GALACTOSE RECOGNITION BY A TETRAMERIC C-TYPE LECTIN, CEL-IV, CONTAINING THE EPN CARBOHYDRATE-RECOGNITION MOTIF

Tomomitsu Hatakeyama[1,4], Takuro Kamiya[1], Masami Kusunoki[2], Sachiko Nakamura-Tsuruta[3], Jun Hirabayashi[3], Shuichiro Goda[1], and Hideaki Unno[1]

From [1]Department of Applied Chemistry, Faculty of Engineering, Nagasaki University, Nagasaki 852-8521, Japan, [2]Research Center for Structural and Functional Proteomics, Institute for Protein Research, Osaka University, Osaka 565-0871, Japan, and [3]Research Center for Medical Glycosciences, National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8568, Japan

Running head: Crystal structure of the C-type lectin CEL-IV

Address correspondence to: Tomomitsu Hatakeyama, Department of Applied Chemistry, Faculty of Engineering, Nagasaki University, 1-14 Bunkyo-machi, Nagasaki 852-8521, Japan. Tel.: +81-95-819-2686; Fax: +81-95-819-2684; E-mail: thata@nagasaki-u.ac.jp.

CEL-IV is a C-type lectin isolated from a sea cucumber, *Cucumaria echinata*. This lectin is composed of four identical C-type carbohydrate-recognition domains (CRDs). X-ray crystallographic analysis of CEL-IV revealed that its tetrameric structure was stabilized by multiple interchain disulfide bonds among the subunits. Although CEL-IV has the EPN motif in its carbohydrate-binding sites, which is known to be characteristic of mannose-binding C-type CRDs, it showed preferential binding of galactose and *N*-acetylgalactosamine. Structural analyses of CEL-IV/melibiose and CEL-IV/raffinose complexes revealed that their galactose residues were recognized in an inverted orientation compared with mannose-binding C-type CRDs containing the EPN motif, by the aid of a stacking interaction with the side chain of Trp79. Changes in the environment of Trp79 induced by binding to galactose were detected by changes in the intrinsic fluorescence and UV-absorption spectra of WT CEL-IV and its site-directed mutants. The binding specificity of CEL-IV toward complex oligosaccharides was analyzed by frontal affinity chromatography using various pyridylamino-sugars, and the results indicated preferential binding to oligosaccharides containing Galβ1-3/4(Fucα1-3/4)GlcNAc structures. These findings suggest that the specificity for oligosaccharides may be largely affected by interactions with amino acid residues in the binding site other than those determining the

monosaccharide-specificity.

Animal lectins play important roles in various molecular and cellular recognition processes (1). They have been categorized into several families. Among these families, C-type lectins, which recognize carbohydrates depending on the presence of Ca^{2+} ions, are widely distributed in various organisms, and contain C-type carbohydrate-recognition domains (CRDs[5]) that function as carbohydrate-binding modules in cooperation with other distinct domains (2, 3). C-type lectins are known to play important roles in immunity in vertebrates, but they also appear to function in self-defense systems in invertebrates (4).

We have been investigating three Ca^{2+}-lectins, CEL-I, III, and IV, from a sea cucumber *Cucumaria echinata*. While CEL-I and CEL-IV belong to the C-type lectins, CEL-III is a unique ricin-type (R-type) lectin that has strong hemolytic and cytotoxic activities. CEL-III is composed of a single polypeptide chain with two CRDs (domains 1 and 2) and one oligomerization domain (domain 3). Although domains 1 and 2 adopt a β-trefoil fold that is similar to the B-chain of the toxic lectin ricin from the castor bean, they recognize specific carbohydrates via Ca^{2+} ions in carbohydrate-binding sites in a similar manner to the C-type CRDs (5). After binding to cell surface carbohydrates, CEL-III forms ion-permeable pores by oligomerization through domain 3 in the membrane.

On the other hand, CEL-I and CEL-IV are composed of two and four C-type CRDs, respectively, that are linked by interchain disulfide bonds. The crystal structure of CEL-I (6) revealed that its extremely high specificity for GalNAc (1000-fold higher than that for galactose) was caused by multiple interactions of the acetamido group of GalNAc with several amino acid residues surrounding it. Although CEL-I has no pore-forming activity, unlike CEL-III, it shows cytotoxic activity, especially toward HeLa, MDCK, and XC cells (7). In addition, CEL-I was found to induce the secretion of various cytokines, such as TNF-α and granulocyte colony-stimulating factor, from macrophage cell line RAW264.7 cells (8). Since these effects of CEL-I on cells are at least partly inhibited by the addition of GalNAc, binding of CEL-I to GalNAc-containing carbohydrate chains on the cells seems to be involved in these activities.

The other C-type lectin from *C. echinata* is CEL-IV, which is a homotetramer of C-type CRDs. This lectin shows Gal/GalNAc-specificity, especially for α-galactosides, whereas CEL-I and CEL-III show higher specificity for β-galactosides (9). In general, C-type CRDs contain carbohydrate-recognition motifs comprised of three amino acid residues, Gln-Pro-Asp (QPD) or Glu-Pro-Asn (EPN), which are considered to determine their galactose- or mannose-specificities, respectively (10). In fact, CEL-I has a QPD motif in its carbohydrate-binding site (residues

101–103) in accordance with its Gal/GalNAc-specificity. On the other hand, CEL-IV has an EPN motif (residues 113–115), but shows galactose-specificity. Similar discrepancies concerning the carbohydrate-recognition motif and specificity have been observed in the case of the tunicate lectin TC14 (11), which contains a Glu-Pro-Ser (EPS) motif, similar to EPN, but recognizes galactose residues.

In the present study, we determined the crystal structures of CEL-IV and its complexes with two α-galactoside-containing carbohydrates, melibiose and raffinose. The results indicate that CEL-IV recognizes galactose residues in a very different orientation compared with CEL-I with the aid of a stacking interaction with a tryptophan residue. The results further reveal that CEL-IV has higher specificity for L-fucose-containing oligosaccharide structures such as Galβ1-3/4(Fucα1-3/4)GlcNAc.

EXPERIMENTAL PROCEDURES

Purification of native CEL-IV—CEL-IV was purified from *C. echinata* by affinity chromatography using carbohydrate-immobilized columns essentially as reported previously (9). However, in this study, we used a column with raffinose (Galα1-6Glcα1-2βFru) as an immobilized carbohydrate for the initial separation of CEL-IV, because this lectin shows relatively a high binding specificity for α-galactoside-containing carbohydrates. A raffinose-Cellulofine column was prepared immobilizing raffinose to Cellulofine gel (Seikagaku Kogyo) using the cross-linking reagent divinyl sulfone (12). The crude extract from *C. echinata* was applied to the raffinose-Cellulofine column in TBS (10 mM Tris-HCl pH 7.5, 0.15 M NaCl) containing 10 mM $CaCl_2$, and the adsorbed proteins (mostly CEL-IV) were eluted with 20 mM EDTA in TBS. The protein was further purified by gel filtration on a Sephacryl S-200 column (2.5×60 cm; GE Healthcare) in TBS.

Crystallization and Data Collection—Crystallization of CEL-IV was carried out by the sitting-drop vapor diffusion method. The protein solution (10 mg/ml, 2–4 μl) in TBS containing $CaCl_2$ was mixed with the same volume of reservoir solution (0.1 M HEPES pH 7.5, 4.3 M NaCl), and subjected to sitting-drop vapor diffusion at 20°C. X-ray data collection was performed on beamlines BL6A, NW12, and BL-17A at the High Energy Acceleration Research Organization (Tsukuba, Japan). Diffraction images were indexed and integrated using the program HKL2000 (13), and processed using the CCP4 programs Scala and Truncate (14). The data collection statistics are summarized in Table 1.

Structure Determination and Refinement—The crystal structure of CEL-IV was solved by the molecular replacement method using the C-type lectin CEL-I (Protein Data Bank code: 1WMY) (6) as a search model. Molecular replacement was performed using the program Phaser (15). The model was refined using the

program Refmac (16) from the CCP4 suite. Manual fitting of the model was carried out using the program Coot (17). The quality of the final model was checked using a Ramachandran plot and the model geometry was analyzed with the program Procheck (18). The refinement statistics are summarized in Table 1. The figures for the protein models were drawn using the program PyMOL (19). The atomic coordinates and structural factors of the native CEL-IV (code 3ALS) and CEL-IV complexed with melibiose (code 3ALT) and raffinose (code 3ALU) have been deposited in the Protein Data Bank.

Expression of Recombinant CEL-IV—The genes for site-directed mutants of CEL-IV (W79Y and W79H) were constructed by PCR with corresponding 30-bp primers containing the mutation site using a plasmid containing wild-type recombinant CEL-IV (20). The amplified DNA fragments were cloned into *Escherichia coli* JM109 using the pGEM-T vector (Promega), and the nucleotide sequences were confirmed by DNA sequencing. The inserted genes were digested with NdeI and BamHI, and then ligated with the pET-3a vector (Novagen) that had been predigested with the same enzymes. The resulting plasmids containing recombinant CEL-IV genes were used for transformation of *E. coli* BL21(DE3)pLysS (Novagen). Induction of protein expression was carried out with 1 mM isopropylthiogalactoside, and the cells were incubated for an additional 5 h at 37°C. The expressed proteins were obtained in inclusion bodies after disruption of the cells by sonication. The inclusion bodies were partially solubilized in 20 mM Tris-HCl buffer (pH 8.5) containing 2 M urea for 30 min at 4°C. After dialysis against TBS containing 10 mM $CaCl_2$, the solubilized proteins were applied to a raffinose-Cellulofine column equilibrated with the same buffer. The adsorbed proteins were eluted with 20 mM EDTA in TBS, and purified by gel filtration on a Sephacryl S-200 column.

