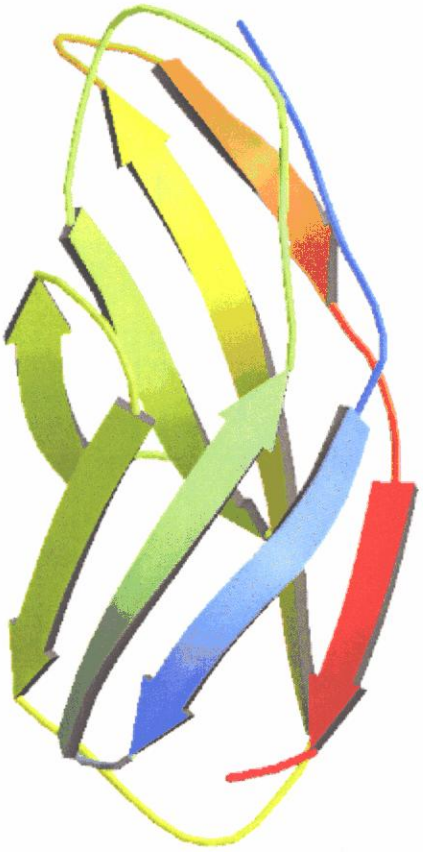
博士論文番号：9981051

NMR Studies of the fibronectin type III domain from *Bacillus circulans* WL-12 chitinase A1

(*Bacillus circulans* WL-12由来キチナーゼA1の
III型フィブロネクチンドメインのNMR研究)



Jee, Jun Goo (池　畯　求)
奈良先端科学技術大学院大学
バイオサイエンス研究科　生体高分子構造学講座

CONTENTS

0. ABSTRACT

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 (FnIIID). Certain extracellular proteins of soil bacteria contain an unusual cluster of FnIIIDs, which exhibits sequence similarity to those of animals. It has been proposed that the bacterial FnIIID had been acquired horizontally from animals. In this thesis, I report the solution structure of the FnIIID of chitinase A1 from *Bacillus circulans* WL-12. To the best of my knowledge, this is the first tertiary structure to be reported for an FnIIID from a bacterial protein. Striking structural similarities to animal counterpart are found. Sequence comparisons with other FnIIIDs from bacterial proteins reveal that the hydrophobic core-forming residues of bacterial FnIIIDs are highly conserved, and thus are under strong evolutionary pressure. The total lack of surface-exposed aromatic residues displays that the role of this FnIIID is different from those of other bacterial β-sandwich domains which function as carbohydrate-binding modules. These results may support the hypothesis that the evolution of the FnIIID in bacterial carbohydrase occurred horizontally. I also conducted NMR relaxation experiments to get the dynamic

information of this FnIIID. The data indicate that the domain experiences anisotropic tumbling and less restrictive internal motion than animal FnIIIDs.

1. INTRODUCTION

The fibronectin type III domain (FnIIID) 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, FnIIIDs are also found in membrane receptor proteins as well as in intracellular proteins (1). X-ray crystallography (2-8) and nuclear magnetic resonance (NMR) spectroscopy (9-12) have been used to elucidate the structures of several animal FnIIIDs, 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 FnIIIDs are still unclear. Each FnIIID comprises domain-intrinsic and domain-specific regions (13). The former, made-up of relatively conserved residues, are responsible for forming the FnIIID scaffold, which comprises a hydrogen-bond network and a hydrophobic core. The scaffold is common to all FnIIID structures, 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 conserved across the FnIIID family. These residues often form the recognition site for the FnIIID of an interacting partner protein (4, 9).

