Characteristic Recognition of N-Acetylgalactosamine by an Invertebrate C-type Lectin, CEL-I, Revealed by X-ray Crystallographic Analysis*

Received for publication, August 3, 2004, and in revised form, August 19, 2004 Published, JBC Papers in Press, August 19, 2004, DOI 10.1074/jbc.M408840200

Hajime Sugawara‡‡, Masami Kusunoki‡‡, Genji Kurisu‡‡, Tokiko Fujimoto, Haruhiko Aoyagi, and Tomomitsu Hatakeyama**

From the ‡‡Research Center for Structural and Functional Proteomics, Institute for Protein Research, Osaka University, 3-2 Tamada-ocho, Suita, Osaka 565-0871, Japan, the ¶Laboratory for Communication Mechanisms, Research Center for Structural and Functional Proteomics, Institute for Protein Research, RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 220-0045, Japan, and the **Department of Applied Chemistry, Faculty of Engineering, Nagasaki University, 14-1 Bunkyo-machi, Nagasaki 852-8521, Japan

CEL-I is a C-type lectin, purified from the sea cucumber Cucumaria echinata, that shows a high specificity for N-acetylgalactosamine (GalNAc). We determined the crystal structures of CEL-I and its complex with GalNAc at 2.0 and 1.7 Å resolution, respectively. CEL-I forms a disulfide-linked homodimer and contains two intramolecular disulfide bonds, although it lacks one intramolecular disulfide bond that is widely conserved among various C-type carbohydrate recognition domains (CRDs). Although the sequence similarity of CEL-I with other C-type CRDs is low, the overall folding of CEL-I was quite similar to those of other C-type CRDs. The structure of the complex with GalNAc revealed that the basic recognition mode of GalNAc was very similar to that for the GalNAc-binding mutant of the mannose-binding protein. However, the acetamido group of GalNAc appeared to be recognized more strongly by the combination of hydrogen bonds to Arg115 and van der Waals interaction with Gln70. Mutational analyses, in which Gln70 and/or Arg115 were replaced by alanine, confirmed that these residues contributed to GalNAc recognition in a cooperative manner.

Carbohydrate-binding proteins (lectins) have been shown to play important roles in molecular recognition processes in various tissues and body fluids. There are increasing numbers of animal lectins (1, 2), which are categorized into several families according to their structural homologies. Various proteins that display carbohydrate-binding activity in a Ca2+-dependent manner are classified into the C-type lectin family (3). They contain C-type carbohydrate-recognition domains (CRDs)1 composed of 110–130 amino acid residues in common. In vertebrate C-type lectins, the C-type CRDs mostly occur as carbohydrate-binding modules linked to other domains with distinct functions. They fall into seven groups (4) and function through cooperation of the CRDs with other individual domains. In contrast, invertebrate C-type lectins are mostly single-domain proteins. One of their possible roles is thought to be inactivating or opsonizing of foreign microorganisms in place of immunoglobulins in vertebrates. There are only limited numbers of C-type CRDs whose tertiary structures have been determined. Mannose-binding protein (MBP) was the first C-type lectin for which a crystal structure was determined (5). This protein contains an N-terminal collagenous domain, followed by a link domain and a C-type CRD. The latter two domains have been expressed in a recombinant form, and the structure as well as various mutant forms have been analyzed by x-ray crystallography (6–10). In addition to MBP, the crystal structures of CRDs complexed with carbohydrates have been solved for DC-SIGN, DC-SIGNR (11), P- and E-selectins (12), tunicate lectin TC14 (13), and rattlesnake venom lectin RSL (14). In addition to these lectins, several C-type lectin-like domains (CTLD) without carbohydrate-binding ability have also been found (2). These lack the essential residues for carbohydrate binding, and instead many are thought to function as receptors for noncarbohydrate ligands. Although these proteins share only slight sequence similarity (~20%), their tertiary structures are basically similar to each other (15).

We previously isolated four Ca2+-dependent galactose/N-acetylgalactosamine-specific lectins (CEL-I, -II, -III, and -IV) from the marine invertebrate Cucumaria echinata (Holothuroidea) (16). Based on their amino acid sequences, CEL-I and CEL-IV are clearly categorized into the C-type lectin family, whereas CEL-III is a novel Ca2+-dependent lectin with strong hemolytic activity as well as cytotoxicity (17–20). CEL-III shows sequence similarity with β-trefoil lectins, such as the B-chains of ricin and abrin (21). The similarity of the three-dimensional structure of CEL-III with these proteins has recently been revealed by x-ray crystallographic analysis (22). CEL-I, which is the smallest lectin in C. echinata, is composed of two identical 16-kDa subunits linked by a single interchain disulfide bond (23). CEL-I has very high specificity for N-acetylgalactosamine (GalNAc), and its affinity for GalNAc is ~1000-fold stronger than that for galactose as judged by hemagglutination inhibition assays. To date, no other C-type lectins with such high specificity for GalNAc have been identified.

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

The atomic coordinates and structure factors (codes 1WMY and 1WMZ) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ Present address: Dept. of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan.

** To whom correspondence should be addressed. Tel.: 81-95-819-2686; Fax: 81-95-819-2684; E-mail: thats@net.nagasaki-u.ac.jp.

