Journal of
Synchrotron Radiation

ISSN 0909-0495

Received 28 June 2007
Accepted 30 October 2007

Trimeric structure and conformational equilibrium of M-ficolin fibrinogen-like domain

Michikazu Tanio,[a] Shin Kondo,[b] Shigetoshi Sugio[b]* and Toshiyuki Kohno[a]*

[a]Mitsubishi Kagaku Institute of Life Sciences (MITILS), 11 Minamiooya, Machida, Tokyo 194-8511, Japan, and [b]Innovation Center Yokohama, Mitsubishi Chemical Corporation, 1000 Kamoshida-cho, Aoba, Yokohama, Kanagawa, Japan. E-mail: shigetoshi@mw.m-kagaku.co.jp, tkohno@mitils.jp

Ficolins are pathogen-recognition molecules in innate immune systems. The crystal structure of the human M-ficolin recognition domain (FD1) has been determined at 1.9 Å resolution, and compared with that of the human fibrinogen γ fragment, tachylectin-5A, L-ficolin and H-ficolin. The overall structure of FD1 is similar to that of the other proteins, although the peptide bond between Asp282 and Cys283, which is in a predicted ligand-binding site, is a normal *trans* bond, unlike the cases of the other proteins. Analysis of the pH-dependent ligand-binding activity of FD1 in solution suggested that a conformational equilibrium between active and non-active forms in the ligand-binding region, involving *cis-trans* isomerization of the Asp282—Cys283 peptide bond, contributes to the discrimination between self and non-self, and that the pK_a values of His284 are 6.1 and 6.3 in the active and non-active forms, respectively.

Keywords: ficolin; innate immunity; conformational equilibrium.

1. Introduction

Innate immune systems play a crucial role as the first line of defense against pathogens, and are present in all multicellular organisms, including humans (Fujita *et al.*, 2004). Ficolins, which are composed of a collagen-like domain at the N-terminus and a fibrinogen-like domain (FBG) at the C-terminus, are one of the most important groups of pattern-recognition molecules in innate immune systems. Ficolins form trimer-based multimers that are mutually linked by intermolecular disulfide bonds at the N-terminal domain (Ohashi & Erickson, 2004). Three human ficolins, M-ficolin in cells and L-ficolin and H-ficolin in serum, have been characterized. The amino acid sequence homologies between M-ficolin and L-ficolin, and between H-ficolin and either L-ficolin or M-ficolin, are 80 and 48%, respectively (Endo *et al.*, 1996; Sugimoto *et al.*, 1998). These ficolins are associated with the mannose binding lectin-associated serine proteases, and the complexes activate the lectin-complement pathway (Fujita *et al.*, 2004; Liu *et al.*, 2005; Frederiksen *et al.*, 2005). The FBGs of ficolins bind to sugars, such as *N*-acetyl-D-glucosamine (GlcNAc) on the pathogen surface (Le *et al.*, 1998; Teh *et al.*, 2000). Interestingly, although the sugars recognized by ficolins also exist on the surface of the host cell, the proteins can discriminate between the host cell (self) and pathogens (non-self). To investigate the detailed mechanism of discrimination between self and non-self by ficolins, we have determined the crystal structure of the human M-ficolin FBG domain (FD1) at 1.9 Å resolution, using synchrotron radiation at beamline BL24XU at the SPring-8 facility in Japan (PDB code 2d39; Tanio *et al.*, 2006, 2007). The crystal structure of FD1, together with its pH-dependent ligand binding activity, provides insight into the discrimination mechanism by ficolins.

