Trivalent Recognition Unit of Innate Immunity System

CRYSTAL STRUCTURE OF TRIMERIC HUMAN M-FICOLIN FIBRINOGEN-LIKE DOMAIN

Received for publication, September 7, 2006, and in revised form, October 23, 2006. Published, JBC Papers in Press, December 4, 2006, DOI 10.1074/jbc.M608627200

Michikazu Tanio1, Shin Kondo5, Shigetoshi Sugio1, and Toshiyuki Kohno1,2
From the 1Mitsubishi Kagaku Institute of Life Sciences, 11 Minamiooya, Machida, Tokyo 194-8511 and the 5ZOEGENE Corporation, 1000 Kamoshida, Aoba, Yokohama 227-8502, Japan

Ficolins are a kind of pathogen-recognition molecule in the innate immune systems. To investigate the discrimination mechanism between self and non-self by ficolins, we determined the crystal structure of the human M-ficolin fibrinogen-like domain (FD1), which is the ligand-binding domain, at 1.9 Å resolution. Although the FD1 monomer shares a common fold with the fibrinogen γ fragment and tachylectin-5A, the Asp-282–Cys-283 peptide bond, which is the predicted ligand-binding site on the C-terminal P domain, is a normal trans bond, unlike the cases of the other two proteins. The trimeric formation of FD1 results in the separation of the three P domains, and the spatial arrangement of the three predicted ligand-binding sites on the trimer is very similar to that of the trimeric collectin, indicating that such an arrangement is generally required for pathogen-recognition. The ligand binding study of FD1 in solution indicated that the recombinant protein binds to N-acetyl-D-glucosamine and the peptide Gly-Pro-Arg-Pro and suggested that the ligand-binding region exhibits a conformational equilibrium involving cis-trans isomerization of the Asp-282–Cys-283 peptide bond. The crystal structure and the ligand binding study of FD1 provide an insight of the self- and non-self discrimination mechanism by ficolins.

Surveillance systems of innate immunity are present in all multicellular organisms and play a crucial role in the first line of defense against pathogens. Ficolins, as well as collectins, are one of the most important groups of pattern recognition molecules in the innate immunity systems (1–7) and have been identified in both vertebrates and invertebrates (6). Ficolins are comprised of a collagen-like domain at the N terminus and a fibrinogen-like domain (FBG), which is the sugar-binding site, at the C terminus (8, 9). Collectins, such as mannose-binding lectin (MBL), lung surfactant protein A, and surfactant protein D, also consist of an N-terminal collagen-like domain and a C-terminal carbohydrate-recognition domain (CRD) that binds to certain carbohydrates such as mannosyl and GlcNAc (9, 10) dependently. The CRD on MBL, surfactant protein A, and surfactant protein D forms a trimeric structure through a triple α-helical coiled-coil at a short neck region between the collagen-like domain and the CRD (10–12). Ficolins also form trimers (8, 13, 14), although the mechanism of trimerization is unclear. Both ficolins and collectins form trimer-based multimers that are N-terminally linked by disulfide bonds (15). Ficolins and MBL also interact with MBL-associated serine proteases, and their complexes activate the lectin complement pathway (6, 16–23).

Ficolins were originally discovered in porcine uterus membrane extracts as transforming growth factor-β-binding proteins (24, 25). In human, L-ficolin and H-ficolin in serum and M-ficolin in cells have been characterized (9, 14, 26–29). L-ficolin (synonymous with ficolin-2 or Ficolin/P35) binds to GlcNAc (29, 30) and GalNAc (8). The binding ability is inhibited by acetylated compounds, indicating that this protein specifically recognizes acetyl groups (31). L-ficolin activates the lectin complement pathway upon binding to lipoteichoic acid, a cell wall component of all Gram-positive bacteria (32). H-ficolin (synonymous with ficolin-3 or Hakata-antigen) has a primary sequence that is 48% identical to that of L-ficolin and binds to GlcNAc, and D-fucose (14). The GlcNAc binding activity of H-ficolin, unlike that of L-ficolin, is not inhibited by acetyl compounds (31). M-ficolin (synonymous with ficolin-1 or Ficolin/P35-related protein) has a primary sequence that is 80 and 48% identical to those of L-ficolin and H-ficolin, respectively (14, 28). M-ficolin binds to GlcNAc, GalNAc, and sialic acid (9, 22), and like L-ficolin, it recognizes acetyl groups (23). Unlike the serum ficolins, M-ficolin has been detected on the surfaces of peripheral blood monocytes and monomeric U937 cells (9, 23, 27) and in secretory granules in the cytoplasm of neutrophils, monocytes, and type II alveolar epithelial cells in the lung (22). An antibody against the recombinant FBG domain of M-ficolin inhibits the phagocytosis of E. coli by U937 cell, suggesting that M-ficolin acts as a monocyte phagocytic receptor or adaptor for recognizing pathogens (9).

