Solution Structure of the Fibronectin Type III Domain from Bacillus circulans WL-12 Chitinase A1*

Jun-Goo Jee†, Takahisa Ikegami†, Masayuki Hashimoto‡, Takeshi Kawabata†, Mitsunori Ikeguchi†, Takeshi Watanabe§, and Masahiro Shirakawa‡***

From the ¥Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, the ¥Department of Applied Biological Chemistry, Faculty of Agriculture, Niigata University, 8050 Ikarashi-2, Niigata 950-2189, the ¥Laboratory for Gene Product Informatics, National Institute of Genetics, 1-111 Yata, Mishima, Shizuoka 411, and the ¥Graduate School of Integrated Science, Yokohama City University, 1-7-29 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

Growing evidence suggests that horizontal gene transfer plays an integral role in the evolution of bacterial genomes. One of the debated examples of horizontal gene transfer from animal to prokaryote is the fibronectin type III domain (FnIII). Certain extracellular proteins of soil bacteria contain an unusual cluster of FnIIIDs, which show sequence similarity to those of animals and are likely to have been acquired horizontally from animals. Here we report the solution structure of the FnIIIId of chitinase A1 from Bacillus circulans WL-12. To the best of our knowledge, this is the first tertiary structure to be reported for an FnIIIId from a bacterial protein. The structure of the domain shows significant similarity to FnIIIIds from animal proteins. Sequence comparisons with FnIIIIds from other soil bacteria proteins show that the core-forming residues are highly conserved and, thus, are under strong evolutionary pressure. Striking similarities in the tertiary structures of bacterial FnIIIs and their mammalian counterparts may support the hypothesis that the evolution of the FnIIIId in bacterial carbohydrases occurred horizontally. The total lack of surface-exposed aromatic residues also suggests that the role of this FnIIIId is different from those of other bacterial β-sandwich domains, which function as carbohydrate-binding modules.

The fibronectin type III domain (FnIIIId) is one of the most common folds in modular proteins. It was initially characterized in fibronectin, and since then has been found in ~2% of all animal proteins. Although most of these are extracellular proteins, FnIIIIds are also found in membrane receptor proteins as well as in intracellular proteins (1). X-ray crystallography (2–8) and NMR spectroscopy (9–12) have been used to elucidate the structures of several animal FnIIIIds, all of which adopt Greek key β-sandwich folds with three and four strands, consisting of 80–100 amino acid residues in total. The functions of many FnIIIIds are still unclear. Each FnIIIId comprises domain-intrinsic and domain-specific regions (13). The former, made up of relatively conserved residues, are responsible for forming the FnIIIId scaffold, which comprises a hydrogen-bond network and a hydrophobic core. The scaffold is common to all FnIIIIds and endows the domain with its mechanical extensibility against tension and its high refolding speed (14, 15). By contrast, the domain-specific regions are formed by exposed residues that are not well conserved across the FnIIIId family. These residues often form the recognition site for the FnIIIId of an interacting partner protein (4, 9).

FnIIIIds have also been found in restricted set of carbohydrases from soil bacteria (16, 17). It is well known that, unlike eukaryotes, bacteria acquire a significant proportion of their genetic diversity through foreign sequences from distantly related organisms (18, 19). In particular, soil bacteria are noted for their mosaic genomes that reflect extensive recombination (20). The domains occur in different locations and in different in the carbohydrases of bacteria; in addition, the bacteria that possess FnIIIIds appear to be broadly distributed between Gram-positive and Gram-negative bacteria. As FnIIIIds appear sporadically in bacterial phylogenetic trees and have a high sequence similarity to those of animals, the presence of this domain in bacteria is regarded as the most convincing example of horizontal gene transfer from animal to prokaryote (17, 21).

The first bacterial FnIIIId to be reported was found in chitinase A1 from Bacillus circulans WL-12 (16), the structure of which is described in this paper. B. circulans WL-12 is a Gram-positive soil bacterium and was identified through its lysing of the cell walls of yeast and fungi. To hydrolyze chitin’s β-1,4-glycosidic-linkages, this bacterium uses three enzymes, chitinase A1 (ChiA1), C1 (ChiC1), and D1 (ChiD1) (22). All three chitinases adopt multidomain structures. ChiA1 consists of an N-terminal catalytic domain, two FnIIIIds, and a C-terminal chitin-binding domain (Fig. 1); ChiD1 is made up of an N-terminal chitin binding domain, an FnIIIId, and a C-terminal catalytic domain; ChiC1 comprises a catalytic domain and a C-terminal portion with no apparent sequence similarity to other known proteins. Among these, ChiA1 is known as the key enzyme of chitin degradation and exhibits the highest enzymatic activity for both insoluble and soluble chitin (23).

Except for the FnIIIIds, the tertiary structures of the domains of ChiA1 have been solved by x-ray crystallography and NMR. The crystal structure of the catalytic domain reveals an (α/β)8-TIM-barrel fold that is common to class 18 glycosyl hydrodrolases (24). Two exposed tryptophan residues (Trp-122 and Trp-134) of CatDChiA1 are thought to play an important role in the hydrolysis of crystalline chitin. We recently reported the solution structure of the chitin-binding domain of ChiA1

* 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.