1 The abbreviations used are: CRD, carbohydrate recognition domain; CTLD, C-type lectin-like domain; MBP, mannose-binding protein; MPD, 2-methyl-2,4-pentanediol; rCEL-I, recombinant CEL-I; DC-SIGN, dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin; DC-SIGNR, DC-SIGN-related molecule; TBS, Tris-buffered saline; r.m.s.d., root mean square deviation.
It is therefore of great interest to clarify the CEL-I-specific carbohydrate-recognition mechanism not only to understand the carbohydrate-recognition mechanism of C-type lectins but also the structural basis to design molecules with novel carbohydrate-specific affinity. We previously obtained single crystals of CEL-I that were suitable for x-ray diffraction studies (24). Recently, we also succeeded in crystallizing the CEL-I-GalNac complex. In this report, we describe x-ray crystallographic analyses of CEL-I and the CEL-I-GalNac complex to elucidate the mechanism of its high specificity for GalNAc along with the site-directed mutagenesis experiments of a recombinant CEL-I (rCEL-I) that was expressed in *Escherichia coli* cells.

**EXPERIMENTAL PROCEDURES**

**Purification of CEL-I**—CEL-I was purified from *C. echinata* according to a previously reported method (19). The proteins extracted from a homogenate of *C. echinata* were applied to a lactose-Cellulofine column equilibrated with 0.15 M NaCl, 10 mM Tris-HCl, pH 7.5, (TBS) containing 10 mM CaCl₂. Adsorbed lectins (CEL-I, -III, and -IV) were eluted with TBS containing 20 mM EDTA. The lectins were then separated using a GalNac-Cellulofine column utilizing the differences in their carbohydrate-binding specificities. After elution of CEL-III with TBS containing 0.1 M lactose, CEL-I and CEL-IV were eluted with TBS containing 20 mM EDTA. CEL-I and CEL-IV were finally separated by gel filtration through Sephadex G-75 in TBS.

**Crystallizations and Data Collection**—The purification and crystallization of native CEL-I have already been reported (24). Briefly, the protein solution (5 mg/ml, 2–4°C) was subjected to hanging or sitting drop vapor diffusion at 20°C. X-ray data collection was performed on beamline BL-18B at the High Energy Acceleration Research Organization, Tsukuba, Japan, using an ADSC Quantum 4R CCD camera (25). The space group was monoclinic C2 with unit cell parameters of *a* = 92.4, *b* = 99.9, *c* = 76.7 Å, and *β* = 136.5°. Reduction of a total of 72,844 reflections from CEL-I crystals yielded 19,924 independent reflections with 93.2% completeness for a resolution range of 52.7–2.0 Å and an overall *R*<sub>merge</sub> of 6.1%. CEL-I-GalNac complex crystals were prepared under essentially the same conditions as the native crystals except for the addition of 1 mM GalNac to the samples. Because the complex crystals were initially formed as spherulites, these were used as seeds for a microseeding method. Single crystals of the CEL-I-GalNac complex grew to sufficient sizes for data collection within one month. Data collection was performed at 100 K using a Rigaku R-AXIS VII imaging plate area detector equipped with a Rigaku FR-E SuperBright rotating-anode generator. Image data were processed by the programs MOSFLM (26) and SCALA (27). The space group and unit cell parameters of the CEL-I-GalNac crystals were determined to be P2₁, and *a* = 39.6, *b* = 52.2, *c* = 136.6, and *β* = 91.7°, respectively. Assuming two molecules of CEL-I, each composed of a homodimer in the asymmetric unit, the value of the Matthews constant, *V*<sub>M</sub> (28), is 2.20 Å<sup>3</sup> Da⁻¹ corresponding to a solvent content of 44.2%. The final data for the CEL-I-GalNac complex crystals comprised 59,176 independent reflections with 96.3% completeness for a range of 33.8–1.7 Å and an overall *R*<sub>merge</sub> of 6.2%. The completeness and *R*<sub>merge</sub> for the highest resolution shell (1.70–1.70 Å) were 84.9 and 16.2%, respectively.

**Structure Determination and Refinement**—The crystal structures of native CEL-I and the CEL-I-GalNac complex were solved by the molecular replacement method. The search model for native CEL-I was based on the human lithostatine structure (PDB code 1JLT) (29), which shares 28% amino acid sequence identity with CEL-I. Molecular replacement calculations were performed using the program AMoRe (30) from the CCP4 suite (26). The model was refined using the program CNS (31) with noncrystallographic symmetry restraints. Five percent of the reflections were set aside for *R*<sub>free</sub> calculations (32). Manual fitting of the model was carried out by the program Xfit (33). The final *R* and *R*<sub>free</sub> factors for all reflections between 52.7- and 2.0-Å resolution for native CEL-I were 0.140 and 0.179, respectively. The structure of the CEL-I-GalNac complex was also solved by the molecular replacement method using the native CEL-I coordinates as a search model. The positions and orientations of GalNac in the complex were clearly shown by the *F*<sub>c</sub> - *F*<sub>c</sub> difference Fourier map with a contour level of 3σ. The model was finally refined by the program REFMAC (34) without non-crystallographic symmetry restraints. The final *R* and *R*<sub>free</sub> factors between 33.8- and 1.7-Å resolution for the CEL-I-GalNac complex were 0.158 and 0.186, respectively. The quality of the final models for native CEL-I were 0.140 and 0.179, respectively. The structure of the CEL-I-GalNAc complex was also solved by the molecular replacement method using the native CEL-I coordinates as a search model. The positions and orientations of GalNAc in the complex were clearly shown by the *F*<sub>c</sub> - *F*<sub>c</sub> difference Fourier map with a contour level of 3σ. The model was finally refined by the program REFMAC (34) without non-crystallographic symmetry restraints. The final *R* and *R*<sub>free</sub> factors between 33.8- and 1.7-Å resolution for the CEL-I-GalNAc complex were 0.158 and 0.186, respectively. The quality of the final models for native and complexed CEL-I was assessed by Ramachandran plots and analysis of the model geometry with the program PROCHECK (35). The crystallographic statistics are summarized in Table I.

