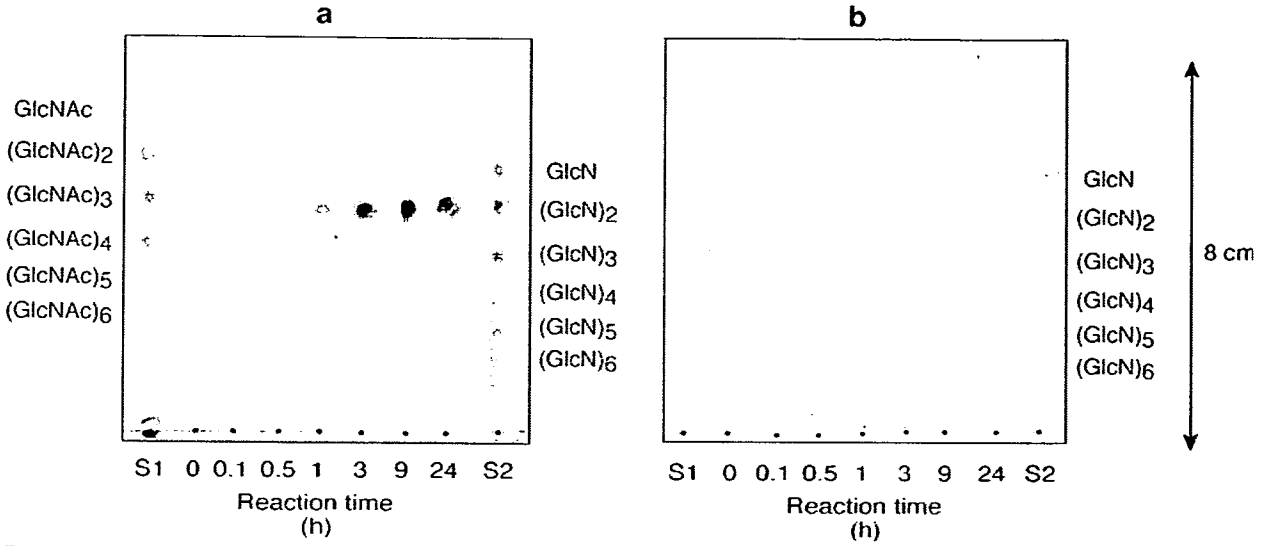

Fig. 1 TLC analysis of oligosaccharide produced by incubating β-chitin with crude enzyme solution. Crude enzyme solution (200 μl) prepared from the supernatant of *V. parahaemolyticus* KN1699 cultures was added to 3.8 ml 20-mM sodium phosphate buffer (pH 7.0) containing 20 mg powdered β-chitin, then the mixture was incubated at 37°C with stirring. After developing the TLC plates, reaction products produced after various incubation times were visualized using the following reagents: **a** phosphomolybdic acid reagent and **b** ninhydrin reagent. *Lane S1* *N*-acetylchitooligosaccharide standards, *lane S2* chitooligosaccharide standards

$J_{1,2}=J_{2,3}$ 8.8 Hz, H-2), 3.31–3.53 (m, 3H, H-3, H-4, and H-5), 3.55–3.59 (m, 1H, H-6a), 3.68 (dd, 1H, $J_{6a,6b}$ 12.4 Hz, H-6b), 4.45 (d, 1H, $J_{1,2}$ 8.0 Hz, H-1). The NMR signals obtained corresponded well to the those of β-D-*N*-acetylglucosaminyl-(1,4)-D-glucosamine (GlcNAc-GlcN) (Ohishi et al. 1997; Tokuyasu et al. 1999). From these results, it became clear that the compound produced from β-chitin by the action of crude enzyme, which was prepared from the supernatant of *V. parahaemolyticus* KN1699 cultures, is GlcNAc-GlcN.

This heterodisaccharide seemed to be produced by the reactions of both chitinase (*Pa*-Chi) and chitin oligosaccharide deacetylase (*Pa*-COD), which were secreted into the culture medium by *V. parahaemolyticus* KN1699. Then, to confirm this supposition, we purified these enzymes and analyzed their reactions.

Purification of the enzymes

Purification of *Pa*-Chi: The *Pa*-Chi was purified in four steps from 1 l of culture fluid (Table 3). Each column chromatographic step produced a single peak showing chitinase activity. The purified enzyme gave a single band on both SDS-PAGE (Fig. 2a) and native-PAGE (Fig. 2b), indicating that the enzyme protein is in a high state of purity. *Pa*-Chi was purified 9.67-fold with 31.2% recovery of initial total activity. The specific activity of the purified enzyme towards powdered β-chitin was 2.32 U/mg of protein.

Table 3 Purification of *Pa*-Chi and *Pa*-COD from culture fluid of strain KN1699

Purification step	Total activity (U)	Specific activity (U/mg of protein)	Yield (%)	Fold
Pa-Chi[a]				
$(NH_4)_2SO_4$ precipitation	34.4	0.24	100	1
DEAE-Toyopearl 650M (first)	18.2	1.61	52.9	6.71
DEAE-Toyopearl 650M (second)	14.9	2.13	43.2	8.88
Toyopearl HW-55F	10.7	2.32	31.2	9.67
Pa-COD[b]				
$(NH_4)_2SO_4$ precipitation	1.95	0.01	100	1
DEAE Sepharose FF	1.04	3.05	53.3	305
DEAE-Toyopearl 650M	0.76	9.50	39.0	950
Phenyl Sepharose HP	0.62	31.0	31.8	3,100

[a] *Pa*-Chi was purified from 1 l of culture fluid.
[b] *Pa*-COD was purified from 2 l of culture fluid.

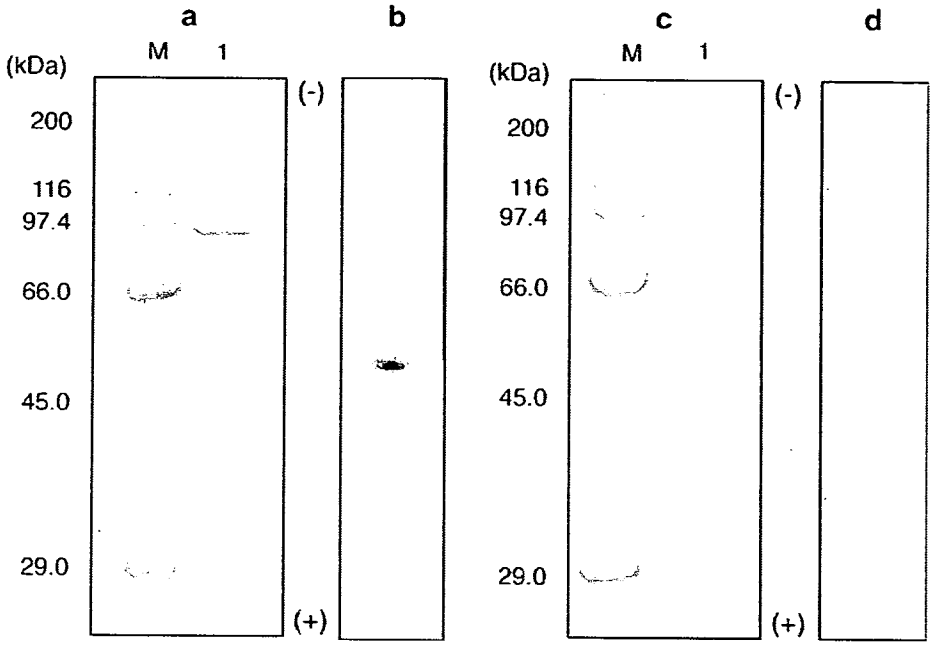

Fig. 2 PAGE of purified enzymes. **a** and **c** are the results from SDS-PAGE of *Pa*-Chi and *Pa*-COD, respectively. **b** and **d** are the results from native-PAGE of *Pa*-Chi and *Pa*-COD, respectively. *Lane M* contains molecular mass standards (Sigma). *Lane 1* of **a** and of **c** contain purified *Pa*-Chi and *Pa*-COD, respectively

Purification of *Pa*-COD: The *Pa*-COD was purified in four steps from 2 l of culture fluid (Table 3). Each column chromatographic step produced a single peak showing deacetylase activity. The purified enzyme gave a single band on both SDS-PAGE (Fig. 2c) and native-PAGE (Fig. 2d), indicating that the enzyme protein is in a high state of purity. *Pa*-COD was purified 3,100-fold with 31.8% recovery of initial total activity. The specific activity of the purified enzyme towards $(GlcNAc)_2$ was 31 U/mg of protein. Although the amount of the deacetylase protein in the culture supernatant was very small, its specific activity was significantly higher than that of *Pa*-Chi.

Properties of the enzymes

The molecular masses of *Pa*-Chi and *Pa*-COD were estimated by SDS-PAGE to be 92 kDa and 46 kDa, respectively, based on molecular mass standards (Sigma). The N-terminal amino acid sequences of *Pa*-Chi and *Pa*-COD were APTAPSVDMYGSNNLQFSKIELAMET and

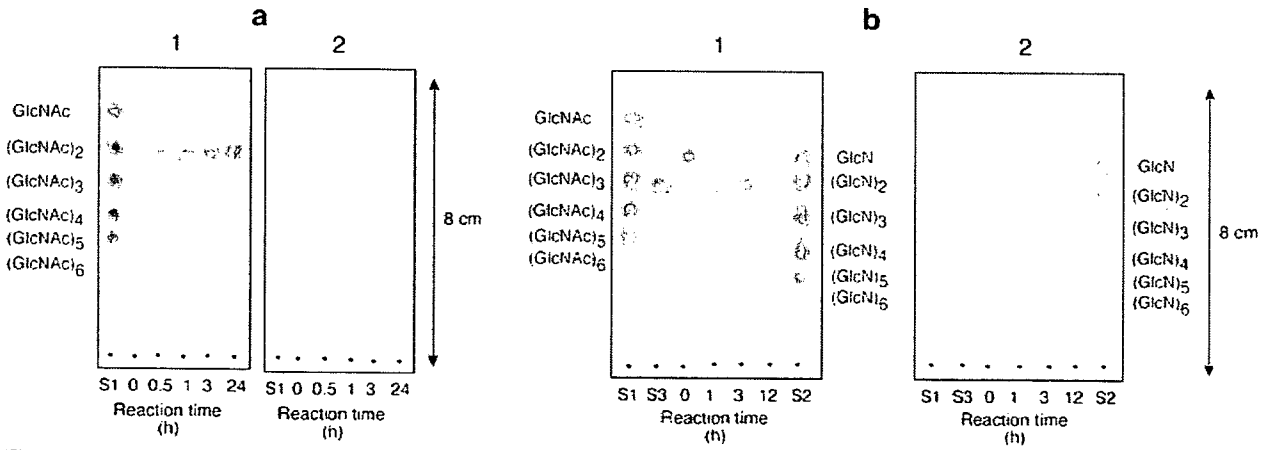

Fig. 3 TLC analysis of oligosaccharides produced by the reaction of *Pa*-Chi and *Pa*-COD. Purified *Pa*-Chi solution (250 µl, 30 mU) was added to 750 µl of 20-mM sodium phosphate buffer (pH 7.0) containing 5 mg of powdered β-chitin, then the mixture was incubated at 37°C with stirring. TLC results of the reaction mixture with *Pa*-Chi are shown in **a**. Purified *Pa*-COD solution (40 µl, 3.12 mU) was added to 40 µl of 20-mM sodium phosphate buffer (pH 7.0) containing 0.4 mg of $(GlcNAc)_2$, then the mixture was incubated at 37°C with stirring. TLC results of the reaction mixture with *Pa*-COD are shown in **b**. Reaction products on the TLC plate after various incubation times were visualized using the following reagents: **a**-1 and **b**-1, phosphomolybdic acid reagent; **a**-2 and **b**-2, ninhydrin reagent. *Lane S1* *N*-acetylchitooligosaccharide standards, *lane S2* chitooligosaccharide standards, *lane S3* GlcNAc-GlcN

Table 4 Substrate specificities of *Pa*-Chi and *Pa*-COD of strain KN1699

Substrate	Specific activity (U/mg of protein)
(*Pa*-Chi)	
Powdered α-chitin	0.8
Powdered β-chitin	2.3
Colloidal chitin	1.5
(*Pa*-COD)	
GlcNAc	n.d.
$(GlcNAc)_2$	31
$(GlcNAc)_3$	6.5
$(GlcNAc)_4$	n.d.
$(GlcNAc)_5$	n.d.
$(GlcNAc)_6$	n.d.

n.d. Not detectable activity

QTDTKGTIYLTFDDGPINASIDVINV, respectively. *Pa*-Chi and *Pa*-COD were most active at pH 8.0 and pH 8.5–9.0, respectively, and the optimum reaction temperatures were 50–55°C for *Pa*-Chi and 45°C for *Pa*-COD. *Pa*-Chi was stable at pH values between 5.0 and 10 (37°C, 60 min) and at temperatures below 45°C (pH 7.0, 30 min). *Pa*-COD was stable at pH values between 6.0 and 10 (37°C, 60 min) and at temperatures below 40°C (pH 7.0, 30 min).

The TLC mobility results indicate that $(GlcNAc)_2$ is the single oligosaccharide produced by the hydrolytic action of *Pa*-Chi on β-chitin (Fig. 3a-1). The product did not react with ninhydrin (Fig. 3a-2), indicating that *Pa*-Chi hydrolyzed β-chitin to $(GlcNAc)_2$. Reaction of *Pa*-Chi with $(GlcNAc)_6$ also gave only $(GlcNAc)_2$ as the final product (data not shown). The specific activity of *Pa*-Chi towards powdered β-chitin was higher than towards colloidal chitin and powdered α-chitin (Table 4). The TLC results with *Pa*-COD indicate that this enzyme converts $(GlcNAc)_2$ to a saccharide that is more polar than $(GlcNAc)_2$ (Fig. 3b-1) and possesses a free amino group (Fig. 3b-2). We confirmed the structure of this product to be GlcNAc-GlcN by ^1H NMR analysis (data not shown). It was confirmed that, of the oligosaccharides examined, *Pa*-COD showed activity not only for $(GlcNAc)_2$ but also some still lower activity for $(GlcNAc)_3$ (Table 4). Figure 4 shows TLC result of the reaction product from $(GlcNAc)_3$ by *Pa*-COD. The ESIMS spectra of this product corresponded to $[M-H]^-$ and $[M+H]^+$ species at *m/z* of 584 and 586, respectively, indicating that this compound is a trisaccharide consisting of one GlcN and two GlcNAc.