** To whom correspondence should be addressed. Tel.: 81-45-508-7213; Fax: 81-45-508-7361; E-mail: shira@tsurumi.yokohama-cu.ac.jp.

† The abbreviations used are: FnIIIId, fibronectin type III domain; ChiA1, B. circulans WL-12 chitinase A1; HSQC, heteronuclear single quantum correlation; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlated spectroscopy; r.m.s.d., root mean square deviation; ChiN, N-terminal domain of chitinase A from S. marcescens.
(ChBD\textsubscript{ChiA1}), which can bind to chitin substrates in an insoluble or crystalline state (25). Although the overall topology of ChBD\textsubscript{ChiA1} is similar to that of the cellulose-binding domain from a bacterial cellulase, the location of exposed hydrophobic residues that are proposed to be important for substrate binding differs between the chitin- and cellulose-binding domains, indicating a difference in the mode of substrate binding.

The catalytic- and chitin-binding domains are tethered by two FnIIIDs (each comprising 86 residues). The N-terminal FnIID (\textsuperscript{1}FnIID\textsubscript{ChiA1}; residues Ala-464 to Thr-549) and the other FnIID (\textsuperscript{2}FnIID\textsubscript{ChiA1}; residues Ala-559 to Thr-644) are linked by a short sequence (9 residues) and share 74.4\% sequence identity with each other. The highest sequence identity between \textsuperscript{2}FnIID\textsubscript{ChiA1} and an animal FnIID is 34\%. It has been reported that the deletion of \textsuperscript{2}FnIID\textsubscript{ChiA1} has no impact on the chitin binding activity of ChiA1, but causes a significant decrease in the colloidal chitin hydrolyzing activity (26). The natural function of these FnIIIDs remains, however, unclear.

Although there has been a rapid increase in the amount of structural and functional information on animal FnIIIDs during the past several years, the complete lack of three-dimen-

sional structural information for bacterial FnIIIDs has obscured the evolutionary relationship of FnIIIDs. The precise structural organization of the FnIID is crucial to our understanding of how the domain functions, as well as to the identification of evolutionary relationships. Here we report the solution structure of \textsuperscript{2}FnIID\textsubscript{ChiA1} solved by multidimensional NMR spectroscopy. To the best of our knowledge, this is the first report of an FnIID structure from a bacterial protein. We also model the structure of \textsuperscript{2}FnIID\textsubscript{ChiA1} by using a homology modeling algorithm. These structures provide additional evidence in support of the hypothesis that the fibronectin type III domains of bacteria and animals may be related by a horizontal gene transfer process.

**EXPERIMENTAL PROCEDURES**

**Sample Preparation**—Recombinant \textsuperscript{2}FnIID\textsubscript{ChiA1} was obtained by expressing a His-tagged protein in *Escherichia coli* BL21(DE3) cells, followed by affinity purification on a Ni\textsuperscript{2+} chelating column, cleavage with Factor Xa, and final purification by gel filtration. After cleavage with Factor Xa, vector-derived His and Met residues remain in the N terminus as His-557/Met-558-\textsuperscript{2}FnIID\textsubscript{ChiA1}. Uniform \textsuperscript{15}N-labeling was achieved by growing the bacteria in M9 minimal medium containing \textsuperscript{15}NH\textsubscript{4}Cl as the sole nitrogen source; for uniform \textsuperscript{15}N- and \textsuperscript{13}C-labeled samples, \textsuperscript{13}C-glucose was used as the sole carbon source. Typical yields were 2 mg of pure protein/liter of bacterial culture. Most NMR experiments were performed with 1.0–1.5 mM \textsuperscript{2}FnIID\textsubscript{ChiA1} samples at 310 K and pH 6.5 (20 mM potassium phosphate, 50 mM KCl, 2 mM Pefabloc\textsuperscript{sc}) using H\textsubscript{2}O/D\textsubscript{2}O 9:1 (v/v) as the solvent. Homonuclear two-dimensional NOESY and TOCSY, \textsuperscript{1}H-\textsuperscript{15}N HSQC, \textsuperscript{1}H-\textsuperscript{13}C HSQC, 3D \textsuperscript{1}H-\textsuperscript{15}N TOCSY-HSQC, and a series of triple-resonance experiments incorporating pulsed field gradients, water flip-back pulses, and sensitivity enhancement when amide protons were detected in the F3 dimension: CBCACOHN, CBCANH, HNCO, HCONH, and HCHC
textsuperscript{2}TOSCY spectra (27). The two-dimensional NOESY (150-ms mixing time), three-dimensional \textsuperscript{1}H-\textsuperscript{15}N-\textsuperscript{13}C NOESY-HSQC (100- and 150-ms mixing times), and three-dimensional \textsuperscript{1}H-\textsuperscript{15}N-\textsuperscript{13}C NOESY-HSQC (150-ms mixing time) spectra were also used to derive distance restraints for structure determination. To identify the hydrogen-bond pattern, a \textsuperscript{3}J\textsubscript{HNCO} HNCO experiment was performed (28). The data on DMX 500 and DRX 500 machines were acquired with 512\times (HN) \times 22\times (N) \times 54\times (C) (complex \textsuperscript{*}) points and 16 scans (CBCACONH); 512\times (HN) \times 22\times (N) \times 50\times (C), 16 scans (CBCANH); 512\times (HN) \times 22\times (N) \times 105\times (C), 8 scans (HCCH-TOCSY); 512\times (HN) \times 22\times (N) \times 64\times (C), 32 scans (CCONH); 512\times (HN) \times 22\times (N) \times 40\times (H), 32 scans (HCONH); 512\times (H) \times 128\times (H) \times 30\times (C), 8 scans (HCCH-TOCSY); 512\times (HN) \times 30\times (N) \times 100\times (H), and 8 scans (3D \textsuperscript{1}H-\textsuperscript{15}N TOCSY-HSQC). The data at DRX 800 were acquired with 1024\times (HN) \times 30\times (N) \times 35\times (C), 32 scans (\textsuperscript{3}J\textsubscript{HNCO} HNCO); 1024\times (HN) \times 32\times (N) \times 135\times (H), 8 scans (3D \textsuperscript{1}H-\textsuperscript{15}N NOESY-HSQC); 1024\times (H) \times 50\times (C) \times 128\times (H), and 8 scans (3D \textsuperscript{1}H-\textsuperscript{13}C NOESY-HSQC). All data were processed with the program NMRPipe (29) and analyzed with the program NMRView (30).