### Table I

**Refinement statistics**

|                  | Native | GalNAc complex |
|------------------|--------|----------------|
| Resolution (Å)   | 53–2.0 | 34–1.7         |
| Protein molecules| 11     | 2              |
| Protein atoms    | 2268   | 4536           |
| Number of water molecules | 247 | 400           |
| Number of reflections | 19,924 | 59,176       |
| Working set      | 1012   | 3012           |
| Test set for *R*<sub>free</sub> | 14.0 | 16.0          |
| *R*<sub>work</sub> (%) | 13.8 | 15.8          |
| *R*<sub>free</sub> (%) | 17.9 | 18.6          |
| r.m.s.d. from standard geometry | 0.022 | 0.016         |
| Bond length (Å)  | 1.70   | 1.60           |
| Ramachandran plot| 86.7   | 86.9           |
| Residues in most favored regions (%) | 13.3 | 13.1          |
| Residues in generously allowed regions (%) | 0 | 0             |

**FIG. 1. Stereo view of the ribbon model of CEL-I (protomer A).** The α-helices, β-strands, loops, and intramolecular disulfide bonds are shown in purple, blue, gray, and green, respectively. Bound calcium ions are shown by cyan spheres. Secondary structures were determined by the program PROMOTIF (45). Fig. 1 was drawn by the program MOLSCRIPT (46) and rendered by the program Raster3D (47).
used to transform *E. coli* BL21(DE3)pLysS (Novagen). The proteins were expressed as inclusion bodies, which were solubilized with 20 mM Tris-HCl buffer, pH 8.5, containing 8 M urea and 1% 2-mercaptoethanol. The solubilized protein was purified by chromatography through a DEAE-Cellulofine column in 20 mM Tris-HCl, pH 8.5, containing 4 M urea and elution with a NaCl gradient from 0 to 0.5 M. The rCEL-I was pooled and dialyzed against TBS, and the active protein was separated by affinity chromatography through a GalNAc-Cellulofine column (18). DNAs encoding mutant proteins, in which Arg115 and/or Gln70 were replaced by Ala (R115A, Q70A and Q70A/R115A), were prepared by polymerase chain reaction using two oligonucleotides (30mers) containing the mutation site and the plasmid containing wild-type rCEL-I DNA as a template. The mutant proteins were purified by the same procedure used for the wild-type protein. The N-terminal sequences of these proteins were confirmed using a Shimadzu PPSQ-21 protein sequencer.

**Hemagglutination Assay—**Serial 2-fold dilutions of a sample (50 μl) were mixed with the same volume of a 5% (v/v) suspension of rabbit erythrocytes in the wells of round bottomed microtiter plates. Incubation was performed in TBS containing 10 mM CaCl₂. The extent of agglutination was examined visually after incubation for 1 h at room temperature. The hemagglutinating activity was expressed as a titer, i.e. the reciprocal of the highest dilution producing detectable agglutination. Hemagglutination inhibition assays were performed by incubating 50-μl aliquots of the protein solutions (titer 2) containing various concentrations of carbohydrates with the same volume of a 5% (v/v) suspension of rabbit erythrocytes in TBS containing 10 mM CaCl₂.

**RESULTS AND DISCUSSION**

The crystal structure of CEL-I was solved by the molecular replacement method using human lithostathine, which contains a CTLD (29), as a search model. Native CEL-I is composed of one disulfide-linked homodimer (protomers A and B) in each asymmetric unit (Fig. 1). As shown in other C-type CRDs, the three-dimensional structure of CEL-I consists of two parts, a lower part composed of two α-helices (α1 and α2) and four β-strands (β0, β1, β1′, and β5), and an upper part composed of four β-strands (β2, β2′, β3, and β4), the latter of which contains the calcium-binding sites (Figs. 1 and 2). Regardless of the low sequence similarity (16–30%) with other C-type CRDs and CTLDs, structural comparison with known three-dimensional structures by DALI (37) revealed that CEL-I is very similar to long form CTLDs, such as lithostathine (Z = 21.6, r.m.s.d. of 1.7 Å for 128 Ca atoms), the H1 subunit of the asialoglycoprotein receptor (Z = 19.6, r.m.s.d. of 1.3 Å for 118 Ca atoms), and tetraneckten (Z = 19.5, r.m.s.d. of 2.1 Å for 132 Ca atoms). CEL-I also shows similarity with short form CTLDs, such as MBP-A (Z = 17.5, r.m.s.d. of 2.1 Å for 123 Ca atoms) and tunicate lectin TC14 (Z = 15.8, r.m.s.d. of 2.2 Å for 117 Ca atoms). Compared with other CRDs, CEL-I has an
Crystal Structure of a GalNAc-specific C-type Lectin, CEL-I