Discussion

Extracellular chitinase and chitin oligosaccharide deacetylase derived from *V. parahaemolyticus* were purified, and their properties were clarified for the first time. The chitinase, *Pa*-Chi, of the strain KN1699, which was isolated from dry beach soil and identified in the present study, was confirmed to hydrolyze chitin and $(GlcNAc)_6$ to produce $(GlcNAc)_2$. These facts indicate that *Pa*-Chi is an *exo-N,N,*-diacetylchitobiohydrolase-like enzyme. The chitin oligosaccharide deacetylase, *Pa*-COD, of this strain was confirmed to catalyze the hydrolysis of acetamide group of reducing end GlcNAc residue of $(GlcNAc)_2$. From these findings, it became clear that GlcNAc-GlcN (main hydrolysis product) is produced from chitin by the cooperative hydrolytic reactions of both *Pa*-Chi and *Pa*-COD present in the supernatant of *V. parahaemolyticus* KN1699 cultures.

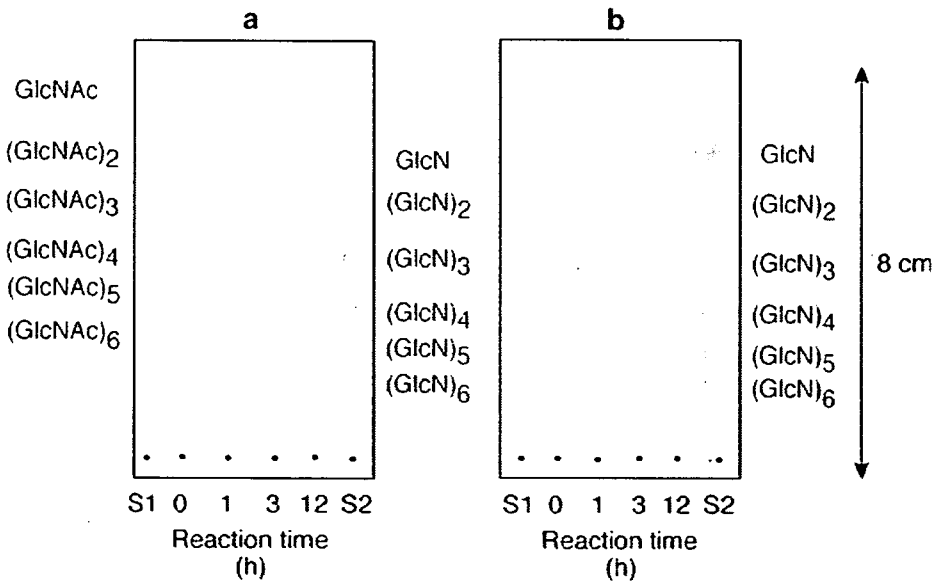

Fig. 4 TLC analysis of oligosaccharide produced from $(GlcNAc)_3$ by the reaction of *Pa*-COD. Purified *Pa*-COD solution (40 μl, 3.12 mU) was added to 40 μl of 20-mM sodium phosphate buffer (pH 7.0) containing 0.4 mg of $(GlcNAc)_3$, then the mixture was incubated at 37°C with stirring. After developing the TLC plates, reaction products produced after various incubation times were visualized using the following reagents: **a** phosphomolybdic acid reagent and **b** ninhydrin reagent. *Lane S1* *N*-acetylchitooligosaccharide standards, *lane S2*, chitooligosaccharide standards

Several chitinases from the genus *Vibrio* were isolated, and their properties were elucidated. The molecular mass (92 kDa) of *Pa*-Chi did not correspond to those of chitinases isolated from other *Vibrios* such as *Vibrio* sp. (63 kDa; Ohtakara et al. 1979), *Vibrio alginolyticus* TK-22 (66 kDa; Murao et al. 1992), *Vibrio* sp. P-6-HC (100 kDa; Takahashi et al. 1993), *V. alginolyticus* H-8 (C1; 81 kDa, C3; 68 kDa; Ohishi et al. 1996) and *Vibrio carchariae* (63–66 kDa; Suginta et al. 2000). Comparison of N-terminal amino acid sequence (26 residues) of *Pa*-Chi to the Basic Local Alignment Search Tool for Protein (BLASTP) protein sequence database showed 100% identity with those of GH family 18 chitinases from *V. parahaemolyticus* RIMD 2210633 (GenBank accession number BA000032) and *V. alginolyticus* H-8 (GenBank accession number AB055155). These facts indicate that *Pa*-Chi belongs to GH family 18.

Two extracellular chitin oligosaccharide deacetylases with different molecular masses (DA1, 48 kDa; DA2, 46 kDa) have been isolated from *V. alginolyticus* H-8 (Ohishi et al. 1997). These enzymes catalyzed the hydrolysis of acetamide group of reducing end GlcNAc residue of $(GlcNAc)_2$, the same as *Pa*-COD. Except for DA1, DA2, and *Pa*-COD, microbial chitin oligosaccharide deacetylases showing this function have not been isolated. The molecular mass of *Pa*-COD is the same as that of the DA2. Comparison of N-terminal amino acid sequence (26 residues) of *Pa*-COD to the BLASTP showed 100% identity with those of CE family 4 deacetylases from *V. parahaemolyticus* RIMD 2210633 (GenBank accession number BA000031) and *V. alginolyticus* H-8, (DA1; GenBank accession number AJ292005). Moreover, from the BLASTP results, the N-terminal amino acid sequence of *Pa*-COD was confirmed to show high homologies with those of CE family 4 deacetylases from other *Vibrios*. A short peptide LTFDDGP (NodB homology domain), which is one of the conserved sequences commonly existing in CE family 4 enzymes (Caufrier et al. 2003), was observed also in the N-terminal sequence of *Pa*-COD. These facts indicate that *Pa*-COD belongs to CE family 4. From some properties, it is suggested that *Pa*-COD is similar to the deacetylases (DA1 and DA2) from *V. alginolyticus* H-8. However, DA1 and DA2 were confirmed to show the activities only against $(GlcNAc)_2$ in chitin oligosaccharides used [$(GlcNAc)_{2-6}$] (Ohishi et al. 1997), whereas *Pa*-COD showed significant activity not only against $(GlcNAc)_2$ but also against $(GlcNAc)_3$. These facts indicate that the specificities toward chitin oligosaccharides are apparently different between *Pa*-COD and the deacetylases of *V. alginolyticus* H-8.

Partially deacetylated chitin oligosaccharides such as GlcNAc-GlcN may have useful physiological functions. Production of chitinases that produce $(GlcNAc)_2$ from chitin was confirmed in many kinds of microorganisms, and they are useful for the production of this homodisaccharide from chitin. On the other hand, reports with chitin oligosaccharide deacetylase, which produce GlcNAc-GlcN from $(GlcNAc)_2$, are very few. *Pa*-COD is precious enzyme for the production of this heterodisaccharide. However, this enzyme is present only in low levels in *V. parahaemolyticus* KN1699 cultures. Therefore, for the large-scale production of GlcNAc-GlcN, efforts are underway to clone the *Pa*-COD gene and overexpress it in *Escherichia coli* cells.

Acknowledgment This work was supported by a grant from the 21st Century Center of Excellence (COE) Program of the Ministry of Education, Science, Sports, and Culture (Japan) to promote advanced scientific research.

References

Caufrier F, Martinou A, Dupont C, Bouriotis V (2003) Carbohydrate esterase family 4 enzymes: substrate specificity. Carbohydr Res 338:687–692

Dahiya N, Tewari R, Hoondal GS (2006) Biotechnological aspects of chitinolytic enzymes: a review. Appl Microbiol Biotechnol 71:773–782

Dische Z, Borenfreund E (1950) A spectrophotometric method for the microdetermination of hexosamines. J Biol Chem 184:517–522

Farmer III JJ, Janda JM, Brenner FW, Cameron DN, Birkhead KM (2005) Genus I. *Vibrio* Pacini 1854, 411^{AL}. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds) Bergey's manual of systematic bacteriology vol 2: the proteobacteria part B: the gammaproteobacteria, 2nd edn. Springer, Berlin Heidelberg New York, pp 494–546

Gao XD, Katsumoto T, Onodera K (1995) Purification and characterization of chitin deacetylase from *Absidia coerulea*. J Biochem (Tokyo) 117:257–263

Imoto T, Yagashita K (1971) A simple activity measurement of lysozyme. Agric Biol Chem 35:1154–1156

Kafetzopoulos D, Martinou A, Bouriotis V (1993) Bioconversion of chitin to chitosan: purification and characterization of chitin deacetylase from *Mucor rouxii*. Proc Natl Acad Sci USA 90:2564–2568

Kobayashi M, Watanabe T, Suzuki S, Suzuki M (1990) Effect of *N*-acetylchitohexaose against *Candida albicans* infection of tumor-bearing mice. Microbiol Immunol 34:413–426

Murao S, Kawada T, Itoh H, Oyama H, Shin T (1992) Purification and characterization of a novel type of chitinase from *Vibrio alginolyticus* TK-22. Biosci Biotechnol Biochem 56:368–369

Ohishi K, Yamagishi M, Ohta T, Suzuki M, Izumida H, Sano H, Nishijima M, Miwa T (1996) Purification and properties of two chitinases from *Vibrio alginolyticus* H-8. J Ferment Bioeng 82:598–600

Ohishi K, Yamagishi M, Ohta T, Motosugi M, Izumida H, Sano H, Adachi K, Miwa T (1997) Purification and properties of two deacetylases produced by *Vibrio algionolyticus* H-8. Biosci Biotechnol Biochem 61:1113–1117

Ohtakara A, Mitsutomi M, Uchida Y (1979) Purification and some properties of chitinase from *Vibrio* sp. J Ferment Technol 57:169–177

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Scigelova M, Crout DHG (1999) Microbial β-*N*-acetylhexosaminidases and their biotechnological applications. Enzyme Microb Technol 25:3–14

Shimahara K, Takiguchi Y (1988) Preparation of crustacean chitin. Methods Enzymol 161:417–423

Suginta W, Robertson PAW, Austin B, Fry SC, Fothergill-Gilmore LA (2000) Chitinases from *Vibrio*: activity screening and purification of chiA from *Vibrio carchariae*. J Appl Microbiol 89:76–84

Suzuki K, Mikami T, Okawa Y, Tokoro A, Suzuki S, Suzuki M (1986) Antitumor effect of hexa-*N*-acetylchitohexaose and chitohexaose. Carbohydr Res 151:403–408

Takahashi M, Tsukiyama T, Suzuki T (1993) Purification and some properties of chitinase produced by *Vibrio* sp. J Ferment Bioeng 75:457–459

Tokoro A, Tatewaki N, Suzuki K, Mikami T, Suzuki S, Suzuki M (1988) Growth-inhibitory effect of hexa-*N*-acetylchitohexaose and chitohexaose against Meth-A solid tumor. Chem Pharm Bull 36:784–790

Tokoro A, Kobayashi M, Tatewaki N, Suzuki K, Okawa Y, Mikami T, Suzuki S, Suzuki M (1989) Protective effect of *N*-acetyl chitohexaose on *Listeria monocytogenes* infection in mice. Microbiol Immunol 33:357–367

Tokuyasu K, Ohnishi-Kameyama M, Hayashi K (1996) Purification and characterization of extracellular chitin deacetylase from *Colletotrichum lindemuthianum*. Biosci Biotechnol Biochem 60:1598–1603

Tokuyasu K, Ono H, Hayashi K, Mori Y (1999) Reverse hydrolysis reaction of chitin deacetylase and enzymatic synthesis of β-D-GlcNAc-(1-4)-GlcN from chitobiose. Carbohydr Res 322:26–31

Tsigos I, Bouriotis V (1995) Purification and characterization of chitin deacetylase from *Colletotrichum lindemuthianum*. J Biol Chem 270:26286–26291

Tsigos I, Martinou A, Kafetzopoulos D, Bouriotis V (2000) Chitin deacetylases: new, versatile tools in biotechnology. Trends Biotech 18:305–312

Tsukada K, Matsumoto T, Aizawa K, Tokoro A, Naruse R, Suzuki S, Suzuki M (1990) Antimetastatic and growth-inhibitory effects of *N*-acetylchitohexaose in mice bearing Lewis lung carcinoma. Jpn J Cancer Res 81:259–265

Biotechnol Lett (2007) 29:1209–1215
DOI 10.1007/s10529-007-9386-6

ORIGINAL RESEARCH PAPER

Production of a recombinant chitin oligosaccharide deacetylase from *Vibrio parahaemolyticus* in the culture medium of *Escherichia coli* cells

Kazunari Kadokura · Yusuke Sakamoto · Kaori Saito · Takanori Ikegami · Takako Hirano · Wataru Hakamata · Tadatake Oku · Toshiyuki Nishio

Received: 21 February 2007 / Accepted: 21 March 2007 / Published online: 4 May 2007

Abstract An open reading frame (ORF) encoding chitin oligosaccharide deacetylase (*Pa*-COD) gene and its signal sequence was cloned from the *Vibrio parahaemolyticus* KN1699 genome and its sequence was analyzed. The ORF encoded a 427 amino acid protein, including the 22 amino acid signal sequence. The deduced amino acid sequence was highly similar to several bacterial chitin oligosaccharide deacetylases in carbohydrate esterase family 4. An expression plasmid containing the gene was constructed and inserted into *Escherichia coli* cells and the recombinant enzyme was secreted into the culture medium with the aid of the signal peptide. The concentration of the recombinant enzyme in the *E. coli* culture medium was 150 times larger than that of wild-type enzyme produced in the culture medium by *V. parahaemolyticus* KN1699. The recombinant enzyme was purified to homogeneity from culture supernatant in an overall yield of 16%. Substrate specificities of the wild-type and the recombinant enzymes were comparable.

K. Kadokura · Y. Sakamoto · K. Saito · T. Ikegami · T. Hirano · W. Hakamata · T. Oku · T. Nishio (✉)
Department of Biological Chemistry, College of Bioresource Sciences, Nihon University, 1866 Kameino, Fujisawa, Kanagawa 252-8510, Japan
e-mail: nishio@brs.nihon-u.ac.jp

Keywords Chitin oligosaccharide deacetylase · *Escherichia coli* · Recombinant enzyme · Signal peptide · *Vibrio parahaemolyticus*

Introduction

Various oligosaccharides exhibit physiologically useful functions. Many of these oligosaccharides are prepared by the enzymatic degradation of biomass polysaccharides or by enzymatic conversion of oligosaccharides produced by higher plants. New oligosaccharides with potential therapeutic activities are currently being developed; of particular interest to us are the physiological properties of oligosaccharides obtained by the hydrolysis of chitin, a β-(1,4)*N*-acetyl-D-glucosamine (GlcNAc) polymer. To date, it has been confirmed that hexa-*N*-acetylchitohexaose, $(GlcNAc)_6$, exhibits antitumor (Suzuki et al. 1986; Tokoro et al. 1988; Tsukada et al. 1990) and antimicrobial activity (Tokoro et al. 1989; Kobayashi et al. 1990) in mice by enhancing the immunological defense system. Chitin is one of the most abundant biomass polysaccharides, composing the shells of crustaceans such as crab and shrimp, the exoskeletons of insects, and the cell walls of fungi. Various enzymes involved in chitin hydrolysis [i.e., chitinase (Chi), β-*N*-acetylhexosaminidase, chitin deacetylase, and chitin oligosaccharide deacetylase (COD)] are known.