The ligand-binding site on the fibrinogen-like domain of ficolins is not yet known. However, the crystal structures of the human fibrinogen γ fragment (33) and of the fibrinogen-like domain of Tachypleus tridentatus tachylectin-5A (TL5A) (34) have shown that the P domain contributes to ligand binding. In angiopeitin-2, it was found that the P domain of the fibrino-
Crystal Structure of M-ficolin Fibrinogen-like Domain

gen-like domain relates to receptor binding (35). These findings suggest that the P domain of ficolins includes the sugar-binding site. Interestingly, although both host cells and pathogens have the same sugars recognized by ficolins and collectins on their surface, these proteins can discriminate between pathogens and the host cell. Here, we present the first crystal structure of the human M-ficolin FBG domain (FD1; 26.8 kDa/monomer). The crystal structure reveals how the three FBG domains form a trimer and provides a structural basis for understanding how these proteins discriminate between self and non-self.

EXPERIMENTAL PROCEDURES

**Purification and Crystallization of FD1**—FD1, comprising residues 115–326 of M-ficolin, with the C-terminal peptide containing a c-myc epitope and a His6 tag, was overexpressed in the yeast *Pichia pastoris*, purified to homogeneity, and crystallized as described previously (36). Crystals were obtained at 20 °C by mixing 0.5 µl of the protein solution (8 mg ml⁻¹ FD1, 8 mM Tris·HCl, pH 8.0, 80 mM NaCl, 4 mM CaCl₂, and 20 mM GlcNAc) with 0.5 µl of the reservoir solution (100 mM MES, pH 5.6, 320 mM Li₂SO₄, and 17% (w/v) polyethylene glycol 4000) (36).

**Structure Determination and Model Refinement**—Two diffraction data sets (180 and 60 images at χ = 0° and −60°, respectively) were collected from the crystal, using radiation at beamline BL24XU of SPring-8 (wavelength 0.8266 Å) at 20 °C. Structure factors were calculated using the program HKL2000 (37), and the two scaled data sets were combined using the CCP4 program suite (38), revealing that Rmerge is 6.3%. The crystal structure was solved with the program EPMR (39), using the crystal structure of TL5A (Protein Data Bank entry: 1JC9 (34)) as a search model. Crystallographic refinement was carried out with the program CNS (40) interspersed with manual model building with XtalView (41). The current atomic model gave Rwork and Rfree factors 20.9 and 24.0%, respectively, against all reflections in the resolution range of 30–1.9 Å (see Table 1). PROCHECK (42) indicates that 85.2% of the residues in the trimer fall in the most favorable regions of φ and ψ conformational space with none in the disallowed conformations. Figures of atomic representations and electron density maps were prepared using PyMOL.

**Measurement of Sugar Binding Activity**—The purified FD1 was diluted to 0.1 mg ml⁻¹ in 200 µl of Tris buffer (50 mM Tris·HCl, pH 8.0, 150 mM NaCl, and 5 mM CaCl₂) with or without 10 mM EDTA, 50 mM dithiothreitol, or 50 mM Gly-Pro-Arg-Glu-GlcNAc. The reaction mixture was incubated with 67 µl of GlcNAc-agarose beads (Sigma; 1 µl of maximally binds to about 0.7 µg of FD1 in the Tris buffer at 4 °C) in a MicroSpin column (Amersham Biosciences) at 20 or 4 °C for 2 h. After washing with the Tris buffer, the bound FD1 was eluted from the column with 200 µl of the Tris buffer containing 0.5 M GlcNAc. The eluted samples were analyzed with SDS-PAGE and visualized by Coomassie Blue staining.