2. Trimeric structure of FD1

Diffraction-quality FD1 crystals were obtained at 293 K by mixing 0.5 µl protein solution (8 mg ml^{-1} FD1, 8 m*M* Tris-HCl pH 8.0, 80 m*M* NaCl, 4 m*M* $CaCl_2$ and 20 m*M* GlcNAc) and 0.5 µl reservoir solution [100 m*M* MES pH 5.6, 320 m*M* Li_2SO_4 and 17%(*w*/*v*) PEG 4000] (Tanio *et al.*, 2006). The crystal structure of FD1 was solved by molecular replacement using the crystal structure of tachylectin-5A (PDB code 1jc9; Kairies *et al.*, 2001). The structure contains one trimeric form of FD1 in the asymmetric unit (Tanio *et al.*, 2007). The overall structure of the FD1 monomer is similar to that of tachylectin-5A (Kairies *et al.*, 2001) and the human fibrinogen γ fragment (Pratt *et al.*, 1997; Yee *et al.*, 1997), and consists of three domains (A, B and P; Fig. 1), as in the other two proteins. The amino acid sequence of FD1 is 49% and 45% identical to that of the FBG domain of tachylectin-5A and the human fibrinogen γ fragment, respectively. The crystal structure revealed that the hydrophobic side-chains of Phe127 and Leu128 are important for trimer formation by FD1 (Tanio *et al.*, 2007). Indeed, the replacement of Phe127 with Thr in FD1 drastically decreased the monomer–monomer interaction in solution, as judged by a dynamic light-scattering experiment (data not shown). In spite of the addition of GlcNAc to the protein solution for crystallization, no obvious electron density for the ligand was observed, probably because FD1 at pH < 6 shows low GlcNAc-binding activity (Fig. 2*a*). The ligand-binding study of FD1 in solution revealed that the ligand-binding site is located near the Ca^{2+} binding site, the Cys270—Cys283 disulfide bond, and an unidentified group with a pK_a of 6.2 (Tanio *et al.*, 2007). The group with a pK_a of 6.2 was tentatively assigned to His284. The predicted GlcNAc-binding sites on the trimeric FD1 form the vertices of an equilateral triangle, with

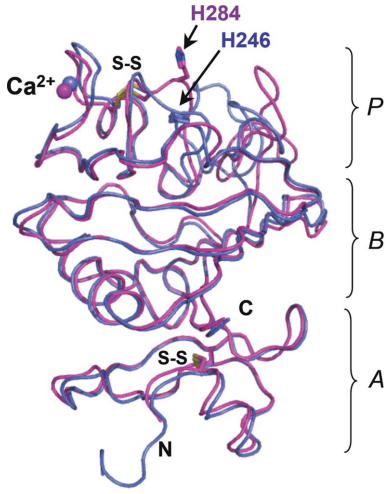


Figure 1
Comparison of the main chain traces between FD1 (magenta) and the FBG domain of L-ficolin (Garlatti *et al.*, 2007). Each Ca^{2+} ion is shown as a sphere. S–S indicates the conserved disulfide bonds in the two proteins. The conserved histidine residues (H284 in FD1 and H246 in L-ficolin) are shown as sticks. The root-mean-square deviation between the two proteins is 0.73 Å for 190 Cα positions.

each side approximately 50 Å in length. Interestingly, the sugar-binding sites of collectins also form the vertices of a similar-sized triangle on each trimeric carbohydrate recognition domain, although the FD1 and collectin structures are quite different. 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, and that a triangle of this size is a basic trivalent recognition unit for general pathogens.

3. Conformational equilibrium of the ligand-binding site of FD1

We previously developed a model describing the conformational equilibrium at the P domain between the active (A) and non-active (N) states at any pH, from an analysis of the pH-dependent GlcNAc-binding activity of FD1 in solution (Tanio *et al.*, 2007). In this model the equilibrium depends on the state of a group with a pK_a of 6.2, which is probably the His284 side-chain. To obtain further information about the group in terms of the pH dependency, we propose the following model [Fig. 2(*a*), inset scheme], which is equivalent to our previous model (Tanio *et al.*, 2007), but distinguishes the pK_a value of the group between the active and non-active states. The relative fraction of the active form of FD1 in this model is described by the following equation,

$$f_{\rm active}({\rm pH}) = \frac{1 + 10^{n({\rm pH} - {\rm p}K_{\rm A})}}{1 + K_2 + (1 + K_1)10^{n({\rm pH} - {\rm p}K_{\rm A})}}, \quad (1)$$

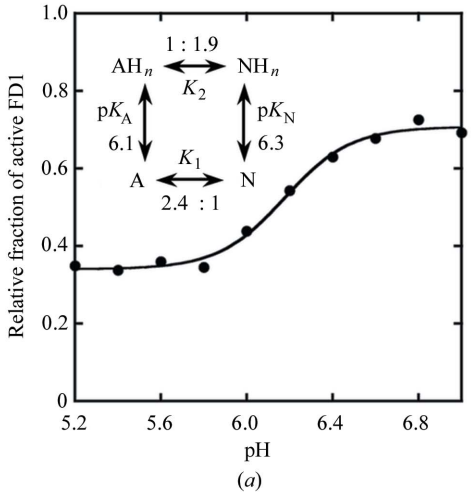
$$= 1 - \frac{1 + 10^{n({\rm pH} - {\rm p}K_{\rm N})}}{1 + K_2^{-1} + (1 + K_1^{-1})10^{n({\rm pH} - {\rm p}K_{\rm N})}}, \quad (2)$$