In a previous paper, we reported that treating powdered chitin with a crude enzyme solution prepared from the supernatant of *Vibrio parahaemolyticus* KN1699 cultures yielded the heterodisaccharide, β-*N*-acetyl-D-glucosaminyl-(1,4)-D-glucosamine (GlcNAc-GlcN), as the primary chitin degradation product (Kadokura et al. 2007). This study confirmed that GlcNAc-GlcN was produced from chitin by cooperative hydrolytic reactions involving both chitinase (*Pa*-Chi) and chitin oligosaccharide deacetylase (*Pa*-COD) in the crude enzyme solution. *Pa*-Chi is an *exo*-*N*,*N*-diacetylchitobiohydrolase-like enzyme that hydrolyzes chitin to produce di-*N*-acetylchitobiose $(GlcNAc)_2$, and *Pa*-COD is an enzyme that catalyzes the hydrolysis of the acetamide bond on the reducing end GlcNAc residue of $(GlcNAc)_2$.

To investigate the physiological functions of GlcNAc-GlcN, we attempted gram-scale production of this unique heterodisaccharide. However, the productivity of *Pa*-COD from *V. parahaemolyticus* KN1699 was extremely low (Kadokura et al. 2007). Therefore, we cloned the *Pa*-COD gene in order to overproduce recombinant *Pa*-COD (*Pa*-rCOD) by *Escherichia coli*. Using recombinant enzyme preparation for oligosaccharide production on a practical scale will require excluding most impurities produced by *E. coli* cells. In order to simplify purification of the recombinant enzyme, we engineered the protein to be secreted into the culture medium of the transformed *E. coli* cells by attaching a signal peptide to *Pa*-rCOD. Several reports dealing with deacetylases describe hydrolysis of the acetamide bond on the non-reducing end GlcNAc residue of $(GlcNAc)_2$ to produce β-D-glucosaminyl-(1,4)-*N*-acetyl-D-glucosamine (GlcN-GlcNAc) (John et al. 1993; Tanaka et al. 2004; Tokuyasu et al. 1997). However, aside from *Pa*-COD, only two deacetylases (DA1 and DA2 from *V. alginolyticus* H-8) are known to produce GlcNAc-GlcN from $(GlcNAc)_2$ (Ohishi et al. 1997). Although the DA1 gene has been cloned and sequenced (Ohishi et al. 2000), overproduction of the product using an expression vector has not yet been attempted.

Here we report the results of cloning the open reading frame (ORF) construct of the *Pa*-COD gene and a signal sequence, the overproduction and secretion of *Pa*-rCOD by transformed *E. coli* cells, and the purification of *Pa*-rCOD from the *E. coli* culture medium.

Materials and methods

Chemicals

GlcNAc, D-glucosamine (GlcN) and chitin oligosaccharides $[(GlcNAc)_{2-6}]$ were purchased from Seikagaku Kogyo. Artificial seawater was prepared using the Sealife (Marine Tech) salt mixture. All other chemicals were analytical grade.

Gene cloning

Vibrio parahaemolyticus KN1699 was grown at 28°C with shaking (135 rpm) in 10 ml of half-strength artificial seawater containing 1% (w/v) peptone and 0.1% (w/v) yeast extract. After cultivating for 15 h, the cells were harvested by centrifugation at 6,000*g* for 10 min, then chromosomal DNA was isolated from the cells using an Isoplant II kit (Nippon Gene). The ORF containing the *Pa*-COD gene and signal sequence was amplified by PCR (Ex *Taq* DNA polymerase, 1× Ex *Taq* buffer, dNTP mixture; Takara), using 10 ng of the genomic DNA and 10 pmol synthetic oligonucleotide primers, 5′-GCG CAACCGGTTGCGCTATTCGTGAACAG-3′ (forward primer) and 5′-CCAGTTGGATGCAAGCCTT GTCACCTCAC-3′ (reverse primer). After 30 amplification cycles (denaturation at 94°C for 1 min; annealing at 60°C for 1 min; elongation at 72°C for 1.5 min), the mixture was incubated at 72°C for 20 min. PCR products were isolated from agarose gel slices using the Geneclean II kit (Q•Biogene) and subcloned into pGEM-T Easy vector (Promega) by TA-cloning, yielding the plasmid pVP-COD1. *E. coli* DH5α was transformed with this plasmid and the transformants were selected by blue-white selection on LB medium-agar plates supplemented with 50 μg ampicillin/ml, 50 μl 100 mM IPTG, and 100 μl 20 mg X-Gal/ml. After cultivating the transformants for 16 h at 37°C in LB medium supplemented with 50 μg ampicillin/ml, the cells were harvested by centrifugation (10,000*g*, 5 min, 4°C). The pVP-COD1 plasmid was isolated from the *E. coli* cells using QIAprep Spin Miniprep Kit (Qiagen) and was used

for nucleotide sequence analysis of the *Pa*-COD gene.

Nucleotide sequence analysis was performed by the dideoxynucleotide method (Sanger et al. 1977). The nucleotide sequence of the gene was determined in both orientations using a ThermoSequenase Fluorescence-labeled primer cycle sequencing kit (GE Healthcare Bio-science) and an automated DNA sequencer (DSQ-2000L; Shimadzu).

Construction of plasmid for enzyme expression

Amplification of the ORF containing the ribosome-binding sequence, signal sequence, and the *Pa*-COD gene was performed by PCR (Ex *Taq* DNA polymerase, 1 x Ex *Taq* buffer, dNTP mixture; Takara) using 18 ng pVP-COD1 as a template and 10 pmol synthetic oligonucleotide primers, 5′-*CCCAAGCTT* AGGAATCGAAAATGAAATTAAATAAACTG-3′ (forward primer) and 5′-*CCGCTCGAG*TTATAGAG GTGTGAACAAGG-3′ (reverse primer) (underlined letters designate the *Hin*dIII site of the forward primer and the *Xho*I site of the reverse primer. Italic letters designate added non-complimentary nucleotides). After 30 amplification cycles (denaturation at 94°C for 1 min; annealing at 60°C for 1 min; elongation at 72°C for 1.5 min), the mixture was incubated at 72°C for 20 min. PCR products were isolated from agarose gel slices using the Geneclean II kit, were digested with *Hin*dIII (Toyobo) and *Xho*I (Nippon gene), then the resulting DNA fragment was ligated into the *Hin*dIII and *Xho*I sites of pET-21(+) (Novagen) to yield the pVP-COD2 plasmid. *E. coli* BL21(DE3) was transformed with this plasmid and the transformants were selected on LB medium-agar plates supplemented with 50 μg/ml ampicillin.

Production and purification of recombinant enzyme

Escherichia coli BL21(DE3) harboring the pVP-COD2 plasmid was grown at 37°C in 10 ml LB medium supplemented with 50 μg ampicillin/ml until an OD_{600} of 0.5. The culture was then diluted into 200 ml fresh medium and cultivated with shaking (135 rpm) at 37°C until an OD_{600} of 0.6. IPTG was added to 1 mM, and the culture was incubated with shaking for an additional 24 h. The *E. coli* cells were removed from the culture broth by centrifugation (10,000*g*, 10 min, 4°C). Proteins in the culture supernatant were precipitated by adding $(NH_4)_2SO_4$ (80% saturation) and were collected by centrifugation (10,000*g*, 15 min, 4°C). The resulting precipitate was dissolved in 20 ml 20 mM sodium phosphate buffer (pH 7.0), and the solution was dialyzed against the same buffer to afford crude enzyme solution. This enzyme solution was loaded on a DEAE-Toyopearl 650 M (Tosoh) column (30 cm × 1.6 cm) pre-equilibrated with the same buffer, and enzyme was eluted with a linear gradient of 0–0.4 M NaCl in the same buffer (total volume: 400 ml). The active fractions were collected, dialyzed against 20 mM sodium phosphate buffer (pH 7.0), and concentrated to 10 ml by diaflow filtration using an Amicon PM-10 membrane.

Homogeneity of the purified enzyme was confirmed by SDS-PAGE. Proteins in the polyacrylamide-gel were stained by Coomassie Brilliant Blue R250 (Tokyo Kasei). Protein concentrations in the enzyme solutions were determined by Lowry's method using bovine serum albumin (Sigma) as a standard. The *N*-terminal amino acid sequences of the purified enzyme were determined using a Perkin Elmer Biosystems model Procise 49× HT protein sequencer.

Assay of enzyme activity

The assay mixture consisted of 75 μl enzyme solution and 425 μl 1 mM $(GlcNAc)_2$ in 20 mM sodium phosphate buffer (pH 7.0). The assay was conducted at 37°C. The enzymatic reaction was stopped by heating the reaction mixture at 100°C in a dry bath for 5 min, and the amount of GlcN residue produced by the enzymatic reaction was determined by the method of Dische and Borenfreund (1950) using GlcN as a standard. One unit of COD activity was defined as the amount of enzyme required to produce 1 μmol GlcN residue per minute under the assay conditions.

Oligosaccharide analysis

The oligosaccharide produced by the enzymatic reaction was purified as previously described (Kadokura et al. 2007). The structure of the oligosaccharide was characterized by ^1H NMR spectrometry using D_2O as a solvent, and by mass

spectrometry (MS). ^{1}H NMR spectra were recorded with a Varian Mercury 400 spectrometer at 20°C. Mass spectra were obtained with a Waters Micro-Mass ZQ instrument under positive or negative ion electron-spray ionization (ESI) conditions.

Results and discussion

Cloning and sequencing of the *Pa*-COD gene

The *N*-terminal amino acid sequence (26 residues) of *Pa*-COD showed 100% correspondence with that estimated from the nucleotide sequence of the *V. parahaemolyticus* RIMD2210633 putative COD gene (GeneBank accession number: BA000031) (Makino et al. 2003). We believed that nucleotide sequences in the vicinity of COD genes in *V. parahaemolyticus* KN1699 and RIMD2210633 genomic DNA are similar. Therefore, PCR forward and reverse primers for the sub-cloning of the *Pa*-COD gene were designed from upstream and downstream nucleotide sequences of the putative COD gene in genomic DNA of strain RIMD2210633. The PCR product was sub-cloned into the vector by TA-cloning and its nucleotide sequence was analyzed (Fig. 1). The ORF (DDBJ accession number: AB275387) consisting of 1281 bp was found in the PCR product and encoded 427 amino acid residues. A putative ribosome-binding sequence, AGGA, was found upstream, 7 bp from the start codon (ATG). The putative −35 and −10 regions were TTACTA and TTAATTT, respectively. These sequences are the same as those in the corresponding region of the *V. alginolyticus* H-8 genome (Ohishi et al. 2000).

N-Terminal sequence analysis of *Pa*-COD (Kadokura et al. 2007) showed that the signal sequence corresponds to the 22 *N*-terminal amino acids. The recombinant enzyme consists of 405 amino acids, and its molecular weight was calculated to be 44,715 Da. This value is in good agreement with that obtained by SDS-PAGE analysis of *Pa*-COD (Kadokura et al. 2007). The deduced amino acid sequence of *Pa*-rCOD showed high homology with those of CE family 4 CODs from *Vibrionaceae* bacteria. Three amino acids (290, 394 and 396) were different between *Pa*-rCOD and the putative COD (405 amino acids) from *V. parahaemolyticus* RIMD2210633. Four amino acids (173, 290, 394 and 396) were different between *Pa*-rCOD and COD from *V. alginolyticus* H-8 (DA1, 405 amino acids; GeneBank accession number: AJ292005) (Ohishi et al. 2000). Moreover, *Pa*-rCOD showed 83% identity with the putative deacetylase from *V. vulnificus* CMC-6 and *V. vulnificus* YJ016, and 82% identity with the putative deacetylase from *V. cholerae* El Tor N16961.

Production and purification of *Pa*-rCOD

The practical use of *Pa*-rCOD will require simple procedures to purify the enzyme from crude enzyme preparations. Therefore, to avoid intermixing large amounts of other proteins derived from transformed *E. coli* cells, we wanted to engineer a version of the recombinant enzyme that would be secreted into the *E. coli* culture medium. This was accomplished with the aid of a signal peptide. The ORF containing the ribosome-binding sequence, the signal sequence and the *Pa*-COD gene was amplified by PCR and inserted into the pET-21(+) vector to make expression plasmid pVP-COD2. *E. coli* cells harboring this plasmid were cultivated and the production of *Pa*-rCOD was induced by the addition of IPTG. Figure 2 shows the time course of the COD activities in both *E. coli* cell lysates and the culture medium. Addition of IPTG initiated production of recombinant COD in the *E. coli* cells, while its secretion into the culture medium did not begin for at least 3 h. After 12 h cultivation, total COD activity in the culture medium peaked and was at the same level as that in *E. coli* lysates. SDS-PAGE results (Fig. 3) show the time course of recombinant protein production in both *E. coli* cells (Fig. 3A) and in the culture medium (Fig. 3B). The protein, with a molecular mass of about 45 kDa, was confirmed to be the main product in the *E. coli* culture medium. Other proteins secreted by the *E. coli* cells were present in very low concentrations, indicating that the conditions used in the present investigation are favorable for the purification of the recombinant enzyme. The amount of COD activity in the *E. coli* culture medium was 150 times larger than in the culture medium of *V. parahaemolyticus* KN1699.

