interactions in enumeration and characterisation of cellulolytic bacteria from the bo-


24) Parker, J. D., Rabinovitch, P. S., and Burmer, G. C., Targeted gene walking polymerase chain reac-


25) Sakka, K., Shimamuki, T., and Shimada, K., Nucleotide sequence of celC207 encoding endo-


27) Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J., Protein measurement with the Folin


