THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 35, Issue of August 27, pp. 37133–37141, 2004
Printed in U.S.A.

Crystal Structure of the Hemolytic Lectin CEL-III Isolated from the Marine Invertebrate *Cucumaria echinata*

IMPLICATIONS OF DOMAIN STRUCTURE FOR ITS MEMBRANE PORE-FORMATION MECHANISM*

Received for publication, April 12, 2004, and in revised form, June 7, 2004
Published, JBC Papers in Press, June 11, 2004, DOI 10.1074/jbc.M404065200

Tatsuya Uchida‡, Takayuki Yamasaki§, Seiichiro Eto§, Hajime Sugawara¶, Genji Kurisu‡, Atsushi Nakagawa‡, Masami Kusunoki‡‖, and Tomomitsu Hatakeyama§**

From the ‡Research Center for Structural and Functional Proteomics, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, the §Department of Applied Chemistry, Faculty of Engineering, Nagasaki University, 1-14 Bunkyo-machi, Nagasaki 852-8521, and ¶Laboratory for Communication Mechanisms, RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

CEL-III is a Ca^{2+}-dependent and galactose-specific lectin purified from the sea cucumber, *Cucumaria echinata*, which exhibits hemolytic and hemagglutinating activities. Six molecules of CEL-III are assumed to oligomerize to form an ion-permeable pore in the cell membrane. We have determined the crystal structure of CEL-III by using single isomorphous replacement aided by anomalous scattering in lead at 1.7 Å resolution. CEL-III consists of three distinct domains as follows: the N-terminal two carbohydrate-binding domains (1 and 2), which adopt β-trefoil folds such as the B-chain of ricin and are members of the $(QXW)_3$ motif family; and domain 3, which is a novel fold composed of two α-helices and one β-sandwich. CEL-III is the first Ca^{2+}-dependent lectin structure with two β-trefoil folds. Despite sharing the structure of the B-chain of ricin, CEL-III binds five Ca^{2+} ions at five of the six subdomains in both domains 1 and 2. Considering the relatively high similarity among the five subdomains, they are putative binding sites for galactose-related carbohydrates, although it remains to be elucidated whether bound Ca^{2+} is directly involved in interaction with carbohydrates. The paucity of hydrophobic interactions in the interfaces between the domains and biochemical data suggest that these domains rearrange upon carbohydrate binding in the erythrocyte membrane. This conformational change may be responsible for oligomerization of CEL-III molecules and hemolysis in the erythrocyte membranes.

A number of lectins have been found in tissues and body fluids in various animals. They play important roles in the recognition of specific carbohydrate chains. Animal lectins are categorized into several groups based on their structures (1). Among them, Ca^{2+}-dependent lectins, such as those containing C-type carbohydrate-recognition domains (2), include various proteins that exhibit carbohydrate binding activity depending on the concentration of Ca^{2+} (3). Four Ca^{2+}-dependent and Gal/GalNAc-specific lectins (CEL-I–IV, respectively) have been isolated from the body fluid of the sea cucumber, *Cucumaria echinata* (4). Among them, CEL-I and CEL-IV belong to the C-type lectin family (5, 6), whereas CEL-III is a novel lectin that exhibits strong hemolytic and cytotoxic activities (4, 7–9). These activities are mediated by binding of the protein to the specific carbohydrate chains on the target cells, followed by the formation of ion-permeable pores in the cell membrane through oligomerization of the protein. After formation of the pores, erythrocytes are ruptured by colloid osmotic shock. CEL-III oligomers of ~270 kDa were found in the erythrocyte membranes lysed by CEL-III (10). This suggests that a CEL-III oligomer in the membrane is composed of six monomers. Most interestingly, when incubated with β-galactosides (*e.g.* lactose), CEL-III also oligomerized in an aqueous solution with an alkaline pH under high ionic strength conditions (10). This oligomer had the same size on SDS-PAGE as that formed in the lysed erythrocyte membrane. This suggests that the binding of specific carbohydrates is important in facilitating oligomerization of CEL-III. Binding to specific carbohydrate chains on the target cell membranes probably triggers conformational change of CEL-III, and this leads to its oligomerization followed by insertion into the cell membrane. This assumption is also supported by the fact that CEL-III underwent secondary structural change accompanied by an increase of its surface hydrophobicity (10) after the binding of carbohydrates and subsequent oligomerization.

The amino acid sequence of CEL-III was determined by sequence analysis of its cDNA and peptide fragments derived from enzymatic cleavage (11). The results suggested that CEL-III is composed of three portions (domains 1–3) based on the similarity of its internal sequence. Most interestingly, domains 1 and 2, which include cysteines involved in disulfide bonds, exhibited definite homology with the B-chains of the toxic lectins, ricin from *Ricinus communis* (12) and abrin from *Abrus precatorius* (13). The B-chains of these proteins are known to be lectin subunits, which are composed of two carbohydrate-bind-

* This work was supported by Grant-in-aid for Scientific Research (14560073 (to T. H.), from the Japan Society for the Promotion of Science and Protein 3000 program (to A. N.) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. This work was conducted under the Cooperative Research Program of the Institute for Protein Research, Osaka University. 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 (code 1VCL) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‖ To whom correspondence may be addressed: Research Center for Structural and Functional Proteomics, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan. Tel.: 81-6-6879-8638; Fax: 81-6-6879-8636; E-mail: kusunoki@protein.osaka-u.ac.jp.

** To whom correspondence may be addressed: Dept. of Applied Chemistry, Faculty of Engineering, Nagasaki University, Bunkyo-machi 1-14, Nagasaki, 852-8521, Japan. Tel.: 81-95-819-2686; Fax: 81-95-819-2684; E-mail: thata@net.nagasaki-u.ac.jp.

ing domains (CBDs),[1] whereas their A-chains exert *N*-glycosidase activity and inactivate 50 S subunits of ribosomes in eukaryotic cells. In contrast, no proteins have been found as a homologue of domain 3 of CEL-III. Because domain 3 includes the hydrophobic region around residues 322–349, it was assumed responsible for self-oligomerization and insertion into the target cell membrane in the course of the hemolytic process. This is also supported by experiments using individual domain fragments produced by partial digestion by using proteases and the recombinant proteins expressed in *Escherichia coli* cells (14). In these experiments, domains 1 and 2 exhibited carbohydrate binding activity, whereas domain 3 showed a tendency to oligomerize immediately after production. These results strongly suggest that hemolysis by CEL-III proceeds through cooperative action of its functional domains; the binding of domains 1 and 2 to carbohydrate chains on the cell surface triggers conformational change of CEL-III, which promotes association of domain 3, followed by its insertion into the cell membrane. A similar pore-forming mechanism is also proposed in the cases of bacterial pore-forming toxins (15–17).

In this report, we describe the crystal structure of CEL-III at 1.7 Å resolution. From the determined structure, three distinct domains are revealed as expected from the amino acid sequence. Domains 1 and 2 were each found to adopt a β-trefoil structure similar to the ricin B-chain. CEL-III and the B-chain of ricin belong to the $(QXW)_3$ family proteins (18). In contrast to the cases of the other β-trefoil lectins, domains 1 and 2 of CEL-III bound five Ca^{2+} and two Mg^{2+} ions, and Ca^{2+} may be associated with recognition of specific carbohydrates. On the other hand, domain 3 contains characteristic β-sheets and two α-helices. These α-helices are located at the interface between carbohydrate-binding domains (1 and 2) and domain 3, and are assumed to make conformational change upon binding to the cell surface carbohydrates. These results suggest a probable structural change of CEL-III to form membrane-associated oligomers.