Pa-rCOD was purified from the culture supernatant (obtained by centrifuging 200 ml culture broth of *E. coli* cells harboring pVP-COD2) by $(NH_4)_2SO_4$ precipitation and ion-exchange column

Fig. 1 Nucleotide sequence of the *Pa*-COD gene and deduced amino acid sequence. Putative promoter elements are underlined twice. The ribosome-binding sequence is enclosed by a square. The signal sequence is underlined

```
-180 CGCAACCGGTTGCCGTATTCGTGAACAGAATCTCACTTTAGAAAATAAGGTCTATTTTTTAATAATTATTTAGGTCTAGGATAAAAGTCA -91
 -90 GAGAATTAAGGAAGATTACGAATTTACATTTACAACTATCAATATATTACTACTGACAACGTCATTATTAATTTCGAAAAGGAATCGAAA  -1

   1 ATGAAATTAAATAAACTGGCTATTGCAACGCTTGTAAGTGCCGCTCTATCTCAATATGCATTTGCCCAAACTGACACCAAAGGAACCATT  90
       M  K  L  N  K  L  A  I  A  T  L  V  S  A  A  L  S  Q  Y  A  F  A  Q  T  D  T  K  G  T  I

  91 TATCTGACGTTTGATGATGGCCCAATCAACGCCTCAATTGACGTCATTAATGTGCTAAATCAAGAAGAAGTAAAAGCGACGTTTTACTTT 180
       Y  L  T  F  D  D  G  P  I  N  A  S  I  D  V  I  N  V  L  N  Q  E  E  V  K  A  T  F  Y  F

 181 AATGCGTGGCACCTAGATGGTATTGGTGATGAAAACGAAGACAGAGCGTTAGAGGCACTAAAACTGGCGTTGGATAGCGGCCACATCGTC 270
       N  A  W  H  L  D  G  I  G  D  E  N  E  D  R  A  L  E  A  L  K  L  A  L  D  S  G  H  I  V

 271 GCAAACCACAGTTATGACCATATGGTTCACAACTGTGTTGAAGAATTTGGCCCAAACAGTGCGGCAGAATGTAATGCAACGGGTGATCAC 360
       A  N  H  S  Y  D  H  M  V  H  N  C  V  E  E  F  G  P  N  S  A  A  E  C  N  A  T  G  D  H

 361 CAGATCAACTCTTATCAAGATCCTGCCTACGATGCCTCGATGTTTGCAGAAAACTTGTCAGTACTAGAAAAATATTTGCCGAACATTACG 450
       Q  I  N  S  Y  Q  D  P  A  Y  D  A  S  M  F  A  E  N  L  S  V  L  E  K  Y  L  P  N  I  T

 451 AGCTACCCAAACTATAAAGCGAATGAGTTTGCTCGTTTGCCGTATACCAATGGTTGGCGCGTTACGAAAGACTTCAAAGCGGACGGCTTA 540
       S  Y  P  N  Y  K  A  N  E  F  A  R  L  P  Y  T  N  G  W  R  V  T  K  D  F  K  A  D  G  L

 541 TGTGCCACGTCCGATGATCTTAAACCTTGGGAGCCTGGCTATGCATGTGATACGGCAAATCCATCCAACAGTGTAAAAGCAGCCATCGCG 630
       C  A  T  S  D  D  L  K  P  W  E  P  G  Y  A  C  D  T  A  N  P  S  N  S  V  K  A  A  I  A

 631 GTTCAAAATATCCTAGCGAATAATGGCTATCAAACTCACGGTTGGGATGTGGATTGGGCCCCTGAAAACTGGGGCATTGCGATGCCAGCA 720
       V  Q  N  I  L  A  N  N  G  Y  Q  T  H  G  W  D  V  D  W  A  P  E  N  W  G  I  A  M  P  A

 721 AACAGCCTTACAGAAGCTGAACCATTCCTTGGGTATGTCGATTCCGCGCTAAATACTTGTGCGCCTACGACAATTAACCCTATCAACTCC 810
       N  S  L  T  E  A  E  P  F  L  G  Y  V  D  S  A  L  N  T  C  A  P  T  T  I  N  P  I  N  S

 811 AAAGCACAAGAGTTCCCATGTGGTACACCTTTGCATGCCGATAAAGTTATTGTACTAACTCACGAGTTCCTGTTTGAAGACGGTAAACGT 900
       K  A  Q  E  F  P  C  G  T  P  L  H  A  D  K  V  I  V  L  T  H  E  F  L  F  E  D  G  K  R

 901 GGCATGGGTGCAACTCAAAACCTACCGAAGCTCACGAAGTTTATCCAGCTAGCCAAACAGGCTGGTTATGTCTTCGATACCATGGATAAC 990
       G  M  G  A  T  Q  N  L  P  K  L  T  K  F  I  Q  L  A  K  Q  A  G  Y  V  F  D  T  M  D  N

 991 TACACACCGAACTGGCAAGTTGGTAACAACTACAGCGCCGGCGACTACGTTCTACACCTCGGTACGGTATATCAAGCAGTAACAAGCCAT 1080
       Y  T  P  N  W  Q  V  G  N  N  Y  S  A  G  D  Y  V  L  H  L  G  T  V  Y  Q  A  V  T  S  H

1081 ACAGCGCAACAAGATTGGGCTCCATCACCAACATCTAGCTTGTGGACGAATGCGGATCCAGCTACTAATTGGACTCAGAATGTTTCATAC 1170
       T  A  Q  Q  D  W  A  P  S  P  T  S  S  L  W  T  N  A  D  P  A  T  N  W  T  Q  N  V  S  Y

1171 AAACAAGGCGATGTGGTGACTTATCAAGGTTTGCGTTACCTAGTTAATGTACCGCACGTATCTCAAGCAGACTGGAGTCCGAGCTCACAG 1260
       K  Q  G  D  V  V  T  Y  Q  G  L  R  Y  L  V  N  V  P  H  V  S  Q  A  D  W  S  P  S  S  Q

1261 AACACCTTGTTCACAGCTCTATAAGCTGTTCCCTACTATCATCATGTAGTAGTTTGCGCGAGGCATACCCTCGCGCTTTTTTCTTTTGAA 1350
       N  T  L  F  T  A  L  *

1351 ACGCGGCAAATCGGTCTTCAACGCGAATCTCCAAACACACGTAAATCTAAGCACATAAACACCCCAAACGACGAAAGGCCTCTATAAAAG 1440
1441 CGACCTTTCTATGTATGGTACTGATAAGTGAGGTGACAAGGCTTGCATCCAACTGG                                    1496
```

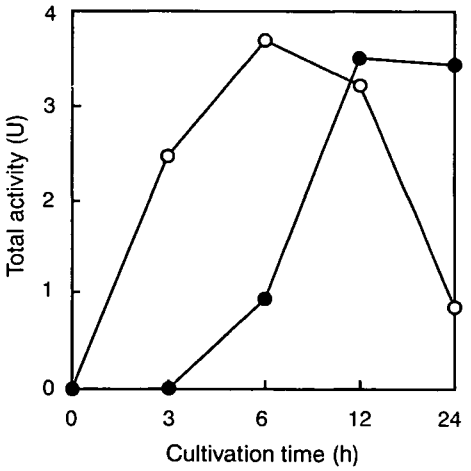

Fig. 2 Time course of *Pa*-rCOD production after addition of IPTG to the transformed *E. coli* culture. Aliquots (7 ml) of the culture broth were withdrawn at the time points indicated and centrifuged (10,000*g*, 10 min) to harvest the cells and obtain the culture supernatant. Pelleted *E. coli* cells were suspended in 7 ml of phosphate buffered saline, then cell membranes were disrupted by sonication (30 s × 6, 4°C) to produce the cell lysate. The culture supernatant was dialyzed against 20 mM sodium phosphate buffer (pH 7.0) for 16 h at 4°C. The COD activities in the cell lysate and the dialyzed culture supernatant were assayed according to the method described in the text. ○, Total COD activity in the *E. coli* cell lysate; ●, total COD activity in dialyzed culture supernatant

chromatography using DEAE-Toyopearl 650 M resin (Table 1). Purified *Pa*-rCOD gave a single band of molecular mass 45 kDa on SDS-PAGE, the same as a purified sample of wild-type enzyme, *Pa*-COD (Fig. 4). *Pa*-rCOD was purified 150-fold with a total recovery of 16% using this procedure.

The *N*-terminal amino acid sequence of *Pa*-rCOD was determined to be QTDTKGTIYLTFDDGPINA-SIDV, showing that the signal peptide region of the expression product was removed as the protein crossed the *E. coli* cell membrane.

Substrate specificity of *Pa*-rCOD

Pa-rCOD showed hydrolysis activity towards two chitin oligosaccharides: $(GlcNAc)_2$, and lower activity toward $(GlcNAc)_3$ (Table 2). The specific activities of *Pa*-rCOD against these oligosaccharides were comparable with those of *Pa*-COD. ^{1}H NMR spectra and ESIMS spectra of the disaccharide produced by treating $(GlcNAc)_2$ with *Pa*-rCOD correspond to those previously obtained with GlcNAc-GlcN (Kadokura et al. 2007), indicating that *Pa*-rCOD hydrolyzes the acetamide bond on the reducing-end GlcNAc residue of $(GlcNAc)_2$. These results indicate

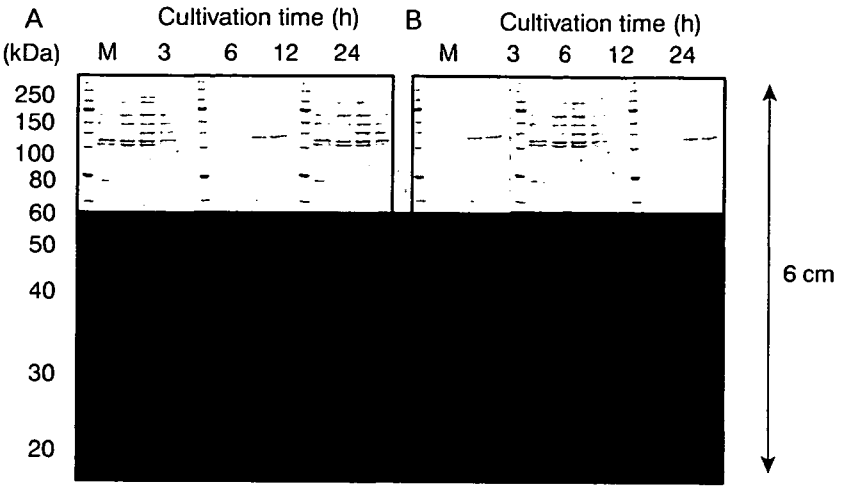

Fig. 3 SDS-PAGE results of lysate and culture supernatant (prepared as described in Fig. 2) of transformed *E. coli* cells. (**A**) cell lysate. (**B**) culture supernatant. Lane M shows molecular mass standards (Promega)

Table 1 Purification of *Pa*-rCOD from culture fluid of *E. coli*

Purification step	Total activity (U)	Specific activity (U/mg of protein)	Yield (%)	Fold
Culture fluid[a]	58.5	0.053	100	1
$(NH_4)_2SO_4$ precipitation	33.4	0.71	57.0	14
DEAE-Toyopearl 650 M	9.3	25.2	16.0	478

[a] Recombinant COD was purified from 200 ml of culture fluid

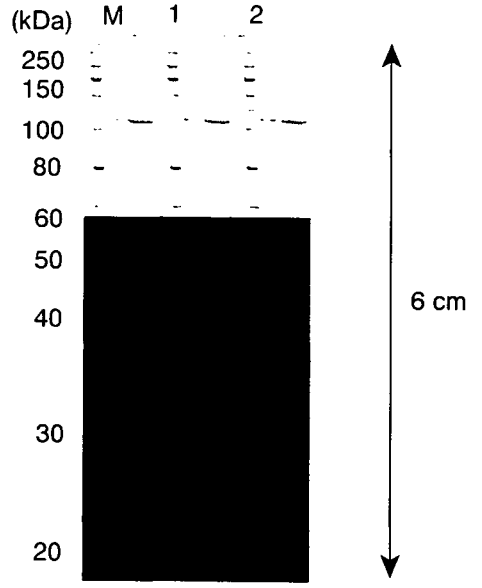

Fig. 4 SDS-PAGE results of purified enzymes. Lane 1 and 2 contain purified *Pa*-COD and *Pa*-rCOD, respectively. Lane M shows molecular mass standards (Promega)

that the substrate specificity of *Pa*-rCOD is the same as that of the wild-type enzyme, *Pa*-COD.

In conclusion, recombinant COD from *V. parahaemolyticus* KN1699 has been over-produced in the culture medium of transformed *E. coli* cells, purified, and shown to have substrate specificities comparable to the wild-type enzyme. In this investigation we succeeded in making the transformed *E. coli* cells produce 150 times more recombinant COD than *V. parahaemolyticus* KN1699 cells can produce wild-type COD. Moreover, by exporting the recombinant COD to the culture medium of the transformed *E. coli* cells with the aid of a signal peptide, the protein could be purified easily. CODs from the soil bacterium *Rhizobium melitoti* (John et al. 1993) and the archaeon *Thermococcus kodakaraensis* KOD1 (Tanaka et al. 2003) have been cloned and over-produced in transformed *E. coli* cells. It has been confirmed that chitin deacetylase from the mold *Colletorichum lindemuthianum* can be produced in the culture medium of transformed *E. coli* cells with the aid of a chitinase signal peptide from *Streptomyces lividans* (Tokuyasu et al. 1999). These recombinant enzymes catalyze the hydrolysis of the acetamide bond on the non-reducing GlcNAc residue of $(GlcNAc)_2$ to produce GlcN-GlcNAc. The present

Table 2 Substrate specificity of *Pa*-rCOD

Substrate	Specific activity[a] (U/mg of protein)
GlcNAc	n.d.
$(GlcNAc)_2$	25
$(GlcNAc)_3$	7.5
$(GlcNAc)_4$	n.d.
$(GlcNAc)_5$	n.d.
$(GlcNAc)_6$	n.d.

[a] Assay was performed according to the method described in the text. n.d., no detectable activity

report is the first to describe the overproduction of a recombinant COD that produces GlcNAc-GlcN from $(GlcNAc)_2$.

Pa-Chi, which we isolated from the culture supernatant of *V. parahaemolyticus* KN1699, is useful for the production of $(GlcNAc)_2$ from α- or β-chitin (Kadokura et al. 2007). Using *Pa*-Chi and *Pa*-rCOD, future studies will attempt the gram-scale production of GlcNAc-GlcN, and will investigate its physiological functions.

Acknowledgements This work was supported by Grants from Nihon University and the 21st Century Center of Excellence (COE) Program of the Ministry of Education, Science, Sports, and Culture (Japan) to promote advanced scientific research.