Fluorescence Measurements—Fluorescence spectra of the proteins were measured at 25°C in TBS containing 10 mM $CaCl_2$ using an F-3010 fluorescence spectrophotometer (Hitachi), with excitation at 290 nm to avoid any influence of the fluorescence of tyrosine residues. The effects of addition of carbohydrates were measured by titrating a carbohydrate solution containing the same concentration of the protein.

UV-absorption Spectra—UV-absorption spectra of the proteins were measured at 25°C in TBS containing 10 mM $CaCl_2$ in a U-2000 spectrophotometer (Hitachi). The changes upon binding of carbohydrates to the proteins were recorded in their difference spectra, which were created by subtracting the spectra of the protein from the spectra of the protein/carbohydrate mixture by placing the solutions in the reference and sample cells, respectively.

Frontal Affinity Chromatography—Frontal affinity chromatography was performed using an automated system as previously described

(21, 22) to examine the affinity of CEL-IV for various oligosaccharides. Briefly, CEL-I, CEL-III, and CEL-IV were immobilized to *N*-hydroxysuccinimide-activated Sepharose 4B Fast Flow (GE Healthcare). Pyridylamino (PA)-oligosaccharides were applied to the lectin-immobilized columns (2.0×10 mm, 31.4-μl column volume) in 10 mM Tris-HCl (pH 7.4) containing 0.14 M NaCl and 10 mM $CaCl_2$, and eluted with a flow rate of 0.125 ml/min at 25°C. The elution of the PA-oligosaccharides was monitored by the fluorescence at 380 nm after excitation at 310 nm. The affinities of the PA-oligosaccharides were evaluated by the retardation of the front of their elution using the following equation:

$$K_a=(V–V_0)/Bt$$

where K_a is the association constant, $(V–V_0)$ is the retardation of the elution volume, and Bt is the effective ligand content, which is dependent on the amount of immobilized lectin on the column.

In the case of CEL-I, the *K*a was able to be calculated using a Bt value estimated from the concentration-dependence of the binding of *p*-nitrophenyl-α-D-GalNAc as a standard labeled saccharide. However, because of the lack of suitable standard saccharides, the affinities of the PA-oligosaccharides for binding to CEL-III and CEL-IV were only expressed as $(V–V_0)$ values.

RESULTS AND DISCUSSION

Crystallization and Structure Determination of CEL-IV—Crystallization of CEL-IV was carried out by the sitting-drop vapor diffusion method at 20°C using a reservoir solution composed of 0.1 M HEPES (pH 7.5) and 4.3 M NaCl. Although native CEL-IV formed crystals very rapidly, the crystals diffracted X-rays with relatively poor resolution (3.0 Å or lower) using synchrotron radiation. This may be partly because of their high solvent content (approximately 80%). On the other hand, CEL-IV/melibiose and CEL-IV/raffinose complex crystals gave higher resolutions (2.5 and 1.65 Å, respectively). Therefore, the crystal structures of these complexes were able to be solved by the molecular replacement method using the structural data of the CRD of the other *C. echinata* lectin CEL-I (6) as a search model. After initial phasing, four protomers of CEL-IV were placed to construct a tetrameric CEL-IV molecule. In each protomer, one Ca^{2+} ion was found in the long loop region (3), which constitutes the carbohydrate-binding sites of many C-type CRDs. Although several C-type CRDs contain more than one Ca^{2+} ion, CEL-IV has only one Ca^{2+} ion. In the CEL-IV/melibiose and CEL-IV/raffinose complex crystals, electron densities of the bound carbohydrates were observed near this Ca^{2+} ion. Therefore, the models of these carbohydrates were fitted with electron density maps, and further refinements were performed. The data collection and refinement statistics are shown in Table 1.

Overall Structure of CEL-IV—Fig. 1 shows

the overall structure of the CEL-IV/raffinose complex. As shown in the figure, CEL-IV consists of four subunits that are each comprised of a C-type CRD. In many CRDs, two intrachain disulfide bonds (C1–C4 and C2–C3 in Fig. 2) are widely conserved (3). However, CEL–IV lacks a C2–C3 disulfide bond. This is also the case for CEL-I, suggesting that this disulfide bond is not necessarily required to maintain the CRD fold. CEL-I and CEL-IV contain additional disulfide bonds at C0–C0', which is characteristic of long-form CRDs (3). In the tetrameric structure of CEL-IV, the four subunits are further held together by interchain disulfide bonds among Cys1, Cys41, and Cys151 (Fig. 1*B*). Dimer units of the subunits (A–C and B–D pairs) are linked by two disulfide bonds between Cys41 and Cys151, and these pairs are further linked by disulfide bonds formed by the N-terminal Cys1 residues between the A–D and B–C pairs. Although these Cys1–Cys1 disulfide bonds were confirmed in the electron density map of the CEL-IV/raffinose complex (Fig. 1*C*), they were not clearly seen in the maps of the native CEL-IV and CEL-IV/melibiose complex. This suggests that the Cys1–Cys1 disulfide bonds are rather flexible, and not fixed at particular positions between the subunits. The quaternary structure of CEL-IV is further stabilized in the A–B and C–D pairs through noncovalent interactions between the loops of residues 38–48 (Fig. 2, *IP-1 region*) and the C-terminal residues 152–157. These regions correspond to additional sequences compared with other CRDs (Fig. 2). There is another insertion (IP-2; residues 70–78) in CEL-IV. This region has a role to place Trp79 in an appropriate position so that it can interact with the hydrophobic face of a galactose residue, as described below. A bound carbohydrate molecule was seen in each subunit in the complex crystals with raffinose and melibiose. These molecules are recognized through coordinate bonds with Ca^{2+} ions as well as hydrogen bond networks with nearby amino acid residues. As shown in Fig. 1A, the four carbohydrate-binding sites of the CEL-IV tetramer are oriented apart from one another. Since one of the likely functions of invertebrate lectins is to aggregate foreign microorganisms by binding to their surface carbohydrates, thereby neutralizing and opsonizing them for phagocytosis, such orientations of the binding sites may be preferable. Some of the dimeric C-type lectins from invertebrates, including CEL-I, are also known to have quaternary structures with the carbohydrate-binding sites oriented in opposite directions (23).

While the fold of the CEL-IV protomer resembles the other C-type CRDs, as representatively shown for *C. echinata* lectin CEL-I (6), tunicate (*Polyandrocarpa misakiensis*) lectin TC14 (11), and rat mannose-binding lectin MBL (MBP-A) (24) (Fig. 3), CEL-IV has two extra loops, which correspond to the IP-1 and IP-2 regions. These loops play important roles in the formation of the tetrameric structure and carbohydrate-binding

site as mentioned above. Although other C-type CRDs contain two or three Ca^{2+} ions, only one Ca^{2+} ion was found in each protomer of CEL-IV. This Ca^{2+} ion corresponds to the Ca-2 position in other CRDs (3), and makes direct interactions with the OH groups of bound carbohydrates.

Carbohydrate-Binding Site Structure of CEL-IV—Fig. 4 shows the carbohydrate-binding sites of the CEL-IV/raffinose and CEL-IV/melibiose complexes. The bound Ca^{2+} ion is held by coordinate bonds with the side chain oxygens of Glu113, Asn115, Asn116, and Asp136, and the main chain carbonyl oxygen of Asp136. The hydroxyl groups at the 3- and 4-positions of the galactose residue of both carbohydrates form coordinate bonds with the Ca^{2+} ion. In addition, the 3-OH group of the galactose residue forms two hydrogen bonds with the side chains of Glu113 and Asn115. This carbohydrate-recognition mode basically conforms to those of the other C-type CRDs. On the other hand, there are no hydrogen bonds between the 4-OH group of the galactose residue and other amino acid residues. Instead, the 6-OH group of the galactose residue makes a hydrogen bond with Gln82. Binding of raffinose and melibiose was further stabilized by a stacking interaction between the hydrophobic face of the galactose residue and the side chain of Trp79. The acetal oxygens in the rings of the glucose and fructose residues in raffinose also make hydrogen bonds with the NH group of Trp79 and the OH group of Tyr129, respectively. On the other hand, in the CEL-IV/melibiose complex, the glucose residue of melibiose forms a hydrogen bond between its 4-OH group and the NH-group of Trp79, resulting in an inverted orientation compared with the glucose residue in the CEL-IV/raffinose complex.

Fig. 5 shows a comparison of the carbohydrate-binding modes of CEL-IV and. CEL-I, TC14, and MBL. The orientation of the galactose residue in melibiose bound to CEL-IV is similar to that of galactose bound to TC14, in which the 3-OH group is hydrogen-bonded with the glutamate and serine residues in the EPS motif (Glu86-Pro-87-Ser88). On the other hand, the relative orientations of the bound carbohydrates are inverted by nearly 180º in the case of the CEL-I/GalNAc complex, in which the glutamine and aspartate residues in the QPD motif (Gln101-Pro102-Asp103) recognize the 4-OH group of *N*-acetylgalactosamine. Since the amide NH group of the side chain of Asn/Gln and the carboxyl group of Asp/Glu act as hydrogen donor and acceptor in hydrogen bonding, respectively, the QPD and EPN motifs are thought to specify the orientation of the hydroxyl groups at the 3- or 4-positions of bound carbohydrates (10, 11). In fact, the relative orientations of the 3-OH group of the galactose residues bound to CEL-IV and TC14 are similar to that of the 4-OH group of methyl α-mannoside bound to MBL containing the EPN motif (Glu185-Pro186-Asn187), which accounts for the galactose recognition via the

EPN(S) motif in CEL-IV and TC14. The binding of the galactose residue is also stabilized by a stacking interaction with Trp79 in CEL-IV and Trp100 in TC14. Such stabilization of galactose binding by aromatic side chains has been observed in various galactose-recognizing lectins, including the other *C. echinata* lectin CEL-III (5). In the case of CEL-I, the binding of GalNAc is stabilized by Trp105 located near the C6 atom of GalNAc. On the other hand, MBL has His189 at a similar position, which stabilizes the binding of mannose. For MBL and CEL-I, the orientations of the hydroxyl groups at the 3-positions of mannose and GalNAc are specified by hydrogen bonds with Glu193/Asn205 and Glu109/Asn123, respectively. Galactose bound to TC14 forms hydrogen bonds between its 4-OH group and Asp107 as well as a water molecule (not shown in Fig. 5) in the binding site. In contrast, the 4-OH group of galactose bound to CEL-IV does not form any hydrogen bonds with amino acid residues. However, the hydrogen bond between the 6-OH group and Gln82 appears to stabilize the binding of the galactose residue in addition to the stacking interaction with Trp79.