EXPERIMENTAL PROCEDURES

Purification of CEL-III—Specimens of *C. echinata* were collected from the Sea of Genkai, Fukuoka, Japan. CEL-III was purified from the body fluid of *C. echinata* by chromatography using lactose-Cellulofine, GalNAc-Cellulofine, and Sephadex G-75 columns, essentially as described previously (4). Protein concentrations were determined from the molar absorption coefficients at 280 nm calculated from the amino acid compositions of the proteins.

Crystallization—Initial crystallization screening experiments for CEL-III (5.0 mg/ml) were carried out at both 4 and 20 °C using the hanging-drop vapor-diffusion method. Two microliters of 5.0 mg/ml CEL-III in 10 mM Tris/HCl buffer, pH 7.5, 150 mM NaCl, and 10 mM $CaCl_2$ were mixed with an equal volume of reservoir solution. Thin crystals were obtained at 4 °C using 30% (w/v) polyethylene glycol 8000, 100 mM sodium cacodylate, pH 6.5, and 200 mM sodium acetate as a reservoir solution. The optimized crystallization was done at 20 °C with a reservoir solution of 12% (w/v) polyethylene glycol 8000, 100 mM BisTris/NaOH, pH 6.5, and 200 mM magnesium acetate. Crystals (~0.05 × 0.2 × 0.3 mm in size), large enough for x-ray diffraction analysis, were grown within 1 week by microseeding. Heavy atom derivatives were obtained by soaking native crystals for 5 days in the mother liquor containing 10 mM lead acetate. The crystals were soaked in the reservoir solution supplemented by 20% (v/v) ethylene glycol as a cryoprotectant. Cryoprotected crystals were flash-cooled in liquid nitrogen before data collection.

X-ray Diffraction Data Collection—Diffraction images from native crystals were collected using synchrotron radiation on beamline BL6A at the Photon Factory (Tsukuba, Japan) at 100 K using a CCD detector, Quantum 4R (ADSC, Poway, CA). X-ray diffraction images from a lead derivative crystal were collected on beamline BL44XU at SPring-8 (Hyogo, Japan) at 100 K using an imaging plate detector, DIP6040 (MAC Science, Japan). Both sets of images were indexed and integrated using the program MOSFLM (19) and processed using the CCP4 programs (20) SCALA and TRUNCATE. Data collection statistics are summarized in Table I.

Structure Determination and Refinement—We assumed that the asymmetric unit has two CEL-III molecules based on a Matthews coefficient (21) (V_M of 2.3 Å^3/Da). Native Patterson maps showed a strong peak at $u = 0.5$, $v = 0.08$, $w = 0.5$, indicating a noncrystallographic translation without rotation between the two molecules in the asymmetric unit. Binding sites of four lead atoms were determined by inspection of Harker sections in the difference Patterson maps. Heavy atom parameters were refined using the program MLPHARE (20) and then using the program SHARP (22); phases were calculated by single isomorphous replacement with anomalous scattering. Additional four lead atom positions were determined using difference Fourier techniques. The resultant phases at 2.7 Å with a figure of merit of 0.42 were improved by density modification using the programs SOLOMON (23) and RESOLVE (24). Molecular boundaries of CEL-III were defined using the programs O (25), MAPMAN (26), and MAMA (27) to produce mask data, which then were used to refine the noncrystallographic symmetry by maximizing correlation between the electron density maps of the two molecules in the asymmetric unit using the program IMP (26). Subsequently, the program DM (20) was used to refine and extend phases to 1.7 Å resolution using noncrystallographic symmetry. The resultant electron density maps were well defined for the whole molecules, and the model was built using the program O. The model was refined using the programs REFMAC5 (28) and ARP/wARP (29). Noncrystallographic symmetry restraints, imposed in the early stage of refinement, were removed in the later stage. At the final stage of refinement, we noticed conflicting amino acid residues between sequences derived from the electron density maps and from the chemical analysis. Hence, 13 residues were replaced as described under "Results." The quality of the model was checked using the programs PROCHECK (30) and WHATCHECK (31). The *R*-factors were further improved with TLS refinement (20).

RESULTS

Structure Determination—CEL-III samples were purified from body fluid of *C. echinata*. Crystals were obtained by the hanging-drop vapor-diffusion method. The crystal data are shown in Table I. A molecular replacement method with the B-chains of ricin or abrin as a search model, having 12 and 16% sequence identities with CEL-III, respectively, did not give good results. Hence, the crystal structure of CEL-III was determined by single isomorphous replacement with anomalous scattering of lead. Statistics related to data collection, phasing, and refinement are summarized in Table I. Electron density maps enabled us to reliably trace all 432 residues. The N-terminal residue was assigned to pyroglutamate according to the published sequence (11). By using omit electron density maps, we found conflicts between our inferred sequences of 13 residues and sequences reported previously (11). These residues were replaced as follows: N15S, Y16F, H36Y, M48I, L92Q, K95R, A97T, V122I, I146V, E173D, Q204S, S340T, and I404V. The reason for such discrepancies is not clear. However, since some differences have been found between protein sequence and that deduced from cDNA, as reported previously (11), these discrepancies may have arisen from microheterogeneity of this protein. In this study, we used the sequence deduced from the electron density map, because they were unambiguously identified. Substitution of BisTris buffer for sodium cacodylate buffer in the crystallization solution was critical for producing better crystals. A BisTris molecule was observed in a cavity of the CEL-III molecule, which may contribute to stabilization of the crystal structure. The two CEL-III molecules (A and B) in the asymmetric unit were essentially the same with an r.m.s. Cα discrepancy of 0.55 Å for 432 residues; residues 356–361 were noticeably different between the two molecules because of their flexibility. Eight peaks with medium to high electron densities per molecule of CEL-III were found in the electron density map. Five Ca^{2+}, two Mg^{2+}, and one Cl^- ion per

[1] The abbreviations used are: CBD, carbohydrate-binding domain; BisTris, bis(2-hydroxyethyl)imino-tris(hydroxymethyl)-methane; r.m.s., root mean square; PDB, Protein Data Bank.

TABLE I
Data collection, phasing, and refinement statistics

	Native	$Pb(CH_3COO)_2$
Data collection and processing statistics		
Space group	$P2_1$	$P2_1$
Unit cell dimension, Å, °	a = 52.4, b = 65.4, c = 126.0, β = 98.2	a = 52.4, b = 64.9, c = 125.7, β = 98.4
Wavelength, Å	1.0000	0.90000
Resolution,[a] Å	26.2–1.70 (1.79–1.70)	45.5–2.70 (2.85–2.70)
Observed reflections	339,219	91,705
Redundancy	3.7	4.0
$\langle I/\sigma I \rangle$	6.9	8.1
Completeness, %	100.0 (100.0)	99.8 (100.0)
R_{merge},[b] %	6.3 (25.7)	6.9 (13.4)
Phasing statistics		
Resolution, Å		45.5–2.7
R_{iso},[c] %		20.5 (23.8)
No. heavy atom sites		8
R_{Cullis},[d] % (centric/acentric)		80.0/75.5
Phasing power[e] (centric/acentric)		1.1/1.3
FOM[f] (centric/acentric)		0.26/0.42
Refinement statistics		
Resolution, Å	26.2–1.7	
Protein atoms	6626	
BisTris buffer	14	
Calcium	10	
Magnesium	4	
Chloride	2	
Water molecules	846	
R_{cryst},[g] %	16.5	
R_{free},[h] %	20.1	
r.m.s. deviation of bond length, Å	0.011	
r.m.s. deviation of bond angle,	1.4	
Ramachandran plot		
In most favored regions, %	90.3	
In allowed regions, %	9.4	
In generously allowed regions, %	0.0	
In disallowed regions, %	0.3	