References

Dische Z, Borenfreund E (1950) A spectrophotometric method for the microdetermination of hexosamines. J Biol Chem 184:517–522

John M, Rohrig H, Schmidt J, Wieneke U, Schell J (1993) Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. Proc Natl Acad Sci USA 90:625–629

Kadokura K, Rokutani A, Yamamoto M, Ikegami T, Sugita H, Itoi S, Hakamata W, Oku T, Nishio T (2007) Purification and characterization of *Vibrio parahaemolyticus* extracellular chitinase and chitin oligosaccharide deacetylase involved in the production of heterodisaccharide from chitin. Appl Microbiol Biotechnol (in press)

Kobayashi M, Watanabe T, Suzuki S, Suzuki M (1990) Effect of *N*-acetylchitohexaose against *Candida albicans* infection of tumor-bearing mice. Microbiol Immunol 34:413–426

Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, Tagomori K, Iijima Y, Najima M, Nakano M, Yamashita A, Kubota Y, Kimura S, Yasunaga T, Honda T, Shinagawa H, Hattori M, Iida T (2003) Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. Lancet 361:743–749, DOI:10.1016/S0140-6736(03)12659-1

Ohishi K, Yamagishi M, Ohta T, Motosugi M, Izumida H, Sano H, Adachi K, Miwa T (1997) Purification and properties of two deacetylases produced by *Vibrio algionolyticus* H-8. Biosci Biotechnol Biochem 61:1113–1117

Ohishi K, Murase K, Ohta T, Etoh H (2000) Cloning and sequencing of the deacetylase gene from *Vibrio alginolyticus* H-8. J Biosci Bioeng 90:561–563, DOI:10.1016/S1389-1723(01)80041-4

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Suzuki K, Mikami T, Okawa Y, Tokoro A, Suzuki S, Suzuki M (1986) Antitumor effect of hexa-*N*-acetylchitohexaose and chitohexaose. Carbohydr Res 151:403–408, DOI:10.1016/S0008-6215(00)90359-8

Tanaka T, Fukui T, Fujiwara S, Atomi H, Imanaka T (2004) Concerted action of diacetylchitobiose deacetylase and exo-β-D-glucosaminidase in a novel chitinolytic pathway in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J Biol Chem 279:30021–30027, DOI:10,1074/jbc.M314187200

Tokoro A, Tatewaki N, Suzuki K, Mikami T, Suzuki S, Suzuki M (1988) Growth-inhibitory effect of hexa-*N*-acetylchitohexaose and chitohexaose against Meth-A solid tumor. Chem Pharm Bull 36:784–790

Tokoro A, Kobayashi M, Tatewaki N, Suzuki K, Okawa Y, Mikami T, Suzuki S, Suzuki M (1989) Protective effect of *N*-acetyl chitohexaose on *Listeria monocytogenes* infection in mice. Microbiol Immunol 33:357–367

Tokuyasu K, Ono H, Ohnishi-Kameyama M, Hayashi K, Mori Y (1997) Deacetylation of chitin oligosaccharides of dp 2-4 by chitin deacetylase from *Colletotrichum lindemuthianum*. Carbohydr Res 303:353–383, DOI:10.1016/S0008-6215(97)00166-3

Tokuyasu K, Kaneko S, Hayashi K, Mori Y (1999) Production of a recombinant chitin deacetylase in the culture medium of *Escherichia coli* cells. FEBS Lett 458:23–26, DOI:10.1016/S0014-5793(99)01113-8

Tsukada K, Matsumoto T, Aizawa K, Tokoro A, Naruse R, Suzuki S, Suzuki M (1990) Antimetastatic and growth-inhibitory effects of *N*-acetylchitohexaose in mice bearing Lewis lung carcinoma. Jpn J Cancer Res 81:259–265

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 38, pp. 28246–28255, September 21, 2007

Structural Basis for Recognition of High Mannose Type Glycoproteins by Mammalian Transport Lectin VIP36*[S]

Received for publication, April 11, 2007, and in revised form, June 25, 2007 Published, JBC Papers in Press, July 25, 2007, DOI 10.1074/jbc.M703064200

Tadashi Satoh[‡], Nathan P. Cowieson[§1], Wataru Hakamata[¶], Hiroko Ideo[‖], Keiko Fukushima[‖**], Masaaki Kurihara[¶], Ryuichi Kato[‡], Katsuko Yamashita[‖**], and Soichi Wakatsuki[‡2]**

*From the [‡]Structural Biology Research Center, Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan, [§]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia, [¶]Division of Organic Chemistry, National Institute of Health Sciences (NIHS), Tokyo 158-8501, Japan, [‖]Innovative Research Initiatives, Tokyo Institute of Technology, Yokohama 226-8503, Japan, and **Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Tokyo 101-0062, Japan*

VIP36 functions as a transport lectin for trafficking certain high mannose type glycoproteins in the secretory pathway. Here we report the crystal structure of VIP36 exoplasmic/luminal domain comprising a carbohydrate recognition domain and a stalk domain. The structures of VIP36 in complex with Ca^{2+} and mannosyl ligands are also described. The carbohydrate recognition domain is composed of a 17-stranded antiparallel β-sandwich and binds one Ca^{2+} adjoining the carbohydrate-binding site. The structure reveals that a coordinated Ca^{2+} ion orients the side chains of Asp^{131}, Asn^{166}, and His^{190} for carbohydrate binding. This result explains the Ca^{2+}-dependent carbohydrate binding of this protein. The Man-α-1,2-Man-α-1,2-Man, which corresponds to the D1 arm of high mannose type glycan, is recognized by eight residues through extensive hydrogen bonds. The complex structures reveal the structural basis for high mannose type glycoprotein recognition by VIP36 in a Ca^{2+}-dependent and D1 arm-specific manner.

In eukaryotic cells, post-translational modification of secreted proteins and intracellular protein transport between organelles are ubiquitous features. One of the most studied systems is the *N*-linked glycosylation pathway in the synthesis of secreted glycoproteins (1–3). The *N*-linked glycoproteins are subjected to diverse modifications and are transported through the endoplasmic reticulum (ER)[3] via the Golgi apparatus to their final destinations inside and outside of the cell. Incorporation of the cargo glycoproteins into the transport vesicles is mediated by transmembrane cargo receptors, which have been identified as intracellular lectins. For example, mannose 6-phosphate receptor (4) functions as a cargo receptor for lysosomal proteins in the *trans*-Golgi network, whereas ER-Golgi intermediate compartment (ERGIC)-53 (5, 6) and its yeast orthologs Emp46/47p (7) are transport lectins for glycoproteins that are transported out of the ER.

VIP36, vesicular-integral protein of 36 kDa, was originally isolated from Madin-Darby canine kidney cells as a component of detergent-insoluble, glycolipid-enriched complexes containing apical marker (8). Confocal and immunoelectron microscopic experiments have suggested that VIP36 is distributed by either the pre-Golgi secretory pathway (9–11) or post-Golgi pathway (8, 12). Furthermore we showed that VIP36 is involved in intracellular transport, in the secretion of glycoproteins (*e.g.* clusterin) from polarized Madin-Darby canine kidney cells (13), and in the secretion of α-amylase from rat parotid glands (14). Taken together, VIP36 appears to play significant roles not only in vesicular transport from the ER to the Golgi complex but also from the *trans*-Golgi network to the plasma membrane.

The exoplasmic/luminal domain of VIP36 as well as the luminal domain of ERGIC-53 and Emp46/47p share homology with L (leguminous)-type lectins and are thus identified as carbohydrate recognition domains (CRDs). It has been shown that ERGIC-53 interacts with glycoproteins carrying high mannose type glycan by endo-β-*N*-acetylglucosaminidase H treatment (15–17) and binds glycoproteins in a Ca^{2+}- and pH-dependent manner (18). We have previously found that VIP36 has high affinity for high mannose type glycans containing Man-α-1,2-Man residues in $Man_{7-9}(GlcNAc)_2$-Asn peptides (19). Kamiya *et al.* (20) recently reported in detail the carbohydrate binding properties of VIP36 assayed by frontal affinity chromatography. This work suggested the Ca^{2+} dependence of carbohydrate binding and the specificity for the D1 arm, Man-α-1,2-Man-α-1,2-Man residues, of high mannose type glycans (corresponding to Man(D1)-Man(C)-Man(4); Fig. 1). In addition, using a flow cytometry-based method, it was also demonstrated that

* This work was supported in part by the Protein 3000 project, by Grant-in-aid for Young Scientists (B) 17790097 from The Ministry of Education, Culture, Sports, Science and Technology of Japan, and by research grants for research on human immunodeficiency virus/AIDS from The Ministry of Health and Labor Sciences of Japan. 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.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1–4.

The atomic coordinates and structure factors (code 2DUO, 2DUP, 2DUQ, 2DUR, and 2E6V) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] Supported by a fellowship from the Australian Synchrotron Research Program.

[2] To whom correspondence should be addressed. Tel.: 81-29-864-5631; Fax: 81-29-879-6179; E-mail: soichi.wakatsuki@kek.jp.

[3] The abbreviations used are: ER, endoplasmic reticulum; ERGIC, ER-Golgi intermediate compartment; CRD, carbohydrate recognition domain; GST, glutathione *S*-transferase; SPR, surface plasmon resonance; MES, 4-morpholineethanesulfonic acid; ConA, concanavalin A.

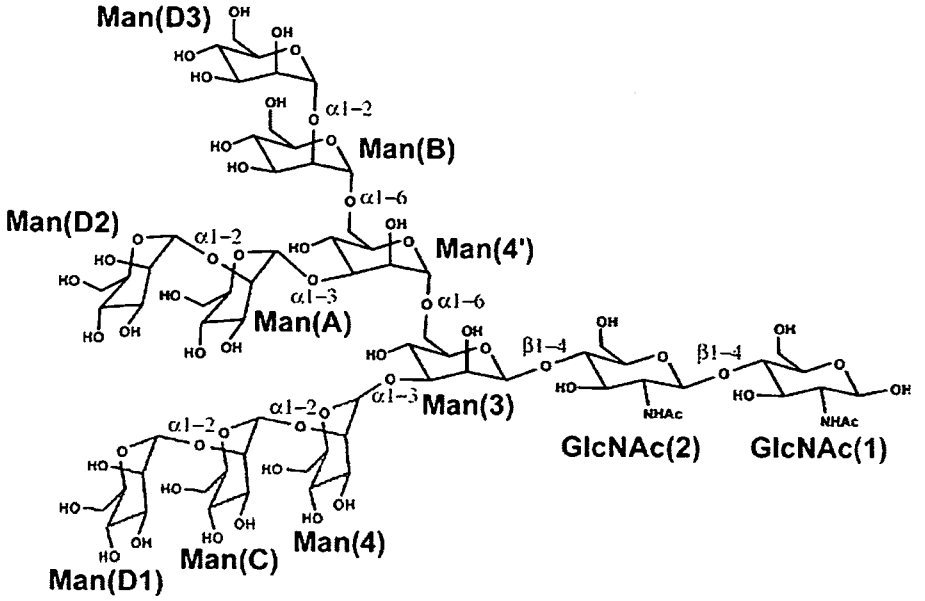

FIGURE 1. **Chemical structures of $Man_9(GlcNAc)_2$.** The individual carbohydrate residues of $Man_9(GlcNAc)_2$ are labeled. The D1 arm of $Man_9(GlcNAc)_2$ is colored in *green*.

VIP36 binds glycoproteins carrying high mannose type glycans (21). These observations suggested that VIP36 is involved in the transport of glycoproteins via high mannose type glycans.

Crystal structures of the CRD of rat ERGIC-53 in the absence and presence of Ca^{2+} have been determined, confirming its structural similarity to the L-type lectins (22, 23). In these reports, it was shown that the putative ligand-binding site of ERGIC-53 is similar to the mannose-binding site of the L-type lectins. Very recently, we reported the crystal structures of the CRD of Ca^{2+}-independent K^+-bound Emp46p and the metal-free form of Emp47p (24). To date, however, no structures of transport lectins in complex with high mannose type glycans have been determined. To investigate the structural basis of the mechanism of high mannose type glycoprotein recognition by VIP36, we determined crystal structures of the exoplasmic/luminal domain of VIP36 alone and in complex with Ca^{2+} and mannose, Man-α-1,2-Man (termed Man_2, which corresponds to Man(D1)-Man(C), Man(C)-Man(4), Man(D2)-Man(A), or Man(D3)-Man(B) of $Man_9(GlcNAc)_2$; Fig. 1), and Man-α-1,2-Man-α-1,3-Man-β-1,4-GlcNAc (termed $Man_3GlcNAc$, which corresponds to Man(C)-Man(4)-Man(3)-GlcNAc(2); Fig. 1).

EXPERIMENTAL PROCEDURES

Synthesis of Man-α-1,2-Man-α-1,2-Man, Man-α-1,2-Man-α-1,3-Man, and Man-α-1,2-Man-α-1,6-Man—Couplings of phenyl 3,4,6-tri-*O*-benzyl-α-D-thiomannopyranoside (i), phenyl 2,4,6-tri-*O*-benzyl-α-D-thiomannopyranoside (ii), and phenyl 2,3,4-tri-*O*-benzyl-α-D-thiomannopyranoside (iii) having hydroxyl groups at the C-2, C-3, and C-6 positions (25) and 1,2-di-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranose (iv) (26) were performed under conditions well established for α-mannosidation (trimethylsilyl trifluoromethanesulfonate/CH_2Cl_2) to give phenyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-thiomannopyranoside (v), phenyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→3)-2,4,6-tri-*O*-benzyl-α-D-thiomannopyranoside (vi), and phenyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→6)-2,3,4-tri-*O*-benzyl-α-D-thiomannopyranoside (vii) (27), respectively. Subsequent deacetylation of the mannobioses (v, vi, and vii) gave phenyl 3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-thiomannopyranoside (viii), phenyl 3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→3)-2,4,6-tri-*O*-benzyl-α-D-thiomannopyranoside (ix), and phenyl 3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→6)-2,3,4-tri-*O*-benzyl-α-D-thiomannopyranoside (x) (27), respectively. Introduction of the non-reducing end of the mannose residue to the mannobioses (viii, ix, and x) using 1-*O*-acetyl-2,3,4,6-tetra-*O*-benzyl-α-D-mannopyranose (xi) (28) was performed using the same α-mannosidation method to give phenyl 2,3,4,6-tetra-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-thiomannopyranoside (xii), phenyl 2,3,4,6-tetra-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→3)-2,4,6-tri-*O*-benzyl-α-D-thiomannopyranoside (xiii), and phenyl 2,3,4,6-tetra-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→6)-2,3,4-tri-*O*-benzyl-α-D-thiomannopyranoside (xiv), respectively. Finally complete deprotection of synthesized mannotriose derivatives (xii, xiii, and xiv) afforded α-D-mannopyranosyl-(1→2)-α-D-mannopyranosyl-(1→2)-α-D-mannopyranose, α-D-mannopyranosyl-(1→2)-α-D-mannopyranosyl-(1→3)-α-D-mannopyranose, and α-D-mannopyranosyl-(1→2)-α-D-mannopyranosyl-(1→6)-α-D-mannopyranose, respectively. These mannotrioses were isolated on a COSMOSIL Sugar-D column (Nacalai Tesque) using an isocratic solvent composed of 65% MeCN and 35% H_2O. NMR and MS spectra of these compounds were in good agreement with those reported for closely related compounds (29, 30).

Preparation of $Man_3GlcNAc$ and $Man_6(GlcNAc)_2$-Asn—$Man_3GlcNAc$ was prepared from urine of a mannosidosis patient as described previously (31). Briefly 10 ml of urine containing 10 mg of creatinine was concentrated to 1 ml and centrifuged for 20 min at 3,000 rpm. The supernatant was subjected to Bio-Gel P-4 (200–400 mesh) column chromatography (2.6 × 97 cm). The column was eluted with water containing 0.002% phenylmercuric nitrate, and the hexose content in each fraction was analyzed with phenol-sulfuric acid reagent. Fractions between $Man_2GlcNAc$ and $Man_4GlcNAc$ were pooled, sequentially subjected to Bio-Gel P-4 (under 400 mesh) column chromatography (2 × 100 cm) at 55 °C, and eluted with distilled water by monitoring with a refractometer. The fraction corresponding to $Man_3GlcNAc$ was collected, and the struc-

ture of $Man_3GlcNAc$ was identified as Man-α-1,2-Man-α-1,3-Man-β-1,4-GlcNAc by methylation analysis and sequential exoglycosidase digestion using Manα1→2-specific *Aspergillus saitoi* α-mannosidase, jack bean α-mannosidase, and snail β-mannosidase. The yield was 1100 nmol. $Man_6(GlcNAc)_2$-Asn was prepared from ovalbumin as described previously (32).