For the investigation of the pH-dependent GlcNAc binding activity, a 300–µl aliquot of 0.05 mg ml⁻¹ FD1 in MES buffer (50 mM MES, 150 mM NaCl, and 5 mM CaCl₂) at several pHs (between 5.2 and 7.0) was applied to a GlcNAc column (34 µl of resin) at 4 °C for 2 h. The bound protein was eluted with 300 µl of the Tris buffer containing 0.5 M GlcNAc at pH 8.0. The concentrations of the applied and eluted FD1 were estimated from the absorption at 280 nm (molar absorption coefficient: 48,700 cm⁻¹ M⁻¹), and the relative fractions of the eluted FD1 at each pH were evaluated using the KaleidaGraph non-linear fitting program (Synergy Software).

**RESULTS**

**Overall Structure of FD1**—The crystal structure of FD1 was solved with molecular replacement and was refined to 1.9 Å resolution (Table 1). The structure contains one trimeric form of FD1 per asymmetric unit (Fig. 1A). The quality of electron density allowed the assignment of 209 residues (Glu and Phe derived from the EcoRI site, residues 115–295 and 301–326 of M-ficolin) to monomer A, 200 residues (Phe from the EcoRI site, residues 115–293, 301–307, and 314–326) to monomer B and 212 residues (residues 115–326) to monomer C. Although GlcNAc was added to the protein solution for crystallization, no obvious electron density for the ligand was observed. Therefore, we concluded that the FD1 structure determined here corresponds to the ligand-free form. Two disulfide bonds, between Cys-118 and Cys-146 and between Cys-270 and Cys-283, were observed in each FD1 monomer. The root-mean-square deviations among the three monomers in the trimer are 0.6 Å for 199–205 equivalent Ca positions. The overall structure of the FD1 monomer is similar to those of TL5A (34) and the fibrinogen γ fragment (33, 43), and like these other two proteins, it consists of three domains (A, B, and P; Fig. 1B). The primary sequence of FD1 is 49 and 45% identical to those of the FBG domain of TL5A and the fibrinogen γ fragment, respectively. The root-mean-square deviation for the α-carbon atoms in FD1 is ~1.1 Å as compared with TL5A (~184 Ca positions in residues 45–262) and ~1.3 Å as compared with the fibrinogen γ

| Table 1 | Data collection and refinement statistics |
|---------|-----------------------------------------|
| χ = 0°  | χ = −60°                                 |
| Data collection |                                  |
| Space group | P2₁                                  |
| Cell dimensions |                                    |
| a, b, c (Å) | 55.16, 117.45, 55.19                   |
| α, β, γ (%) | 90.00, 99.88, 90.00                    |
| Resolution (Å) | 30.19 (1.97–1.19)                   |
| Rmerge (%) | 6.2 (25.6)                             |
| Rfree (%) | 6.7 (28.1)                             |
| Completeness (%) | 34.8 (6.8)                          |
| Redundancy | 23.5 (4.5)                             |
| Rwork (%) | 98.8 (98.3)                            |
| Rfree (%) | 65.6 (67.0)                            |
| Bond lengths (Å) | 0.0056                               |
| Bond angles (°) | 1.33                                 |

Note: Rwork = ΣhI [I(h)−〈I(h)〉]/ΣhI 〈I(h)〉, Rfree was calculated by using random 5% of the reflection data that was omitted in the refinement.
Crystal Structure of M-ficolin Fibrinogen-like Domain

FIGURE 1. Molecular structure of FD1. A, top (left) and side (right) views of the FD1 trimer, with each monomer colored differently. The Ca$^{2+}$ ions are shown as orange spheres. B, a ribbon diagram of the FD1 monomer backbone of monomer C colored by domains: A in blue (residues 115–154), B in red (residues 155–243 and 317–326), and P in cyan (residues 244–316). Ca$^{2+}$ is shown as a yellow sphere.