Fig. 4. Homodimeric structures of CEL-I and TC14. A, CEL-I molecule. B, TC14 molecule. The colors are the same as those described in the legend to Fig. 1. In the top figures, pseudo 2-fold axes pass through the center of the two protomers perpendicular to the paper. The bottom figures were obtained by a 90-degree rotation around the horizontal line at the center of the top figures. Fig. 4 was drawn by the program MOLSCRIPT (46) and rendered by the program Raster3D (47).

insertion (residues 36–41) between α1 and β1′ containing a 3₁₀ helix (residues 37–39) (Fig. 3).

There are two intramolecular disulfide bonds in CEL-I (Cys²⁻Cys¹⁴ and Cys¹³⁵–Cys¹⁴⁵). The bond between Cys¹ and Cys¹⁵ is equivalent to a widely conserved disulfide bond among CTLDs, whereas that between Cys² and Cys¹⁴ is only observed in long form C-type CRDs (38). However, CEL-I does not have a disulfide bond connecting β3 and the loop between β4 and β5, which is well conserved in CTLDs. In CEL-I, the corresponding cysteine residues are replaced by Tyr¹¹ and Ala¹²⁷ (Fig. 3). The space resulting from the lack of this disulfide bond is filled by Tyr¹¹ι, which may contribute to the stabilization of the protein structure instead of the corresponding disulfide bond.

The two identical protomers of CEL-I are linked by an interchain disulfide bond between the Cys²⁸ residues (Fig. 4A). In addition to this disulfide bond, there are several interactions between the two subunits around the N-terminal (residues 1–4), C-terminal (residues 137–140), and insertion loop (residues 36–41) (Fig. 5). Eight hydrogen bonds are present between the two protomers (Gln²⁻Gln², Gln²⁻Gly¹, Glu¹⁴₀-Thr¹⁵, Tyr¹⁴ (protomer A)–Thr¹⁴ (protomer B) and Thr¹⁹ (protomer A)–Ser¹⁹ (protomer B)). In particular, Gln² appears to be especially important for dimeric interactions, because the Oε₁ and Nε₂ atoms of its side chain in one subunit make hydrogen bonds to the main-chain amide group of Gln² and the main-chain carbonyl group of Gly¹ in the other subunit, respectively (Fig. 5A). There are also hydrophobic interactions between the two subunits through Pro⁴, Leu⁹⁹, Val¹⁴¹, Leu¹³⁷, and Phe¹⁵⁹ (Fig. 5B). Among the available C-type lectin crystal structures, only tunicate lectin TC14 has a homodimeric structure composed of two C-type CRDs (38). The two subunits of the TC14 homodimer associate with each other through noncovalent interactions between β-strands and α-helices (Fig. 4B), because two N-terminal β-strands (β1) make an anti-parallel β-sheet, and two α-helices (α2) associate by hydrophobic interactions. Recently, another C-type CRD linked by intermolecular disulfide bonds has been reported in the decameric structure of rattlesnake venom lectin RSL (14). Meanwhile, the coagulation factor-binding proteins also form a disulfide-linked dimer composed of CTLDs, which have long loops to make dimeric interactions (40, 41).

Three Ca²⁺ ions (designated CA1, CA2, and CA3) are present in each subunit (Fig. 2). They are bound in a similar way to those of MBP-A (6), lung surfactant protein D (42), DC-SIGN, and DC-SIGNR (11). CA1 is coordinated by the side chains of Asp⁷, Glu⁸¹, Asn¹⁰⁴, and Asp¹¹⁰, the main-chain carbonyl oxygen of Glu¹⁰⁹, and one molecule of water (Fig. 2A). CA2 is coordinated by the side chains of Glu¹⁰⁹, Asp¹¹⁰, Asp¹¹³, and Asp¹¹⁴ and the main-chain carbonyl oxygen of Asp¹²⁴ (Fig. 2B). In addition, two hydroxyl groups of MPD, used as a precipitant for crystallization, coordinate with CA2. They also form hydrogen bonds with the side chains of Glu¹⁰⁹, Asp¹¹⁰, Glu¹⁰⁹, and Asn¹²³, and the main-chain carbonyl oxygen of Asp¹²⁴. CA3 shares the side chains of Glu⁸¹ and Asp¹¹⁰ with CA1, and the other ligands for CA3 are the main-chain carbonyl group of Gly⁹⁸ in the symmetry-related protomer B (only for CA3 in protomer A) and three water molecules (Fig. 2A).