Protein Expression and Purification—The DNA fragment for residues 51–301, which correspond to the CRD and part of the stalk domain of canine VIP36, was cloned into the BamHI and EcoRI sites of pGEX4T-1 plasmid (GE Healthcare). The recombinant VIP36 was expressed in *Escherichia coli* BL21(DE3). Cells were harvested after induction with 0.1 mM isopropyl β-D-thiogalactoside (Wako) for 15 h at 20 °C and lysed by sonication in phosphate buffered saline buffer containing 2 mM $CaCl_2$. The cell lysate was loaded on a glutathione-Sepharose 4B column (GE Healthcare). The glutathione *S*-transferase (GST) fusion protein was eluted by glutathione (Wako) and cleaved by thrombin protease (GE Healthcare). The cleaved proteins were passed through a glutathione-Sepharose 4B column to remove GST protein and further purified by a benzamidine-Sepharose 4FF column (GE Healthcare) to remove the thrombin protease. The purified protein was dialyzed against 10 mM MES (pH 6.5) and 2 mM $CaCl_2$.

Crystallization and X-ray Data Collection—Initial crystallization conditions were screened using the Large Scale Protein Crystallization and Monitoring System (PXS) (33). The crystallization conditions of VIP36 in its Ca^{2+}-bound form were obtained in a buffer containing 18 mg ml^{-1} protein, 15% (w/v) polyethylene glycol 4000, 1.5 M NaCl, and 0.1 M MES (pH 6.5) with incubation at 277 K for 4 days. For the metal free-form, the Ca^{2+}-bound crystal was soaked with this buffer containing 10 mM EDTA to remove Ca^{2+}. The crystal of Ca^{2+}.Man-bound VIP36 was obtained by soaking the Ca^{2+}-bound crystal with the buffer containing 50 mM D-mannose (Sigma). The Ca^{2+}.Man_2-bound VIP36 was co-crystallized in a buffer containing 10 mg ml^{-1} protein, 3.4 mM 2α-mannobiose (Sigma), 5% (w/v) polyethylene glycol 4000, 0.3 M $MgCl_2$, and 0.1 M MES (pH 6.5) with incubation at 277 K for 1 week. On the other hand, the Ca^{2+}.$Man_3GlcNAc$-bound VIP36 was co-crystallized in a buffer containing 10 mg ml^{-1} protein, 3.4 mM $Man_3GlcNAc$, 10% (w/v) polyethylene glycol 4000, and 0.4 M imidazole malate (pH 6.0) with incubation at 277 K for 3 weeks. Despite the extensive co-crystallization with $Man_3GlcNAc$ the electron density map did not show any interaction with the GlcNAc moiety; only the Man_3 portion was recognized by VIP36. Data sets of the metal-free and Ca^{2+}.Man_2- and Ca^{2+}.$Man_3GlcNAc$-bound forms were collected under cryogenic conditions with crystals soaked with a cryoprotectant buffer containing 20% (v/v) glycerol. The Ca^{2+}-, Ca^{2+}.Man-bound crystals were soaked with a buffer containing 2.5 M LiCl instead of 1.5 M NaCl for data collection under cryogenic conditions. The diffraction data were processed using HKL2000 (34). The metal-free and Ca^{2+}-, Ca^{2+}.Man-, and Ca^{2+}.Man_2-bound crystals belong to space group *C*2 with two molecules per asymmetric unit. In contrast, the Ca^{2+}.$Man_3GlcNAc$-bound crystal belongs to space group $P2_12_12_1$ with five molecules per asymmetric unit. The crystallographic parameters of VIP36 are shown in Table 1.

Structure Determination and Refinement—The crystal structure of VIP36 was solved by the molecular replacement method using the program MOLREP (35) with the Ca^{2+}-bound ERGIC-53 (Protein Data Bank code 1R1Z) (23) as a search model. The refinement procedures were carried out with Crystallography and NMR System (CNS) (36) and REFMAC5 (37). Model fitting to the electron density maps was performed manually using Coot (38). The stereochemical quality of the final models was assessed by PROCHECK (39). Final refinement statistics are summarized in Table 1. Figures were prepared using GRASP (40), LIGPLOT, (41), and PyMOL (42).

Computer-aided Model Building—The model of VIP36·$Man_8(GlcNAc)_2$-Asn complex was built using coordinates of well ordered high mannose type glycans on glycoprotein crystal structures (human pancreatic α-amylase (Protein Data Bank code 1BSI), *Erythrina corallodendron* lectin (Protein Data Bank code 1LTE), and exo-(1,3)-β-glucanase (Protein Data Bank code 1H4P)) and mannosyl ligand-bound VIP36 structures. The corresponding glycan residues were superimposed on each other, and appropriate coordinates were used as follows: Asn-GlcNAc(1), Protein Data Bank code 1BSI; GlcNAc(2), Protein Data Bank code 1LTE; Man(3)-Man(4)-Man(C), $Man_3GlcNAc$-bound VIP36; Man(D1), Man_2-bound VIP36; Man(4')-(Man(A))-Man(B)-Man(D3), Protein Data Bank code 1H4P; VIP36, Man_2-bound form. Based on the above model and human salivary α-amylase (Protein Data Bank code 1SMD) structures, the complex model of VIP36·rat salivary α-amylase carrying $Man_8(GlcNAc)_2$ was built. The salivary α-amylase was docked onto the VIP36·$Man_8(GlcNAc)_2$-Asn complex model through superimposition with an *N*-glycosylation site (Asn^{461}) of the salivary α-amylase and the asparagine residue of the high mannose type glycan bound with VIP36.

Surface Plasmon Resonance (SPR) Analysis—SPR measurements were carried out at 25 °C using a Biacore 2000 (Biacore) equipped with a CM5 sensor chip. GST-VIP36 (residues 51–322) was purified by affinity chromatography using glutathione-Sepharose 4B and benzamidine-Sepharose 4FF columns. The purified protein was immobilized on the flow cell using the amine coupling method. Various concentrations of mannotrioses (Man-α-1,2-Man-α-1,2-Man, Man-α-1,2-Man-α-1,3-Man, and Man-α-1,2-Man-α-1,6-Man) and $Man_6(GlcNAc)_2$-Asn in sample buffer (50 mM HEPES (pH 6.0) and 1 mM $CaCl_2$) were injected over the flow cells at a flow rate of 20 μl/min using HBS-P buffer (10 mM HEPES (pH 7.4), 150 mM NaCl, and 0.005% surfactant P20) as the running buffer.

RESULTS

Crystallization and Overall Structure of Exoplasmic/Luminal Domain of VIP36—The exoplasmic/luminal domain (residues 51–301) of VIP36, corresponding to the CRD and part of the short stalk domain, was crystallized. Despite extensive crystallization screening, diffraction quality crystals of the CRD (residues 51–278) alone could not be produced. The crystal structure of the exoplasmic/luminal domain of Ca^{2+}-bound VIP36 was solved by the molecular replacement method using the Ca^{2+}-bound ERGIC-53 CRD (Protein Data Bank code

TABLE 1
Data collection and refinement statistics of VIP36

	Data set				
	VIP36	VIP36·Ca^{2+}	VIP36·Ca^{2+}·Man	VIP36·Ca^{2+}·Man_2	VIP36·Ca^{2+}·Man_3GlcNAc
Crystallographic data					
Space group	*C*2	*C*2	*C*2	*C*2	$P2_12_12_1$
Unit cell					
a/*b*/*c* (Å)	171.0/45.2/117.1	170.1/45.4/116.1	171.2/45.0/117.0	171.2/45.5/117.4	57.2/151.2/177.1
$\alpha/\beta/\gamma$ (°)	90.0/132.6/90.0	90.0/131.5/90.0	90.0/131.9/90.0	90.0/132.7/90.0	90.0/90.0/90.0
Data processing statistics					
Beam line	PF-AR NW12A	PF-AR NW12A	PF-BL5A	PF-BL5A	PF-BL5A
Wavelength (Å)	1.0000	1.0000	1.0000	1.0000	1.0000
Resolution (Å)[a]	50-2.10 (2.18-2.10)	50-1.80 (1.86-1.80)	50-1.80 (1.86-1.80)	50-1.65 (1.71-1.65)	50-2.50 (2.59-2.50)
Total reflections	128,308	228,854	191,856	283,509	350,729
Unique reflections	38,945	62,420	62,207	80,482	54,228
Completeness (%)[a]	93.7 (78.9)	97.8 (97.1)	96.1 (84.5)	98.4 (91.1)	99.9 (100.0)
R_{merge} (%)[a]	8.3 (31.4)	5.9 (38.5)	9.9 (28.2)	5.1 (28.6)	13.3 (37.6)
$I/\sigma(I)$[a]	13.4 (3.7)	14.6 (3.1)	9.1 (2.9)	14.4 (3.4)	8.8 (5.8)
Refinement statistics					
Resolution (Å)	20-2.10	20-1.80	20-1.80	20-1.65	20-2.50
R_{work}	22.5	20.5	20.6	19.9	22.1
R_{free}	27.8	24.1	24.5	22.8	27.9
r.m.s.[b] deviation from ideal values					
Bond length (Å)	0.012	0.012	0.011	0.011	0.013
Angle distance (Å)	1.34	1.33	1.33	1.32	1.41
Ramachandran plot (%)					
Most favored	87.5	88.1	89.0	88.5	86.6
Additionally allowed	11.8	11.4	10.5	10.8	13.3
Generously allowed	0.7	0.5	0.5	0.7	0.1
Disallowed	0	0	0	0	0
Number of molecules and atoms					
Protein atoms	3,913	3,957	3,974	4,071	9,617
Water molecules	194	421	429	418	66
Ca^{2+} ions	0.5	2	2	2	5
Cl^- ions	4	11	13	8	
Glycerol atoms	18			12	
Sugar atoms			24	46	80
Average *B*-values (Å^2)					
Protein atoms	32.3/46.6	21.2/28.1	24.4/29.5	22.9/32.0	24.0/28.9/23.8/26.5/36.7
Water molecules	36.5	29.9	32.9	33.4	20.6
Ca^{2+} ions	33.4	23.5	21.8	20.0	25.4
Cl^- ions	44.0	29.4	32.0	33.5	
Glycerol atoms	47.9			33.1	
Sugar atoms			23.0	27.6	39.5

[a] Values in parentheses are for the highest resolution shell.
[b] Root mean square.

1R1Z) (23) as a search model. The crystal belongs to space group *C*2 with two molecules (A and B) per asymmetric unit. Both VIP36 molecules are related by ~2-fold symmetry, forming a pseudo-homodimer. However, gel filtration data demonstrate that this protein is monomeric in solution (supplemental Fig. 1).

To obtain crystals of mannose-bound VIP36, the Ca^{2+}-bound crystals were soaked in a solution containing mannose in molar excess. The Man_2-bound form was obtained by co-crystallization. To obtain the metal-free form, the Ca^{2+}-bound crystal was soaked with buffer containing 10 mM EDTA. Crystallization of VIP36 in the absence of Ca^{2+} was not successful. Following treatment with EDTA the Ca^{2+} is completely absent from molecule B, whereas approximately half of the Ca^{2+} ions are removed from molecule A. In the *C*2 crystal form, crystallization of VIP36 in complex with longer oligomannoses was not successful due to the crystal packing around the ligand-binding site. To find other crystal forms, we therefore carried out crystallization screening in the presence of Ca^{2+} and longer oligomannoses, Man-α-1,2-Man-α-1,2-Man, Man-α-1,2-Man-α-1,3-Man, Man_3GlcNAc, and $Man_6(GlcNAc)_2$-Asn. Diffraction quality crystals were obtained from co-crystallization with the Man_3GlcNAc alone. The crystal belongs to space group $P2_12_12_1$ with five molecules (A, B, C, D, and E) per asymmetric unit. The structure has the Man-α-1,2-Man-α-1,3-Man moiety in molecules A and B, whereas the GlcNAc moiety is disordered. In contrast, only one mannose residue is ordered in molecule C, and all the carbohydrate residues are disordered in molecules D and E. The dimer interfaces of the $P2_12_12_1$ crystal form are different from that of the *C*2 crystal form, suggesting that VIP36 is monomeric.

The CRD of VIP36 has an overall globular shape composed of a β-sandwich of two antiparallel β-sheets and is composed of 17 β-strands and three 3_{10} helices, each with a single turn (Fig. 2*A*). The β-sandwich comprises a seven-stranded (β2-β5-β14-β7-β8-β9-β10) concave β-sheet and a six-stranded (β1a,b-β15-β6-β11-β12-β13a) convex β-sheet in a variation of the jelly roll fold. The β-strands are numbered according to the secondary structure numbering scheme of ERGIC-53 (22). A β-hairpin (strands β3 and β4) is inserted between β2 and β5. Residues Cys^{202} (strand β10) and Cys^{239} (strand β13a) form a disulfide bond. The stalk domain is composed of a β-strand (β16), a short α-helix (H4), and one turn of a 3_{10} helix (H5). The β16 forms a β-sheet with β13b on the face of the protein opposite to the ligand-binding site between the concave and convex β-sheets. The stalk domain (residues

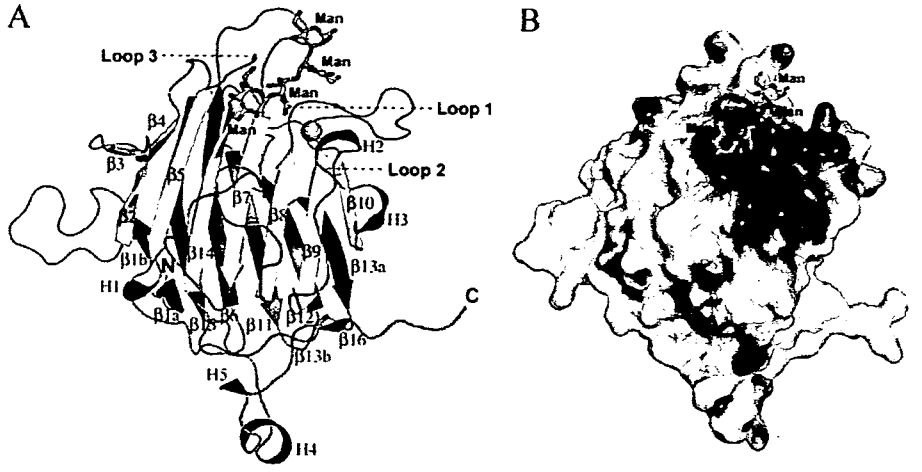

FIGURE 2. **Overall structure of the exoplasmic/luminal domain of VIP36.** Ribbon models of the VIP36 (Man_2-bound form, molecule A) are shown in *A*. The secondary structures are highlighted (β-strands belonging to the concave β-sheets, *yellow*; β-strands belonging to convex β-sheets, *blue*; β-strands belonging to β-hairpin, *cyan*; β-strands belonging to the short β-sheet formed between the stalk domain and one of the loops of the CRD, *magenta*; helices, *red*), and the loops of the CRD and stalk domain are colored *gray* and *green*, respectively. The bound Ca^{2+} is shown as a *pink sphere*. The bound oligomannoses are superimposed from the VIP36 complex structures with Man-α-1,2-Man and Man-α-1,2-Man-α-1,3-Man and are shown as a *green stick* model. The reducing-end mannose residue in the Man_2-bound form is omitted because its position is almost the same as that of the $Man_3GlcNAc$-bound form. Positions of Loops 1, 2, and 3, which are bound to the oligomannose, are indicated. The surface models (*B*) are shown in the same orientations as in *A* and colored according to the electrostatic surface potential (*blue*, positive; *red*, negative; scale from −10 to +10 *kT/e*).