where K_1 = [N]/[A] and K_2 = [NH_n]/[AH_n] are the equilibrium constants between the active and non-active forms in the deprotonated and protonated states, respectively, and n is the number of protons involved in the transition from the protonated (AH_n and NH_n) to deprotonated (A and N) forms. p$K_{\rm A}$ and p$K_{\rm N}$ are the pK_a of the group related to the pH dependency in the active and non-active states, respectively. The curve fitting by equations (1) and (2) revealed that K_1, K_2, p$K_{\rm A}$, p$K_{\rm N}$ and n are 0.4, 1.9, 6.1, 6.3 and 2.8, respectively [Fig. 2(*a*), solid line, coefficient of determination r^2 = 0.99]. These values, except for p$K_{\rm A}$ and p$K_{\rm N}$, are identical to those of our previous model (Tanio *et al.*, 2007). Although the p$K_{\rm A}$ value (6.1) is slightly lower than the p$K_{\rm N}$ value (6.3), this difference is significant. The relationship between p$K_{\rm A}$ and p$K_{\rm N}$ is described by the following equation,

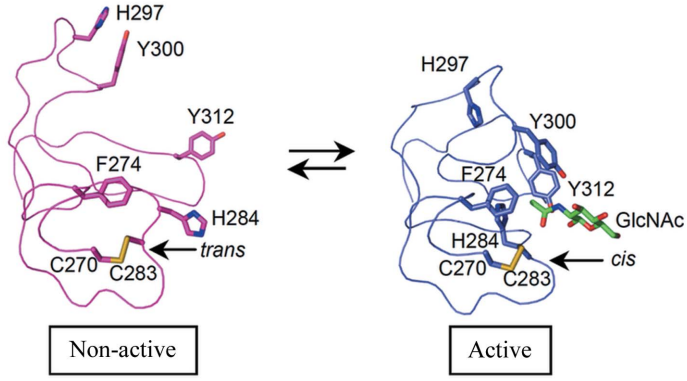
$$K_{\rm N}^n = \frac{[{\rm N}][{\rm H}]^n}{[{\rm NH}_n]} = \frac{K_1[{\rm A}][{\rm H}]^n}{K_2[{\rm AH}_n]} = \frac{K_1}{K_2}K_{\rm A}^n$$

$$\Leftrightarrow {\rm p}K_{\rm A} - {\rm p}K_{\rm N} = \frac{1}{n}\log\frac{K_1}{K_2}. \quad (3)$$

Therefore, if $K_1 < K_2$, then p$K_{\rm A}$ < p$K_{\rm N}$ in this model. Since the increase in the GlcNAc-binding activity of FD1 associated with increased pH results in $K_1 < K_2$ (Fig. 2*a*), the situation of p$K_{\rm A}$ < p$K_{\rm N}$ is corollary. This means that the group related to the pH-dependent



(*a*)



(*b*)

Figure 2
(*a*) pH-dependent GlcNAc binding activity of FD1 at 277 K. The relative fractions, calculated as the eluted FD1 per total amount of protein applied to the GlcNAc column, are plotted against the pH (Tanio *et al.*, 2007). The solid line is the regression curve, as calculated from equations (1) or (2) (see text). The calculation revealed that K_1, K_2, p$K_{\rm A}$, p$K_{\rm N}$ and n are 0.41 ± 0.03, 1.94 ± 0.10, 6.06 ± 0.04, 6.30 ± 0.04 and 2.81 ± 0.46, respectively. The inset is the scheme of the equilibrium model proposed here. (*b*) Proposed conformational equilibrium between the active (slate) and non-active (magenta) forms of the P domain of FD1. In the active form the peptide bond between Asp282 and Cys283 is in the *cis* configuration. The putative model of the active conformation complexed with GlcNAc was built based on the crystal structure of tachylectin-5A complexed with GlcNAc (Kairies *et al.*, 2001).

activity of FD1 exists in different environments between the active and non-active states; that is, the conformational difference exists in the ligand-binding region between the two states. If the group is a basic residue such as histidine, then its side-chain in the active state should be located within a more hydrophobic environment than that in the non-active state.

