PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Crystal structure of a novel dimer form of FlgD from *P. aeruginosa* PAO1

Hua Zhou,[1,2†] Miao Luo,[1,2†] Xuefei Cai,[1] Jian Tang,[1,2] Siqiang Niu,[1,2] Wenlu Zhang,[1] Yuan Hu,[1] Yibing Yin,[2] Ailong Huang,[1] Dacheng Wang,[3] and Deqiang Wang[1,2,3*]

[1] Key Laboratory of Molecular Biology on Infectious Disease, Chongqing Medical University, YiXueYuanlu-1, Chongqing 400016, China

[2] Key Laboratory of Diagnostic Medicine, Department of Laboratory Medicine, Chongqing Medical University, YiXueYuanlu-1, Chongqing 400016, China

[3] National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, China

Key words: FlgD; hook cap; crystal structure; homodimer; interface.

INTRODUCTION

The bacterial flagellum, the organ of motility for many species of bacteria, is a complex nanostructure protein machine which requires about 50 proteins for regulation and assembly.[1,2] The flagellum consists of three major sections: the filament, the hook, and the basal body.[3] The filament, serving as a propeller, is a helical assembly composed of thousands of copies of the flagellin (FliC) protein. The hook, connecting the filament to the basal body, is a short, highly curved tube consisting of multiple copies of the hook protein FlgE. The appropriate length and flexibility of the hook appear to be important for its functioning as a protein "universal joint."[4]

The flagellar hook protein FlgE has been studied by electron microscopy as well as by biochemical and physicochemical methods.[5–7] The overall length of the hook is well regulated and is typically found to be 55 ± 6 nm long. However, mutations in FlgD, which functions in protein export and flagellar assembly, result in an increased length of the flagellar hook.[8] FlgD serves as the hook scaffolding protein and is also considered to be the hook-capping protein and the basal body rod-modification protein. With the aid of FliK, FlgD regulates the assembly of the hook cap structure by controlling hook monomer polymerization and preventing the leakage of hook monomers into the medium as well playing an important role in ensuring correct hook length.[2,8] Studies of flagellar biosynthesis in *Escherichia coli* and *Salmonella enterica* suggest that the hook cap is a transient intermediate structure in hook biosynthesis, and FlgD is not found in mature flagella.[2] Because FlgD is required for hook regulation and assembly and therefore plays a critical role in bacterial pathogenesis, understanding its regulation, and activities has become an important area of research.

To date, only the crystal structure of the C-terminal domain (residues 87–220) of FlgD from *Xanthomonas campestris* (XcFlgD; PDB code 3c12) has been determined. This structure comprises a novel hybrid fold consisting of a tudor-like domain and a fibronectin type III domain. XcFlgD forms three types of dimers in the crystal.[9] Efforts in our laboratory to study the structure of FlgD from *P. aeruginosa* PAO1 (PaFlgD) have yielded an alterative arrangement of the FlgD homodimer that differs substantially from that previous described. The observation that the PaFlgD homodimer can exist in alterative configurations may have ramifications for the hook cap assembly.

Additional Supporting Information may be found in the online version of this article.

Chinese National Natural Science Foundation; Grant numbers: 30600101, 30770481, 30970563; Grant sponsors: Natural Science Foundation of Chongqing; Grant numbers: 2006BB5275, 2009BB5413

[†]Hua Zhou and Miao Luo contributed equally to this work.

Deqiang Wang, Key Laboratory of Molecular Biology on Infectious Disease, Chongqing Medical University, YiXueYuanlu-1, Chongqing 400016, People's Republic of China. E-mail: wangd@ustc.edu

Received 18 February 2011; Revised 29 March 2011; Accepted 31 March 2011
Published online 20 April 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.23058

METHODS

Cloning, expression, and purification

Purification and crystallization of the FlgD from *P. aeruginosa* PAO1 (PaFlgD) were carried out as described earlier.[10] The *PaFlgD* gene fragments corresponding to residues 1–235 and 93–235 (PaFlgD and PaFlgD-C, respectively) were amplified using *P. aeruginosa* PAO1 genomic DNA as the template. The two PCR products were cloned into the *Nde*I/*Xho*I sites of pET28a to give pET28FlgD-His6 and pET28FlgD-C-His6, respectively. Protein was expressed in *E coli* BL21 (DE3) grown in LB at 310 K. When the optical absorbance of the culture (OD600) reached 0.4-0.6, IPTG was added to 0.5 m*M*, and the temperature was lowered to 293 K with expression allowed to occur overnight.