There are two CEL-I homodimer molecules (the subunit structure is designated 1A/1B and 2A/2B for each molecule) in the asymmetric unit of the CEL-I-GalNAc complex crystal. Four bound GalNAc molecules coordinate to CA2 with their 3-OH and 4-OH in place of the two hydroxyl groups of MPD found in the native CEL-I crystal (Fig. 6A). These hydroxyl groups also form hydrogen bonds with the side chains of Glu¹⁰⁹, Asp¹¹⁰, Glu¹⁰⁹, and Asn¹²³. This binding mode is basically consistent with those of the carbohydrate complexes of MBP-C (9) and the GalNAc-binding mutant of MBP (10) (Fig. 6B). Although the indole rings of Trp¹⁰⁵ in protomers 1B and 2B are flipped relative to those in protomers 1A and 2A, because of the hydrogen-bonding with the main-chain carbonyl oxygen of Glu¹⁰⁹ in the symmetry-related molecule of protomers 1B and 2B, the Trp¹⁰⁵ residues in all protomers make van der Waals contact with the C-6 atom of GalNAc, thereby stabilizing the binding of GalNAc. Tryptophan residues have been found to stabilize the binding of carbohydrates in the Gal/GalNAc-recognizing C-type CRDs (7, 13). However, the interaction between Trp¹⁰⁵ and the C-6 atom of GalNAc appears to be relatively weak in the case of CEL-I, compared with that of Trp¹⁰⁵ in the GalNAc-binding mutant of MBP, which makes van der Waals contacts with the C-3, C-4, C-5, and C-6 carbon atoms of GalNAc (Fig. 6B). In the mutant MBP, five residues (192–196) inserted as a glycine-rich loop (7) appear to lead to closer contact between Trp¹⁰⁵ and GalNAc.

In the case of the CEL-I-GalNAc complex, there are two hydrogen bonds between the guanidium group of Arg¹¹⁵ and the carbonyl oxygen of the acetamide group of GalNAc (Fig. 6C). An arginine residue (Arg¹²²) corresponding to Arg¹¹⁵ in CEL-I has been found in starfish (Asterina pectinifera) lectin, which also specifically binds to GalNAc (43). In addition, the...
Crystal Structure of a GalNAc-specific C-type Lectin, CEL-I

The affinity binding of GalNAc to CEL-I. CEL-I binds to these interactions are assumed to be closely related to the high affinity of the bound GalNAc and Asp83 in protomer 2A. In protomers 1A and 2A, the terminal NH2 group of the side chain of Arg115 makes van der Waals contact with Gln70 as mentioned above. Therefore, it appears that Gln70 may play a similar role to Asn123, which is essential as a ligand for CA2, in the GalNAc-binding mutant of MBP. On the other hand, other galactose and galactose-containing carbohydrates showed only slight changes in their inhibitory profiles, compared with the wild-type. This confirms that Arg115 and Gln70 make significant contributions to the recognition of GalNAc. In particular, Gln70 appears to be more important than Arg115 despite hydrogen bonds between the latter and GalNAc. As shown in Table II, the Q70A/R115A mutant still had the same affinity for Gal as well as a moderately high affinity for GalNAc. This indicates that Gln70 and Arg115 have little influence on the recognition of the galactose moiety and also suggests that other residue(s), such as Asn123, may be involved in recognizing the acetamido group of GalNAc in addition to Gln70 and Arg115.

A GalNAc-binding mutant of MBP was constructed based on sequence comparison with rat hepatic lectin, which shows higher affinity for GalNAc than for Gal (44), and its complex structure with GalNAc was reported (10). Its preferential affinity for GalNAc compared with that for Gal is achieved by van der Waals contact between the methyl group of the acetamido group of GalNAc and the Ce-1 and Ne-2 atoms of His202, which was introduced in place of threonine, the corresponding amino acid in rat hepatic lectin. This MBP mutant attained a 60-fold higher affinity for GalNAc than for Gal, indicating that such van der Waals interactions are very important. As shown in Fig. 6, B and C, the orientation of GalNAc bound to CEL-I is very similar to that in the GalNAc-binding mutant of MBP. In the case of CEL-I, the methyl group of the acetamide of GalNAc makes van der Waals contact with Gln70 as mentioned above. Therefore, it appears that Gln70 may play a similar role to His202 in the GalNAc-binding mutant of MBP. On the other hand, Gln70, which is located at the corresponding position to His202 in the MBP mutant, may have the potential to interact with the methyl group of GalNAc, although a mutational analysis investigating this was not performed in the present study.

CEL-I exhibits strong cytotoxicity (36), which suggests an actual biological role for this protein as a defense toxin against predators. The toxicity is inhibited in the presence of GalNAc suggesting that it is mediated by binding to specific carbohydrate chains on the cell surface. One probable mechanism is that the binding of CEL-I to cell surface carbohydrate chains triggers intracellular signaling, leading to cell death. An identification of the natural carbohydrate ligands for CEL-I that are present on the target cell surface seems to be important for understanding the cytotoxicity at the molecular level. Further investigation of the involvement of the residues around the carbohydrate-binding site of CEL-I using recombinant proteins should provide important clues regarding the recognition mechanism for natural ligands on the target cell surface.

Acknowledgments—We thank Dr. N. Sakabe of the Foundation for Advancement of International Science and Drs. N. Watanabe, M. Suzuki, and N. Igarashi of the National Laboratory for High Energy Physics for their kind help in the data collection at the Photon Factory. This study was performed under the Cooperative Research Program of the Institute for Protein Research, Osaka University.

