Amino Acid Sequence and Carbohydrate-binding Analysis of the N-acetyl-D-galactosamine-specific C-Type Lectin, CEL-I, from the Holothuroidea, Cucumaria echinata

Tomomitsu HATAKEYAMA,1† Noriaki MATSUO,1 Kouhei SHIBA,1 Shoichi NISHINOHARA,2 Nobuyuki YAMASAKI,2 Hajime SUGAWARA,3 and Haruhiko Aoyagi1

1Department of Applied Chemistry, Faculty of Engineering, Nagasaki University, Bunkyo-machi 1-14, Nagasaki 852-8521, Japan
2Laboratory of Biochemistry, Faculty of Agriculture, Kyushu University, Hakozaki 6-10-1, Fukuoka 812-8581, Japan
3Metabolic Function Research Group, RIKEN Plant Science Center, Hirosawa 2-1, Wako, Saitama 351-0198, Japan

Received August 27, 2001; Accepted September 25, 2001

CEL-I is one of the Ca2+-dependent lectins that has been isolated from the sea cucumber, Cucumaria echinata. This protein is composed of two identical subunits held by a single disulfide bond. The complete amino acid sequence of CEL-I was determined by sequencing the peptides produced by proteolytic fragmentation of S-pyridylethylated CEL-I. A subunit of CEL-I is composed of 140 amino acid residues. Two intrachain (Cys3-Cys14 and Cys31-Cys135) and one interchain (Cys36) disulfide bonds were also identified from an analysis of the cystine-containing peptides obtained from the intact protein. The similarity between the sequence of CEL-I and that of other C-type lectins was low, while the C-terminal region, including the putative Ca2+ and carbohydrate-binding sites, was relatively well conserved. When the carbohydrate-binding activity was examined by a solid-phase microplate assay, CEL-I showed much higher affinity for N-acetyl-D-galactosamine than for other galactose-related carbohydrates. The association constant of CEL-I for p-nitrophenyl N-acetyl-D-galactosaminide (NP-GalNAc) was determined to be 2.3 × 104 M⁻¹, and the maximum number of bound NP-GalNAc was estimated to be 1.6 by an equilibrium dialysis experiment.

Key words: C-type lectin; carbohydrate; amino acid sequence; peptide

Among the various carbohydrate-binding proteins (lectins) in animals,10 those requiring Ca2+ to bind carbohydrates, called C-type lectins, have varying carbohydrate-binding specificity and constitute an important group.21 Those for vertebrate C-type lectins have been classified into seven categories,23 in which selectin, collectin, proteoglycan core proteins and several cell surface receptor proteins are included. They play important roles in the molecular recognition process as carbohydrate-binding modules in combination with other distinct domains. Among the C-type lectins, mannos-binding protein (mannan-binding protein; MBP) in mammalian serum is the one most studied in respect of the relationship between its structure and carbohydrate-recognition ability.40 This protein is involved in activation of the complement system in response to microbial infection.5 In contrast to vertebrate C-type lectins, those in invertebrates are generally composed only of the C-type carbohydrate recognition domain (CRD), and are assumed to participate in the host defense system in the place of immunoglobulins in vertebrates.

Four Ca2+-dependent lectins (CEL-I, II, III and IV) have been purified from the Holothuroidea, Cucumaria echinata, and characterized.6 Among these, CEL-III (47.5 kDa) is a unique lectin which showed hemolytic as well as cytotoxic activities.7 These activities were found to be mediated by oligomerization of the protein in the target cell membranes, which led to the formation of membrane pores.8,9 Among the various carbohydrate-binding proteins (lectins) in animals,10 those requiring Ca2+ to bind carbohydrates, called C-type lectins, have varying carbohydrate-binding specificity and constitute an
ments.\textsuperscript{11} This protein is composed of four identical subunits with a molecular mass of 17 kDa, held together with disulfide bonds between them. While CEL-III is not a C-type lectin, CEL-IV shows relatively low but definite homology with the C-type lectin family. It seems that many invertebrates have C-type lectins in the body fluids which play important roles in the molecular recognition process. However, information concerning the structure of C-type lectins in invertebrates is still very limited.