Cells were then harvested, resuspended in ice-cold buffer A (20 mM Tris–HCl at pH 8.0 and 300 m*M* NaCl), and then lysed by sonication. The cell debris was removed by centrifugation and the resulting soluble fraction applied to Ni^{2+} NTA affinity resin (Qiagen). After washing (40 mM imidazole), the protein was eluted from the resin with 200 m*M* imidazole (in buffer A). The eluent was then applied to DEAE Sepharose Fast Flow column (Amersham Bioscience) equilibrated with 20 m*M* Tris–HCl pH 8.0. Recombinant proteins were eluted with a linear gradient of 0–500 m*M* sodium chloride in 20 m*M* Tris–HCl pH 8.0. PaFlgD eluted as a single monomeric peak and was concentrated to 20 mg mL^{-1}. Selenomethionine-substituted FlgD protein (SeMet-FlgD) was produced in *E. coli* strain B834(DE3) in a inorganic media supplemented with L-selenomethionine and purified using the above described method.

Oligomeric state determination

Size exclusion chromatography was performed with a 16/60 Superdex 200 column (GE Healthcare) equilibrated to and eluted with 50 m*M* Tris-HCl pH 8.0 and 200 m*M* NaCl. Protein molecular weight standards used were: albumin bovine V (66.2 kDa), chicken egg albumin (44.2 kDa), chymotrypsinogen A (24.5 kDa), and lysozyme (14.4 kDa); the void volume was determined with Blue Dextran (GE Healthcare). The peak elution volumes were used for the calculation of the standard curve equation Log MW = −0.01712Ve + 6.05667 (R^2 value for the curve-fit = 0.93352). PaFlgD and PaFlgD-C were diluted to 1 mg mL^{-1} in the corresponding buffer and then loaded onto the column. The estimated molecular weights of PaFlgD and PaFlgD-C were calculated from the appropriate standard curve using the protein's peak elution volume. Under these conditions, PaFlgD and PaFlgD-C were eluted as single peaks at volumes 77.6 mL and 88.8 mL, corresponding to estimated molecular weights of 53.4 kDa and 33.8 kDa, respectively.

Crystallization, data collection, and structure determination

The initial crystallization conditions of both the native and SeMet-substituted PaFlgD (residues 1–235) were determined using commercial protein crystallization screens (Hampton Research). Ultimately, suitable crystals for diffraction experiments were grown within 3 days at 293 K using vapor diffusion and a reservoir solution containing 18% (w/v) polyethylene glycol 6000, 0.1*M* MES pH 6.6.[10] For data collection, crystals were equilibrated against reservoir solution adjusted to 20% (w/v) glycerol and then immediately placed in a 100 K nitrogen gas stream. A multiple wavelength anomalous dispersion (MAD) data set was collected from a single crystal of SeMet-derivative PaFlgD protein at 100 K on beamline 3W1A of Beijing Synchrotron Radiation Facility (BSRF). Data were collected at two wavelengths (λ_{peak} = 0.9790 Å and λ_{edge} = 0.9793 Å), and the data were processed and scaled to 2.35 Å using the HKL2000 software package.[11]

Data collection statistics are presented in Table I. The SeMet-substituted PaFlgD belongs to the space group I222, with unit-cell parameters a = 114.83, b = 118.38, and c = 118.30 Å. Using 2.6Å MAD data and the program SOLVE, 12 selenium atoms in the asymmetric unit were located, and the initial phases were calculated.[12] Initial phases were then improved by density modification and ~50% of the residues in the model were automatically built using the program RESOLVE.[13] Further

Table I
X-ray Data Collection and Refinement Statistics for PaFlgD

Space group	I222	
Cell dimensions (Å)		
a	114.828	
b	118.379	
c	118.301	
	Peak	Edge
Wavelength (Å)	0.9790	0.9793
Resolution (Å)[a]	50–2.35 (2.43–2.35)	50–2.35 (2.43–2.35)
Total reflections	828,120	824,966
Unique reflections	65,067	64,907
Completeness (%)[a]	99.9 (99.2)	99.9 (99.0)
Multiplicity	7.1 (6.4)	7.2 (6.4)
I/Sigma	5.9	5.9
Rmerge (%)[a,b]	6.9 (38.9)	6.4 (38.3)
Refinement		
Resolution (Å)	50–2.35	
R_{free}[c]	26.8%	
R_{work}[c]	21.1%	
Bond length rmsd (Å)[d]	0.0222	
Bond angle rmsd (°)[d]	1.996	