FIGURE 2. Comparison of the structures among FD1, TL5A, and the human fibrinogen γ fragment. A, the main chain of FD1 (magenta) superimposed on those of TL5A (Protein Data Bank code 1JC9 (34); blue) and the human fibrinogen γ fragment (Protein Data Bank code 3FIB (33); slate). Each Ca$^{2+}$ of the proteins is shown as a sphere. B, focus of the conserved disulfide bond (S–S bond) in the P domain of the three proteins from the region shown by the red arrow in panel A.

is identical to those of TL5A (34) and the fibrinogen γ fragment (33, 43). These aspartic acid residues (Asp-262 and Asp-264) are also conserved in the other two human ficolins. The electron density at the Ca$^{2+}$ in monomer B was weaker than those of the other monomers, and no coordinating water molecules were found around the loop. Therefore, we assigned this position to a water molecule, instead of a Ca$^{2+}$. In monomer C, one of the water molecules coordinating the Ca$^{2+}$ was replaced by the carbonyl oxygen atom of Gly-250 of monomer B from a neighboring trimer in the crystal. The electron density of the Ca$^{2+}$ was relatively weak in monomer C, and thus, its occupancy was assumed to be 0.5 at the refinement stage. Crystal packing may cause these differences of Ca$^{2+}$ bindings between the three monomers because the Ca$^{2+}$-binding sites in monomer A and C, but not that in monomer B, face the neighboring trimers in the crystal.

Trimeric Formation of FD1.—The trimerization of ficolins is thought to be mediated by the N-terminal collagen-like region (13, 14). However, we found that the recombinant FD1, which lacks the collagen-like region, forms a trimer in solution as judged by a dynamic light scattering experiment in 0.1–1 mg ml$^{-1}$ FD1 solution (apparent molecular mass 86.3–88.9 kDa, data not shown). This finding indicates that the trimerization of M-ficolin is accomplished not only by the N-terminal collagen-like region but also by the FBG domain. Our crystal structure of FD1 clarified the details of FBG domain trimerization (Fig. 4). The A and B domains, but not the P domain, contribute to the trimerization. The buried surface areas at the monomer–monomer interfaces are about 500 Å$^2$. Fig. 4B shows the interface of the interactions between monomers A and C. The main chains of Phe-127 and Ser-129 of one monomer form hydrogen bonds with those of Gly-180 and Gln-182 of another monomer. The hydrophobic side chains of Phe-127 and Leu-128 of one mon-
omer are buried in a hydrophobic pocket, mainly composed of Trp-131, Leu-142, Gly-178, Gly-180, Phe-186, and Leu-188 of another monomer. The phenyl ring of Phe-127 is also in contact with the carboxyl oxygen atoms of Gln-177 and Leu-188 and the polar side chain of Asn-192 of the neighboring monomer. The carboxyl oxygen atom of Gln-177 is close to the center of the phenyl ring of Phe-127 and probably prevents its release from the hydrophobic pocket. The C=O...π interaction (44) between these groups might contribute to preventing this release. A salt bridge between Asp-149 of one monomer and Arg-170 of another monomer further stabilizes the trimerization.

Sugar Binding and Novel Peptide Binding Activities—A GlcNAc-agarose affinity assay revealed that the GlcNAc binding activity of FD1 was enhanced at lower temperature (4 °C) and inhibited in the presence of EDTA (Fig. 5A). The elution of the EDTA-treated FD1 was restored by the addition of CaCl2 (data not shown), clearly indicating that M-ficolin requires Ca2+ for GlcNAc-binding. Since the Ca2+-binding site is located on the P domain (Fig. 1B), this domain is involved in ligand binding. Dithiothreitol treatment of FD1 reduced the binding activity (Fig. 5A). Since FD1 contains only two disulfide bonds (Cys-118–Cys-146 on the A domain and Cys-270–Cys-283 on the P domain), the GlcNAc-binding site would be located near the Cys-270–Cys-283 disulfide bond. Interestingly, the peptide Gly-Pro-Arg-Pro (GPRP), which mimics the N-terminal sequence of the fibrin α-chain and inhibits the polymerization of fibrin, prevents the binding of FD1 to GlcNAc-agarose by 50 mM GPRP peptide (data not shown) indicates that GlcNAc and GPRP share the same binding site on FD1. Although the biological significance of GPRP binding of FD1 is unknown, it suggests that M-ficolin might be involved not only in pathogen-recognition but also in blood coagulation and/or fibrinolysis.