[a] Numbers in parentheses are for the highest shell.
[b] $R_{merge} = 100\ \Sigma|I - \langle I \rangle|/\Sigma I$, where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections.
[c] $R_{iso} = 100\ \Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$, where $|F_P|$ and $|F_{PH}|$ are the observed native and heavy atom derivative structure factor amplitudes, respectively.
[d] $R_{Cullis} = 100\ \Sigma||F_{PH} \pm F_P| - F_H(calc)|/\Sigma|F_{PH} \pm F_P|$, where $|F_H|$ is the structure factor amplitude for the heavy atom.
[e] Phasing power = $\langle |F_H| \rangle / E$, where E is the lack of closure error.
[f] FOM (figure of merit) = $\langle \Sigma P(\alpha) e^{i\alpha} / \Sigma P(\alpha) \rangle$, where α is the phase, and $P(\alpha)$ is the phase probability distribution.
[g] $R_{cryst} = 100\ \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[h] R_{free} was calculated by using 5% of randomly selected reflections that were excluded from the refinement.

molecule were assigned in order to flatten difference Fourier maps (Table II). Fig. 1 shows an omit electron density map for a Ca^{2+} ion, surrounding residues, and two water molecules (subdomain 2β, as described below). We describe the molecular structure of CEL-III referring only to molecule A of CEL-III. The final and free R-factors were 16.5 and 20.1%, respectively. Asp-373 was found in the disallowed region in the Ramachandran plot, but its model fitted well to an omit map.

Molecular Structure of CEL-III—Sequence analysis (11) and biochemical data (14) indicated that CEL-III consists of two N-terminal CBDs and a C-terminal oligomerization domain containing two hydrophobic regions, the former being homologous to the B-chains of ricin and abrin. The crystal structure of CEL-III supported its three-domain architecture as shown in Fig. 2, *A* and *B*. The two CBDs, domains 1 (residues 1–149) and 2 (residues 150–283), share a similar fold with an r.m.s. Cα discrepancy of 0.77 Å between the corresponding 115 residues. Domains 1 and 2 show 33.8% sequence identity for 145 residues. Domains 1 and 2 each have a β-trefoil fold with a pseudo 3-fold rotation axis (32) and are divided into subdomains 1α (1–50), 1β (51–101), 1γ (102–149), 2α (150–195), 2β (196–241), and 2γ (242–283); residue ranges are in parentheses. Subdomains 1α through 2γ have lengths of 42–51 residues and resemble each other (Fig. 3), with an averaged pairwise r.m.s. Cα discrepancy of 0.62 Å for the corresponding 29 residues. As expected from the internal sequence similarities (11), the structures of subdomains are very similar, except for subdomain 1β. Although domains 1 and 2 of CEL-III resemble the B-chain of ricin (Fig. 4), the relative arrangement of the two domains of CEL-III is different from those of ricin. When domain 1 of CEL-III is superimposed on that of the B-chain of ricin, domain 2 of CEL-III deviates by a 60° clockwise rotation from that of the B-chain of ricin. The C-terminal domain, domain 3 (residues 284–432), comprises two α-helices (H8 and H9) and an eight-stranded β-sandwich. Domain 3 is a unique structure, and it has no structurally similar proteins in the Protein Data Bank (PDB) using the DALI search program (33). The two β-sheets (one consisting of β27 and β29; the other consisting of β25, β26, β28, β30, β31, and β32) of the β-sandwich in domain 3 are held together by hydrophobic interactions. Two α-helices, H8 and H9, lie on the surface of the β-sandwich in parallel with the β-sandwich surface and perpendicular to the longest β-strand β30. A C-terminal helical loop of 25 residues is stabilized by two disulfide bonds as shown in Fig. 2, *A* and *B*. All 24 cysteine residues form 12 disulfide bonds, none of which is an interdomain disulfide bond in CEL-III. Two Mg^{2+} each lie on the pseudo 3-fold axes of the two β-trefoil domains, and two and three Ca^{2+} ions were found in domains 1 and 2, respectively. They were assigned based on difference Fourier maps, coordination numbers, bond distances, and ligand atoms consisting exclusively of oxygen (Figs. 1 and 5*A* and Table II). It seems

TABLE II
Electron density heights and coordination data for ions

Ion	Ion-binding site	Molecule A[a]				Molecule B[a]			
		Electron density[b]	*B*-factor	Bond distance	No. ligands	Electron density[b]	*B*-factor	Bond distance	No. ligands
		e/Å^3	Å^2	Å		*e*/Å^3	Å^2	Å	
Ca^{2+}	Subdomain 1α	1.5	33.8	2.3	7	1.2	43.1	2.2	6
	Subdomain 1β[c]								
	Subdomain 1γ	1.2	44.8	2.4	7	1.2	45.2	2.4	6
	Subdomain 2α	0.92	57.8	2.2	6	1.1	44.2	2.1	6
	Subdomain 2β	1.5	31.6	2.2	6	1.7	28.9	2.1	6
	Subdomain 2γ	1.8	27.2	2.4	7	1.8	26.2	2.4	7
Mg^{2+}	Center of domain 1	0.98	38.3	2.4	6	1.0	41.5	2.4	6
	Center of domain 2	1.3	25.5	2.4	6	1.4	21.8	2.4	6
Cl^-		0.98	45.3	3.2	3	1.0	44.8	3.1	3

[a] Molecules A and B are two molecules in the asymmetric unit.
[b] These values refer to difference Fourier maps, in which the *F*(000) was included.
[c] No electron density peaks corresponding to Ca^{2+} were observed.

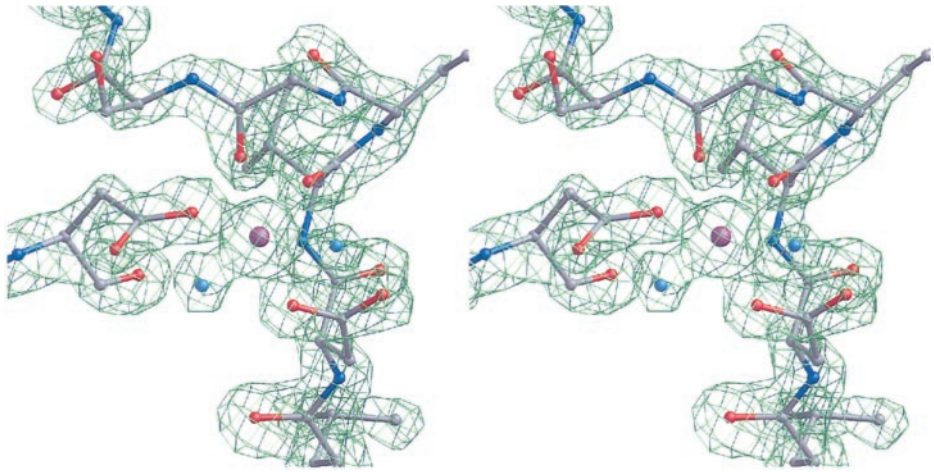

FIG. 1. **Stereo pair of the omit electron density map contoured at 4.2σ for a Ca^{2+}-binding site (subdomain 2β).** The residues, Ca^{2+} (*purple*), and water molecules (*cyan*) depicted in the figure were omitted in the calculation of the map. The figure was produced using the program BOBSCRIPT (49).

possible that the binding of two Mg^{2+} ions at the pseudo 3-fold axes of the β-trefoil domains resulted from the relatively high concentration of Mg^{2+} (200 mM) compared with that of Ca^{2+} (10 mM) under the crystallization conditions in this study, and these positions might be occupied by Ca^{2+} in the absence of a high concentration of Mg^{2+}. In each subdomain, except for subdomain 1β, the Ca^{2+} is coordinated by the side chain oxygen atoms of two Asp residues (*e.g.* Asp-256 and Asp-276 for subdomain 2γ in Fig. 5*A*), the main chain carbonyl groups of Ile/Val and Gly residues (*shaded cyan* in Fig. 5*B*; *e.g.* Val-257 and Gly-259 in Fig. 5*A*), and two or three water molecules (shown in Fig. 5*A*). Table II shows the electron density heights and temperature factors of plausible Ca^{2+}, Mg^{2+}, and Cl^- for the two CEL-III molecules in the asymmetric unit. The descending order of heights for five Ca^{2+} ions in the subdomains is almost the same for molecules A and B as shown in Table II, namely 2γ, 2β, 1α, 1γ, and 2α. The distribution of temperature factors for α-carbons along the peptide chain is very similar for the two molecules in the asymmetric unit. The temperature factors of domains 2 and 3 as a whole are low, whereas those of domain 1 are relatively high. It is also noted that the temperature factors of the front part of the molecule in Fig. 2*A* are high compared with those of the distant part.