[a]Values in parentheses are for the highest-resolution shells
[b]$R_{merge} = \sum(I - \langle I\rangle)/\sum I$. where I is the observed intensity and $\langle I\rangle$ is the statistically weighted average intensity of multiple symmetry related observation.
[c]R_{factor}: $R = \sum||F_{calc}| - |F_{obs}|/\sum|F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure factors, respectively. R_{free}, calculated the same as R_{factor}, but from a test set containing 5% of data excluded from the refinement calculation.
[d]rmsd: root-mean-square deviation from ideal geometry.

model building was performed manually using the program WinCoot, and the refinement was performed with REFMAC5 of the CCP4 suite.[14,15] The stereochemical quality of the final model of PaFlgD was ascertained using PROCHECK.[16] Data collection and structural refinement statistics are listed in Table I. Ribbon cartoons and surface representations were generated using PyMOL.[17] Because we failed to see electron density for the N-terminal residues (residues 1–101), we were only able to determine the atomic coordinates and structural factors for the C-terminal region of PaFlgD, which have been deposited in the PDB database with accession code 3OSV.

Dimer interface analysis

Interface residues were defined as surface residues that lost relative surface accessible areas (RSAs) on complex formation. Surface residues were defined as those for which RSA $\geq$ 5%.[18] The relative accessible surface area of each residue in a protein in the data set was calculated using NACCESS, which implements the Lee and Richards algorithm, with a probe sphere of radius 1.4 Å.[19,20] Using these criteria, we obtained the interface residues and the surface area of FlgD homodimer buried upon dimer formation.

RESULTS AND DISCUSSION

Structure solution and model quality

Full length PaFlgD was initially crystallized by hanging drop vapor diffusion. Crystals that diffracted to a resolution 2.35 Å were obtained after 3 days. However, when electron density was calculated using this data, only the C-terminal residues (102–236) could be clearly defined. Additionally, SDS-PAGE gel experiments confirmed that the crystal was a fragment protein with molecular weight 15 kDa and about 10 kDa segments had been degraded (unpublished data). Interestingly, the N-terminus of FlgD from *Xanthomonas campestris* (XcFlgD), corresponding to residues 84–220, was also disordered in that crystal structure (Kuo *et al.*[9]). The N-terminal region of FlgD, which is much more flexible and more susceptible to proteolytic cleavage than the C-terminal region, might serve as a secretory signal to help the flagellar protein arrive at target location. This has been shown to be the case for other flagellar proteins.[9,21]

The X-ray structure of PaFlgD has been determined by MAD phasing. After refinement to 2.35 Å resolution against the native data, the final model exhibits an R_{work} of 21.1% and an R_{free} of 26.8%. The r.m.s.d. from ideal values of bond lengths and bond angles were 0.009 Å and 1.774°, respectively [Fig. 1(A)]. Of all the residues in the final refined model, 87.3% have main chain torsion angles in the most favored regions and only one residue, Ile-134 in chain B, which is well defined in the experimental electron density, has torsion angles in the disallowed region. There are four molecules in the asymmetric unit of the crystal. The model comprises residues 103–212 and 216–236 in chain A, residues 102–210 and 215–236 in chain B, residues 102–211 and 215–236 in chain C, and residues 103–211 and 215–235 in chain D. Residues that are not included in the model are disordered. In addition to the four protein chains, the model contains 191 water and 12 glycerol molecules. Data collection and refinement statistics are summarized in Table I.

To assess the degree of structural similarity, we compared the structure of PaFlgD with the structure of its homolog from *Xanthomonas campestris*.[9] PaFlgD superimposes onto the XcFlgD structure with a calculated r.m.s.d. of 2.6 Å for 129 aligned Cα atoms. The two proteins share 35% sequence identity in the superimposed region (Supporting Information Fig. S1). In detail, the conformation is almost identical for domain I (tudor domain) and domain II (Fn III domain) between the two hook cap proteins, except for some loop regions likely altered by amino acids insertions and deletions. Within each of these domains, the most noticeable conformational differences are observed in the tudor domain. XcFlgD contains an additional β-strand at the C-terminus, whereas corresponding residues at the C-terminus of PaFlgD are flexible and form a loop. Additionally, the N-terminal region of XcFlgD (residues 87–95) forms a 3_{10} helix, which is not found in PaFlgD [Fig. 1(B)].