A comparison of the structure of FD1 with the ligand-binding structures of tachylectin-5A (Kairies *et al.*, 2001), the human fibrinogen γ fragment (Pratt *et al.*, 1997) and two other ficolins (L and H) (Garlatti *et al.*, 2007) shows that the predicted ligand-binding site of FD1 is in good agreement with the ligand-binding sites of the other four proteins. However, the conformation of the ligand-binding site of FD1 is different from those of the other four proteins (Fig. 1, Tanio *et al.*, 2007). Interestingly, the side-chains of the conserved histidine residues on the ligand-binding sites in tachylectin-5A (His220), the human fibrinogen γ fragment (His340) and L-ficolin (His246), which correspond to His284 of FD1, are apparently buried in the surface of each protein, while that of His284 of FD1 is exposed to the solvent from the protein surface (Fig. 1). Such different environments around the histidine residues would cause the different pK_a values of the residues between FD1 and the other three proteins, and therefore His284 of FD1 is a strong candidate for the group with the pK_A of 6.1 and the pK_N of 6.3. In this connection, the peptide bond between Asp282 and Cys283, neighboring His284, is a normal *trans* bond in FD1, while the corresponding regions of the other four proteins are an unusual *cis* peptide bond which is necessary for ligand binding (Pratt *et al.*, 1997; Kairies *et al.*, 2001; Garlatti *et al.*, 2007). From these findings we propose a model of conformational equilibrium at the P domain of FD1, involving *cis-trans* isomerization of the Asp282—Cys283 peptide bond (Fig. 2*b*).

Although the conformational equilibrium of FD1 would cause weak affinity for the ligand, it is crucial to prevent self-recognition. The low ligand density on the surfaces of self-cells also contributes to the prevention of self-recognition. On the other hand, the M-ficolin multimer, which is an assemblage of the trivalent recognition unit (Tanio *et al.*, 2007), has an advantage in recognizing surfaces with high ligand density composed of pathogen-associated molecular patterns. Therefore, the multimerization based on the trivalent recognition unit is an important property for recognizing non-self, whereas the weak affinity of each molecule is an essential feature for preventing self-recognition. This strategy of the innate immunity systems provides a structural basis for understanding the mechanism of discrimination between self and non-self, and for designing selective drugs for targets, such as general pathogens and aberrantly glycosylated cancer cells.

This work was supported in part by a grant from the National Project on Protein Structural and Function Analyses.

References

Endo, Y., Sato, Y., Matsushita, M. & Fujita, T. (1996). *Genomics*, **36**, 515–521.
Frederiksen, P. D., Thiel, S., Larsen, C. B. & Jensenius, J. C. (2005). *Scand. J. Immunol.* **62**, 462–473.
Fujita, T., Matsushita, M. & Endo, Y. (2004). *Immunol. Rev.* **198**, 185–202.
Garlatti, V., Belloy, N., Martin, L., Lacroix, M., Matsushita, M., Endo, Y., Fujita, T., Fontecilla-Camps, J. C., Arlaud, G. J., Thielens, N. M. & Gaboriaud, C. (2007). *EMBO J.* **26**, 623–633.
Kairies, N., Beisel, H. G., Fuentes-Prior, P., Tsuda, R., Muta, T., Iwanaga, S., Bode, W., Huber, R. & Kawabata, S. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 13519–13524.
Le, Y., Lee, S. H., Kon, O. L. & Lu, J. (1998). *FEBS Lett.* **425**, 367–370.
Liu, Y., Endo, Y., Iwaki, D., Nakata, M., Matsushita, M., Wada, I., Inoue, K., Munakata, M. & Fujita, T. (2005). *J. Immunol.* **175**, 3150–3156.
Ohashi, T. & Erickson, H. P. (2004). *J. Biol. Chem.* **279**, 6534–6539.
Pratt, K. P., Cote, H. C., Chung, D. W., Stenkamp, R. E. & Davie, E. W. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 7176–7181.
Sugimoto, R., Yae, Y., Akaiwa, M., Kitajima, S., Shibata, Y., Sato, H., Hirata, J., Okochi, K., Izuhara, K. & Hamasaki, N. (1998). *J. Biol. Chem.* **273**, 20721–20727.
Tanio, M., Kondo, S., Sugio, S. & Kohno, T. (2006). *Acta Cryst.* F**62**, 652–655.
Tanio, M., Kondo, S., Sugio, S. & Kohno, T. (2007). *J. Biol. Chem.* **282**, 3889–3895.
Teh, C., Le, Y., Lee, S. H. & Lu, J. (2000). *Immunology*, **101**, 225–232.
Yee, V. C., Pratt, K. P., Cote, H. C., Trong, I. L., Chung, D. W., Davie, E. W., Stenkamp, R. E. & Teller, D. C. (1997). *Structure*, **5**, 125–138.