Fig. 5. Dimeric interactions of CEL-I. Hydrophilic (A) and hydrophobic (B) interactions connecting the two subunits of CEL-I. Pseudo 2-fold axes pass through the center of the two protomers. Only the residues forming each hydrophilic and hydrophobic interaction are shown. The subunits in pink and blue represent protomer A and protomer B, respectively. Oxygen and nitrogen atoms of the illustrated side chains and the backbone atoms are shown in red and blue, respectively. Hydrogen bonds are indicated by yellow dots. Fig. 5 was drawn by the program PyMOL (49).

acetamido group of GalNAc also makes van der Waals contacts with the side chains of Gln70 and Asn123 (Fig. 6C). Therefore, these interactions are assumed to be closely related to the high affinity binding of GalNAc to CEL-I. CEL-I binds to the β-anomer of GalNAc, except for protomer 1B which preferentially binds the α-anomer because of a hydrogen bond between 1-OH of the bound GalNAc and Asp83 in protomer 2A. In protomers 1A and 2A, the terminal NH2 group of the side chain of Arg115 also makes a hydrogen bond with 1-OH of GalNAc, which preferentially stabilizes the binding of the β-anomer of GalNAc (Fig. 6C). Because Asn123 is essential as a ligand for CA2, we constructed mutants of rCEL-I, in which Arg115 and/or Gln70 were replaced by alanine (R115A, Q70A, and Q70A/R115A), to evaluate the contributions of these residues to GalNAc recognition. As reported previously, the wild-type rCEL-I expressed in E. coli cells exhibited a very similar, although slightly lower affinity, carbohydrate-binding specificity to the native protein in hemagglutination inhibition assays (36). Therefore, we evaluated the carbohydrate-binding specificity of the CEL-I mutants through the inhibition of hemagglutination by several simple carbohydrates. As shown in Table II, hemagglutination was inhibited by galactose-related carbohydrates in the wild-type as well as the mutant rCEL-I. In particular, GalNAc exhibited much stronger inhibition than the other carbohydrates (more than 250-fold in the case of the wild-type). However, inhibition by GalNAcs decreased to 1/8, 1/32, and 1/64 for the R115A, Q70A, and Q70A/R115A mutants, respectively, compared with the wild-type. On the other hand, other galactose and galactose-containing carbohydrates showed only slight changes in their inhibitory profiles, compared with the wild-type. This confirms that Arg115 and Gln70 make significant contributions to the recognition of GalNAc. In particular, Gln70 appears to be more important than Arg115 despite hydrogen bonds between the latter and GalNAc. As shown in Table II, the Q70A/R115A mutant still had the same affinity for Gal as well as a moderately high affinity for GalNAc. This indicates that Gln70 and Arg115 have little influence on the recognition of the galactose moiety and also suggests that other residue(s), such as Asn123, may be involved in recognizing the acetamido group of GalNAc in addition to Gln70 and Arg115.

Inhibition by GalNAc decreased to 1/8, 1/32, and 1/64 for the R115A, Q70A, and Q70A/R115A mutants, respectively, compared with the wild-type. On the other hand, other galactose and galactose-containing carbohydrates showed only slight changes in their inhibitory profiles, compared with the wild-type. This confirms that Arg115 and Gln70 make significant contributions to the recognition of GalNAc. In particular, Gln70 appears to be more important than Arg115 despite hydrogen bonds between the latter and GalNAc. As shown in Table II, the Q70A/R115A mutant still had the same affinity for Gal as well as a moderately high affinity for GalNAc. This indicates that Gln70 and Arg115 have little influence on the recognition of the galactose moiety and also suggests that other residue(s), such as Asn123, may be involved in recognizing the acetamido group of GalNAc in addition to Gln70 and Arg115.

A GalNAc-binding mutant of MBP was constructed based on sequence comparison with rat hepatic lectin, which shows higher affinity for GalNAc than for Gal (44), and its complex structure with GalNAc was reported (10). Its preferential affinity for GalNAc compared with that for Gal is achieved by van der Waals contact between the methyl group of the acetamido group of GalNAc and the Ce-1 and Ne-2 atoms of His202, which was introduced in place of threonine, the corresponding amino acid in rat hepatic lectin. This MBP mutant attained a 60-fold higher affinity for GalNAc than for Gal, indicating that such van der Waals interactions are very important. As shown in Fig. 6, B and C, the orientation of GalNAc bound to CEL-I is very similar to that in the GalNAc-binding mutant of MBP. In the case of CEL-I, the methyl group of the acetamide of GalNAc makes van der Waals contact with Gln70 as mentioned above. Therefore, it appears that Gln70 may play a similar role to His202 in the GalNAc-binding mutant of MBP. On the other hand, Gln70, which is located at the corresponding position to His202 in the MBP mutant, may have the potential to interact with the methyl group of GalNAc, although a mutational analysis investigating this was not performed in the present study.

CEL-I exhibits strong cytotoxicity (36), which suggests an actual biological role for this protein as a defense toxin against predators. The toxicity is inhibited in the presence of GalNAc suggesting that it is mediated by binding to specific carbohydrate chains on the cell surface. One probable mechanism is that the binding of CEL-I to cell surface carbohydrate chains triggers intracellular signaling, leading to cell death. An identification of the natural carbohydrate ligands for CEL-I that are present on the target cell surface seems to be important for understanding the cytotoxicity at the molecular level. Further investigation of the involvement of the residues around the carbohydrate-binding site of CEL-I using recombinant proteins should provide important clues regarding the recognition mechanism for natural ligands on the target cell surface.