CEL-I is the smallest Ca\textsuperscript{2+}-dependent lectin (32 kDa) in \textit{C. echinata} which is composed of two identical subunits linked by a single disulfide bond.\textsuperscript{6} CEL-I has been suggested to have high specificity for N-acetyl-d-galactosamine (GalNAc), compared with other \textit{C. echinata} lectins. Therefore, structural details of this lectin would provide valuable information about how C-type lectins recognize different carbohydrates. We describe in this report a determination of the amino acid sequence of CEL-I as well as its carbohydrate-binding properties.

**Materials and Methods**

Lysylendopeptidase, thermolysin, 4-vinylpyrididine, tri-n-butylphosphine, citraconic anhydride and ammonium 7-fluorobenzo-2-oxa-1,3-diazole-4-sulfonate (SBD-F) were obtained from Wako Pure Chemicals (Osaka, Japan). Trypsin and \( p \)-nitrophenyl N-acetyl-\( \beta \)-d-galactosaminide (NP-GalNAc) were from Sigma. All other reagents were of analytical grade.

**Purification of CEL-I.** Specimens of \textit{C. echinata} were collected in the Sea of Genkai (Fukuoka, Japan). CEL-I was purified from the protein extract of \textit{C. echinata} by using column chromatography on lactosyl-Cellulofine, GalNAc-Cellulofine, and Sephadex G-75, essentially as reported earlier.\textsuperscript{8} Chromatography was conducted at 4°C. The purified protein was dialyzed against 10 mM Tris-HCl containing 0.15 M NaCl at pH 7.5 (TBS) and was stored frozen at \(-30°C\) until needed.

**Pyridylethylation of the half-cystine residues.** S-Pyridylethylation was achieved by the method of Hermodson et al.\textsuperscript{12} The protein was dissolved in 300 \( \mu \)l of a 0.5 M Tris-HCl buffer (pH 8.5) containing 7 M guanidine hydrochloride and 10 mM EDTA. To this solution, 1 \( \mu \)l of 4-vinylpyrididine and 2 \( \mu \)l of tri-n-butylphosphine were added. After reacting for 4 h at room temperature, the resulting pyridylethyl (PE)-CEL-I was dialyzed against distilled water to remove the reagents.

**Digestion and peptide separation.** PE-CEL-I was digested in 200 \( \mu \)l of an NH\textsubscript{4}HCO\textsubscript{3} solution with 5–10 \( \mu \)g of proteolytic enzymes at 37°C for 3–4 h. The resulting peptides were separated by reverse-phase HPLC, using a Hitachi L-6200 HPLC system with a column of Wakosil 5C4 (4 \( \times \) 150 mm). The peptides were eluted by a linear gradient of acetonitrile in 0.1% trifluoroacetic acid. Selective digestion at the arginine residues by trypsin was done after citraconylation of PE-CEL-I.\textsuperscript{11} After the tryptic digestion, the citraconyl groups were removed in 10% formic acid for 1 h at room temperature, and the peptides were separated by using reverse-phase HPLC.

**Sequence analysis.** The amino acid sequences of the protein and peptides were analyzed with a Shimadzu PPSQ-10 protein sequencer. The mass of the peptides was confirmed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (PerSeptive Voyager and Finnigan Mat Vision 2000) operated in the linear mode. Bovine serum albumin and bovine insulin were used as external standards.

**Determination of the disulfide bonds.** The intra- and interchain disulfide bonds in CEL-I were identified by an analysis of the cystine-containing peptides obtained from the tryptic digestion of intact CEL-I. The cystine residues were detected by the fluorescent reagent, SBD-F, which emits fluorescence at 515 nm upon reaction with the cysteines formed by reduction with tri-n-butylphosphine.\textsuperscript{13}

**Comparison of the carbohydrate-binding activity by a solid-phase microplate assay.** The affinity of CEL-I for different carbohydrates was compared by a solid-phase microplate assay.\textsuperscript{14,15} A CEL-I solution in TBS-10 mM CaCl\textsubscript{2} was placed in the wells of a microplate (96-well amino-type, Sumitomo Bakelite, Tokyo, Japan) which had been previously coated with different carbohydrates with divinylsulfone. After 1 h at 20°C, the protein solution was removed, and the wells were washed several times with the same buffer. The protein bound to the surface of the carbohydrate-coated wells was detected with a colloidal gold solution. The binding of CEL-I is expressed as the absorbance at 620 nm measured by a microplate reader.