Although Trp79 in CEL-IV plays a similar role to Trp100 in TC14 for galactose binding, their positions in the amino acid sequence are not equivalent, as shown in Fig. 2. This suggests that they are not closely related in the evolutionary lineage of C-type lectins. In fact, the sequence identity between CEL-IV and TC14 is relatively low (20.7%). Fig. 6 illustrates the difference in the positions of these residues. Trp79 of CEL-IV is located at the end of the IP-2 loop, whereas Trp100 of TC14 is at the end of β-strand β2. These facts suggest that the similarity in the galactose recognition modes of CEL-IV and TC14 is an example of convergent evolution. Other lectins with the EPN motif showing galactose-specificity are also present in some marine organisms such as AJL-2 from the skin mucus of the Japanese eel (*Anguilla japonica*) (25) and the eggs of the shishamo smelt (*Spirinchus lanceolatus*) (26). In these cases, tryptophan residues are not conserved at the positions corresponding to Trp79 in CEL-I and Trp100 in TC14.

Interactions Between CEL-IV and Specific Carbohydrates—Since Trp79 makes a stacking interaction with the hydrophobic face of the galactose residue, it was examined using the intrinsic fluorescence changes, which mostly arise from tryptophan residues in the protein. As shown in Fig. 7*A*, when melibiose was added to the CEL-IV solution, the maximum fluorescence of CEL-IV shifted to a shorter wavelength, and its intensity increased. These observations suggested that binding of melibiose changed the environment of tryptophan residue(s) in the protein. Binding of a specific carbohydrate was also detected by changes in the UV-absorption spectra of CEL-IV. As shown in Fig. 8*A*, when melibiose was added to the protein solution, increases in the

positive maxima at 293 and 286 nm were observed, which arose from a red shift of the UV-absorption spectra. Along with the fluorescence spectral changes, these findings suggest that the environment of tryptophan residue(s), possibly Trp79, became more nonpolar (27, 28). These spectral changes increased as the concentration of the carbohydrate increased (Fig. 9), and produced saturation curves. To clarify whether such spectral changes are caused by Trp79, site-directed mutants of CEL-IV, in which Trp79 was replaced by tyrosine (W79Y) or histidine (W79H) residues were prepared and their fluorescence and UV spectra were examined. As shown in Figs. 7*B*, 7*C*, and 8*B*, the W79Y and W79H mutants showed little changes in their fluorescence spectra when melibiose was added, while W79Y exhibited a positive maximum at 289 nm in the UV-difference spectra, which was probably caused by an interaction with Tyr79, judging from the wavelength (28, 29). These results strongly suggest that the changes in the fluorescence and UV-absorption spectra of CEL-IV were caused by an interaction between melibiose and Trp79 located in the carbohydrate-binding site. Since the fluorescence of the tyrosine residue was very weak, binding of melibiose to W79Y was not fully observed (Fig. 7*B*). Association constants for the binding of the carbohydrates to WT CEL-IV and W79Y were calculated from the fluorescence and UV-absorption changes. As shown in Fig. 9, the changes were well approximated to the curves based on the equations $\Delta F=\Delta F_{max}[S]/(K_d+[S])$ or $\Delta A=\Delta A_{max}[S]/(K_d+[S])$, where K_d is the dissociation constant, [*S*] is the carbohydrate concentration, ΔF_{max} is the maximal difference in the fluorescence, and ΔA_{max} is the maximal difference of the UV absorption. The obtained K_a (= $1/K_d$) values are listed in Table 2. The association constants calculated from the changes in the fluorescence and UV-absorption spectra showed relatively good agreement for binding of both GalNAc and melibiose to WT CEL-IV, confirming that they reflect the changes in the identical tryptophan residue, Trp79. Furthermore, the UV-difference spectra of W79Y with positive peaks at 289 nm indicate that the tyrosine residue can substitute for Trp79, but with lower affinity, suggesting the importance of the more hydrophobic nature of tryptophan compared with that of tyrosine. In addition, the bulkiness of tryptophan may be advantageous in carbohydrate binding by eliminating water molecules from the carbohydrate-binding site. The importance of a tryptophan residue in galactose binding was also observed in the case of galactose-binding MBL (MBP-A) mutants, in which the EPN motif had been replaced by the QPD motif. In these mutants, an additionally introduced tryptophan residue instead of His189 (see MBL/α-Me-Man in Fig. 5) effectively enhanced the galactose binding, while phenylalanine at the same position had less effects (30). Regarding the CEL-IV mutants, the

carbohydrate-binding ability of W79H appeared to be even weaker when its amount of binding to the affinity column (GalNAc-Cellulofine) was compared with that of W79Y (data not shown). It is likely that the smaller and more hydrophilic nature of the histidine side chain is not suitable for providing sufficient force to stabilize galactose binding.

Oligosaccharide Specificity of CEL-IV Analyzed by Frontal Affinity Chromatography—Fig. 10 shows the carbohydrate-binding specificities of CEL-IV in comparison with those of CEL-I and CEL-III for various oligosaccharides analyzed by frontal affinity chromatography. In these analyses, fluorescently labeled PA-oligosaccharides (Fig. 11) were applied to columns containing the immobilized lectins, and the retardation of their elution ($V–V_0$) caused by interactions between the oligosaccharides and the lectins were measured to determine their affinities. As shown in Fig. 10, CEL-IV exhibited relatively high affinity for oligosaccharide nos. 731 (Fucα1-2Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4Glc) and 910 (Fucα1-2Galβ1-4(Fucα1-3)GlcNAc). Oligosaccharide nos. 419, 420, 726, 727, and 730 also showed appreciable affinities. These oligosaccharides commonly contain Galβ1-3/4(Fucα1-3/4)GlcNAc structures. Therefore, it seems reasonable to assume that CEL-IV preferentially binds to oligosaccharides with these structures. Although oligosaccharide no. 721, which also has this structure, showed no affinity, this may be ascribed to the additional GalNAc residue at the nonreducing end linked to the Gal residue by an α1-3 linkage. On the other hand, the other C-type lectin CEL-I exhibited considerably different preferences for binding. This lectin strongly bound to oligosaccharide nos. 702, 708, 716, and 717, which contain GalNAcβ1-3/4 structures, and weakly bound to oligosaccharide nos. 719, 720, and 721, which contain GalNAcα1-3 linkages. These observations are consistent with the results obtained from hemagglutination assays (9), in which GalNAc showed the highest hemagglutination inhibition, and also revealed a preference for β-GalNAc. CEL-III is a unique R-type lectin with a β-trefoil fold (31, 32) containing Ca^{2+} ions in the carbohydrate-binding domains. This lectin showed the highest affinity for oligosaccharide no. 729, and moderately high affinity for several *N*-linked glycans. These oligosaccharides contain Galβ1-3/4GlcNAc structures. These findings show that the oligosaccharide specificities of *C. echinata* lectins are apparently different, although they are Gal/GalNAc-specific in terms of monosaccharide specificities. Such differences in the oligosaccharide specificities may be related to the diversity of their target molecules.

Diversity of Carbohydrate-Recognition by C-type Lectins—Although the exact functions and targets of the C-type lectins, CEL-I and CEL-IV, in *C. echinata* are still unclear, it is very likely that these lectins play defensive roles against invading microorganisms or

predators. In addition to the role as opsonins for phagocytosis, it seems also possible that they may directly act as toxic substances, as does CEL-III. It is well known that various lectins are present in the seeds of plants, and that some of them exhibit strong toxicity. For example, ricin from the castor bean (*Ricinus communis*) and abrin from *Abrus precatorius* (33) exert strong toxicities via enzymatic activities as *N*-glycosidases that cleave off a particular adenine base in the 28S rRNA of eukaryotic cells (34). On the other hand, even binding of some lectins to carbohydrate chains on the cell surface often causes damage to cell functions, leading to various biological consequences such as apoptosis (35-37). For self-defense of animals via chemical substances, such as toxins, recognition of a wide variety of carbohydrates on foreign cells may be advantageous to exert their biological effects.