Comparison of Carbohydrate-binding Domains—Structure-based alignment of CBDs supplemented by the program ClustalW (34) was shown for β-trefoil and $(QXW)_3$ (18) proteins of known structure in Fig. 5*B*. All of these proteins have a subdomain structure in common. It is composed mainly of the following: two β-strands (β22 and β23 for subdomain 2γ of CEL-III in Fig. 5*A*); a loop (between β22 and β23 for 2γ); a 3_{10} helix (H7 for 2γ) or a kink; and residues forming the hydrophobic core of β-trefoil domains as shown in *green letters* (Val-257 and Phe-279 for 2γ). The positions of two cysteine residues (*shaded yellow* in Fig. 5*B*; *e.g.* Cys-254 and Cys-271 in Fig. 5*A*) and one 3_{10} helix (written in *red letters* in Fig. 5*B*) are conserved among the six subdomains of CEL-III. Most interestingly, only CEL-III is Ca^{2+}-dependent among the β-trefoil proteins shown in Fig. 5*B*. The two Asp residues as indicated by *vertical arrows* in Fig. 5*B*, corresponding to Asp-256 and Asp-276 in Fig. 5*A*, are strictly conserved and coordinate Ca^{2+} in CEL-III except for subdomain 1β. The motif of $(QXW)_3$ (shown as *boxes* in Fig. 5*B*) is found commonly in galactose-specific β-trefoil lectins and is also observed in the subdomains of CEL-III, although Trp at the third position in the motif is replaced by Ile for 1α, and Phe for 1γ, 2α, and 2γ. In the complex structures of ricin, ebulin from *Sambucus ebulus* (35), ML-I from *Viscum album* (36), and xylanase from *Streptomyces olivaceoviridis* E-86 (37), the first Gln residues in the motif, which do not bind sugars directly, contribute to stabilize the local structure of carbohydrate-binding sites (Fig. 6*C*). As in other proteins of Fig. 5*B*, this Gln residue of CEL-III (Gln-277 in Fig. 5*A*) is the fourth residue in the 3_{10} helix (Fig. 5*B*) and makes hydrogen bonds with main chain atoms (Fig. 5*A*). It is likely that the Gln residues are essential for maintaining galactose-binding sites.

DISCUSSION

Carbohydrate-binding Domains—CEL-III is the first crystal structure of Ca^{2+}-dependent animal lectins with two β-trefoil domains, which were originally found in the B-chain of ricin. The β-trefoil domains of ricin and similar proteins have the characteristic sequence of $(QXW)_3$ and are referred to as the

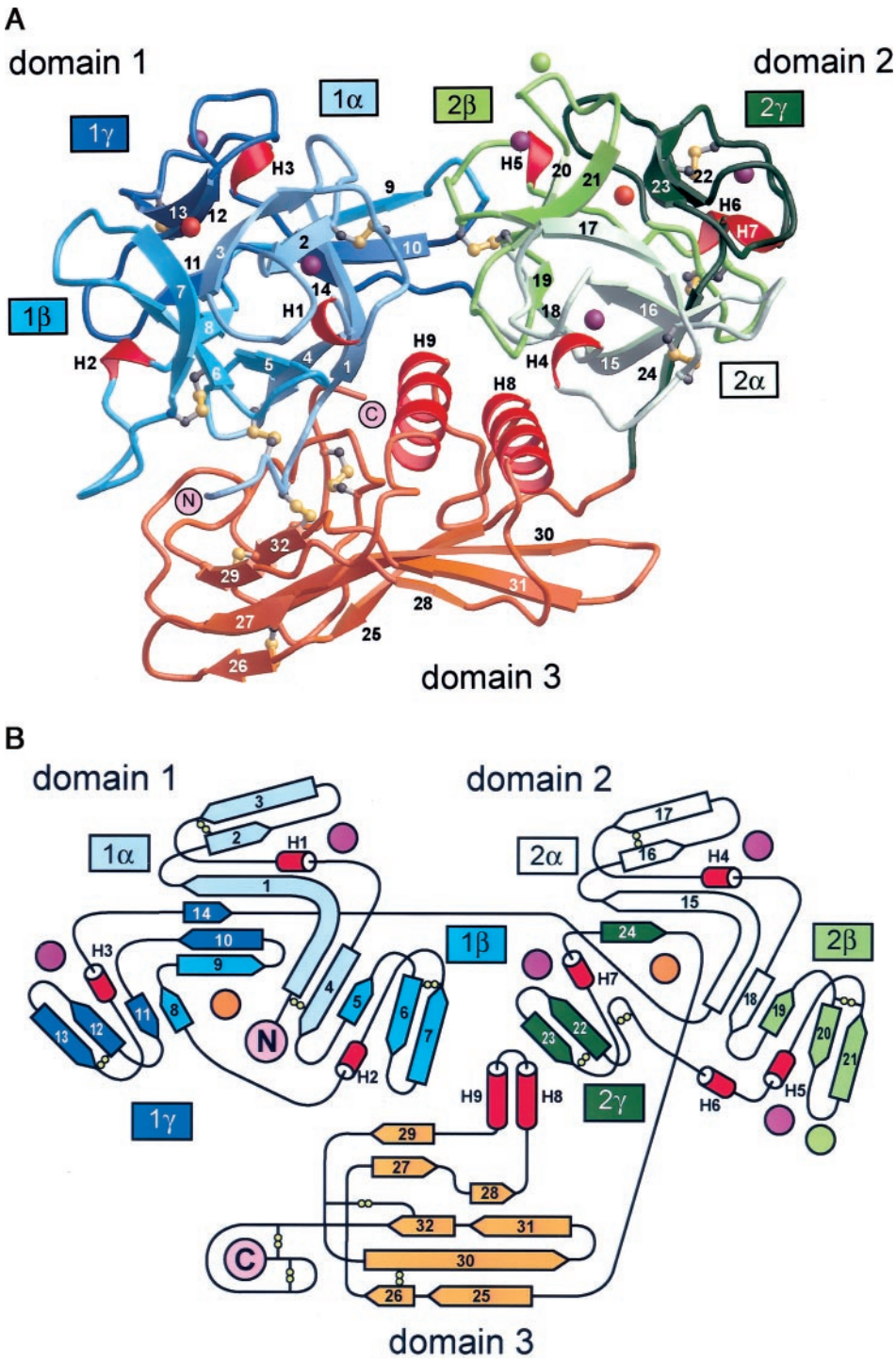

FIG. 2. **Three-dimensional structure and topology diagram of CEL-III.** *A*, ribbon model of a CEL-III molecule. Domains 1–3 are colored *blue, green*, and *orange*, respectively. The color gradation is shown from bright (*e.g.* subdomain α) to dark (subdomain γ) colors along the polypeptide chain from the N to the C termini in domains 1 and 2. α-Helices and 3_{10} helices are colored *red*. Cys residues are depicted as *ball-and-stick* models (*yellow* indicates sulfur atoms, and *gray* indicates carbon atoms). Calcium, magnesium, and chloride ions are shown as *purple, orange*, and *green balls*, respectively. Numbers *1–32* indicate the serial numbers of β-strands, β1–β32. *H1–H7* designate 3_{10} helices and *H8-H9* α-helices. *Numbers* and *Greek letters* in *boxes* indicate subdomains. Figs. 2*A*, 3–5*A*, 6*A–C*, and 7*B–C* were made using the programs MOLSCRIPT (50) and RASTER3D (51). *B*, topology diagram of CEL-III. Color coding is the same as in *A*. Disulfide bonds are shown as *double yellow circles*.