The electrostatic potential surface of PaFlgD displays prominent asymmetry. As shown in Figure 1(C), a negatively charged groove stretches along the top loop region of Fn-III domain. The five aspartic acid residues in the groove (Asp120, Asp123, Asp144, Asp145, and Asp171) give rise to the electro-negative potential of that surface. Interestingly, a similarly charged groove was observed in the structure of XcFlgD.[9] However, a corresponding negatively charged region is not observed in the Fn III domain in human fibronectin proteins,[22,23] implying that this feature of FlgD proteins may be functionally important as a binding site for proteins and/or small molecules during flagellar hook biosynthesis.

The novel dimer form of PaFlgD

FlgD-linked flagellar proteins such as FlgE and FliC have been reported to form either linear or helical filaments. It seems possible that serving as a hook cap that is known to associate with homo-oligomers, FlgD might also self-assemble into a stable and functional quaternary structure necessary for the regulation of the flagellar hook. To characterize the oligomeric state of recombinant PaFlgD in solution, size exclusion chromatography experiments were performed using both PaFlgD and PaFlgD-C proteins. In these experiments, elution peaks were observed at volumes

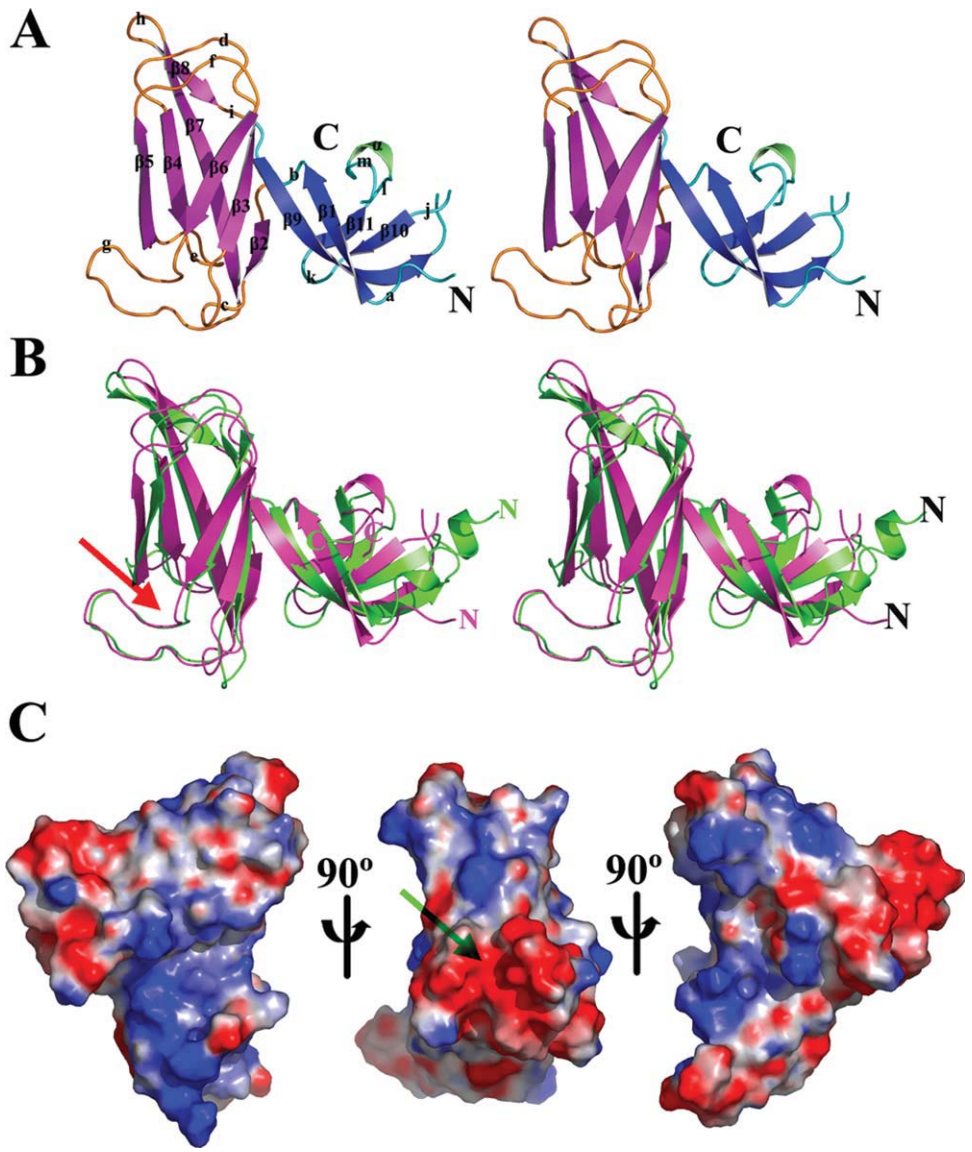

Figure 1