Recently, regenerating gene family proteins RegIII and RegIV, which adopt a C-type lectin-like fold, were found to bind to carbohydrates with EPN or DPQ motifs located in a distinct region from the Ca^{2+}-binding sites of canonical C-type lectins (38, 39). They recognize peptidoglycan (RegIII) or mannan/heparin (RegIV) in a Ca^{2+}-independent manner, and exert bactericidal activity. Although the detailed structural data for the binding mechanisms of these proteins remain to be elucidated, these findings suggest that such motifs of three consecutive amino acids (EPN and DPQ), composed of hydrogen donor and acceptor residues flanking a proline residue, could play versatile roles in the carbohydrate-recognition of C-type lectins even without bound Ca^{2+}. These Reg family proteins recognize polysaccharides, probably through interactions with relatively large portions of their carbohydrate-binding sites. In the case of RegIV, it has been suggested that different conformers around the binding site would enable a wide range of binding functions with the same molecule (39). For recognition of oligosaccharides or polysaccharides, multiple interactions with sites other than the specificity-determining tripeptide motifs may be very important. As shown in Fig. 4 the orientation of the galactose moiety bound to CEL-IV via the EPN motif appears to lead favorable binding to α-galactoside-containing oligosaccharides through increased interactions between carbohydrates and the protein, which may also be associated with the unexpected binding specificity for oligosaccharides containing α-fucoside. The versatile roles of the C-type CRDs in the recognition of various target carbohydrates seem to be based on their potential to utilize combinations of the limited binding motifs and structural diversity of the long loop regions, supported by the stable core structure of the common fold composed of α-helices and β-sheets.

Acknowledgment—We are grateful to Junko Kominami for her technical assistance in frontal affinity chromatography analysis.

REFERENCES

1. Kilpatrick, D.C. (2002) *Biochim. Biophys. Acta* **1572**, 187-197
2. Drickamer, K. (1999) *Curr. Opinion Struct. Biol.* **9**, 585-590
3. Zelensky, A.N. and Gready, J.E. (2005) *FEBS J.* **272**, 6179-6217
4. Liu, Y.C., Li, F.H., Dong, B., Wang, B., Luan, W., Zhang, X.J., Zhang, L.S. and Xiang, J.H. (2007) *Mol. Immunol.* **44**, 598-607
5. Hatakeyama, T., Unno, H., Kouzuma, Y., Uchida, T., Eto, S., Hidemura, H., Kato, N., Yonekura, M. and Kusunoki, M. (2007) *J. Biol. Chem.* **282**, 37826-37835
6. Sugawara, H., Kusunoki, M., Kurisu, G., Fujimoto, T., Aoyagi, H. and Hatakeyama, T. (2004) *J. Biol. Chem.* **279**, 45219-45225
7. Kuramoto, T., Uzuyama, H., Hatakeyama, T., Tamura, T., Nakashima, T., Yamaguchi. K. and Oda, T. (2005) *J. Biochem.* (Tokyo) **137**, 41–50
8. Yamanishi, T., Yamamoto, Y., Hatakeyama, T., Yamaguchi, K. and Oda, T. (2007) *J. Biochem.* (Tokyo) **142**, 587–595
9. Hatakeyama, T., Kohzaki, H., Nagatomo, H. and Yamasaki, N. (1994) *J. Biochem.* (Tokyo) **116**, 209-214
10. Drickamer, K. (1992) *Nature* **360**, 183-186
11. Poget, S.F., Legge, G.B., Proctor, M.R., Butler, P.J., Bycroft, M. and Williams, R.L. (1999) *J. Mol. Biol.* **290**, 867-879
12. Hatakeyama, T., Nagatomo, H. and Yamasaki, N. (1995) *J. Biol. Chem.* **270**, 3560-3564
13. Otwinowski, Z. and Minor, W. (1997) *Methods in Enzymology* **276**, 307-326
14. Collaborative Computational Project, N.4. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760-763
15. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. and Read, R.J. (2005) *Acta Crystallogr. D Biol. Crystallogr.* **61**, 458-464
16. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255
17 Emsley, P. and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132
18. Laskowski, R., MacArthur, M., Moss, D. and Thornton, J. (1993) *J. App. Cryst.* **26**, 283-291
19. DeLano, W.L. (2002) *The PyMOL Molecular Graphics Systems,* DeLano Scientific, San Carlos, CA
20. Hatakeyama, T., Hozawa, T., Hirotani, I., Tsuda, N., Kusunoki, M. and Shiba, K. (2006) *Biochim. Biophys. Acta* **1760**, 318-325
21. Hirabayashi, J. (2004) *Glycoconj. J.* **21**, 35-40

22. Nakamura, S., Yagi, F., Totani, K., Ito, Y. and Hirabayashi, J. (2005) *FEBS J.* **272**, 2784-2799
23. Gourdine, J.-P., Cioci, G., Miguet, L., Unverzagt, C., Silva, D.V., Varrot, A., Gautier, C., Smith-Ravin, E.J. and Imberty, A. (2008) *J. Biol. Chem.* **283**, 30112-30120
24. Ng, K.K., Kolatkar, A.R., Park-Snyder, S., Feinberg, H., Clark, D.A., Drickamer, K. and Weis, W.I. (2002) *J. Biol. Chem.* **277**, 16088-16095
25. Tasumi, S., Ohira, T., Kawazoe, I., Suetake, H., Suzuki, Y. and Aida, K. (2002) *J. Biol. Chem.* **277**, 27305-27311
26. Hosono, M., Sugawara, S., Ogawa, Y., Kohno, T., Takayanagi, M. and Nitta, K. (2005) *Biochim. Biophys. Acta* **1725**, 160-173
27. Cowgill, R.W. (1967) *Biochim Biophys Acta* **133**, 6-18
28. Yanari, S. and Bovey, F.A. (1960) *J. Biol. Chem.* **235**, 2818-2826
29. Yamasaki, N., Hatakeyama, T. and Funatsu, G. (1985) *J. Biochem.* (Tokyo) **98**, 1555-1560
30. Iobst, S.T. and Drickamer, K. (1994) *J. Biol. Chem.* **269**, 15512-15519
31. Uchida, T., Yamasaki, T., Eto, S., Sugawara, H., Kurisu, G., Nakagawa, A., Kusunoki, M. and Hatakeyama, T. (2004) *J Biol Chem* **279**, 37133-37141
32. Hazes, B. (1996) *Protein Sci.* **5**, 1490-1501
33. Olsnes, S. (2004) *Toxicon* **44**, 361-370
34. Endo, Y., Tsurugi, K., Yutsudo, T., Takeda, Y., Ogasawara, T. and Igarashi, K. (1988) *Eur. J. Biochem.* **171**, 45-50
35. Liu, B., Xu, X.C., Cheng, Y., Huang, J., Liu, Y.H., Liu, Z., Min, M.W., Bian, H.J., Chen, J. and Bao, J.K. (2008) *BMB Rep.* **41**, 369-375
36. SundarRaj, S., Soni, C. and Karande, A.A. (2009) *Mol. Immunol.* **46**, 3411-3419
37. Kuroda, J., Yamamoto, M., Nagoshi, H., Kobayashi, T., Sasaki, N., Shimura, Y., Horiike, S., Kimura, S., Yamauchi, A., Hirashima, M. and Taniwaki, M. (2010) *Mol. Cancer Res.* **8**, 994-1001
38. Lehotzkya, R.E., Partch, C.L., Mukherjee, A.,, Cash, H.L., Goldman, R.E., Gardner, K.H. and Hooper, L.V. (2010) *Proc. Natl. Acad. Sci. USA*, **107**, 7722-7727
39. Ho, M.R., Lou, Y.C., Wei, S.Y., Luo, S.C., Lin, W.C., Lyu, P.C. and Chen, C. (2010) *J. Mol. Biol.* **402**, 682–695
40. Hatakeyama, T., Ohuchi, K., Kuroki, M. and Yamasaki, N. (1995) *Biosci. Biotechnol. Biochem.* **59**, 1314-1317
41. Hatakeyama, T., Matsuo, N., Shiba, K., Nishinohara, S., Yamasaki, N., Sugawara, H. and Aoyagi, H. (2002) *Biosci. Biotechnol. Biochem.* **66**, 157-163
42. Suzuki, T., Takagi, T., Furukohri, T., Kawamura, K. and Nakauchi, M. (1990) *J. Biol. Chem.*

265, 1274-1281

43. Drickamer, K., Dordal, M.S. and Reynolds, L. (1986) *J. Biol. Chem.* **261**, 6878-6887

FOOTNOTES

This work was conducted under the Cooperative Research Program of Institute for Protein Research, Osaka University, and was supported by a Grant-in-Aid for Scientific Research (20580100) from the Japan Society for the Promotion of Science.

The atomic coordinates and structural factors of the native CEL-IV and CEL-IV complexed with melibiose and raffinose have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/) under codes 3ALS, 3ALT, and 3ALU, respectively.

[2]Present address: Graduate School of Medicine and Engineering, University of Yamanashi, Takeda, Kofu 400-8511, Japan.

[4]To whom correspondence should be addressed: Tomomitsu Hatakeyama, Department of Applied Chemistry, Faculty of Engineering, Nagasaki University, 1-14 Bunkyo-machi, Nagasaki 852-8521, Japan. Tel: +81-95-819-2686; Fax: +81-95-819-2684; E-mail: thata@nagasaki-u.ac.jp.

[5]The abbreviations used are: CRD, carbohydrate-recognition domain; PA, pyridylamino.