$(QXW)_3$ family (18). CEL-III also has $(QXW)_3$ motifs in its domains 1 and 2, although the third residues are replaced by Ile or Phe in some subdomains as described above. The $(QXW)_3$ family includes not only lectins but also various enzymes, toxins, and cell surface receptors. In these proteins, domains having $(QXW)_3$ motifs are thought to function as carbohydrate-binding modules to assist other distinct domains. In CEL-III, besides Gln-Xaa-Trp/Ile/Phe sequences, hydrophobic core residues and disulfide bonds are conserved as in other $(QXW)_3$ proteins shown in Fig. 5*B*. One of the structural features of domains 1 and 2 of CEL-III is the relatively high sequence similarity between their subdomains, and the highest sequence identity is 54% between subdomains 2α and 2β. Such a high similarity between subdomains is reflected in a higher pseudo

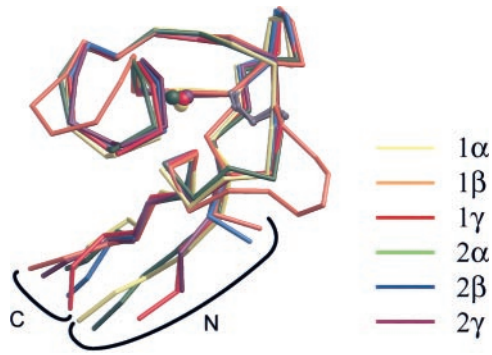

FIG. 3. **Superimposition of Cα traces for six subdomains 1α through 2γ of CEL-III.** Each subdomain was superimposed onto subdomain 1α by the least squares method using 29 Cα atoms. Calcium ions are shown as *spheres*. Disulfide bridges are depicted in *gray*.

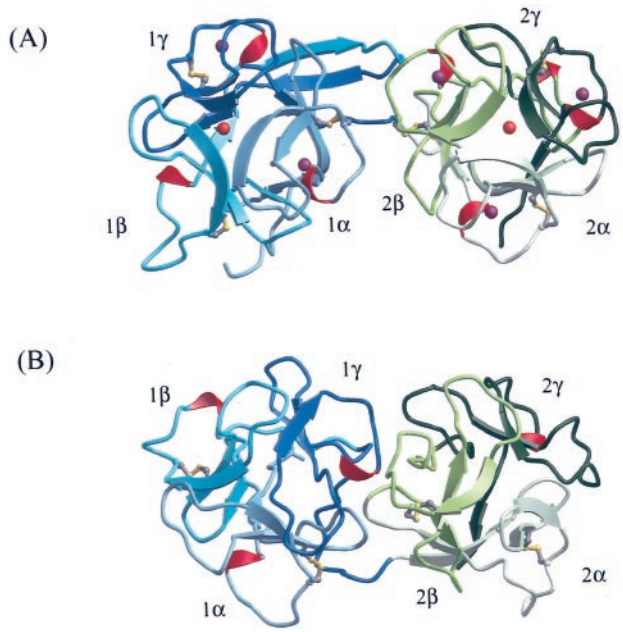

FIG. 4. **Ribbon models of β-trefoil domains for CEL-III (*A*) and the B-chain of ricin (*B*).** The color coding is the same as in Fig. 2*A*.

3-fold symmetry of domains 1 and 2 than those of the ricin B-chain (Fig. 4). In contrast to β-trefoil lectins other than CEL-III in Fig. 5*B*, CEL-III requires Ca^{2+} for its carbohydrate binding ability (4). In the crystal structure of CEL-III, bound Ca^{2+} ions were found to be coordinated by residues located in the putative carbohydrate-binding sites (Fig. 5, *A* and *B*). This suggests that bound Ca^{2+} may be directly involved in carbohydrate binding like C-type lectins and Ca^{2+}-dependent lectins from *Pseudomonas aeruginosa* (38, 39). In the carbohydrate complex crystals of these lectins, 3-OH and 4-OH of the bound carbohydrates, directly coordinate Ca^{2+} in addition to protein ligands. On the other hand, there were also two or three open coordinating sites for a Ca^{2+} located in each subdomain of CEL-III. Water molecules in CEL-III crystals occupied these positions (Figs. 1 and 5*A*). Therefore, it seems possible that CEL-III binds carbohydrate through bonds between OH groups of the carbohydrates and the Ca^{2+} like the above-mentioned Ca^{2+}-dependent lectins (Fig. 6*A*). A hypothetical lactose model was positioned such that the peptide chain of 25 residues in subdomain 2γ of CEL-III was best superimposed onto that of the B-chain of ricin. The 3-OH and 4-OH of galactose in the crystal structure of ricin happened to occupy the positions of water molecules coordinating Ca^{2+} in CEL-III. Alternatively, direct binding of carbohydrate to the protein residues in the putative binding sites could also be possible through replacement of Ca^{2+}. In this case, the Ca^{2+}-coordinating residues, such as Asp-256 and Asp-276, may form hydrogen bonds with OH groups of carbohydrate as illustrated in Fig. 6*B*. However, formation of bifurcated hydrogen bonds of 3-OH in the galactose moiety like the ricin B-chain (Fig. 6*C*) seems rather unfavorable, because these Asp side chains only play a role as hydrogen bond acceptors. Solvent-exposed aromatic residues are known to be involved in Gal/GalNAc binding in C-type lectins (40, 41), as well as β-trefoil lectins listed in Fig. 5*B*. Similar protruding aromatic residues are also found in all subdomains of CEL-III except for 1β (*shaded green* in Fig. 5*B*; *e.g.* Trp-269 in Fig. 5*A*). Because they are in the vicinity of Ca^{2+} ions, they may be involved in the interaction with carbohydrate in the manner of the above-mentioned proteins. Previous biochemical data suggested two carbohydrate-binding sites in CEL-III with equal affinity (42). However, all the subdomains of CEL-III except for subdomain 1β share very similar structural features (Figs. 3 and 5*B*). It is therefore reasonable to infer that all of the five putative binding sites could bind carbohydrate chains, although their affinity may be different. Binding of multiple carbohydrate chains to one β-trefoil domain has been observed in the case of the xylan-binding domain of xylanase (37), whose subdomains (α, β, and γ) share structural features for binding carbohydrates.

Mechanism of Hemolysis—In the process of erythrocyte hemolysis, CEL-III specifically recognizes carbohydrates containing a β-1,4-galactosidic bond on the surface of erythrocytes (4). At first, CEL-III binds to these carbohydrates, and then possibly six molecules of CEL-III oligomerize, leading to the formation of an ion-permeable pore and subsequent osmotic disruption of erythrocytes (10). The estimated diameter of the pore is 17.5 Å at maximum (7) and 10 Å at minimum, considering that ATP (7) and carboxyfluorescein (8) go through the pores. Biochemical data revealed that domain 3 is responsible for oligomerization and insertion of CEL-III into the erythrocyte membrane (14). Our working hypothesis presented so far based on biochemical data is that the binding of CEL-III to carbohydrates triggers secondary structural alterations, leading to the oligomerization of six molecules of CEL-III and membrane insertion. Based on both biochemical and structural information, we propose to extend the four-step mechanism to five steps, in which the domains rearrange between carbohydrate binding and secondary structural alteration steps. We envision this domain rearrangement as a change in the relative positions of domains 1–3 leading to the exposure of domain 3 to the solvent. Most interestingly, there are no disulfide bonds and only a few hydrophobic bonds in the interfaces of domains 1 and 3 and domains 2 and 3. A synthetic peptide fragment corresponding to residues 332–351 (see Fig. 7*A*) showed antibacterial activity (43). Although this region is buried in the CEL-III molecule, it seems possible that this region interacts with target cell membranes once the interior of domain 3 is exposed to the solvent. The secondary structural alterations of CEL-III are supported by spectroscopic experiments (10, 14). The circular dichroism spectral change of domain 3 upon oligomerization revealed a change in secondary structural contents. Fluorescence spectral change of CEL-III upon carbohydrate binding in the presence of Ca^{2+} showed an increase in hydrophobicity of the molecule probably because the hydrophobic interior of domain 3 was exposed to the solvent accompanied by secondary structural alterations. SDS-PAGE experiments revealed an increase in molecular mass attributable to the oligomerization of six molecules (14, 44).