**Structure Calculations**—Initially, structure calculation and NOE peak assignment were performed in an iterative and manual manner using the program DYANA (version 1.5) (31). NOE cross-peak intensities were classified as strong, medium, or weak, and assigned to restraints of 8–3.0, 3.0–1.8, or 1.8–0.5 Å, respectively. On the base of \textsuperscript{3}J\textsubscript{HNCO} HNCO spectrum, hydrogen-bond restraints were applied as 2.5–3.3 Å for N–O pairs and 1.8–2.5 Å for N–O pairs and used from the initial procedure. Backbone torsion angle restraints were derived from \textsuperscript{3}J\textsubscript{HNC}. HMQC-J (27) and TALOS program (32). The backbone \textgreek{phi} angle restraints were used as −65 to 25° for \textsuperscript{3}J\textsubscript{HNC}, < 4.7°, −120° to 40° for \textsuperscript{3}J\textsubscript{HNC}, > 8.5 Hz and < 9.9 Hz, and −120° to 20° for \textsuperscript{3}J\textsubscript{HNC} > 10.0 Hz. The side chain TALOS \textgreek{psi} angle was within the range of −120° to 120°, and the TALOS-derived \textgreek{phi} angle was used for restraints. The torsion angles \textgreek{chi}1 of Tyr, Phe and Trp were estimated from \textsuperscript{3}J\textsubscript{HNC}, \textsuperscript{3}J\textsubscript{HNC}, and \textsuperscript{3}J\textsubscript{HNC}, coupling constants (33). Final refinement employing ambiguous NOE assignments and floating chimera assignments was performed using the program ARIA (version 1.0) and a CNS package as described (34). A total of 150 structures were refined in the last (ninth) iteration, and the 30 lowest energy structures were analyzed using MOLMOL (35), AQUA, and PROCHECK-NMR software (36).

**Homology Modeling and Sequence Alignments**—The structure of \textsuperscript{1}FnIID\textsubscript{ChiA1} was modeled using the \textsuperscript{1}FnIID\textsubscript{ChiA1} structure as a template. \textsuperscript{1}FnIID\textsubscript{ChiA1} shows 74.4\% amino acid sequence identity with \textsuperscript{2}FnIID\textsubscript{ChiA1} and there are no gaps in the sequence. The MODELLER program (37) was used for modeling, and the quality of the model was assessed using PROCHECK. Homologous bacterial sequences of \textsuperscript{2}FnIID\textsubscript{ChiA1} were obtained from the SMART (smart.embl-heidelberg.de) server (38). To date (July 2001), 135 bacterial FnIIIDs from 102 proteins are registered in the SMART server. Among these, we chose only FnIIIDs from the proteins whose functions are clearly identified as bacterial glycosyl hydrolases. Initially there were 83 FnIIIDs from 61 proteins. These sequences were filtered manually to get only one FnIID sequence in the case of a protein with multiple FnIIIDs. Species redundancies were also considered. For comparison, three FnIIIDs from *B. circulans* were added. Finally, 21 FnIID sequences from 20 carbohydrate families were used to multiple sequence alignments using ClustalW with default parameters (39). CHROMA software was used for analyzing and annotating the results from sequence alignments (40).