Acknowledgments—We thank Dr. N. Sakabe of the Foundation for Advancement of International Science and Drs. N. Watanabe, M. Suzuki, and N. Igarashi of the National Laboratory for High Energy Physics for their kind help in the data collection at the Photon Factory. This study was performed under the Cooperative Research Program of the Institute for Protein Research, Osaka University.

Fig. 5. Dimeric interactions of CEL-I. Hydrophilic (A) and hydrophobic (B) interactions connecting the two subunits of CEL-I. Pseudo 2-fold axes pass through the center of the two protomers. Only the residues forming each hydrophilic and hydrophobic interaction are shown. The subunits in pink and blue represent protomer A and protomer B, respectively. Oxygen and nitrogen atoms of the illustrated side chains and the backbone atoms are shown in red and blue, respectively. Hydrogen bonds are indicated by yellow dots. Fig. 5 was drawn by the program PyMOL (49).
FIG. 6. Comparison of the carbohydrate-binding modes of CEL-I and the GalNAc-binding mutant of MBP. A, electron density map around the GalNAc-binding site of CEL-I (protomer 1A) and the mutant MBP around the bound GalNAc. C, overlapped structures of CEL-I (protomer 1A) and residue numbers of CEL-I and the MBP mutant are superimposed. Carbon, calcium atoms, and residue numbers of CEL-I and the mutant MBP around the acetamido group. GalNAc, the calcium ion, and its binding residues in CEL-I and the MBP mutant are shown in green, red, blue, and purple, respectively. B, overlapped structures of CEL-I (protomer 1A) and the mutant MBP around the acetamido group. GalNAc, the calcium ion, and its binding residues in CEL-I and the MBP mutant are shown in green and yellow, respectively. Hydrogen and coordination bonds are indicated by dots. The GalNAc-binding site structure of protomer 2A is similar to this site. Fig. 6 was drawn by the program PyMOL (49).

### TABLE II

| Carbohydrate | Minimum concentration of carbohydrate to inhibit hemagglutination (nM) | Wild-type | R115A | Q70A | Q70A/R115A |
|--------------|------------------------------------------------------------------------|----------|-------|------|------------|
| GalNAc       | 0.0031 0.0244 (8) 0.0977 (32) 0.195 (64)                                |          |       |      |            |
| d-Galactose  | 6.25 25 (4) 6.25 (2) 3.13 (1)                                          |          |       |      |            |
| Lactose      | 3.13 6.25 (2) 6.25 (2) 3.13 (1)                                         |          |       |      |            |
| p-Nitrophenyl| 1.56 3.13 (2) 3.13 (2) ND                                              |          |       |      |            |
| β-d-Galactoside| 12.5 12.5 (1) 12.5 (1) ND                                             |          |       |      |            |
| Melibiose    | 12.5 12.5 (1) 12.5 (1) ND                                             |          |       |      |            |
| Raffinose    | 12.5 12.5 (1) 12.5 (1) ND                                             |          |       |      |            |
| d-Glucose    | >50 >50 >50 ND                                                        |          |       |      |            |
| d-Mannose    | >50 >50 >50 ND                                                        |          |       |      |            |

*ND, not determined. Because of a low yield of Q70A/R115A during purification steps, its binding specificity was examined only with GalNAc and lactose.*