pH-dependent Change of Sugar Binding Activity—The GlcNAc binding activity is pH-dependent (Fig. 5B) and is decreased below pH 6.2. This means that the crystallized FD1 at pH 5.6 represents the structure that facilitates ligand release. Our attempt to explain this sigmoid curve using the Henderson-Hasselbalch equation was unsuccessful (correlation coefficient; r² = 0.92), probably because both the binding and the
non-binding forms coexist at any pH. Therefore, we assumed that FD1 exists in an equilibrium between an active (A) and a non-active (N) state at any pH and that the equilibrium depends on the pK_a of some group(s) but not on the presence of the ligands. The relative fraction of the active form in this model is described by the following equation:

\[ f_{\text{active}} = \frac{(1 + K_1^a)(1 + K_2^a)^{pK_A^a - pH}}{(1 + K_1^a)(1 + K_2^a)(1 + 10^{(pH - pK_A^a)})} \]

where \( K_1^a \) and \( K_2^a \) are the equilibrium constants between the active and non-active forms and \( pK_A^a \) is the number of protons involved in the transition from the protonated to deprotonated forms. The calculation revealed that \( K_1^a, K_2^a, pK_A^a, n = 0.4, 0.1, 6.2, 2.8 \), respectively (goodness of fit: \( r^2 = 0.99 \)). From the \( K_1^a \) and \( K_2^a \) values, it was calculated that 71% and 34% of FD1 are active in the deprotonated and protonated states, respectively.

Our crystal structure of FD1 revealed that the FBG domain forms a trimer without the collagen-like region. This trimerization mechanism determines the spatial arrangement of the three P domains and the resulting separation of the three Ca^{2+}-binding sites of FD1. Since the residues related to trimer formation are conserved in L-ficolin, the interaction between the FBG domains in trimeric H-ficolin might be modified. Indeed, the replacement of Phe-127 with threonine in FD1 drastically decreased the monomer-monomer interaction in solution, as judged by a dynamic light scattering experiment (apparent molecular mass 27.6–42.4 kDa in 0.1–1 mg ml^{-1} protein solution) and the GlcNAc binding activity (data not shown). The latter result suggests that the trimeric formation is necessary for the activity of FD1.

The solution study of the GlcNAc binding activity (Fig. 5) indicated that our crystallization condition of FD1 at pH 5.6 and 20 °C is inadequate to form the ligand-binding complex. Although the active FD1 exists at pH 5.6 in solution (Fig. 5B), the crystal structure of FD1 at this pH shows the ligand-free form, suggesting that the crystalized FD1 in this condition would tend to form the non-active state probably by crystal packing. We tried to crystallize FD1 with GlcNAc at neutral pH, but it was unsuccessful. A search for other crystallization conditions for the ligand-binding form is in progress.

The analysis of the pH-dependent GlcNAc binding activity suggests that FD1 exists in the equilibrium between the active and non-active forms (Fig. 5B). In this equilibrium model, the charged state of the unidentified group(s) with pK_a 6.2 is indirectly related to ligand binding because the two forms coexist at any pH. Instead, it may be concerned with the conformation of the ligand-binding region. Since His-284 is located near the Ca^{2+}-binding site and the Cys-270–Cys-283 disulfide bond (Fig. 3A), this residue is a strong candidate for the unidentified group with pK_a 6.2. The pH dependence of FD1 also suggests that the conformation of the ligand-binding region of M-ficolin is affected by the electrostatic environment at the contact surface. This also means that the surfaces of some pathogens can induce the formation of the active conformation of the FBG domains and thereby increase the avidity for the surface. Indeed, the binding specificity for various bacteria differs among ficolins (9, 20, 22, 29, 32, 45). Such surface-dependent conformational changes would facilitate the discrimination between self and non-self.