**Equilibrium dialysis.** The association constant for binding of specific carbohydrate and the number of binding sites of CEL-I were determined by an equilibrium dialysis experiment with \( p \)-nitrophenyl N-acetyl-\( \beta \)-d-galactosaminide (NP-GalNAc). NP-GalNAc at different concentrations and CEL-I in TBS-10 mM CaCl\textsubscript{2} (250 \( \mu \)l) were placed into each compartment of the microdialysis cells across a semi-permeable membrane. The microdialysis cells were kept at 6°C for 24 h and, after equilibrium had been achieved, aliquots of the solutions were withdrawn from each compartment with a microsyringe. The
concentration of NP-GalNAc was determined by the absorbance at 313 nm, using a molar absorption coefficient of $1.0 \times 10^4 \text{ M}^{-1}$. The concentration of CEL-I was derived by using an absorption coefficient (0.1% (w/v)) of 3.57 at 280 nm which had been calculated from the content of tryptophan, tyrosine, and half-cystine residues. The data were analyzed by a Scatchard plot, using the following equations:

$$
\frac{r}{[S]_{\text{free}}} = -K_a \cdot r + n \cdot K_a
$$

$$
\frac{r}{[S]_{\text{bound}}} / [P]_{\text{total}} = [S]_{\text{bound}}
$$

where $[S]$ is the concentration of NP-GalNAc, $[P]$ is the protein concentration, $K_a$ is the association constant, and $n$ is the maximum number of bound NP-GalNAc per protein molecule.

**Results**

**Determination of the amino acid sequence of CEL-I**

Since CEL-I contained inter- and intrachain disulfide bonds, they were reduced and S-pyridylethylated before enzymatic digestion. The entire amino acid sequence of PE-CEL-I was determined from an N-terminal sequence analysis of PE-CEL-I and its proteolytic fragments, as shown in Fig. 1. Lysylendopeptidase digestion yielded six peptides (K1–K6). These were completely sequenced, except for K1 which contained 78 amino acid residues and was recovered as a precipitate. The sequence of K1 was determined up to Ser50 after its solubilization with trifluoroacetic acid. The rest of the sequence of K1 was determined after digestion with trypsin (KT1 to KT4). The alignment of these peptides was established by sequencing a thermolytic peptide of PE-CEL-I (Th1) and the tryptic peptides of citraconylated PE-CEL-I (T1–T5). The molecular masses of the purified peptides (KT1–KT4 and K2–K6) were confirmed by MALDI-TOF mass spectrometry (Table 1).

**Identification of the disulfide bonds**

The positions of the disulfide bonds in CEL-I were identified by analyzing the cystine-containing peptides obtained from the tryptic digest of intact CEL-I (Table 2). Cystines were detected with the fluorescent reagent, SBD-F, after reduction with tri-$n$-butylphosphine, as described in the Materials and Methods section. Two cystine-containing fractions (CT1 and CT2) were obtained by reverse-phase HPLC, and their sequences were analyzed. CT1 contained the peptide, Asn1-Arg16, indicating the presence of a disulfide bond between Cys3 and Cys4. On the other hand, CT2 contained the peptides, Phe17-Arg42 and Tyr133-Lys136. Since the former peptide contained two half-cystines in the sequence, it was further digested with thermolysin. As a result, two thermolytic peptides, CT2-Th1 (Ala27-

Cys31 and Tyr133-Lys136) and CT2-Th2 (Tyr34-Thr38), were obtained. From their sequences, two disulfide bonds, Cys31-Cys135 and Cys36-Cys36 (between the subunits) were determined. Cys36 was also identified as a cystine residue at the third cycle of the analysis of CT2-Th2 by the protein sequencer. The molecular mass calculated from the complete amino acid sequence data was 32064.0 Da. This value is in good agreement with that obtained experimentally from MALDI-TOF mass spectrometry (32103.0 Da), confirming the validity of the sequence determined here.
Fig. 2. Comparison of the Amino Acid Sequences of CEL-I and Other C-type CRDs.