**REFERENCES**

1. Gabius, H. J. (1997) Eur. J. Biochem. 243, 543–576
2. Dodd, R. B., and Drickamer, K. (2001) Glycoconjug. 11, 71R–78R
3. Drickamer, K. (1988) J. Biol. Chem. 263, 9557–9560
4. Day, A. J. (1994) Biochem. Soc. Trans. 22, 83–88
5. Weis, W. I., Kuhn, R., Fourme, R., Drickamer, K., and Hendrickson, W. A. (1991) Science 254, 1608–1615
6. Weis, W. I., Drickamer, K., and Hendrickson, W. A. (1992) Nature 360, 127–134
7. Isotb, S. T., and Drickamer, K. (1994) J. Biol. Chem. 269, 15512–15519
8. Kolatkar, A. R., and Weis, W. I. (1996) J. Biol. Chem. 271, 6679–6685
9. Ng, K. K., Drickamer, K., and Weis, W. I. (1996) J. Biol. Chem. 271, 663–674
10. Kolatkar, A. R., Leung, A. K., Beck, R., Brossmer, R., Drickamer, K., and Weis, W. I. (1998) J. Biol. Chem. 273, 19502–19508
11. Feinberg, H., Mitchell, D. A., Drickamer, K., and Weis, W. I. (2001) Science 294, 2163–2166
12. Somers, W. S., Tang, J., Shaw, G. D., and Camphausen, R. T. (2000) EMBO J. 19, 487–479
13. Deleted in proof
14. Walker, J. R., Nagar, B., Young, N. M., Hiramta, T., and Rini, J. M. (2004) Biochemistry 43, 3783–3792
15. Swaminathan, G. J., Weaver, A. J., Loegering, D. A., Checkel, J. L., Leonidas, D. D., Gleich, G. J., and Ackhara, K. R. (2001) J. Biol. Chem. 276, 20693–20699
16. Hatakeyama, T., Kohzaki, H., Naganmata, H., and Yamasaki, N. (1994) J. Biochem. (Tokyo) 116, 209–214
17. Oda, T., Tsuru, M., Hatakeyama, T., Naganmata, H., Muramatsu, T., and Yamasaki, N. (1997) J. Biochem. (Tokyo) 121, 560–567
18. Hatakeyama, T., Furukawa, M., Naganmata, H., Yamasaki, N., and Mori, T. (1996) J. Biol. Chem. 271, 16915–16920
19. Hatakeyama, T., Naganmata, H., and Yamasaki, N. (1995) J. Biol. Chem. 270, 3560–3564
20. Oda, T., Shimura, N., Nishio, K., Komatsu, N., Hatakeyama, T., and Muramatsu, T. (1999) J. Biochem. (Tokyo) 125, 713–720
21. Nakano, M., Tabata, S., Sugihara, K., Kouzuma, Y., Kimura, M., and Yamasaki, N. (1999) Biochim. Biophys. Acta 1435, 167–176
22. Uchida, T., Yamasaki, T., Eto, S., Sugawara, H., Kurisu, G., Nakagawa, A., Kusunoki, M., and Hatakeyama, T. (2004) J. Biol. Chem. 279, 37133–37141
23. Hatakeyama, T., Matsu, N., Shiha, K., Nishiohara, S., Yamasaki, N., Sugawara, H., and Aoyagi, H. (2002) Biosci. Biotechnol. Biochem. 66, 157–163
24. Hatakeyama, T., Matsu, N., Aoyagi, H., Sugawara, H., Uchida, T., Kurisu, G., and Kusunoki, M. (2002) Acta Crystallogr. Sect. D 58, 143–144
25. Watanabe, N., Nakagawa, A., Adachi, S., and Sakabe, N. (1995) Rev. Sci. Instrum. 66, 1824–1826
26. Leslie, A. G. W. (1992) Joint CCP4 + ESF-EAMCR Newsletter on Protein Crystallography 26
27. Collaborative Computational Project, Number 4 (1994) Acta Crystallogr. Sect. D 50, 760–763
28. Matthews, B. W. (1968) J. Mol. Biol. 33, 491–497
29. Bertrand, J. A., Pignol, D., Bernard, J. P., Verdie, J. M., Dagorn, J. C., and Fontecilla-Camps, J. C. (1996) EMBO J. 15, 2678–2684
30. Novazza, J. (1994) Acta Crystallogr. Sect. A 50, 157–163
31. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D 54, 905–921
Crystal Structure of a GalNAc-specific C-type Lectin, CEL-I

32. Brunger, A. (1997) Methods Enzymol. 277, 366–396
33. McRee, D. E. (1999) J. Struct. Biol. 125, 156–165
34. Murshudov, G. N., Vagin, A., and Dodson, E. (1997) Acta Crystallogr. Sect. D 53, 240–255
35. Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
36. Hatakeyama, T., Shiha, K., Matsuo, N., Fujimoto, T., Oda, T., Sugawara, H., and Aoyagi, H. (2004) J. Biochem. (Tokyo) 135, 101–107
37. Holm, L., and Sander, C. (1994) Proteins 19, 165–173
38. Drickamer, K. (1999) Curr. Opin. Struct. Biol. 9, 585–590
39. Poget, S. F., Legge, G. B., Proctor, M. R., Butler, P. J., Bycroft, M., and Williams, R. L. (1999) J. Mol. Biol. 289, 867–879
40. Mizuno, H., Fujimoto, Z., Kozumi, M., Kano, H., Atoda, H., and Morita, T. (1997) Nat. Struct. Biol. 4, 438–441
41. Mizuno, H., Fujimoto, Z., Kozumi, M., Kano, H., Atoda, H., and Morita, T. (1999) J. Mol. Biol. 289, 103–112
42. Kjell Håkansson, R., Lim, N. K., Hoppe, H. J., and Reid, K. B. (1999) Structure 7, 255–264
43. Kakiuchi, M., Okino, N., Saeryoshi, N., Ichinose, S., Oseri, A., Kawabata, S., Yamaguchi, K., and Ito, M. (2002) Glycobiology 12, 85–94
44. Iobst, S. T., and Drickamer, K. (1996) J. Biol. Chem. 271, 6686–6693
45. Hutchinson, E. G., and Thornton, J. M. (1996) Protein Sci. 5, 212–220
46. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
47. Merritt, E., and Bacon, D. (1997) Methods Enzymol. 277, 505–524
48. McDonald, I. K., and Thornton, J. M. (1994) J. Mol. Biol. 238, 777–793
49. DeLano, W. L. (2002) The PyMOL Molecular Graphics Systems, DeLano Scientific, San Carlos, CA
50. Barton, G. J. (1993) Protein Eng. 6, 37–40
Characteristic Recognition of N-Acetylglucosamine by an Invertebrate C-type Lectin, CEL-I, Revealed by X-ray Crystallographic Analysis
Hajime Sugawara, Masami Kusunoki, Genji Kurisu, Tokiko Fujimoto, Haruhiko Aoyagi and Tomomitsu Hatakeyama

J. Biol. Chem. 2004, 279:45219-45225.
doi: 10.1074/jbc.M408840200 originally published online August 19, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M408840200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 47 references, 13 of which can be accessed free at http://www.jbc.org/content/279/43/45219.full.html#ref-list-1