The predicted ligand-binding site, which is located near the Cys-270–Cys-283 disulfide bond, is in good agreement with those of TL5A and the fibrinogen γ fragment. However, a comparison of the structures among FD1, TL5A, and the fibrinogen γ fragment shows that the conformational difference exists in the ligand-binding region (Fig. 2). An unusual cis peptide bond is present in the structures of TL5A (Arg-218–Cys-219) (34) and the fibrinogen γ fragment (Lys-338–Cys-339) (33), whereas a normal trans bond is found in the structure of FD1 (Asp-282–Cys-283) (34). Since the cis peptide bond allows the backbone nitrogen atom of the cysteine residue to interact...
with the ligand in TL5A and in the fibrinogen γ fragment, the cis configuration in these proteins is required for their ligand binding. The cis configuration may be more abundant than was previously thought (46), and in many cases, a region with a cis configuration is involved in protein function (47). Considering the equilibrium model of FD1 (Fig. 5B), we predict that the active form has a cis peptide configuration at the Asp-282–Cys-283 peptide bond, which would be required for ligand binding. Since the cationic side chain of His-284 is able to interact with the anionic side chain of Asp-282 (Fig. 3B), the charged state of His-284 might reflect cis-trans isomerization at the Asp-282–Cys-283 bond.

Arrangement of the Sugar-binding Sites on the Trimeric FD1—
The predicted GlcNAc-binding sites on the trimeric FD1 form the vertices of an equilateral triangle, with each side ∼50 Å in length (Fig. 6). Interestingly, the sugar-binding sites of collectins (10–12, 48) also form vertices of a similarly sized triangle on each trimeric CRD, although the FD1 and collectin structures are quite different (Fig. 6). Sheriff et al. (10) suggested that the MBL trimers are the basic recognition unit for branched oligosaccharides on microorganisms. The common spatial arrangement of the ligand-binding sites on each of the proteins suggests that such an arrangement is universally required to recognize pathogen-associated molecular patterns (49) and that a triangle of this size is a basic trivalent recognition unit for non-active conformations. The multimer is therefore ready to bind ligands at any time. However, because the affinity of each FBG domain is weak, the amount of time during which the multimer associates with a surface bearing few ligands (such as the surfaces of host cells) will be short. On the other hand, when the multimer encounters a surface with high ligand density (such as the surfaces of pathogens), most of FBG domains of the multimer will be in contact with ligands. Since some of the FBG domains are active, the association time of the multimer with the surface would drastically increase, resulting in high avidity.

Thus, the behavior of an ensemble average of the recognition molecules is an important property for permitting non-self-recognition, whereas the weak affinity of each molecule is an essential feature for preventing self-recognition. Interestingly, MBL shows weak affinity for ligands with dissociation constants in the millimolar range (51, 52). The importance of the multivalent form for the discrimination was also described in the case of TL5A (34). Therefore, these findings indicate that the innate immunity system uses multimeric recognition molecules with a common spatial array of ligand-binding sites for designing drugs selective against targets such as general pathogens and aberrantly glycosylated cancer cells.

Speculation of the Discrimination Mechanism between Self and Non-self by M-ficolin—
The equilibrium constants between the active and non-active forms (κ1 = 0.4 and κ2 = 1.9, Fig. 5B) suggest that the forward and backward exchange rate constants are similar, and therefore, FD1 alone displays weak affinity for the ligand. On the other hand, the M-ficolin multimer contains a trimeric FD1 assembly with active and non-active FBG domains coexisting simultaneously, and each FBG domain always exchanges between the active and non-active conformations. The multimer is therefore ready to bind ligands at any time. However, because the affinity of each FBG domain is weak, the amount of time during which the multimer associates with a surface bearing few ligands (such as the surfaces of host cells) will be short. On the other hand, when the multimer encounters a surface with high ligand density (such as the surfaces of pathogens), most of FBG domains of the multimer will be in contact with ligands. Since some of the FBG domains are active, the association time of the multimer with the surface would drastically increase, resulting in high avidity.

Thus, the behavior of an ensemble average of the recognition molecules is an important property for permitting non-self-recognition, whereas the weak affinity of each molecule is an essential feature for preventing self-recognition. Interestingly, MBL shows weak affinity for ligands with dissociation constants in the millimolar range (51, 52). The importance of the multivalent form for the discrimination was also described in the case of TL5A (34). Therefore, these findings indicate that the innate immunity system uses multimeric recognition molecules with a common spatial array of ligand-binding sites for designing drugs selective against targets such as general pathogens and aberrantly glycosylated cancer cells.