Alignment of the sequences was achieved with the multiple-sequence alignment program, Clustal W.18) Echinoidin, Anthocidaris crassispina (sea urchin) lectin;19) CEL-IV, C. ehinata lectin;11) SJL-I, Stichopus japonicus lectin;20) BRA-2 and BRA-3, Megabalanus rosa (acorn barnacle) lectins;21) Tunicate, Polyandrocarpa misakiensis (tunicate) lectin;22) MBP-A, rat mannose-binding protein A.23) Shaded parts represent residues conserved in at least seven of the eight proteins. Residues enclosed in circles are Ca²⁺-binding sites of the tunicate lectin and MBP-A.24,25) Residues enclosed in boxes are half-cystine residues forming one of the intrachain disulfide bonds in the respective proteins. Only the residue numbers for CEL-I are indicated. Numbers in parentheses show the residues extending at the N-terminal sides.

Table 1. Molecular Masses of the Peptides Derived from PE-CEL-I

<table>
<thead>
<tr>
<th>Peptide</th>
<th>No. of residues</th>
<th>Theoretical mass (Da)</th>
<th>Observed mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KT1</td>
<td>16</td>
<td>2135.8</td>
<td>2135.4</td>
</tr>
<tr>
<td>KT2</td>
<td>26</td>
<td>3269.3</td>
<td>3262.5</td>
</tr>
<tr>
<td>KT3</td>
<td>20</td>
<td>3242.6</td>
<td>3243.7</td>
</tr>
<tr>
<td>KT4</td>
<td>16</td>
<td>1765.0</td>
<td>1765.6</td>
</tr>
<tr>
<td>K2</td>
<td>13</td>
<td>1533.6</td>
<td>1529.7</td>
</tr>
<tr>
<td>K3</td>
<td>5</td>
<td>565.6</td>
<td>565.1</td>
</tr>
<tr>
<td>K4</td>
<td>36</td>
<td>3936.8</td>
<td>3932.8</td>
</tr>
<tr>
<td>K5</td>
<td>4</td>
<td>647.7</td>
<td>647.0</td>
</tr>
<tr>
<td>K6</td>
<td>4</td>
<td>509.6</td>
<td>506.8</td>
</tr>
</tbody>
</table>

Table 2. Amino Acid Sequences of the Cystine-containing Peptides Derived from Intact CEL-I

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT1</td>
<td>NQ PTDWEAEGDH 16</td>
</tr>
<tr>
<td>CT2</td>
<td>FFNLTLTENAHHE 42</td>
</tr>
<tr>
<td>CT2-Th1</td>
<td>AHHE 133</td>
</tr>
<tr>
<td>CT2-Th2</td>
<td>YS 34</td>
</tr>
</tbody>
</table>

Comparison of the sequence with those of other lectins

Figure 2 shows a comparison of the amino acid sequence of CEL-I with those of other C-type lectins from marine invertebrates, together with rat mannose-binding protein A (MBP-A). As listed in Table 3, the highest similarity was found with echinodin19) (sea urchin) for an identity of 36.4%. Similarity among C-type CRDs is generally low, and even CEL-IV has an identity of only 30.2% with CEL-I. The C-terminal region after position 74 was
Table 3. Similarity between CEL-I and Other C-type Lectins
The values are the percentage identity calculated from the shortest common regions based on the alignment shown in Fig. 2.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEL-IV</td>
<td>30.2</td>
</tr>
<tr>
<td>SJL-I</td>
<td>28.9</td>
</tr>
<tr>
<td>Echinoidin</td>
<td>36.4</td>
</tr>
<tr>
<td>BRA-2</td>
<td>21.8</td>
</tr>
<tr>
<td>BRA-3</td>
<td>28.4</td>
</tr>
<tr>
<td>Tunicate</td>
<td>16.4</td>
</tr>
<tr>
<td>MBP-A</td>
<td>19.3</td>
</tr>
</tbody>
</table>

relatively well conserved. While Cys31-Cys315 corresponds to one of the two disulfide bonds highly conserved among the C-type lectins,2) the other disulfide bond (marked by boxes in Fig. 2) is missing. This was also the case for the other C-type lectins from Holothuroidea such as CEL-IV11) and SJL-I (Stichopus japonicus).20)