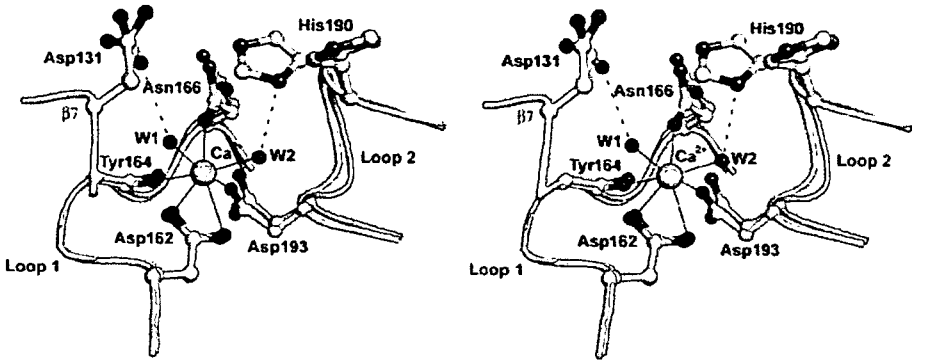

FIGURE 3. **Ca^{2+}-binding site and its conformational changes upon Ca^{2+} binding of VIP36.** The Ca^{2+}-bound (molecule A) and metal-free (molecule B) structures are shown in stereo and colored in *yellow* and *green*, respectively. Residues coordinating Ca^{2+} and those with notable conformational changes are shown in *ball-and-stick* models. Water molecules are labeled *W1* and *W2*. Ca^{2+}-coordinating bonds are *solid lines*, and hydrogen bonds are *dotted lines*.

289–301) contributes to an extensive crystal contact (Fig. 2*A* and supplemental Fig. 2) that explains the successful crystallization of this construct.

Ca^{2+}-binding Site and Its Structural Changes in VIP36 CRD—The $F_o - F_c$ electron density map of VIP36 shows one prominent peak corresponding to a Ca^{2+} ion coordinated between two loops, which are termed Loop 1 (between β8 and β9) and Loop 2 (between β9 and β10). The side-chain oxygen atom of Asp^{162} (Oδ-1 and Oδ-2), Asn^{166} (Oδ-1), and Asp^{193} (Oδ-2); main-chain carbonyl oxygen atoms of Tyr^{164} (O); and two water molecules termed W1 and W2 (O) are coordinated to the Ca^{2+} with distances of 2.4–2.5 Å (Fig. 3).

Upon Ca^{2+} binding, significant conformational changes occur around Loops 1 and 2 (Fig. 3). Large movements of the Ca^{2+}-coordinating and neighboring atoms are observed for the Oδ-2 of Asp^{131}, Oδ-1 of Asn^{166}, and Nδ-1 of His^{190}. The distances are 1.7, 1.6, and 4.0 Å, respectively. The Oδ-2 of Asp^{131} and Nδ-1 atom of His^{190} form hydrogen bonds with Ca^{2+}-coordinating water molecules W1 and W2, respectively, whereas the Oδ-1 of Asn^{166} is directly coordinated with Ca^{2+}. As we will describe further below, these residues bind the carbohydrate moiety in the presence of Ca^{2+} suggesting a mechanism for the Ca^{2+}-dependent carbohydrate binding of VIP36.

Specific Binding of VIP36 to D1 Arm of High Mannose Type Glycan—Previous studies have suggested that VIP36 CRD recognizes D1 arm, Man-α-1,2-Man-α-1,2-Man residues of high mannose type glycans (19, 20). In fact, we observed using SPR analysis that VIP36 has higher affinity for Man-α-1,2-Man-α-1,2-Man oligosaccharide (corresponding to Man(D1)-Man(C)-Man(4) of the D1 arm) than for Man-α-1,2-Man-α-1,3-Man (Fig. 4). No interaction was observed between VIP36 and Man-α-1,2-Man-α-1,6-Man. In addition, we observed that Ca^{2+} ion is required for the interaction between VIP36 and $Man_6(GlcNAc)_2$-Asn (supplemental Fig. 3). The calculated dissociation constant between VIP36 and $Man_6(GlcNAc)_2$-Asn in the presence of 1 mM Ca^{2+} was 70 μM. On the other hand, no interaction was observed in the presence of 5 mM EDTA.

Structure of VIP36 in Complex with Ca^{2+} and Man—In the structure of VIP36 in complex with Ca^{2+} and mannose, the mannose is located in a pocket neighboring the Ca^{2+}-binding site on the concave β-sheet and has well defined electron density (Fig. 5*A*). The mannose-binding site (also called the "primary binding site" hereafter) comprises β7 and Loops 1, 2, and 3. A number of specific contacts can be seen between the protein and the ligand. The mannose is bound by Asp^{131} (Oδ-1 and Oδ-2), Asn^{166} (Nδ-2), His^{190} (Nε-2), Gly^{260} (N), Asp^{261} (N), and Leu^{262} (N) through hydrogen bonds in the complex (Fig. 5*D*). Incidentally the side-chain positions of Asp^{131}, Asn^{166}, and His^{190} in the primary binding site are stabilized by Ca^{2+}.

Structure of VIP36 in Complex with Ca^{2+} and Man_2—The Man-α-1,2-Man residues have extremely well defined electron density in the structure of VIP36·Man_2 complex (Fig. 5*B*). The 4-OH and 6-OH groups of the non-reducing mannose residue make hydrogen bonds with Ser^{96} (Oγ) and Asp^{261} (Oδ-1),

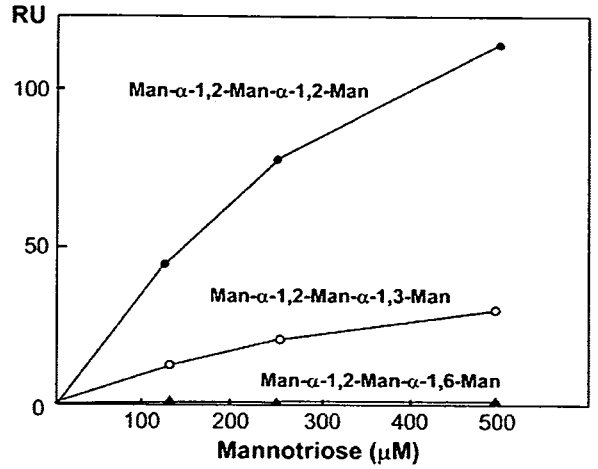

FIGURE 4. **Specific binding of GST-VIP36 to the D1 arm (Man-α-1,2-Man-α-1,2-Man) revealed by SPR analysis.** Three different mannotrioses were added over flow cells at the indicated concentrations. Specific binding of mannotrioses was obtained by subtracting the resonance unit (*RU*) value of the GST immobilized sensor chip from the values of GST-VIP36 immobilized sensor chips. The plots were obtained by subtracting the values measured using the sample buffer without carbohydrates. The dose binding curves were obtained from the resonance unit value at 200 s. *Solid circle*, Man-α-1,2-Man-α-1,2-Man; *open circle*, Man-α-1,2-Man-α-1,3-Man; *solid triangle*, Man-α-1,2-Man-α-1,6-Man.

respectively (Fig. 5*E*). In contrast, the hydrogen bond between the 6-OH group and Asp^{261} (Oδ-1) is not observed in molecule B (supplemental Fig. 4A). The binding site of the reducing-end mannose residue of Man_2 is almost the same as the primary binding site (Fig. 5, *D* and *E*).

Structure of VIP36 in Complex with Ca^{2+} and $Man_3GlcNAc$—In the structure of VIP36 with $Man_3GlcNAc$, the Man_3 moiety is ordered, whereas the GlcNAc moiety is disordered. The Man-α-1,2-Man-α-1,3-Man residues have defined electron density (Fig. 5*C*). The binding site of the non-reducing mannose residue of $Man_3GlcNAc$ is almost the same as the primary binding site (Fig. 5, *D* and *F*). The 6-OH group of the α1–2- and α1–3-linked mannose residue makes hydrogen bonds with Tyr^{164} (Oη) and Asn^{166} (O) (Fig. 5*F*). The α1–3- and β1–4-linked mannose residue is recognized by Asp^{261} (Oδ-1) through a hydrogen bond, whereas the hydrogen bond is not observed in molecule B (supplemental Fig. 4B).

DISCUSSION

Many lectins, such as mannose-binding proteins and the asialoglycoprotein receptors, achieve higher affinity and selectivity through oligomerization of their CRDs (43). For instance, ERGIC-53 and Emp46/47p are known to form oligomeric complexes through the putative coiled-coil region in the stalk domain (15, 44, 45). On the other hand, it has been shown that

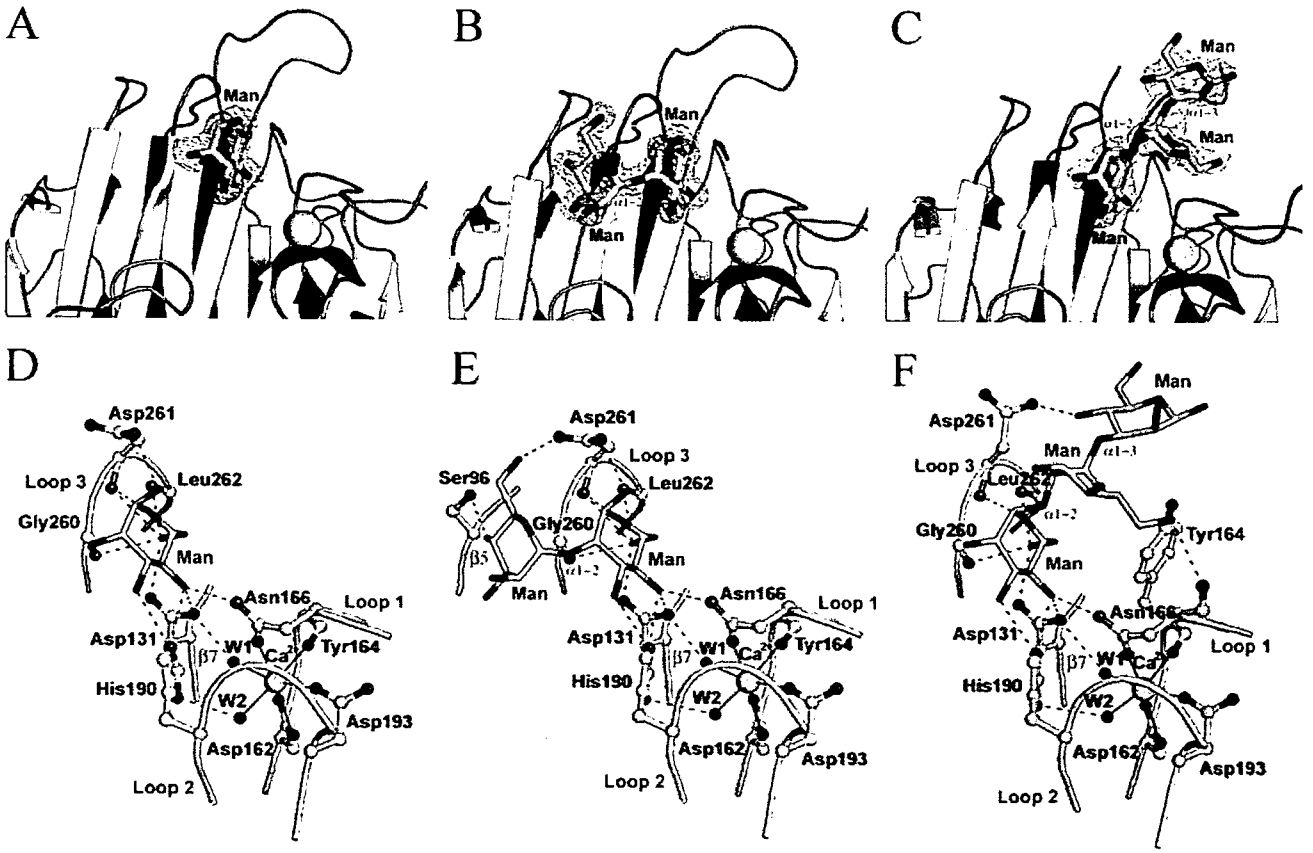

FIGURE 5. **Carbohydrate ligand-binding site of VIP36.** $2F_o - F_c$ electron density map of mannose of the Man-bound form (*A*), Man-α-1,2-Man of the Man_2-bound form (*B*), and Man-α-1,2-Man-α-1,3-Man of the $Man_3GlcNAc$-bound form (*C*) contoured at 1.2σ. Secondary structures are shown as in Fig. 2*A*. *D*, structure of mannose and Ca^{2+}-binding site of VIP36 (molecule A). *E*, structure of Man-α-1,2-Man and Ca^{2+}-binding site of VIP36 (molecule A). *F*, structure of Man-α-1,2-Man-α-1,3-Man and Ca^{2+}-binding site of VIP36 (molecule A). The bound carbohydrate residues are shown as *green stick* models. Residues of VIP36 binding to the ligands are shown in *ball-and-stick* models.