**RESULTS**

**Resonance Assignment, Restraints for Structure Calculation, and Structure Determination**—The sequence-specific NMR backbone assignment was obtained from CBCACONH and CBCANH spectra. Except for the signals from the vector-derived N-terminal His and Met residues, which were not observed in \textsuperscript{15}N\textsubscript{HSQC}, all the backbone signals were assigned (Fig. 2). Side-chain assignment was achieved using mainly CCONH, HCONH, and HCHC
textsuperscript{2}TOSCY spectra. Tyrosine, phenylalanine, and tryptophan side-chain chemical shifts were assigned from two-dimensional NOESY and TOCSY spectra. To help assignment, we used programs written in-house using the chemical shift statistics of the BioMagResBank data base (www.bmr.wisc.edu) and the semi-automatic assignment...
module of NMRView. A nearly complete assignment of 2Fn-
IIIID\textsubscript{ChiA1} was achieved. This result has been deposited in the
BioMagResBank under accession no. 5178. To extract distance
restraints, we used NOESY spectra with 100- and 150-ms
mixing times. No severe spin diffusion effects were found in the
spectrum with a 150-ms mixing time. We also measured
$^{3}$\textsubscript{H}J\textsubscript{HNCO} HNCO spectrum to get hydrogen-bond information
directly. By comparing HNCO and $^{3}$\textsubscript{H}J\textsubscript{HNCO} HNCO spectra, we
identified 23 inter-hydrogen bonds between the
$/\textsubscript{H}$9252-strands (Fig. 3). With 117 torsion angle restraints, these hydrogen-bond
restraints (23*2) were applied from the initial step of structure
calculation. Structure calculation with DYANA was repeated
by adding distance information derived from NOE cross-peaks
identified manually, until an ensemble of the calculated struc-
tures gave the global fold. When the calculated structure con-
verged enough to identify the global fold, the ARIA procedure
was employed for further refinement using ambiguous NOE
cross-peaks. This procedure is essentially the same as that
described previously (34).

The structure of the 2FnIIIID\textsubscript{ChiA1} was determined from the
distance and torsional angle restraints listed in Table I. An
overlay of the final 30 structures shows that the backbone and
hydrophobic side chains have been well defined (Figs. 4 and 5).
The average root mean square deviation (r.m.s.d.) values of
these structures (Ala-559 to Thr-644) are 0.431 and 0.714 Å
for the backbone and all the heavy atoms, respectively. There was
no distance restraint violation of $>$0.3 Å, and no dihedral angle
restraint violation of $>$5 Å. The detailed statistics for the struc-
tures are shown in Table I.

Sequence Alignments of Bacterial FnIIIs—Of the 61
FnIII-containing bacterial carbohydrases that were initially
retrieved from SMART data base, 21 are from \textit{Streptomyces},
and 16 are from \textit{Bacillus} genera. For a comparison of FnIIIs
across genera, we selected only one FnIII from each genus. As
the sequences of FnIIIs from the same enzyme tend to be
more similar to each other than to FnIIIs from other proteins,
we also selected only one FnIII sequence per enzyme. If all
the FnIII sequences were used in the alignment, then the
apparent similarity between them and the proportion of those
from Gram-positive bacteria increased. All data concerning
acronyms, enzymatic function, organism name, Gram sensitiv-
ity, and order/total number are listed in Table II (Fig. 6).

FnIII from \textit{Yersinia enterocolitica} (O68975) exhibited the
lowest sequence identity (32.6%) to 2FnIIIID\textsubscript{ChiA1}. The lengths
of aligned sequences are nearly the same except for those of
\textit{Erwinia chrysanthemi} (PEHX\textsubscript{ERWCH}) and O68975, which
contain $\sim$20 more amino acids in their C\textsubscript{E} loops in comparison
with the others. A comparison of bacterial FnIII sequences
with animal FnIII sequences indicates that the C-terminal 20
residues are more conserved in the bacterial sequences.

Structure Description—The structure of 2FnIIIID\textsubscript{ChiA1} is a can-
nical $\beta$-sandwich structure containing two antiparallel
$\beta$-sheets that are packed face to face (Fig. 4B). One sheet is
composed of three $\beta$-strands (A, B, and E), whereas the other is
composed of four $\beta$-strands (C, C\textsubscript{E}, F, and G). Strand G is divided
into two sections, G1 and G2, as in animal FnIIIs. Three loops
in the direction of N terminus (loops BC, C\textsubscript{E}, and FG) and three
in the direction of the C terminus (loops AB, CC\textsubscript{E}, and EF)
connect the seven $\beta$-strands, respectively (Fig. 4B).

The $\beta$-sandwich scaffold of 2FnIIIID\textsubscript{ChiA1} is stabilized by an
extensive hydrogen-bond network between the $\beta$-strands, and by a hydrophobic core formed by inward-facing residues from the $\beta$-sheets. It is well known that animal FnIIIDs contain several $\beta$-bulge structures on the edge strands A, C', and G. In $^{2}F_{nIII}D_{CHI A1}$, three $\beta$-bulge structures at residues Asn-565/Leu-566 at the beginning of strand A, Thr-569/Ala-570 at middle of strand A, and Ala-601/Thr-602 within strand C' could be unambiguously identified by the direct detection of hydrogen bonds through scalar couplings. This method naturally revealed the hydrogen-bond network that defines the strand topology, as shown in Fig. 3, demonstrating its utility. The hydrogen-bond network shows that strand G is divided into two segments, G1 and G2, as is commonly seen in animal FnIIIDs.

The residues that are conserved in bacterial FnIIIDs and form the hydrophobic core are well defined in the final structures (Fig. 5). The following residues are totally buried in the core, with an solvent-accessible surface area of less than 7%: Pro-563, Ile-576, Leu-578, Trp-580, Ser-583, Tyr-592, Val-594, Ala-609, Ile-611, Phe-622, Val-624, Ala-626, Ser-637, and Val-642. The hydrophobic core is made up of two clusters in which three aromatic residues play a central role. One of the clusters contains Trp-580 and Tyr-592 surrounded by Pro-563, Ser-583, Val-594, Val-624, and Ala-626. These aromatic residues are nearly completely conserved, even in animal FnIIIDs, suggesting that this cluster is integral for maintaining the FnIIIId fold. The other cluster is made up of Phe-662 surrounded by Ile-576, Leu-578, Ile-611, and Val-642. The hydrophobic properties of these residues are also highly conserved across bacterial and animal FnIIIDs (Figs. 5 and 6).