The structure of PaFlgD. **A**: A stereo view of PaFlgD. The secondary structure elements are numbered as in Supporting Information Figure S1. **B**: Superposition of PaFlgD (cyan) and XCFlgD (gray; PDB code 3C12). The N- and C-termini of PaFlgD and XcFlgD are shown in magenta and green, respectively. The red arrow represents the negative charged groove as shown in (**C**). **C**: Electrostatic potential surface of PaFlgD. Saturating red indicates $\Phi < -10$ kiloteslas/e, and saturating blue indicates $\Phi > -10$ kiloteslas/e, $T = 293$ K. The green arrow points to the negative-charged groove.

of 77.6 mL and 88.8 mL, corresponding to estimated molecular weights of about 53.4 kDa and 33.8 kDa for PaFlgD and PaFlgD-C, respectively. Because PaFlgD and PaFlgD-C have calculated molecular masses ~25.7 kDa and 16.5 kDa, respectively, it would appear that both PaFlgD and PaFlgD-C assemble into dimers in solution. These results strongly suggest that dimerization may be required for the proper functioning of the FlgD protein.

Kuo *et al.*[9] observed three types of crystal packing interfaces in XcFlgD crystals [Fig. 2(C)]. These XcFlgD packing interfaces involve the loss of 320Å^2, 320Å^2, and 260 Å^2 of solvent exposed surface areas, which are much smaller than that of typical for a tight dimer interface (~1000 Å^2), implying that the observed interfaces might not be the relevant interactions required for dimerization.[24] It is worthwhile to note that the XcFlgD crystallizes in a high ionic strength solution comprising 1.7*M* sodium acetate which may interfere with dimer formation mediated by predominantly polar/charged surfaces. On the other hand, a high ionic strength crystallization solution would tend to stabilize interactions that are predominantly hydrophobic in nature.