A molecular mechanism of pore formation by bacterial toxins, such as α-hemolysin from *Staphylococcus aureus*, based on

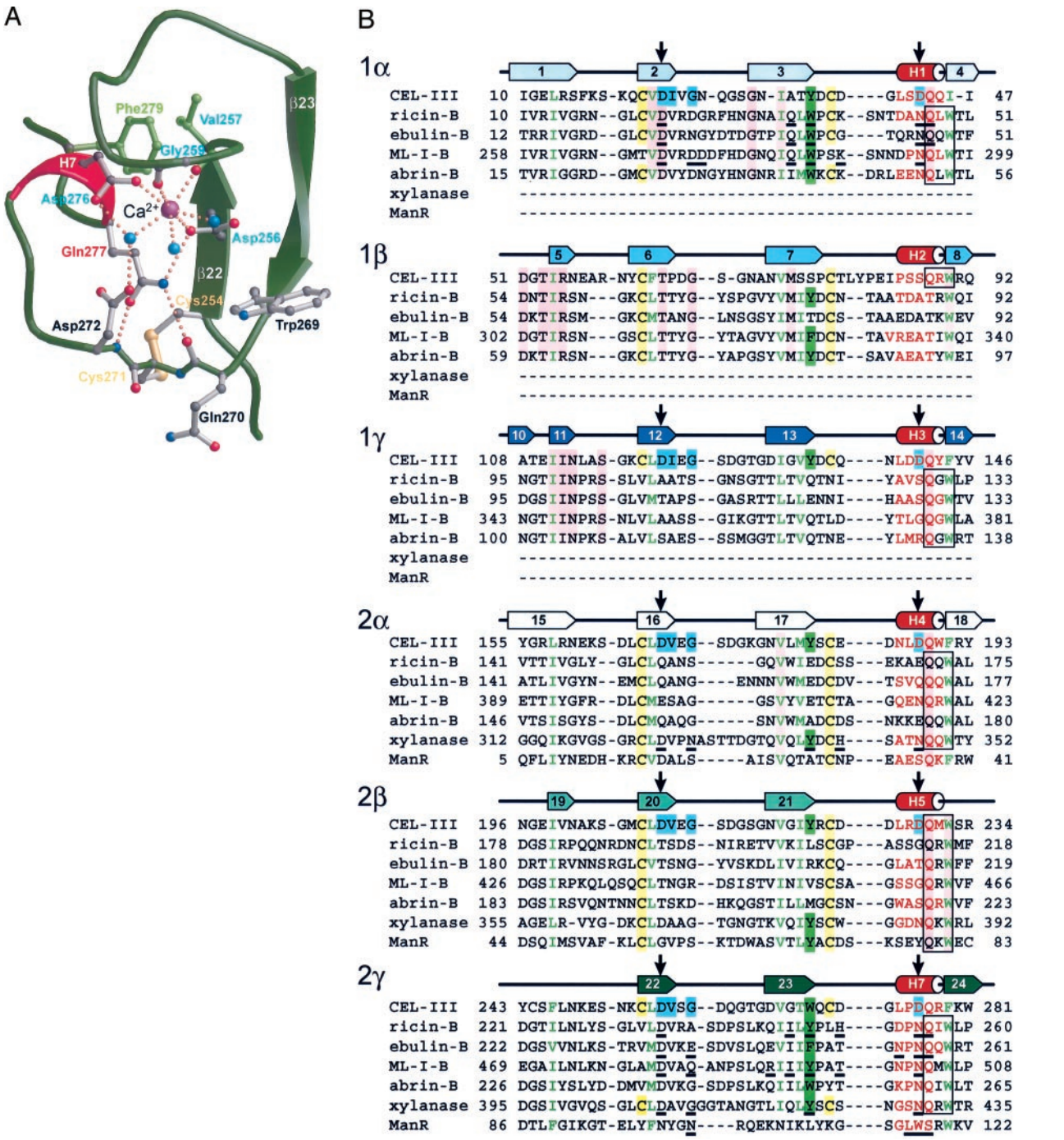

FIG. 5. **Structure and sequence alignment of subdomains.** *A*, ribbon model of subdomain 2γ of CEL-III. Color coding is the same as in Fig. 2, *A* and *B*. Val-257 and Phe-279, contributing to the hydrophobic core of a β-trefoil domain, are colored *green.* Carbon, oxygen, and nitrogen atoms are colored *gray, red,* and *blue*, respectively. Three water molecules coordinating to the Ca^{2+} are shown as *blue balls.* Hydrogen and Ca^{2+}-oxygen bonds are indicated as *orange dotted lines.* Color coding for residues corresponds to that of *B*. *B*, structure-based multiple sequence alignment for six subdomains of CEL-III and $(QXW)_3$ proteins of known structure. Ricin-B stands for the B-chain of ricin from *R. communis* (12) (PDB code 2AAI), ebulin-B for the B-chain of ebulin from *S. ebulus* (35) (PDB code 1HWO), abrin-B for the B-chain of abrin from *A. precatorius* (13) (PDB code 1ABR), ML-I-B for the B-chain of Mistletoe lectin I from *V. album* (36, 52) (PDB code 1M2T), and ManR for mannose receptor from *Mus musculus* (53) (PDB code 1FWU). Xylanase is derived from *S. olivaceoviridis* E-86 (37) (PDB code 1IS0). Secondary structural elements of CEL-III are shown *above* the sequences of CEL-III, whose colors and notations are the same as for Fig. 2, *A* and *B*. *Pickets* and *cylinders* represent β-strands and 3_{10} helices, respectively. Strictly conserved residues are *shaded pink,* and Cys residues forming disulfide bonds are *shaded yellow*. Protruding aromatic residues exposed to solvent are *shaded green.* Residues coordinating Ca^{2+} in CEL-III are *shaded cyan. Vertical arrows* indicate the Asp residues commonly found in the subdomains of CEL-III. Residues forming hydrophobic cores of β-trefoil domains are shown in *green.* Residues in 3_{10} helices are shown in *red.* Motifs $(QXW)_3$ are *boxed.* The residues involved in binding to carbohydrates are *underlined.*

an x-ray structure has been proposed in which seven monomers associate with 7-fold rotational symmetry, forming a pore at the center of the oligomer, concomitantly with conformational change (15, 16). In α-hemolysin, a two-stranded antiparallel β-sheet of ~40 residues of toxin protomer was assumed to be formed by conformational change during the process of oligomerization, leading to the formation of a β-barrel composed of 14 β-strands inserted into the host cell membrane. A similar pore-forming model was proposed for anthrax toxin protective antigen from *Bacillus anthracis* (17). In our model, α-helices,