For animal proteins, some FnIIIDs display a variety of binding modes with other proteins using combinations of the loop regions. For example, human fibronectin binds to integrin through the RGD site on the FG loop and the PXSRN site on the C'E loop, which form interfacial surfaces (4, 9). The importance of the loop sequence for molecular recognition is also demonstrated by an artificially engineered FnIIIId with altered sequences on the BC and FG loops that binds ubiquitin with high affinity (41). In bacterial FnIIIDs, the AB, EF, and FG loops are relatively well conserved in length and sequence. In contrast, the BC, CC', and C'E loops are variable (Fig. 6). The BC loop is stabilized by a hydrogen bond between the mainchain HN group of Ser-583 on the loop and O$_{\gamma}$ of Tyr-592. In 28 of 30 final structures, this HN group and the O$_{\gamma}$ are located within 2.35 A. This bond is also found in some animal FnIIIDs.
Table I
Structural statistics for 2FnIIID_{Chair}

|                | 1FnIIID_{Chair} | 2FnIIID_{Chair} | 1FnIIID_{Chair} - 2FnIIID_{Chair} |
|----------------|-----------------|-----------------|----------------------------------|
| Deviation from restraints | 0.017 ± 5.7 × 10^{-3} Å | 0.19 × 3.9 × 10^{-2} Å | 0.099 ± 3.0 × 10^{-4} Å |
| Hydrogen bond  | 0.25 × 3.9 × 10^{-2} Å | 0.25 × 3.9 × 10^{-2} Å | 0.16 ± 3.0 × 10^{-2} Å |
| Ramachandran analysis by PROCHECK (residues Ala-559 to Thr-644) | | |
| Residues in most favored regions | 74.0% | 74.0% | 74.0% |
| Residues in additional allowed regions | 23.6% | 23.6% | 23.6% |
| Residues in generously allowed regions | 2.4% | 2.4% | 2.4% |
| Residues in disallowed regions | 0.0% | 0.0% | 0.0% |
| Coordinate precision (residues Ala-559 to Thr-644) | | |
| r.m.s.d. of backbone atoms | 0.431 Å | 0.431 Å | 0.431 Å |
| r.m.s.d. of heavy atoms | 0.743 Å | 0.743 Å | 0.743 Å |

The Lennard-Jones van der Waals energy was calculated using CHARMM parameters and was not included in the target function employed in the structure calculations.

None of these structures exhibited distance violations >0.3 Å or dihedral angle violations >5°.

Fig. 4. Solution structure of 2FnIIID_{Chair}. A, stereo view of the superposition of the final 30 lowest energy structures (residues Ala-559 to Thr-644) calculated by ARIA. Regular secondary structure regions are shown in dark colors. B, ribbon diagram of the lowest energy structure. The b-strands (A, B, C, C’, E, F, and G (G1 and G2)) are labeled. All structure figures except for Fig. 7 were prepared using the program MOLMOL (35).

In addition to the loop regions, some exposed residues in the b-sheets play a role in molecular recognition in some proteins. In human growth factor and human tissue factor receptor, the association of successive domains in the FnIIID-FnIIID segment creates charged surfaces around the domain boundaries, which serve as binding sites for their ligands (2, 5). In contrast, the surface of 2FnIIID_{Chair} is mostly surrounded by noncharged residues as shown in Fig. 7A. 2FnIIID_{Chair} has only four negatively charged (Asp-585, Asp-593, Asp-617, and Asp-628) and three positively charged (Lys-625, Lys-627, and Lys-643) residues on the surface, which are sparsely located on the protein surface and do not form a noticeable charged patch.

To analyze the charge status of the 1FnIIID_{Chair}–2FnIIID_{Chair} module, we modeled the structure of 1FnIIID_{Chair}. As there are no gaps in the sequence alignment and a high sequence identity (74.4%) between 2FnIIID_{Chair} and 1FnIIID_{Chair}, homology modeling of the 1FnIIID_{Chair} structure was quite straightforward. The r.m.s.d. between the modeled 1FnIIID_{Chair} and 2FnIIID_{Chair} is 1.173 Å over the Ca coordinates of Ala-559 to Thr-644. Contacts between the hydropobic core-forming residues are nearly identical between the two FnIIIDs. All the backbone angles of the modeled structure, except for those of glycines, are in the allowed regions of Ramachandran plot.

The electrostatic potential map on the surface is drawn in Fig. 7B. The modeled structure of 1FnIIID_{Chair} reveals that charged groups are sparsely distributed on the protein surface with no noticeable charged patches, as seen for 2FnIIID_{Chair}. Modeling of the 1FnIIID_{Chair}–2FnIIID_{Chair} module using the coordinates of 1FnIIID_{Chair} and 2FnIIID_{Chair} at a variety of inter-domain orientations suggests that, unlike human tissue factor and human growth factor receptor, the two domains are unlikely to produce a cluster of charged side chains at the domain boundary (data not shown).