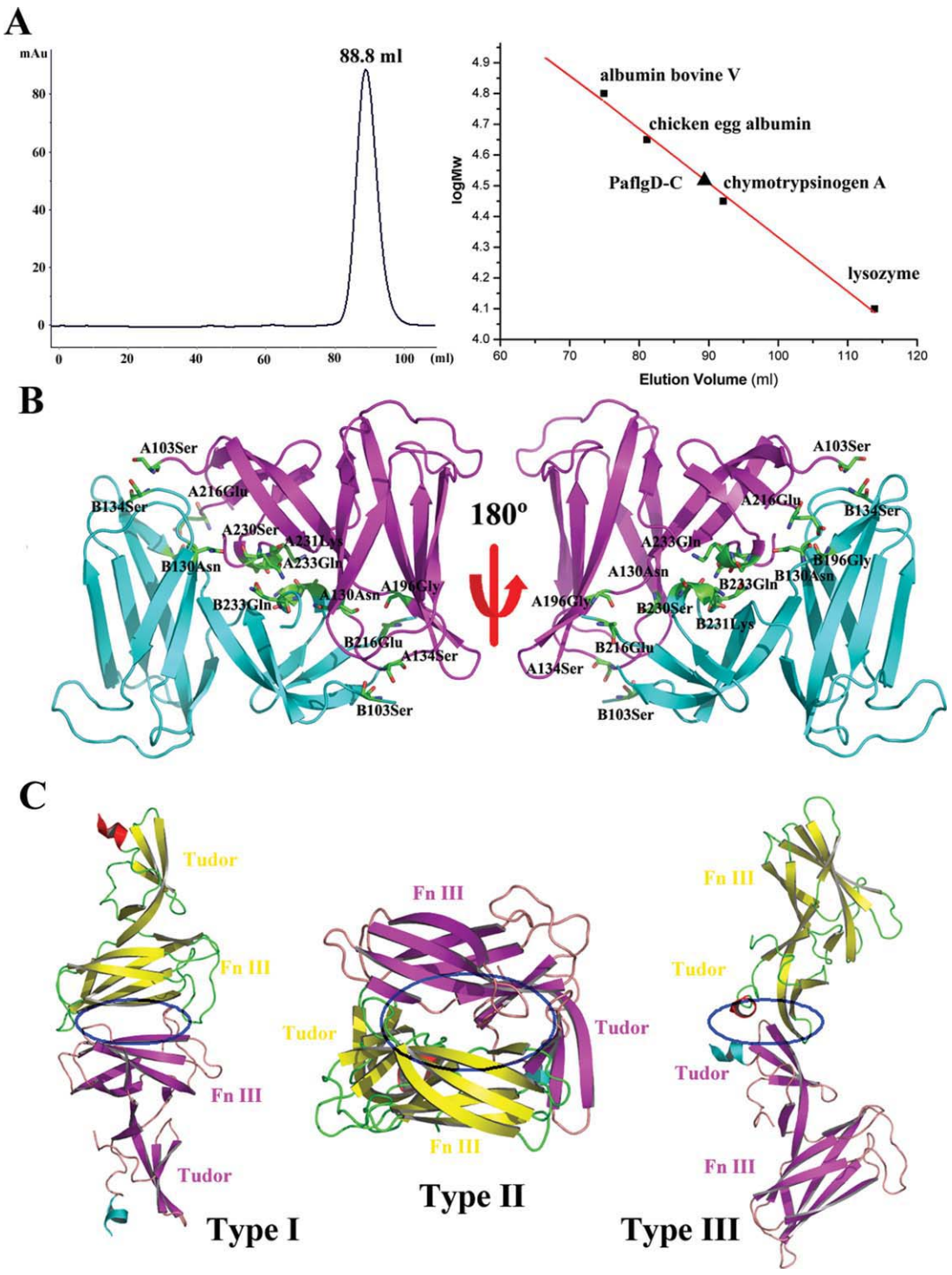

Figure 2

The Dimer of PaFlgD. **A**, PaFlgD (MW ~16.5 kDa) eluted as a single peak at a volume of 88.8 mL, which corresponds to a molecular weight of 33.8 kDa on the calibrated column. Standard protein samples (black data points) used for calibration include albumin bovine V (66.2 kDa), chicken egg albumin (44.2 kDa), chymotrypsinogen A (24.5 kDa), and lysozyme (14.4 kDa). **B**, PaFlgD forms a two-fold symmetric dimer in the crystals (Two-fold axis is into the page; some residues participating in hydrogen bond at the dimer interface are shown in stick representation). **C**, The three types of interfaces between the XC1894 monomer structures in crystal. The dimer interfaces are blue circled for clearance. Type I and III interface is formed mainly from residues located in the Fn-III domain, Type II interface is formed from residues situated in the tudor domain only.[9] [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

PaFlgD crystal structure was determined using crystals grown in under relatively low ionic strength conditions, and the asymmetric unit of the crystals contains four monomers of the hook protein. By examining the crystal packing, we identified an extensive packing interaction that is consistent with a relevant dimer interaction (Fig. 2). Two monomers of PaFlgD pack symmetrically across an interface comprised of extensive hydrophobic and polar contacts, resulting in the burial of 2775 Å^2 of solvent accessible surface upon dimer formation (Note: The aforementioned negatively charged groove does not participate in this interaction). The extensive nature of this interface, which is predominantly hydrophobic (1865 Å^2), is consistent with that expected for a physiologically relevant protein-protein interface (1200–2000 Å^2; Janin *et al.*[24]). The dimer interface of PaFlgD is composed of about 30 (mostly equivalent) residues from each monomer, due to the twofold character of the dimer. To a large extent, the contact surfaces are formed by residues 103 to 105, 108, 111, 113, 130, 132 to 134, 160 to 162, 195 to 196, 198, 200, 208, 210, 216 to 217, and 229 to 235 from each monomer. Approximately half of the surface is composed of highly conserved residues (Supporting Information Fig. S1).

Polar interactions at the dimer interface are largely mediated by residues which are conserved residues or main-chain atoms situated in the loops a, j, m, and helix 1 from the tudor domain and the loops d and β-strand 8 from the Fn-III domain (Fig. 2), which formed a polar area upon 910 Å^2 for dimer formation. Additionally, hydrophobic conserved residues, including Leu105, Ile111, Val162, Val208, Leu210, Leu217, Leu229, Val232, and Ile235, formed a nonpolar area about 1865 Å^2 contributing to the homodimer assemble. Taken together, the structural and solution data suggest that the FlgD homodimer we see in the crystal structure is likely to be the biologically functional unit required for hook cap assembling.