FIGURE LEGENDS

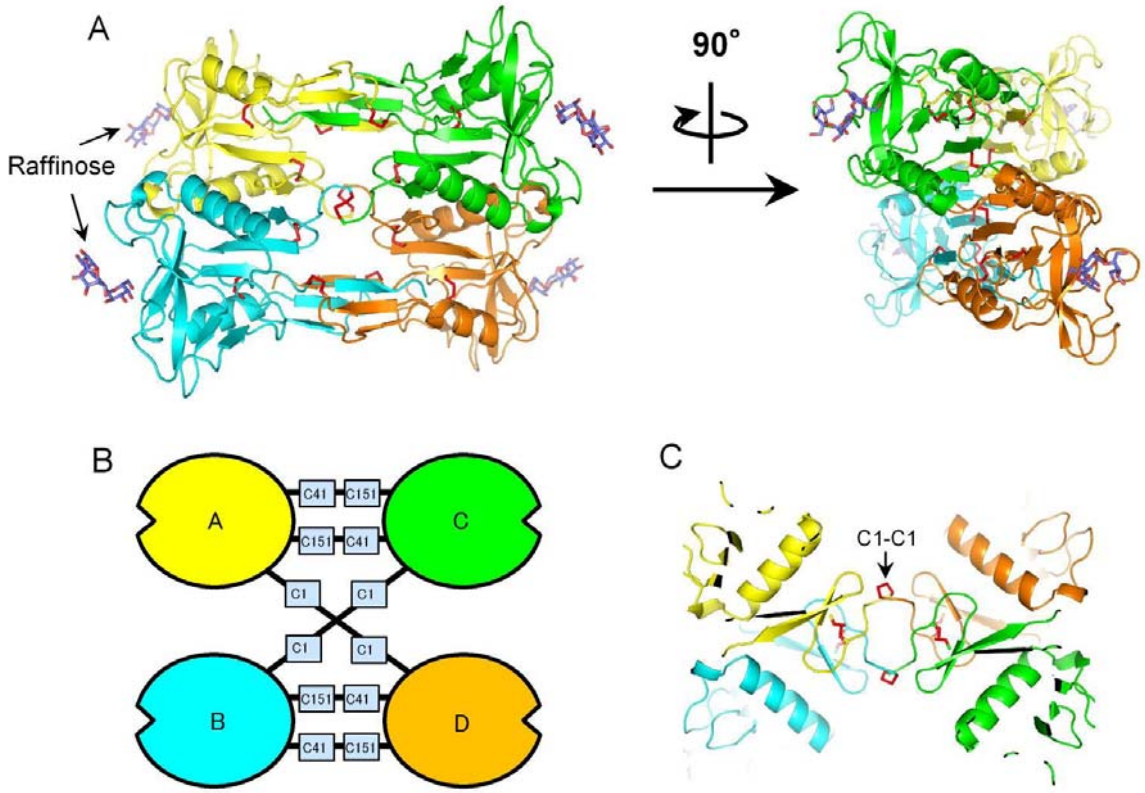

Fig. 1. **Overall structure of the CEL-IV/raffinose complex.** *A*, ribbon model of the homotetrameric structure. The four identical polypeptides (subunits) are indicated in *different colors*. The side chains of cysteine residues are shown as *red sticks*. One raffinose molecule (*stick model*) is bound to each subunit. *B*, positions of the interchain disulfide bonds between the four subunits of CEL-IV. *C*, disulfide bonds linking the N-termini of the subunits.

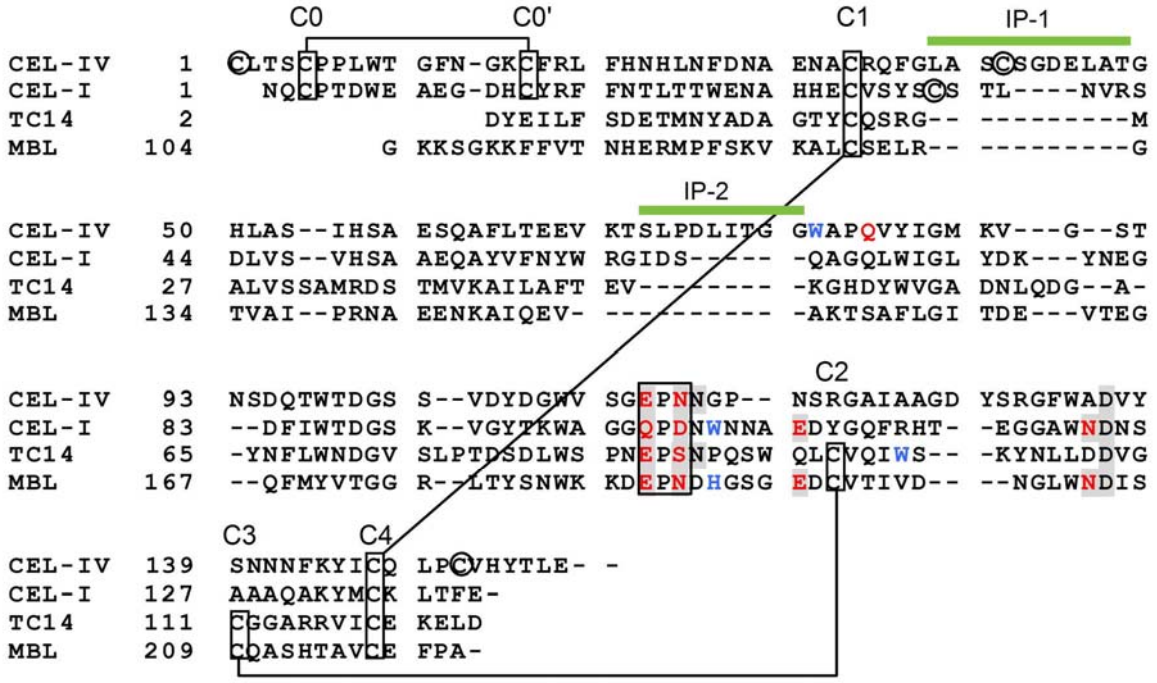

Fig. 2. **Comparison of the amino acid sequences of CEL-IV (40), CEL-I (41), tunicate lectin (TC14) (42), and mouse mannose-binding lectin MBL (MBP-A) (43).** The cysteine residues involved in disulfide bonds are enclosed in *boxes* and designated *C0*, *C0'*, and *C1–C4*, according to a previous report (3). The EPN, QPD, and EPS sequences are also enclosed in a *box*. The residues coordinating the Ca^{2+} ion are *shaded*. The residues forming hydrogen bonds with the galactose residue and stabilizing the binding of carbohydrates through a stacking interaction are shown in *red* and *blue*, respectively. The insertion peptides characteristic of CEL-IV are designated *IP-1* and *IP-2*.

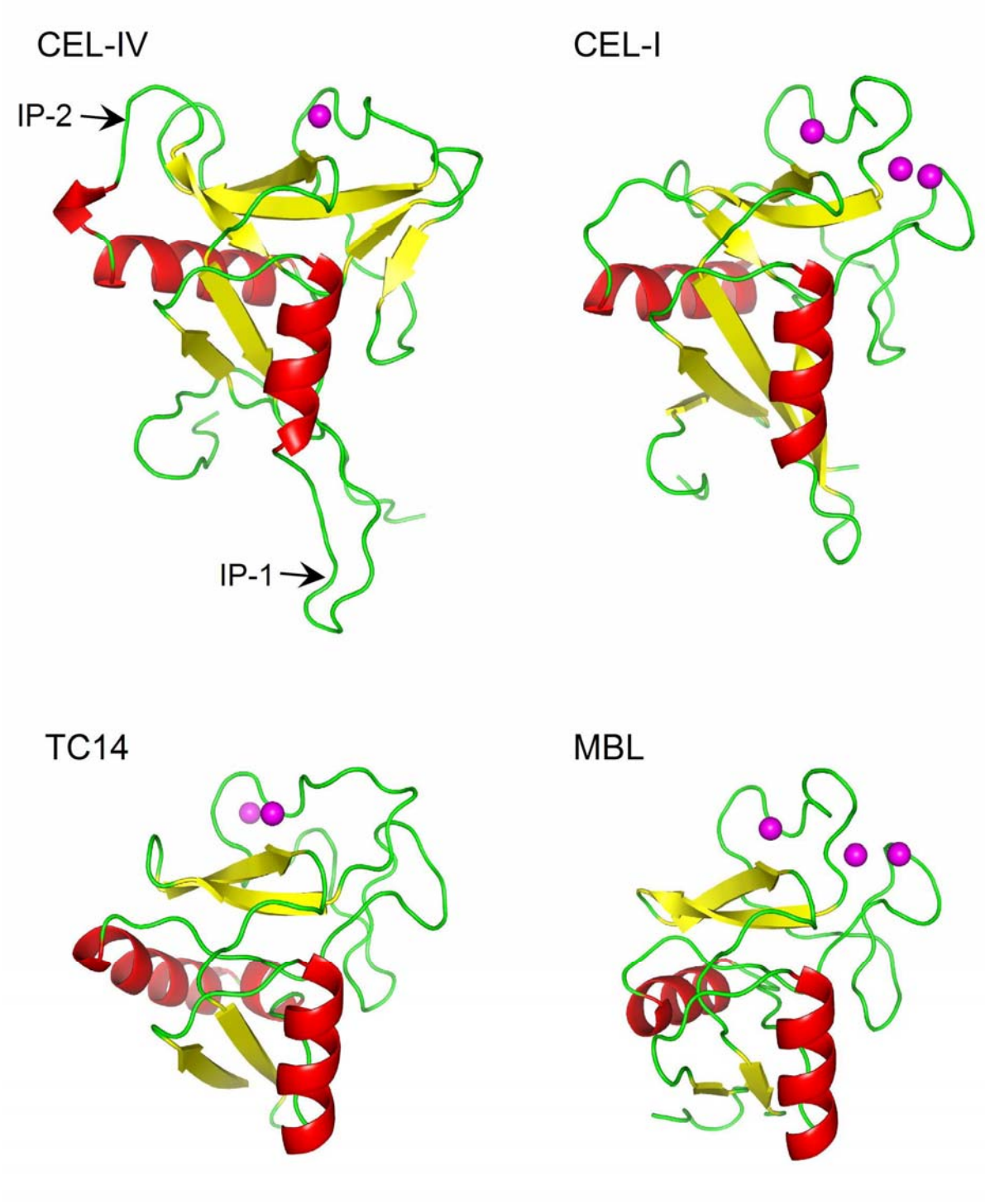

Fig. 3. **Comparison of the main chain structures of CEL-IV, CEL-I, TC14, and MBL.** The α-helices and β-sheets are shown in *red* and *yellow*, respectively. The Ca^{2+} ions are shown as *magenta spheres*. *IP-1* and *IP-2* indicate the insertion peptides shown in Fig. 2.