Low Sequence Complexity—The noteworthy feature of 2FnIIID_{Chair} is its unusual amino acid composition. The sequence of this domain is rich in amino acids with short side chains. It has 20 threonine, 16 alanine, and 10 serine residues, and these three types of amino acid make up over 50% of the residues of this domain. Owing to the biased amino acid composition, if using a BLAST sequence homology search without a filtering option, then proteins with low complexity such as antifreeze
proteins (threonine- and alanine-rich) or mucin (threonine- and 
serine-rich) are retrieved with high probability. Although pro-
tein with low sequence complexity are more likely to generate 
relatively extended structures (42), 2FnIIIDChiA adopts a glob-
ular and compact fold. The molecular mass of 2FnIIIDChiA 
(Ala-559 to Thr-644) is 8423.1 daltons, giving an average mass 
per residue of 97.9 daltons, which is lower than 99% of the 
protein sequences contained in SWISS-PROT (43). This high 
content of small side chains probably caused the relatively 
small difference between backbone and heavy atom r.m.s.d. 
values in the final 30 NMR structures. A high content of small 
side chains is generally found in bacterial FnIIIDs, but not in 
animal FnIIIDs.

Most of the threonine residues are located on the surface of 
2FnIIIDChiA (Fig. 7C) and, thus, have little effect on its global 
fold. Most threonines are not highly conserved in bacterial 
FnIIIDs; instead, small residues (Ala, Cys, Ser, Thr, Asp, Asn, 
Val, Gly, and Pro) occur with a probability greater than 80% at 
the positions corresponding to threonines in 2FnIIIDChiA. 
These residues presumably contribute to the solvent-accessible 
surface of each molecule and, thus, may be important functionally.

**Structural Comparison**—The program DALI was used to 
search protein data bases for structures similar to that of 
2FnIIIDChiA (44). It retrieved 131 structures with Z scores 
greater than 2.0 from the Protein Data Bank. All of the struc-
tures with Z scores larger than 9.0 are FnIIIDs. Other families 
of β-sheet sandwich structures, including the immunoglobulin 
fold, were retrieved with Z scores of 2.0–8.0. The following 
structures were found to be the most similar to that of 2Fn-
IIIDChiA with Z scores larger than 10.0 and a Cα atom r.m.s.d. 
smaller than 2.0 Å: integrin 4 subunit fragment (Protein Data 
Bank code: 1QG3, Z score: 11.9, r.m.s.d.: 1.8 (over 83 residues)) 
(8), tenascin fragment (1QRA, 11.4, 1.6 (81)) (6), fibronectin 
fragment (1FNH, 11.3, 1.7 (81)) (7), fibronectin (1FNF, 11.0, 1.8 
(81)) (4), Drosophila neuroglian (1CFB, 10.7, 1.9 (81)) (3), and 
titin fragment (1BVV, 10.6, 1.9 (84)) (10). Despite the high 
similarity of global fold, the surface electrostatic potentials and 
loop conformations of these proteins are different from those of 
2FnIIIDChiA, and also differ between one another. In particu-

---

**TABLE II**

Bacterial FnIIIDs used for multiple sequence alignment

| Identification | Organism | Enzyme | Gram |
|----------------|----------|--------|------|
| 2Fn3           | B. circulans | Chitinase | +    |
| 1Fn3           | B. circulans | Chitinase | +    |
| CHID_BACCI     | B. circulans | Chitinase | +    |
| Q45494         | K. zopfii  | Chitinase | +    |
| Q62N8         | Acidocorax | PHB depolymerase | -    |
| Q6L3N6         | Leptotrix  | PHB depolymerase | -    |
| 024719         | C. testosteroni | PHB depolymerase | -    |
| PHB_ALCFA      | A. faecalis | PHB depolymerase | -    |
| Q52155         | B. pickettii | PHB depolymerase | -    |
| Q9WX3          | S. marcescens | Chitinase | -    |
| Q9IIH5         | Pseudomonas aeruginosa | Chitinase | -    |
| O30678         | Xanthomonas maltophilia | Chitinase | -    |
| GUXB_CELFI     | C. fimii    | Exoglucanase | +    |
| GUN4_THEFU     | T. fusca    | Endoglucanase | +    |
| Q6L36          | Streptomyces coelicolor | α-Amylase | +    |
| Q66348         | Cytobagga  | Collagenase | -    |
| Q9AJ4          | Ruminococcus albus | Xylanase | +    |
| Q9CE95         | Lactococcus lactis | Chitinase | +    |
| APU_THETU      | T. thermosulfurogenes | Amylopullulanase | +    |
| PEIX_ERKCH     | E. chrysanthemi | Exo-poly-α-β-galacturonosidase | -    |
| O88975         | Y. enterocolitica | Exo-polygalacturonase | -    |
Fig. 6. Multiple sequence alignment of bacterial FnIIIDs. Amino acid residues with a 80% consensus are colored. Classification of the consensus amino acid type is according to that used in program CHROMA as follows: * for positively charged residues (His, Lys, and Arg), † for negatively charged residues (Asp and Glu), a for aromatic residues (Phe, His, Trp, and Tyr), h for hydrophobic residues (Ala, Cys, Phe, His, Ile, Leu, Met, Val, Trp, and Tyr), l for aliphatic residues (Ile, Leu, and Val), ‡ for alcohol residues (Ser and Thr), p for polar residues (Asp, Glu, His, Lys, Asn, Gln, Arg, Ser, and Thr), s for small residues (Ala, Cys, Ser, Thr, Asp, Asn, Val, Gly, and Pro), w for tiny residues (Ala, Gly, and Ser), and uppercase letters for consensus amino acids. The lengths of the considered sequences are given in parentheses on the left, and the orders and total numbers of FnIIIDs in enzymes are in parentheses on the right. The first and last sequence numbers are indicated. 2FnIID_chia1 and 1FnIID_chia1 are abbreviated as 2Fn3 and 1Fn3, respectively. Secondary structure elements of 2FnIID_chia1 are indicated above the alignments. Sequence alignment was carried out by ClustalW (39) and annotated using CHROMA (40).