Carbohydrate-binding properties of CEL-I
The carbohydrate-binding properties of CEL-I were examined by the equilibrium dialysis method and the solid-phase microplate assay. Figure 3 shows the Scatchard plot for the binding of NP-GalNAc to CEL-I as examined by the equilibrium dialysis method. The Scatchard plot shows a linear correlation between \( r \) and \( r/[S]_{\text{free}} \), indicating simple binding equilibrium without cooperativity between the binding sites. From this plot, the association constant and maximum number of bound NP-GalNAc per protein molecule were calculated to be \( 2.3 \times 10^4 \) M\(^{-1} \) and 1.60, respectively. These results indicate that there were two carbohydrate-binding sites with equal affinity on the CEL-I molecule. The affinity for different carbohydrates was compared by the solid-phase assay which used carbohydrate-coated microplates.14,15)

Fig. 4 shows the increase in absorbance at 620 nm arising from the protein bound to the carbohydrates immobilized on the microplate wells. The highest affinity was found for GalNAc among the carbohydrates tested. This is consistent with the results previously obtained from a hemagglutination inhibition assay, in which GalNAc exhibited much stronger inhibition against CEL-I-induced hemagglutination than the other carbohydrates.6)

Discussion
The determined amino acid sequence proved CEL-I to be a C-type lectin. However, the sequence similarity with other C-type lectins was rather low, and even CEL-IV, another C-type lectin of C. echinata, showed only 30.2% of sequence identity with CEL-I. Identical residues were relatively rich in the C-terminal half. This region is known to include the carbohydrate-binding site, as demonstrated in the case of other C-type lectins, whose tertiary structures have been determined.25-30) C-terminal region of CEL-I was also rich in Asp, Asn, Glu, and Gln residues, especially at positions Gln101-Asn125. Since this region corresponds to the Ca\(^{2+}\)-binding site of other C-type lectins whose spatial structure is known, such as tunicate lectin and mannose-binding protein A (Fig. 2), this region is also a possible candidate for the Ca\(^{2+}\)-binding site in CEL-I.

Although the disulfide bond of Cys31-Cys135 in CEL-I corresponds to one of those widely conserved in C-type lectins,2) another conserved disulfide bond located in the C-terminal part (Fig. 2) is missing in the holothuroidea C-type lectins, CEL-I, CEL-IV13)
and SJL-I (S. japonicus). Another feature regarding the disulfide bond is the presence of the interchain bond between Cys36 from each subunit. The tunicate lectin, which is also composed of two identical subunits, has no disulfide bond between the subunits. An X-ray crystallographic analysis of this lectin has revealed that the subunits were held by hydrogen bonds and hydrophobic interaction between $\alpha$-helices and $\beta$-sheets from both subunits. It seems interesting to clarify the contribution of the disulfide bond between the two subunits of CEL-I to maintaining its dimer structure.

The equilibrium dialysis experiments using NP-GalNAc as a carbohydrate ligand, indicates that CEL-I had two carbohydrate-binding sites of equal affinity. This is consistent with the fact that CEL-I is composed of two identical subunits. While generally higher affinity for GalNAc has also been observed for the other C. echinata lectins, CEL-III and CEL-IV, CEL-I showed remarkably stronger binding to GalNAc than to other galactose-related carbohydrates, as revealed by the solid-phase binding assay and the hemagglutination inhibition assay previously reported. Taking into account the probable role of invertebrate soluble lectins in the host defense system, CEL-I may be responsible for the inactivation or removal of foreign microorganisms which are rich in GalNAc on their surfaces.

In spite of the relatively low sequence similarity between C-type CRDs, their basic folding structure may be considerably more similar to each other, as exemplified by tunicate lectin, mannos-binding lectin and E-selectin. In addition to the C-type lectins, some homologous proteins without carbohydrate-binding ability are also known (the C-type lectin superfamily). These facts suggest that C-type CRD is one of the most versatile protein structures to function as binding modules in biological molecular recognition. It therefore seems important to obtain insight into the structural features of C-type CRDs not only for an elucidation of their structure-function relationships, but also for genetically engineering new proteins with the recognition ability of various substances. We have recently succeeded in obtaining single crystals of CEL-I which could provide useful information concerning its tertiary structure through an X-ray crystallographic analysis.

Acknowledgments

We are grateful to Dr. Yoshinobu Kimura (Okayama University) and Dr. Hideki Ohba (National Institute of Advanced Industrial Science and Technology, Kyushu) for the MALDI-TOF mass spectrometric study.

References