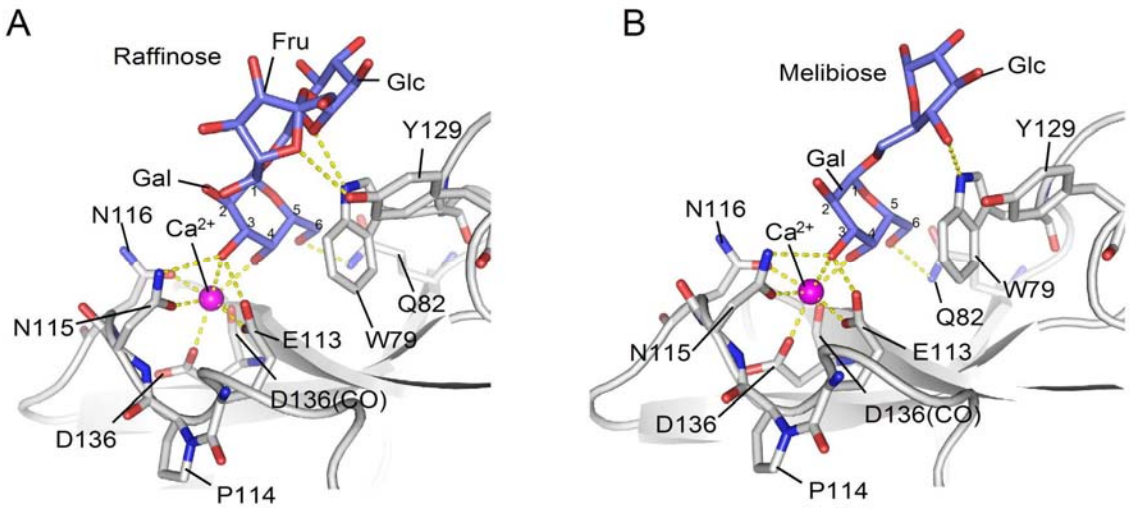

Fig. 4. **Carbohydrate-binding modes to CEL-IV.** *A, B,* carbohydrate-binding modes of raffinose (*A*) and melibiose (*B*) to CEL-IV. The coordinate bonds and hydrogen bonds are shown as *yellow dotted lines*.

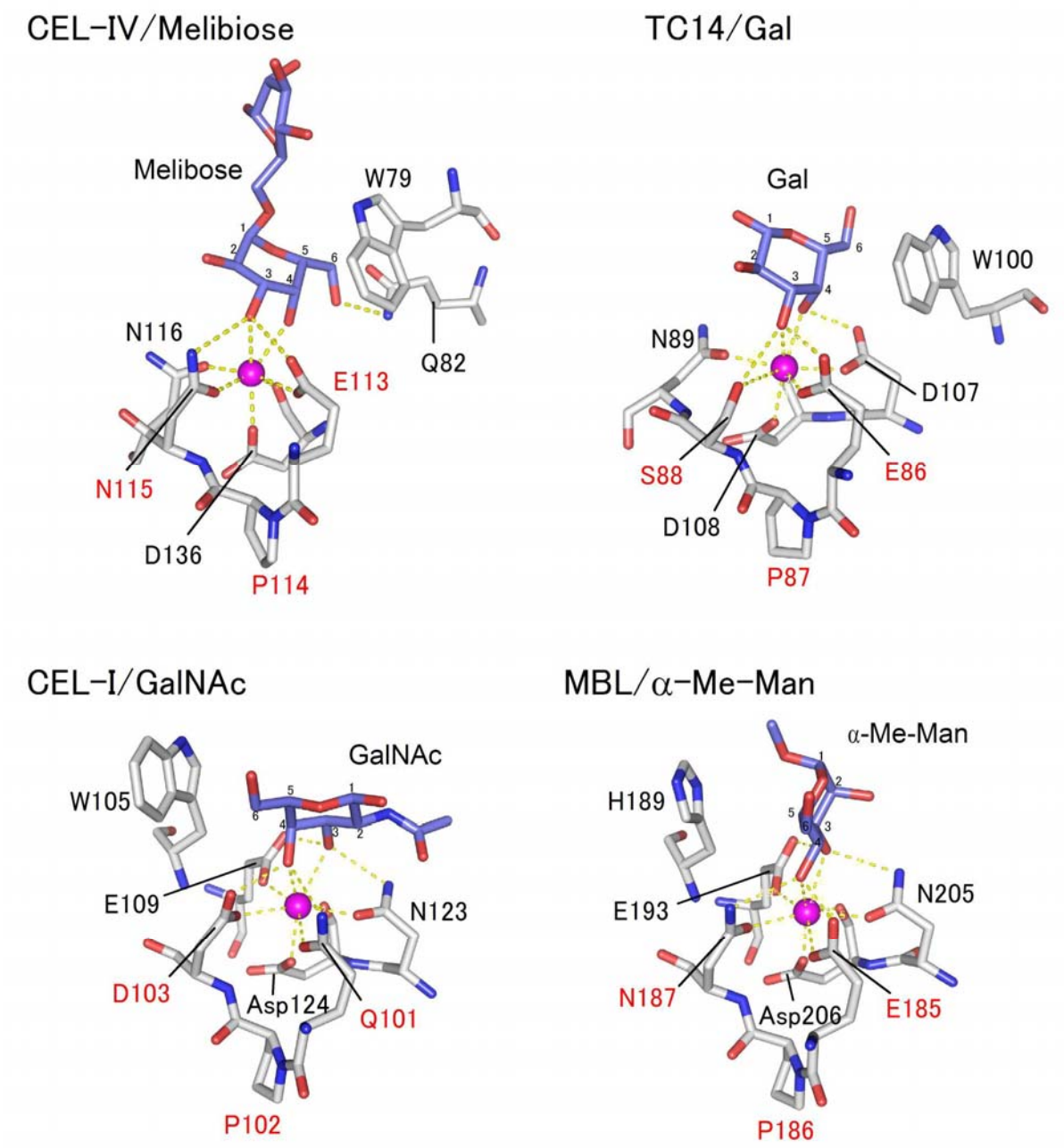

Fig. 5. **Residues involved in the binding of specific carbohydrates to CEL-IV, CEL-I, TC14, and MBL.** The EPN, EPS, and QPD sequences are shown in *red*.

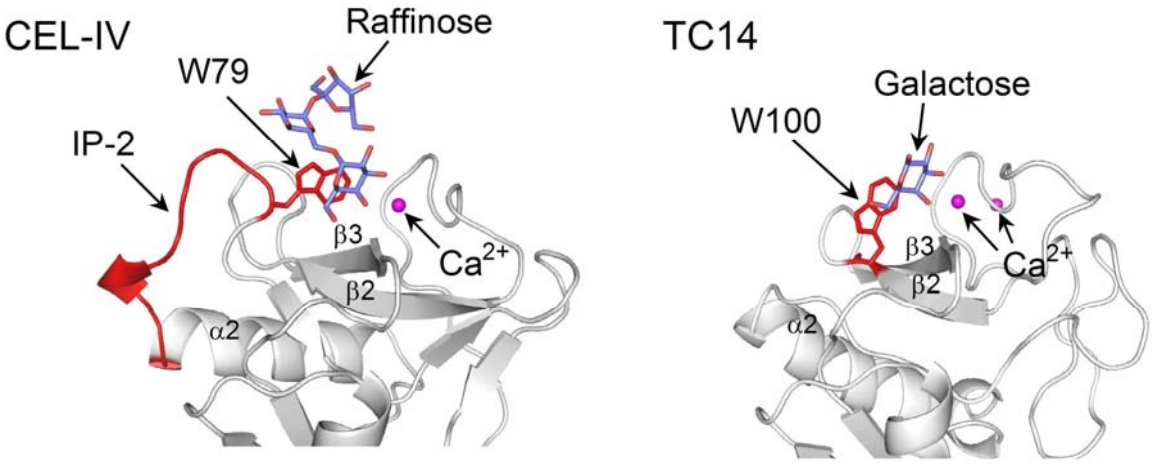

Fig. 6. **Positions of the tryptophan residues involved in stacking interactions with the galactose residues in CEL-IV and TC14.** The tryptophan residues and IP-2 region (CEL-IV) are shown in *red*.