Discussion

The Residues under Evolutional Constraints Are Similar between the Bacterial and Animal FnIIIDs—The NMR solution structure of 1FnIID_chia1 and multiple sequences alignment of bacterial FnIIIDs show that these domains are surprisingly similar each other despite the broad and sporadic distribution of the bacteria containing them. Residues that play important roles for the scaffold formation are totally conserved. Interestingly, the properties of amino acids that are presumably under weak evolutionary pressure, such as residues on loops or exposed on β-sheets, are also preserved, as shown in Figs. 5 and 6. Although we cannot exclude the possibility that the theronines on the surface of 2FnIID_chia1 have specialized functions, such as those of antifreeze proteins (45, 46), it seems more reasonable to think that these domains evolved recently from a domain that was rich in amino acids with small side chains and that these amino acids have not been substituted much, considering the relatively high content of light amino acids in bacterial FnIIIDs. To the best of our knowledge, it is very rare to find bacterial and vertebrate proteins that share such common features.

The principle that protein evolution is determined mainly by constraints on activity, specificity, folding, and stability is generally accepted (47, 48). The key residues that play roles in preserving the nature of the protein (packing, hydrogen bonding, or unusual dihedral angles) are inclined to be strongly conserved in property, if not in identity, and can be used to measure evolutionary distance. From this point of view, if bacterial FnIIIDs correlate with animal FnIIIDs because of horizontal DNA acquisition, the residues under strong evolutionary pressure will be more highly conserved between closer relatives.

As a BLAST search indicates that the sequence of the FnIIIId from titin proteins, which are intracellular members of the FnIIIId-containing family, is most similar to that of 2FnIID_chia1 with an amino acid identity of 34%, we use titin FnIIId as an example. Fortunately, the tertiary structure of a titin module from human cardiac muscle (FnIIId1BPV; Protein Data Bank code, 1BPV) has been reported (10). FnIIId1BPV, which shares a sequence identity of 32% with 2FnIID_chia1 and shows the Ca r.m.s.d. of 1.9 Å over corresponding 81 residues, reveals a very similar hydrophobic-core packing. Of the 14 residues of 2FnIID_chia1 with a solvent-accessible surface area of less than 7%, 11 residues from FnIIId1BPV (including 8 identical residues: Pro-563, Leu-578, Trp-580, Tyr-592, Val-594, Phe-622, Ala-626, and Ser-637 after 2FnIID_chia1 numbering) have the same hydrophobic feature (Fig. 8).

Our argument for a close relationship between FnIIIDs from bacteria and animals is reinforced by the discovery of FnIIIDs in other kingdoms. Powerful searching tools, such as the hidden Markov model approach, have recently revealed new FnIIIDs from yeast and plant (49, 50), despite low sequence similarities with bacterial FnIIIDs (~15% identities with the sequence of 2FnIID_chia1). Of these, the crystal structure of the
These observations indicate that, in terms of sequence and tertiary structure, $^{2}$FnIIID$_{ChiA1}$ are less similar to plant FnIIID$^{4KBP}$ than animal FnIIID$^{1BPV}$. This finding may imply that bacterial $^{2}$FnIIID$_{ChiA1}$ shares a relationship with animal FnIIIDs that is closer than would be expected from the evolutionary distance between animals and bacteria.

**The Structure of $^{2}$FnIIID$_{ChiA1}$ Shows Different Features from Other $\beta$-Sandwich Domains of Bacterial Carbohydrases**—It has been reported that the FnIIIDs of chitinase A1 from *B. circulans* WL-12 are not directly involved in chitin binding. The truncated form of chitinase A1 from *B. circulans* WL-12, lacking either $^{2}$FnIIID$_{ChiA1}$ or both $^{2}$FnIIID$_{ChiA1}$ and $^{2}$FnIIID$_{ChiA1}$, showed nearly the same binding affinity for chitin as the full-length enzyme. By contrast, a significant decrease in chitin hydrolyzing activity was observed for the truncated forms, in proportion to the number of FnIIIDs missing (26).