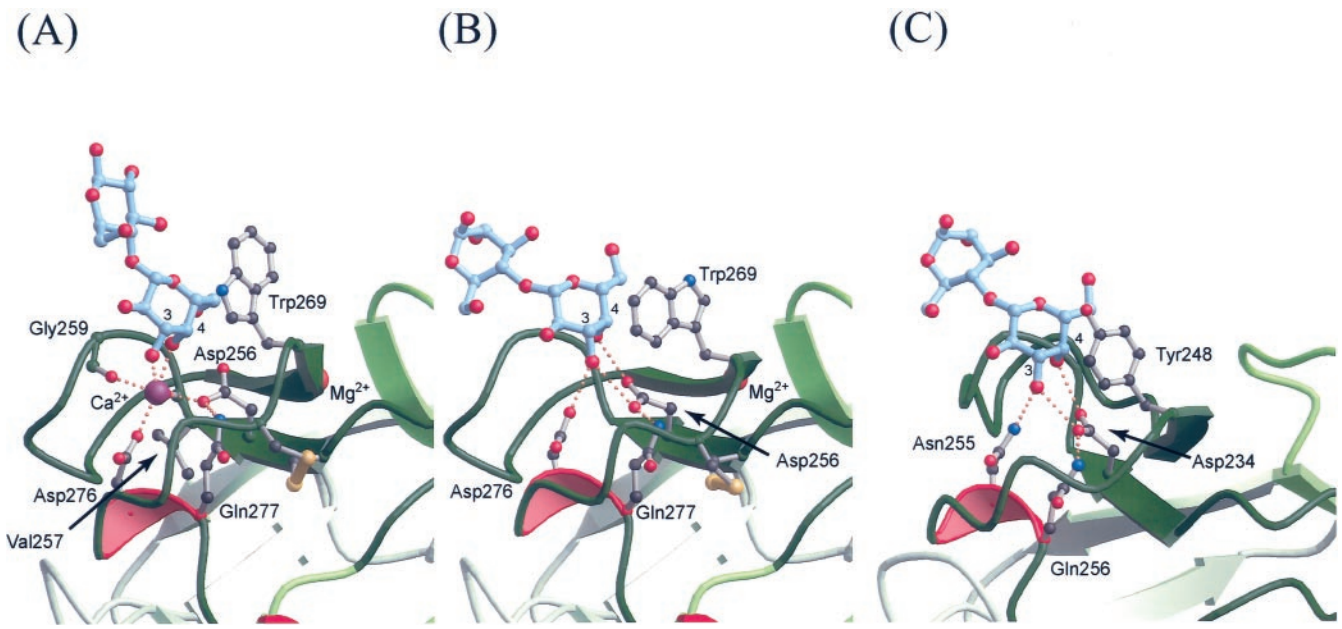

FIG. 6. **Ribbon models of lactose binding to subdomain 2γ of CEL-III and ricin B-chain.** *A*, hypothetical model for CEL-III in the presence of Ca^{2+}. *B*, hypothetical model for CEL-III in the absence of Ca^{2+}. The lactose model was placed in order to mimic the hydrogen network in ricin. *C*, experimentally determined lactose-binding site of ricin. Bonds of lactose molecules are colored *cyan*.

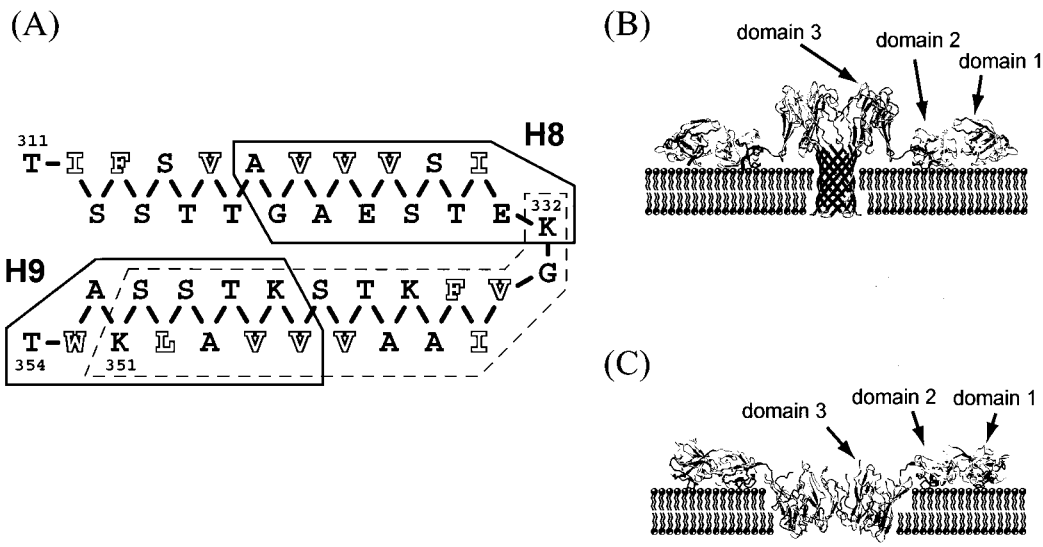

FIG. 7. **Oligomerization and hemolysis models of CEL-III.** *A*, hypothetical model of a two-stranded antiparallel β-sheet, which under certain conditions would be transformed from α-helices, H8 and H9. *Solid lines* enclose the residues in α-helices H8 and H9. *Broken lines* enclose residues Lys-332 to Lys-351 with antibacterial activity. Hydrophobic residues are *letters outlined in black* and have a *white background* (*e.g. I*, *F*, *V*, *W*, and *L*). *B*, hypothetical hemolysis model for secondary structural alteration of α-helices H8 and H9 into a β-barrel and insertion into the cell membrane. Lipid bilayers are schematically depicted. *C*, hypothetical hemolysis model for insertion of the whole domain 3 of CEL-III into the cell membrane. A β-structure thought to be transformed from two α-helices, H8 and H9, is not shown.

H8 and H9, in domain 3 make a conformational change to a two-stranded antiparallel β-sheet of 42 residues, and six β-sheets of CEL-III then associate to form a β-barrel structure in the erythrocyte membrane, as do bacterial toxins (Fig. 7*B*). Fig. 7*A* shows our hypothetical two-stranded antiparallel β-sheet, which under certain conditions would be transformed from two α-helices, H8 and H9, that are present in the crystal structure of CEL-III. When these α-helices rearrange to a β-sheet, many hydrophobic residues align on one side of the β-sheet in this model. The number of residues in our antiparallel β-sheet of two strands is similar to that of α-hemolysin. This is sufficient for penetrating the erythrocyte cell membrane. In our β-barrel model, these hydrophobic residues should face the outside of the β-barrel, consistent with an increase in hydrophobicity and secondary structural alterations described above. Several prediction methods of secondary structure did not show an α-helical structure for α-helices, H8 and H9, suggesting that this region has a tendency for a non-α-helical structure. Six CEL-III monomers associate to form an oligomer by strong noncovalent interactions, as evidenced by almost irreversible association of domain 3 in the solution (14, 44).

Alternatively, another oligomerization model might be presented, in which the domains rearrange upon carbohydrate binding, followed by a transformation of two α-helices into a β-structure and insertion of the whole domain 3 into the cell membrane (Fig. 7*C*). The longest dimension of domain 3 is ~40 Å long, enough to penetrate cell membranes. It was reported that CEL-III induces severe damage to the erythrocyte membrane after relatively long treatments (10), as well as drastic

changes in the shapes of liposomes containing lactosyl ceramide (45), compared with bacterial toxins. In contrast, bacterial pore-forming toxins generally form stable membrane pores without destruction of the membrane, thereby enabling them to be used as membrane-permeabilizing agents (46). This may be reflected in the difference in the extent to which protein is inserted into the lipid membranes between CEL-III and bacterial pore-forming toxins.

Recently, Tateno and Goldstein characterized (47) and crystallized (48) hemolytic and hemagglutinating lectins from *Laetiporus sulfureus*. These lectins are thought to consist of an N-terminal carbohydrate-binding domain containing $(QXW)_3$ motifs and a C-terminal oligomerization domain. There may be similarity in the hemolytic mechanisms between CEL-III and these lectins. To verify our model, we are determining the structures of CEL-III in complexes with carbohydrates and of the oligomerized CEL-III.