Collectively, due to relatively low ionic strength conditions, this study provides novel dimeric form in crystal with 2775 Å^2 of surface area for dimer formation, suggesting the FlgD homodimer functions as the structural elementary unit. Further studies of higher-order FlgD self-association will be needed to elucidate the hook cap assembly.

ACKNOWLEDGMENTS

The authors acknowledge the X-ray facility of the Key Laboratory of Structural Biology of the Chinese Academy of Science and thank Professor Liwen Niu and Maikun Teng, School of Life Science, University of Science and Technology of China. They thank Drs Yuhui Dong and Peng Liu for assistance with MAD data collection on beamline 3W1A of the Beijing Synchrotron Radiation Facility (BSRF). They are also grateful to Dr. David Worthylake in Louisiana State University Health Science Center on homodimer interface analysis and article editing.

REFERENCES

1. Macnab RM. How bacteria assemble flagella. Annu Rev Microbiol 2003;57:77–100.
2. Pallen MJ, Penn CW, Chaudhuri RR. Bacterial flagellar diversity in the post-genomic era. Trends Microbiol 2005;13:143–149.
3. Soutourina OA, Bertin PN. Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol Rev 2003;27:505–523.
4. Aldridge P, Hughes KT. Regulation of flagellar assembly. Curr Opin Microbiol 2002;5:160–165.
5. Samatey FA, Matsunami H, Imada K, Nagashima S, Shaikh TR, Thomas DR, Chen JZ, Derosier DJ, Kitao A, Namba K. Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. Nature 2004;431:1062–1068.
6. Shaikh TR, Thomas DR, Chen JZ, Samatey FA, Matsunami H, Imada K, Namba K, Derosier DJ. A partial atomic structure for the flagellar hook of Salmonella typhimurium. Proc Natl Acad Sci USA 2005;102:1023–1028.
7. Fujii T, Kato T, Namba K. Specific arrangement of alpha-helical coiled coils in the core domain of the bacterial flagellar hook for the universal joint function. Structure 2009;17:1485–1493.
8. Ohnishi K, Ohto Y, Aizawa SI, Macnab RM, Iino T. FlgD is a scaffolding protein needed for flagellar hook assembly in Salmonella typhimurium. J Bacteriol 1994;176:2272–2281.
9. Kuo WT, Chin KH, Lo WT, Wang AHJ, Chou SH. Crystal structure of the C-terminal domain of a flagellar hook-capping protein from Xanthomonas campestris. J Mol Biol 2008;381:189–199.
10. Luo M, Niu S, Yin Y, Huang A, Wang D. Cloning, purification, crystallization and preliminary X-ray studies of flagellar hook scaffolding protein FlgD from Pseudomonas aeruginosa PAO1. Acta Crystallogr Sect F Struct Biol Cryst Commun 2009;65:795–797.
11. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation. Mode Enzymol 1997;276:307–326.
12. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55:849–861.
13. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr 2000;56:965–972.
14. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr 2004;60:2126–2132.
15. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr Sect D Biol Crystallogr 1997;53:240–255.
16. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
17. DeLano WL. The PyMOL user's manual. San Carlos, CA: Delano Scientific; 2004.
18. Valdar V, Thornton JM. Protein–protein interfaces: analysis of amino acid conservation in homodimers. Proteins 2001;42:108–124.
19. Hubbard SJ. NACCESS. Department of Biochemistry and Molecular Biology, University College, London. 1993.
20. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:55379–55400.
21. Namba K. Roles of partly unfolded conformations in macromolecular self-assembly. Genes Cells 2001;6:1–12.
22. Sharma A, Askari JA, Humphries MJ, Jones EY, Stuart DI. Crystal structure of a heparin- and integrin-binding segment of human fibronectin. EMBO J 1999;18:1468–1479.
23. Lundell A. Olin AI, Morgelin M, al-Karadaghi S, Aspberg A, Logan DT. Structural basis for interactions between tenascins and lectican C-type lectin domains: evidence for a crosslinking role for tenascins. Structure 2004;12:1495–1506.
24. Janin J, Rodier F, Chakrabarti P, Bahadur RP. Macromolecular recognition in the Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2007;63:1–8.