A structure resembling that of FnIID has been found in another bacterial carbohydrate, chitinase A from *Serratia marcescens* (53). This enzyme consists of three domains: an N-terminal domain (ChiN), a catalytic domain, and a small ($\alpha+\beta$) domain. The sequence of the catalytic domain is conserved between this enzyme and chitinase A1 from *B. circulans*, and their tertiary structures are similar (r.m.s.d. of 1.25 Å, corresponding Ca atoms), suggesting that they share an identical catalytic mechanism. ChiN adopts a global fold similar to $^{2}$FnIIID$_{ChiA1}$ but with an additional $\beta$-$\alpha$-$\beta$ element between the A and B strands that interacts with the other domain (Fig. 9, B and C). Although a structure-based sequence alignment shows that ChiN has a sequence identity of 19.8%, the arrangement of hydrophobic-core-forming residues in ChiN is different from those in $^{2}$FnIIID$_{ChiA1}$ and other FnIIIDs. Moreover, important differences are found in the surface residues of these two domains. ChiN has adjacently arranged tryptophans (Trp-33 and Trp-69) exposed on a continuous surface with the conserved aromatic residues of catalytic domain (Trp-245 and Phe-232). These residues play important roles in guiding a chitin chain into the catalytic site (54). In contrast, $^{2}$FnIIID$_{ChiA1}$ has no exposed tryptophans. The only aromatic residue with a solvent-accessible surface area of more than 10% is Tyr-620 (13.3%), a residue that is highly conserved throughout animal proteins.

Unlike the bacterial FnIIIDs, which are broadly distributed across both Gram-positive and Gram-negative bacteria, the ChiN domain has been found only in the chitinase of Gram-negative bacteria. Interestingly, all the ChiN domains are located at the N termini of these enzymes. In contrast, the FnIIIDs of bacteria, like animal FnIIIDs, are located at a variety of positions within the proteins. This observation suggests that the FnIIID-containing proteins have arisen through domain shuffling. Recently, another chitinase (ChiC) was identified from *S. marcescens* that has three domains including an FnIID (55). Although its FnIID exhibits a high sequence similarity to $^{1}$FnIIID$_{ChiA1}$ and $^{2}$FnIIID$_{ChiA1}$, the sequence of the catalytic domain is different from that of chitinase A1 from *B. circulans*. As above, this observation implies that these
FnIIIDs have been inserted into the polypeptides by domain shuffling.

Several other β-sandwich architectures besides ChiN and FnIII have been found in bacterial carbohydrases, such as domain N of β-amylase II (56), C-terminal starch-binding domain of β-amylase (57), cellulose-binding domain (58), and xylan-binding domain (59). The secondary structure topologies of these modules are different from that of FnIIID, although their tertiary structures show limited similarity to that of FnIIIDChiA1 with DALI Z scores ranging from 2.6 to 5.4. These domains are capable of binding to carbohydrates and commonly possess surface-exposed aromatic residues that are thought to make contacts with substrate sugar chains. Only the structure of FnIIIDChiA1 does not exhibit such features.

The high degree of conservation of the global structures among the bacterial FnIII-containing carbohydrateases and the lack of known function of the domain and of surface features reminiscent of specific functions in FnIIIChiA1 and FnIIIChiA1 tempt us to postulate that FnIII has the role of a spacer in bacterial carbohydrateases. To degrade insoluble substrate efficiently, catalytic and binding domains must adopt various relative positions. The fact that most FnIIIIs are located between the catalytic and binding domains may support this hypothesis. Moreover, mechanical elasticity, which is an intrinsic property of FnIII, may make the FnIII the best candidate for this type of role.

Acknowledgment—We thank Dr. Chojiro Kojima for helpful discussions.

REFERENCES

1. Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996) Q. Rev. Biophys. 29, 119–167
2. de Vos, A. M., Uitsch, M., and Kossiakoff, A. A. (1992) Science 255, 306–312
3. Huber, A. H., Wang, Y. M., Bieber, A. J., and Bjorkman, P. J. (1994) Neuron 12, 717–731
4. Leahy, D. J., Aukhil, I., and Erickson, H. P. (1996) Cell 84, 155–164
5. Leahy, D. J., Aukhil, I., and Erickson, H. P. (1996) Cell 84, 155–164
6. de Pereda, J. M., Wiche, G., and Liddington, R. C. (1999) EMBO J. 18, 4087–4095
7. Sharma, A., Askari, J. A., Humphries, M. J., Jones, E. Y., and Stuart, D. I. (1999) EMBO J. 18, 4146–4147
8. de Pereda, J. M., Wiche, G., and Liddington, R. C. (1999) EMBO J. 18, 4087–4095
9. Main, A. L., Harvey, T. S., Baron, M., Boyd, J., and Campbell, I. D. (1992) Cell 71, 671–678
Solution Structure of the Fibronectin Type III Domain from *Bacillus circulans* WL-12 Chitinase A1

Jun-Goo Jee, Takahisa Ikegami, Masayuki Hashimoto, Takeshi Kawabata, Mitsunori Ikeguchi, Takeshi Watanabe and Masahiro Shirakawa

*J. Biol. Chem. 2002, 277:1388-1397.*

doi: 10.1074/jbc.M109726200 originally published online October 12, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M109726200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 57 references, 12 of which can be accessed free at [http://www.jbc.org/content/277/2/1388.full.html#ref-list-1](http://www.jbc.org/content/277/2/1388.full.html#ref-list-1)