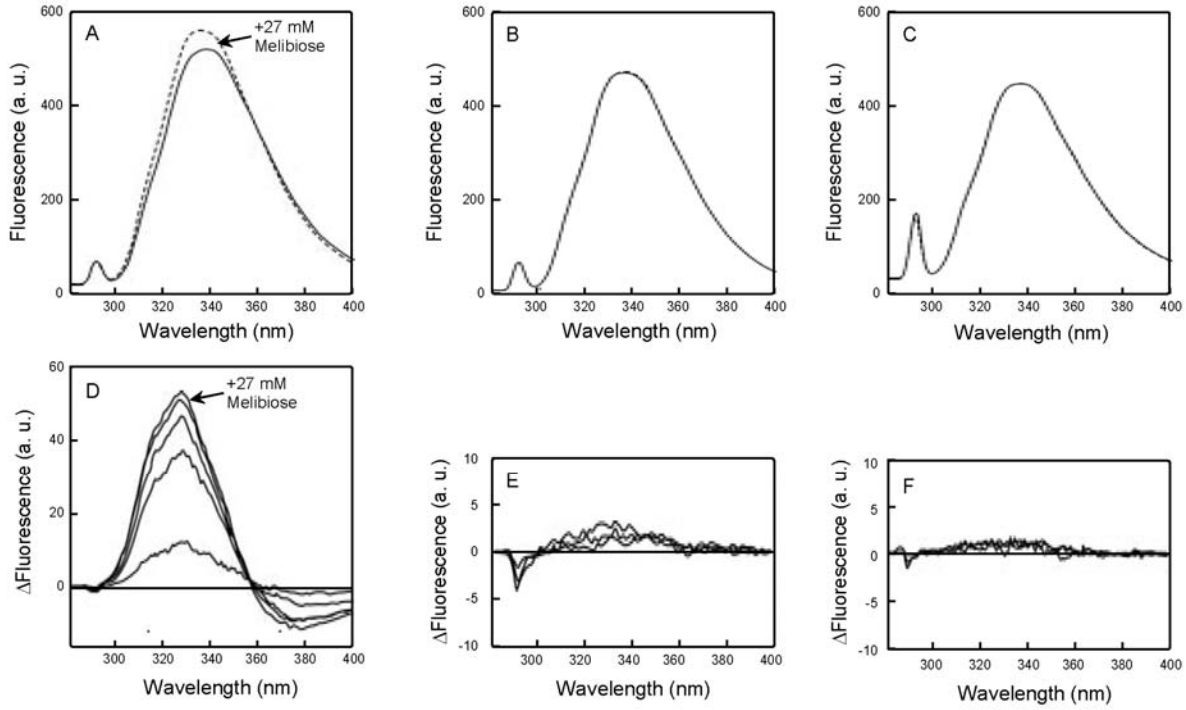

Fig. 7. **Fluorescence changes of CEL-IV and its mutants upon binding of melibiose.** *A–F*, fluorescence spectra in the absence (*solid lines*) and presence (*dashed lines*) of melibiose (*A–C*) and the difference spectra of WT CEL-IV (*A* and *D*), W79Y (*B* and *E*), and W79H (*C* and *F*). The excitation wavelength was 290 nm. *a.u.*, arbitrary units.

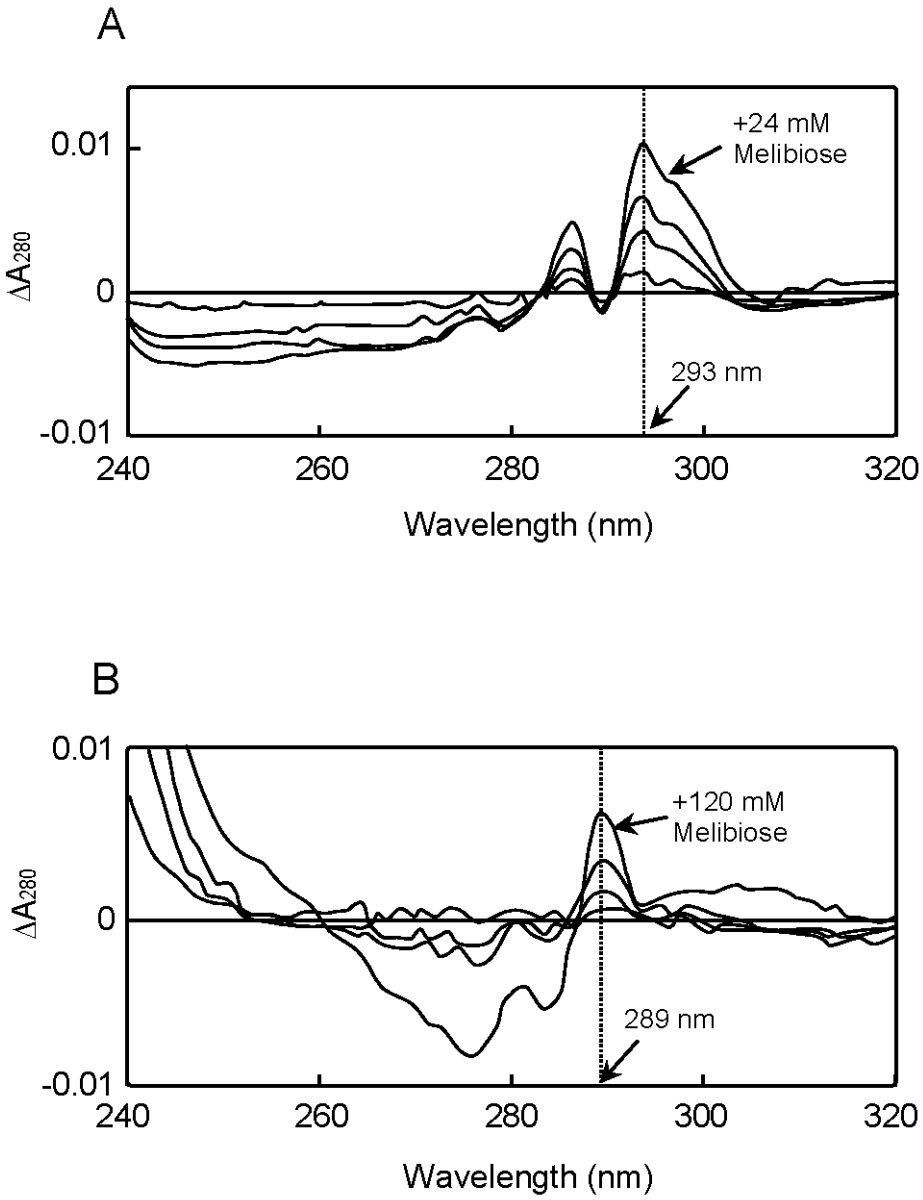

Fig. 8. **UV-difference spectra of WT CEL-IV and its W79Y mutant induced by binding of melibiose.** *A,* WT. *B,* W79Y mutant.

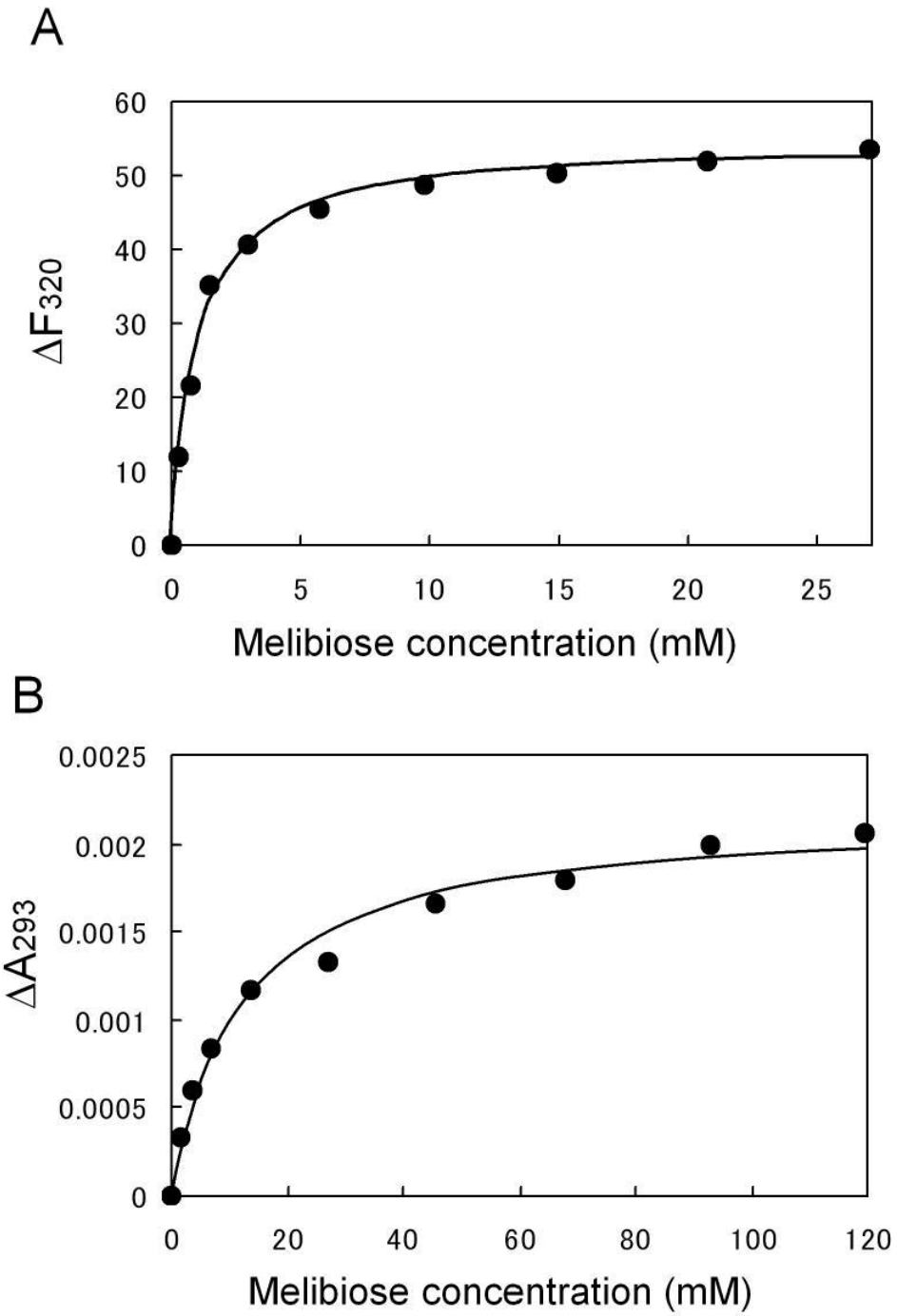

Fig. 9. **Variations of the changes in the fluorescence and UV-difference spectra of CEL-IV upon binding of melibiose.** The changes in the fluorescence intensity at 320 nm and absorbance at 293 nm are plotted against the melibiose concentrations. The *solid lines* show the theoretical curves.