Acknowledgments—We thank Dr. Yoshiaki Kouzuma (Ibaraki University) and Dr. Makoto Kimura (Kyushu University) for useful discussions.

REFERENCES

1. Kilpatrick, D. C. (2002) *Biochim. Biophys. Acta* **1572,** 187–197
2. Drickamer, K. (1988) *J. Biol. Chem.* **263,** 9557–9560
3. Day, A. J. (1994) *Biochem. Soc. Trans.* **22,** 83–88
4. Hatakeyama, T., Kohzaki, H., Nagatomo, H., and Yamasaki, N. (1994) *J. Biochem.* (*Tokyo*) **116,** 209–214
5. Hatakeyama, T., Matsuo, N., Shiba, K., Nishinohara, S., Yamasaki, N., Sugawara, H., and Aoyagi, H. (2002) *Biosci. Biotechnol. Biochem.* **66,** 157–163
6. Hatakeyama, T., Ohuchi, K., Kuroki, M., and Yamasaki, N. (1995) *Biosci. Biotechnol. Biochem.* **59,** 1314–1317
7. Hatakeyama, T., Nagatomo, H., and Yamasaki, N. (1995) *J. Biol. Chem.* **270,** 3560–3564
8. Oda, T., Tsuru, M., Hatakeyama, T., Nagatomo, H., Muramatsu, T., and Yamasaki, N. (1997) *J. Biochem.* (*Tokyo*) **121,** 560–567
9. Oda, T., Shinmura, N., Nishioka, Y., Komatsu, N., Hatakeyama, T., and Muramatsu, T. (1999) *J. Biochem.* (*Tokyo*) **125,** 713–720
10. Hatakeyama, T., Furukawa, M., Nagatomo, H., Yamasaki, N., and Mori, T. (1996) *J. Biol. Chem.* **271,** 16915–16920
11. Nakano, M., Tabata, S., Sugihara, K., Kouzuma, Y., Kimura, M., and Yamasaki, N. (1999) *Biochim. Biophys. Acta* **1435,** 167–176
12. Montfort, W., Villafranca, J. E., Monzingo, A. F., Ernst, S. R., Katzin, B., Rutenber, E., Xuong, N. H., Hamlin, R., and Robertus, J. D. (1987) *J. Biol. Chem.* **262,** 5398–5403
13. Tahirov, T. H., Lu, T.-H., Liaw, Y.-C., Chen, Y.-L., and Lin, J.-Y. (1995) *J. Mol. Biol.* **250,** 354–367
14. Kouzuma, Y., Suzuki, Y., Nakano, M., Matsuyama, K., Tojo, S., Kimura, M., Yamasaki, T., Aoyagi, H., and Hatakeyama, T. (2003) *J. Biochem.* (*Tokyo*) **134,** 395–402
15. Song, L., Hobaugh, M. R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J. E. (1996) *Science* **274,** 1859–1866
16. Olson, R., Nariya, H., Yokota, K., Kamio, Y., and Gouaux, E. (1999) *Nat. Struct. Biol.* **6,** 134–140
17. Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H., and Liddington, R. C. (1997) *Nature* **385,** 833–838
18. Hazes, B. (1996) *Protein Sci.* **5,** 1490–1501
19. Powell, H. R. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 1690–1695
20. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
21. Matthews, B. W. (1968) *J. Mol. Biol.* **33,** 491–497
22. de La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494
23. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **52,** 30–42
24. Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56,** 965–972
25. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
26. Kleywegt, G. J., and Read, R. J. (1997) *Structure* (*Lond.*) **5,** 1557–1569
27. Kleywegt, G. J., and Jones, T. A. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 941–944
28. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255
29. Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 1445–1450
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
31. Hooft, W. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* **381,** 272–272
32. Murzin, A. G., Lesk, A. M., and Chothia, C. (1992) *J. Mol. Biol.* **223,** 531–543
33. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
34. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
35. Pascal, J. M., Day, P. J., Monzingo, A. F., Ernst, S. R., Robertus, J. D., Iglesias, R., Pérez, Y., Férreras, J. M., Citores, L., and Girbés, T. (2001) *Proteins* **43,** 319–326
36. Niwa, H., Tonevitsky, A. G., Agapov, I. I., Saward, S., Pfüller, U., and Palmer, R. A. (2003) *Eur. J. Biochem.* **270,** 2739–2749
37. Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2002) *J. Mol. Biol.* **316,** 65–78
38. Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gautier, C., Pérez, S., Wu, A. M., Gilboa-Garber, N., and Imberty, A. (2002) *Nat. Struct. Biol.* **9,** 918–921
39. Cioci, G., Mitchell, E. P., Gautier, C., Wimmerová, M., Sudakevitz, D., Pérez, S., Gilboa-Garber, N., and Imberty, A. (2003) *FEBS Lett.* **555,** 297–301
40. Kolatkar, A. R., Leung, A. K., Isecke, R., Brossmer, R., Drickamer, K., and Weis, W. I. (1998) *J. Biol. Chem.* **273,** 19502–19508
41. Poget, S. F., Legge, G. B., Proctor, M. R., Butler, P. J. G., Bycroft, M., and Williams, R. L. (1999) *J. Mol. Biol.* **290,** 867–879
42. Sallay, I., Hatakeyama, T., and Yamasaki, N. (1998) *Biosci. Biotechnol. Biochem.* **62,** 1757–1761
43. Hatakeyama, T., Suenaga, T., Eto, S., Niidome, T., and Aoyagi, H. (2004) *J. Biochem.* (*Tokyo*) **135,** 65–70
44. Sallay, I., Toji, S., Nomiyama, K., Kouzuma, Y., Kimura, M., and Yamasaki, N. (2001) *Biosci. Biotechnol. Biochem.* **65,** 1347–1352
45. Hatakeyama, T., Sato, T., Taira, E., Kuwahara, H., Niidome, T., and Aoyagi, H. (1999) *J. Biochem.* (*Tokyo*) **125,** 277–284
46. Bhakdi, S., Weller, U., Walev, I., Martin, E., Jonas, D., and Palmer, M. (1993) *Med. Microbiol. Immunol.* **182,** 167–175
47. Tateno, H., and Goldstein, I. J. (2003) *J. Biol. Chem.* **278,** 40455–40463
48. Mancheño, J. M., Tateno, H., Goldstein, I. J., and Hermoso, J. A. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 1139–1141
49. Esnouf, R. M. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 938–940
50. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
51. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524
52. Krauspenhaar, R., Eschenburg, S., Perbandt, M., Kornilov, V., Konareva, N., Mikailova, I., Stoeva, S., Wacker, R., Maier, T., Singh, T., Mikhailov, A., Voelter, W., and Betzel, C. (1999) *Biochem. Biophys. Res. Commun.* **257,** 418–424
53. Liu, Y., Chirino, A. J., Misulovin, Z., Leteux, C., Feizi, T., Nussenzweig, M. C., and Bjorkman, P. J. (2000) *J. Exp. Med.* **191,** 1105–1115

Crystal Structure of the Hemolytic Lectin CEL-III Isolated from the Marine Invertebrate *Cucumaria echinata*: IMPLICATIONS OF DOMAIN STRUCTURE FOR ITS MEMBRANE PORE-FORMATION MECHANISM

Tatsuya Uchida, Takayuki Yamasaki, Seiichiro Eto, Hajime Sugawara, Genji Kurisu, Atsushi Nakagawa, Masami Kusunoki and Tomomitsu Hatakeyama

J. Biol. Chem. 2004, 279:37133-37141.
doi: 10.1074/jbc.M404065200 originally published online June 11, 2004