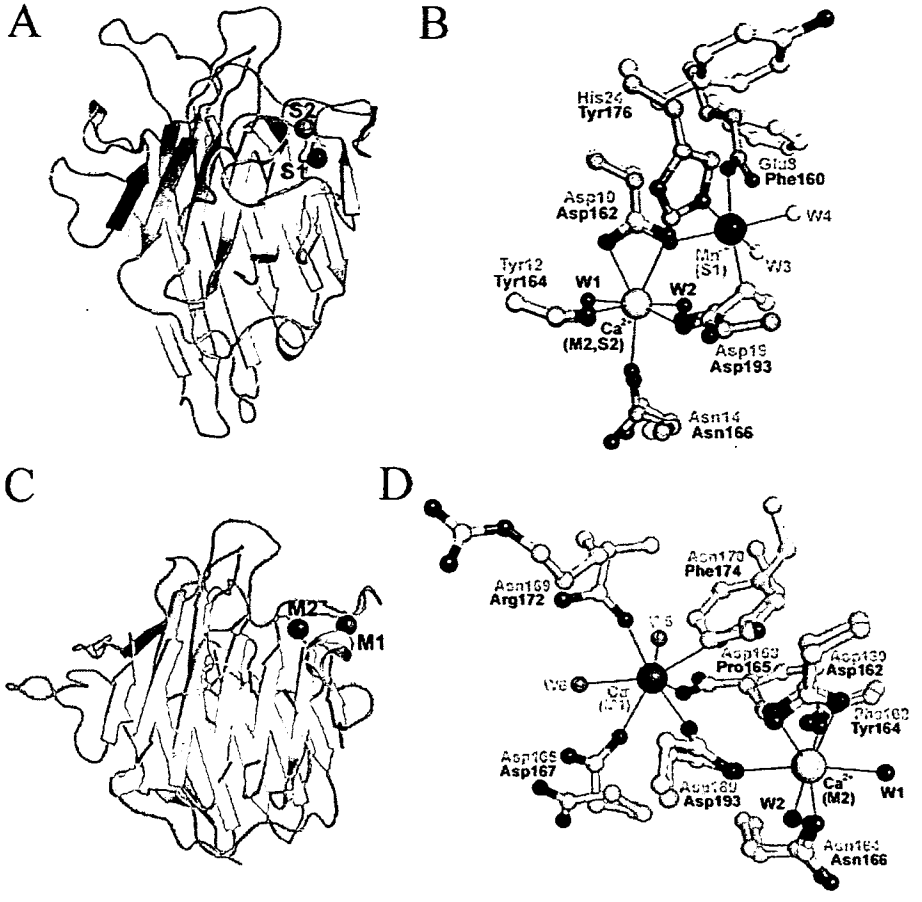

FIGURE 6. *A*, overall structure of ConA monomer. A *purple* ribbon model of ConA (molecule A) is shown. The bound Mn^{2+} (S1) and Ca^{2+} (S2) are shown as *large gray* and *black spheres*, respectively. *B*, comparison between VIP36 and ConA metal-binding site structures. The VIP36 (molecule A) and ConA structures are colored *yellow* and *purple*, respectively. Residues of VIP36 and ConA are labeled in *black* and *purple*, respectively. The Ca^{2+} in VIP36 is shown as a *large pink sphere*. Because the position of Ca^{2+} at the S2 site in ConA is almost the same as in VIP36, it is not shown. Water molecules found in the S1 site of ConA are shown as *small white spheres* and are labeled *W3* and *W4*. *C*, overall structure of ERGIC-53 CRD. A *cyan* ribbon model of ERGIC-53 (molecule A) is shown. The bound Ca^{2+} is shown as *large magenta spheres* (M1 and M2). *D*, comparison between VIP36 (*yellow*) and ERGIC-53 (*cyan*) Ca^{2+}-binding site structures. Because the position of Ca^{2+} at the M2 site in ERGIC-53 is almost the same as in VIP36, it is not shown. Water molecules found in the M1 site of ERGIC-53 are shown as *small orange spheres* and are labeled *W5* and *W6*. Residues involved in the metal binding are shown as *ball-and-stick* models.

no disulfide-linked or stable non-covalent oligomers of recombinant exoplasmic/luminal domain (residues 45–322) or endogenous VIP36 could be detected by cross-linking or sedimentation analysis (46). Furthermore we confirmed that the exoplasmic/luminal domain of VIP36 (residues 51–301 and 51–322) is monomeric in physiological solution by gel filtration analyses (supplemental Fig. 1). Indeed the stalk domain (residues 279–322) of VIP36 is 95–162 residues shorter than those of ERGIC-53, Emp46p, and Emp47p. The portion of the stalk domain of VIP36 included in our construct (residues 279–301) does not form coiled-coil structure (Fig. 2*A* and supplemental Fig. 2). The short stalk domain and the absence of coiled-coil domain suggest that VIP36 may likely function as a monomer.

It is known that leguminous lectins coordinate Mn^{2+} and Ca^{2+} ions, termed S1 and S2, respectively, in their β-sandwich structures (Fig. 6, *A* and *B*) (47). The S1 ion stabilizes the S2-binding site, and the S2 ion fixes the positions of the amino acids that interact with the oligosaccharide ligands. In this study, we showed that VIP36 has a single Ca^{2+} in the CRD structures and that the Ca^{2+} fixes the positions of Asp^{131}, Asn^{166}, and His^{190} that interact with carbohydrate ligands in the primary binding site (Figs. 3 and 5). Specifically significant conformational changes upon Ca^{2+} binding take place around the Ca^{2+} and primary carbohydrate-binding site of VIP36 (Fig. 3). Similar but more pronounced structural changes of the corresponding site upon metal binding were also observed in concanavalin A (ConA) (48) and ERGIC-53 (23). The Ca^{2+} of ConA induces large conformational changes to stabilize the correct geometry of the Ca^{2+}-binding site that comprise the *trans* to *cis* isomerization of the Ala^{207}-Asp^{208} peptide bond accompanied by the formation of the carbohydrate-binding site (48). In VIP36 structures, such isomerization changes were not observed. We also observed that Ca^{2+} is required for interaction between VIP36 and $Man_6(GlcNAc)_2$-Asn by SPR experiments (supplemental Fig. 3). These results explain the Ca^{2+}-dependent carbohydrate binding of VIP36.

It was shown that ERGIC-53 contains two Ca^{2+} ions termed M1 and M2 (Fig. 6, *C* and *D*) and that the M1 ion does not lie at the same site as the S1 ion, although M2 is located at the corresponding S2 site (23). The Ca^{2+} of VIP36 corresponds to the M2 site of ERGIC-53. When the VIP36 and ERGIC-53 structures are superimposed, the M2 metal ions overlay with a separation that is less than 0.2 Å. Although the Ca^{2+} of VIP36 is equivalent to the M2 site of ERGIC-53, the electron density maps of VIP36 show no peak that could be assigned as a metal ion either at the corresponding M1 site or at any other sites within the structure. In the Ca^{2+}-binding site (M2 and S2), the Ca^{2+}-coordinating residues are structurally very well conserved except for Asp^{193} in VIP36 (Fig. 6, *B* and *D*). The corresponding Asp^{19} (ConA) and Asp^{189} (ERGIC-53) residues are coordinated by the two metal ions. In VIP36, the Oδ-1 of Asp^{193} forms a hydrogen bond with the main-chain nitrogen atom of Asp^{167} to stabilize the Ca^{2+}-binding site. As a result, only one Ca^{2+} ion fixes the ligand binding residues in VIP36. Our crystallographic studies also suggest that VIP36 did not bind other divalent cations, neither Mn^{2+} nor Mg^{2+} (data not shown). Loop 1 of VIP36 is two residues longer than that of ERGIC-53. Likewise the residues coordinating the M1 ion in ERGIC-53 (Asp^{163}, Asp^{165},

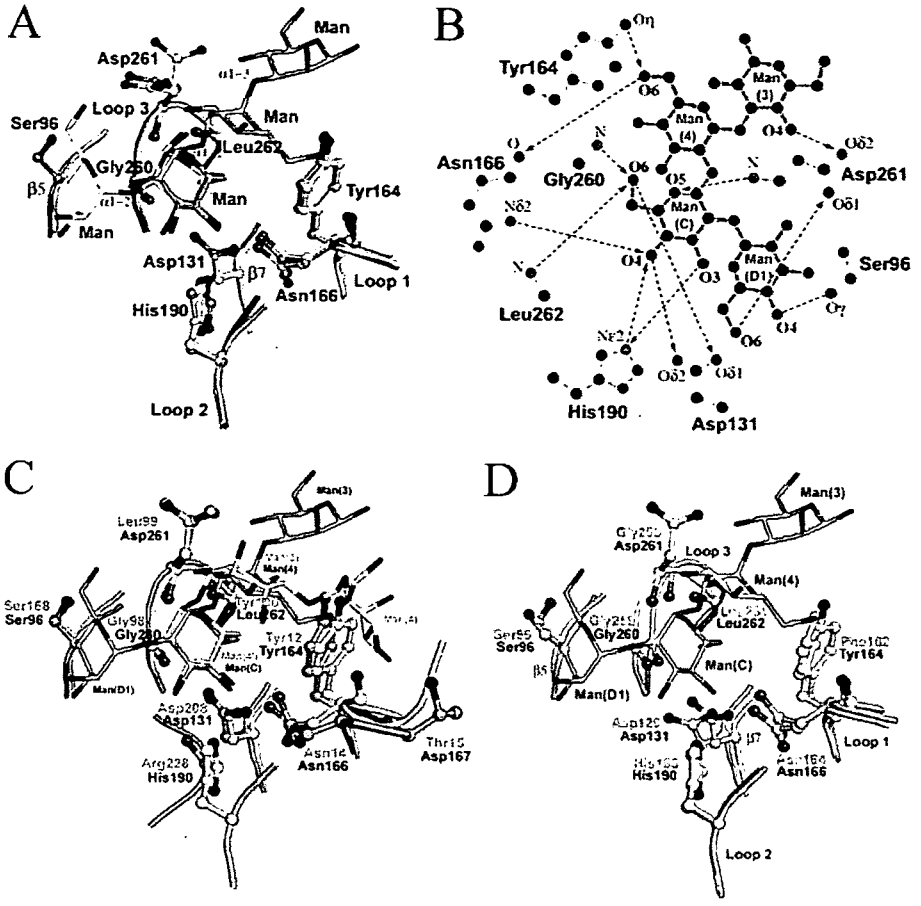

FIGURE 7. *A*, superimposition of Man-, Man_2-, and $Man_3GlcNAc$-bound forms of VIP36. The Man-, Man_2-, and $Man_3GlcNAc$-bound forms of VIP36 (molecule A) are colored in *marine blue*, *white*, and *yellow*, respectively. The bound Man, Man_2, and Man_3 are shown as *magenta*, *white*, and *green stick* models, respectively. *B*, ligand binding diagram. The model of Man-α-1,2-Man-α-1,2-Man-α-1,3-Man (Man_4) is designated as Man(D1)-Man(C)-Man(4)-Man(3), respectively. Hydrogen bonds are indicated with *dotted arrows* pointing in the direction from donor to acceptor. *C*, comparison between the VIP36 (*yellow*) and ConA (*purple*) carbohydrate ligand-binding site. Residues binding to the ligands are shown in *ball-and-stick* models. *D*, comparison between the VIP36 (*yellow*) and ERGIC-53 (*cyan*) carbohydrate ligand-binding site. The model of Man_4 is shown as a *green stick* model. Residues of VIP36 binding to the ligands and the corresponding ones in ERGIC-53 are shown in *ball-and-stick* models.

Asn^{169}, Asn^{170}, and Asp^{189}; shown in green in supplemental Fig. 2) are poorly conserved in VIP36. Taken together, we conclude that VIP36 binds only one Ca^{2+} ion and that the single Ca^{2+} fixes the positions of residues involved in carbohydrate ligand binding.

Our crystallographic studies reveal extensive interactions between VIP36 and Man-α-1,2-Man-α-1,2-Man residues of the D1 arm of high mannose type glycans. Based on the Man-, Man_2-, and $Man_3GlcNAc$-bound structures, a model structure of VIP36 in complex with Man-α-1,2-Man-α-1,2-Man-α-1,3-Man (Man_4) was built and then designated as Man(D1)-Man(C)-Man(4)-Man(3), respectively (Fig. 7). The carbohydrate-binding site is located in a negatively charged pocket (Fig. 2*B*). The extended carbohydrate-binding site comprises β5, β7, and Loops 1, 2, and 3. The mannose residue in the primary binding site corresponds to the Man(C) moiety at the middle of the D1 arm. In addition, Asp^{261} is flexible in mannose recognition and can flip between Man(3) and Man(D1) depending on the local environment (supplemental Fig. 4). The crystallographic results correlate well with the SPR experiments (Fig. 4) and the previous results (19, 20) that VIP36 recognizes Man-α-1,2-Man-α-1,2-Man residues of the D1 arm of high mannose type glycans.

Next the carbohydrate-binding site of VIP36 was compared with that of ConA in complex with Man-α-1,3-(Man-α-1,6-)Man (corresponding to Man(3)-(Man(4′))-Man(4)) (49). Although the carbohydrate binding specificity of VIP36 and ConA is essentially different, the structural conservation of ligand-binding sites between them is observed in not only the Man(C)-binding site but also the Man(D1)- and Man(4)-binding sites of VIP36 (Fig. 7*C*). In the Man(C)-binding site of VIP36, the ligand binding residues are structurally very well conserved except for His^{190} in VIP36 (Fig. 7*C*). The corresponding main-chain nitrogen atom of Arg^{228} in ConA is bound to only the 3-OH group of the mannose residue. In contrast, Nε-2 of His^{190} in VIP36 is bound to the 3-OH and 4-OH groups and acts simultaneously as hydrogen bond donor and acceptor (Fig. 7*B*). In the Man(4)-binding site, although the carbohydrate binding loop conformation of VIP36 is largely different than that of ConA, the side-chain position of Tyr^{164} is very similar to that of Tyr^{12} in ConA.

When the carbohydrate-binding site of VIP36 was compared with the corresponding site of ERGIC-53 in complex with Ca^{2+} (23), the Man_4 binding residues of VIP36 are largely identical to the corresponding residues of ERGIC-53 (Fig. 7*D* and supplemental Fig. 2). The structural conservation suggests that ERGIC-53 also binds the D1 arm moiety of high mannose type glycoproteins, which is consistent with the function of ERGIC-53 as a transport lectin for high mannose type glycoproteins. These contain $Man_8(GlcNAc)_2$ with an intact D1 arm and are transported from the ER by the ER quality control mechanism (2, 3). However, there are some structural differences in the β5, β7, and Loops 1 and 3 regions: (i) in β5 the side-chain orientation of Ser^{96} of ERGIC-53 is dissimilar to that of VIP36, (ii) the side-chain orientation of Asp^{129} in β7 of ERGIC-53 is different than that of the corresponding Asp^{131} of VIP36, (iii) Phe^{162} of ERGIC-53 is replaced by Tyr^{164} in VIP36, and (iv) the Loop 3 of ERGIC-53 is positioned further away from the ligand when compared with VIP36, and Asp^{261} of VIP36 is replaced by Gly^{260} in ERGIC-53. Most significantly, Tyr^{164} and Asp^{261} in VIP36 are better suited than Phe^{162} and Gly^{260} in ERGIC-53 for binding to the Man_4. In ERGIC-53, the