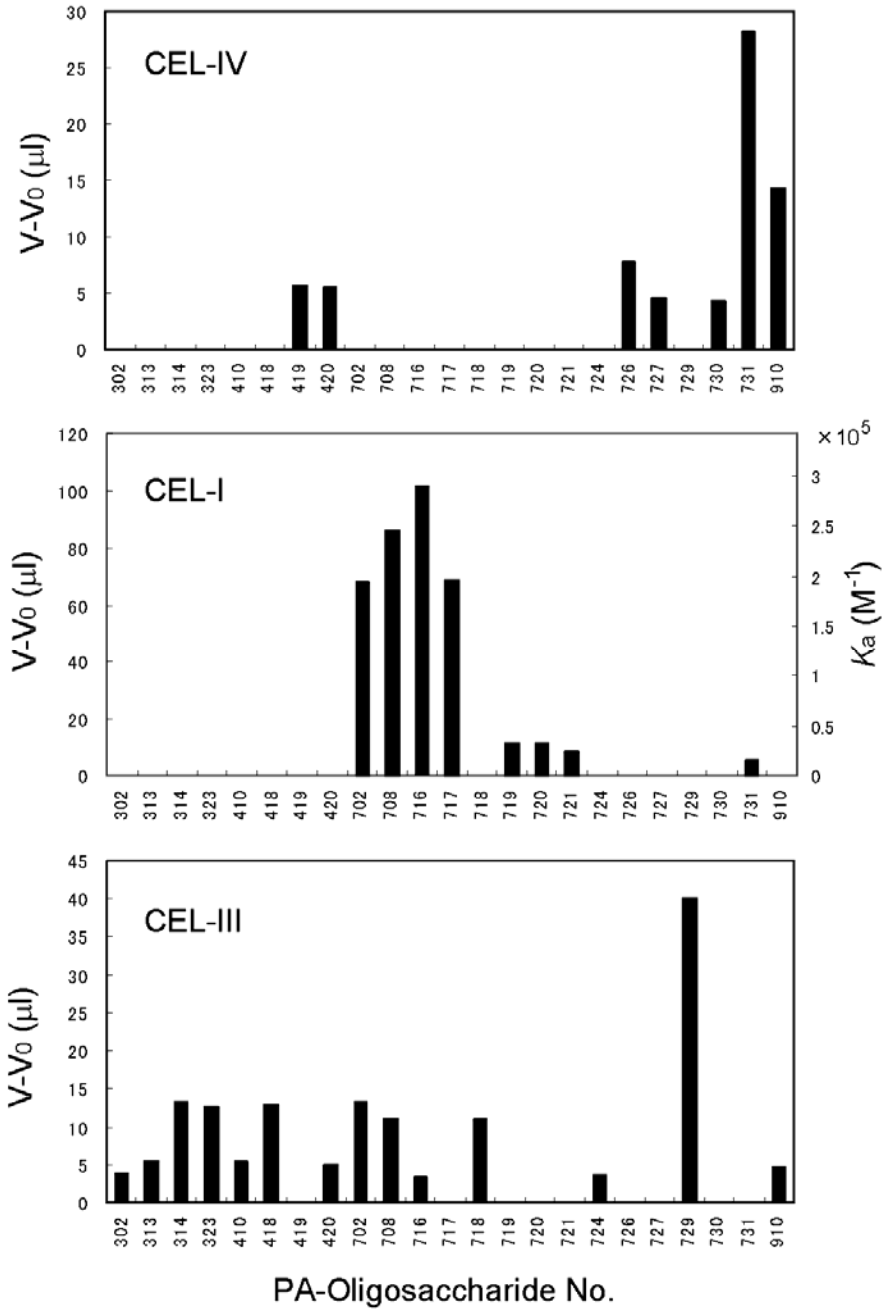

Fig. 10. **Affinities of PA-oligosaccharides for CEL-IV assessed by frontal affinity chromatography.** The analysis was performed using 93 oligosaccharides (22). The affinities of the oligosaccharides are represented by the retardation of their elution ($V–V_0$) values. The oligosaccharides are indicated by *numbers* that correspond to those shown in Fig 11, and are the same as in the previous report (22). In the case of CEL-I, the affinities are also represented as association constant (K_a) values on the *right scale*. Oligosaccharides with $V–V_0$ values of <5 μl (CEL-I) or <3 μl (CEL-III and CEL-IV) were classified as showing no interactions.

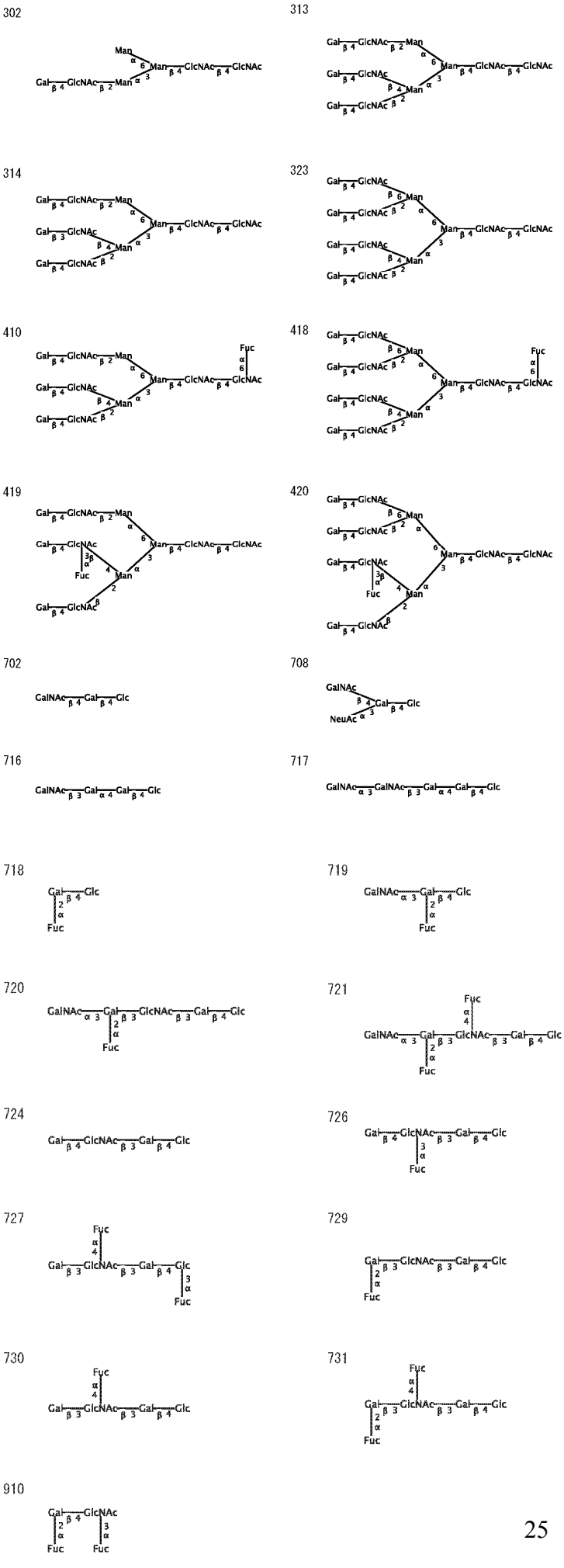

Fig. 11. **Structures of the PA-oligosaccharides used in the frontal affinity chromatography.** The images were drawn using the website GlycomeDB (www.glycome-db.org).

Table 1. Data collection and refinement statistics

	Native	CEL-IV/melibiose complex	CEL-IV/raffinose complex
Data collection			
X-ray source	Photon Factory, BL6A	Photon Factory, NW12	Photon Factory, BL17A
Wavelength (Å)	0.978	1.000	1.000
Resolution (Å)	47.9 - 3.00 (3.16 - 3.00)	41.9 - 2.50 (2.59 - 2.50)	102 - 1.65 (1.71 – 1.65)
Space group	$P6_5$	$P2_1$	$P2_1$
Unit cell (Å)	$a = b = 86.07$, $c = 375.29$ $\alpha = \beta = 90.00°$, $\gamma = 120.00°$	$a = 43.99$, $b = 78.50$, $c = 102.90$ $\alpha = \gamma = 90.00°$, $\beta = 98.41°$,	$a = 41.81$, $b = 78.16$, $c = 102.88$ $\alpha = \gamma = 90.00°$, $\beta = 96.82°$
$I/\sigma(I)$	29.4 (8.6)	30.2 (7.0)	29.4 (2.9)
Redundancy	11.4 (11.1)	3.5 (3.3)	3.6 (3.3)
Completeness (%)	99.9 (99.5)	95.6 (92.4)	95.0 (85.5)
R_{merge}	0.077 (0.298)	0.109 (0.275)	0.049 (0.336)
Total reflections	359264	79693	74625
Unique reflections	31417	22552	266904
Refinement statistics			
Resolution (Å)	43.0 - 3.0	38.1 - 2.5	40.2 - 1.65
R_{work} (%)	19.6	22.9	16.5
R_{free} (%)	21.5	27.9	20.3
Model statistics			
r. m. s. deviation			
Bond length (Å)	0.024	0.016	0.015
Bond angles (deg)	2.129	1.737	1.644

The values in parentheses are for the highest resolution shell.

Table 2. Association constants for the binding of specific carbohydrates to recombinant CEL-IV obtained from the changes in the fluorescence and UV-absorption spectra.

Recombinant CEL-IV	Association constant ($\times 10^3$ M^{-1})			
	GalNAc		Melibiose	
	Fluorescence	UV-absorption	Fluorescence	UV-absorption
WT	6.4	5.3	0.84	0.33
W78Y	n.d.	0.26	n.d.	0.083

n.d.: not determined.