FnIIIDs have also been found in a 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 number in the carbohydrases of bacteria; in addition, the bacteria that possess FnIIIDs appear to be broadly distributed among Gram-positive and Gram-negative bacteria. As FnIIIDs 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 FnIIID to be reported was found in chitinase A1 from *Bacillus circulans* WL-12 (16), the structure of which is described in this thesis. *Bacillus circulans* WL-12 is a Gram-positive soil bacterium and was identified as through its lysis 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 ($CatD_{ChiA1}$), two FnIIIDs, and a

C-terminal chitin-binding domain ($ChBD_{ChiA1}$) (Fig. 1-1); ChiD1 is made up of an N-terminal chitin binding domain, an FnIIID, 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 FnIIIDs, 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 $(\alpha/\beta)_8$-TIM-barrel fold that is common to class-18 glycosyl hydrolases (24). Two exposed tryptophan residues (Trp-122 and Trp-134) of $CatD_{ChiA1}$ play important roles in the hydrolysis of crystalline chitin (25). Recently the solution structure of $ChBD_{ChiA1}$ was reported, which can bind to chitin substrates in an insoluble or crystalline state (26). Although the overall topology of $ChBD_{ChiA1}$ 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 FnIIID ($^{1}FnIIID_{ChiA1}$; residues Ala-464 to Thr-549) and the other FnIIID ($^{2}FnIIID_{ChiA1}$; 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 $^{2}FnIIID_{ChiA1}$ and an animal FnIIID is 34%. It has been reported that the deletion of $^{2}FnIIID_{ChiA1}$ has no impact on the chitin-binding activity of ChiA1, but causes a significant decrease in the colloidal chitin-hydrolyzing activity (27). 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 3D-structural information for bacterial FnIIIDs has obscured the evolutionary relationship of FnIIIDs. The precise structural organization of the FnIIID is crucial to our understanding of how the domain functions, as well as to the identification of evolutionary relationships.

Here I report the solution structure of $^{2}FnIIID_{ChiA1}$ solved by multidimensional NMR spectroscopy. To the best of my knowledge, this is the first report of an FnIIID structure from a bacterial protein. The structure of $^{1}FnIIID_{ChiA1}$ was also calculated using a homology modeling algorithm. I compare the structure of $^{2}FnIIID_{ChiA1}$ with other known structures with evolutionary and functional views.

At present NMR spectroscopy is the only tool to monitor the dynamic features of a protein at residue specific sites and on a broad range of timescales. There have been enormous NMR experiments and many techniques have been developed to investigate dynamic process of protein (28-30). Besides determining the structure of $^{2}FnIIID_{ChiA1}$, I obtained the data of backbone dynamics using ^{15}N relaxation experiments. This result is used to get insight into the feature of $^{2}FnIIID_{ChiA1}$.

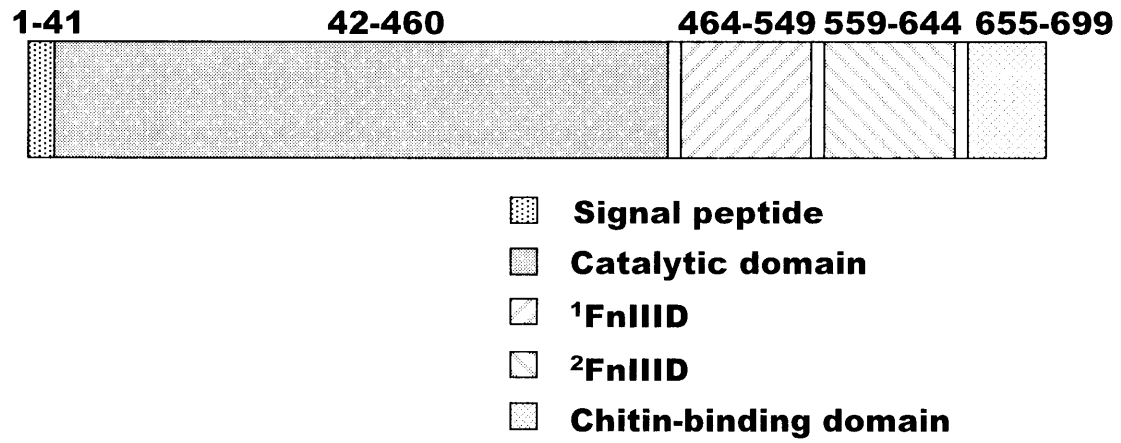


Fig. 1-1. Domain structure of chitinase A1 from *Bacillus cirulans* WL-12.

2. MATERIALS AND METHODS

2.1 Sample Preparation

Recombinant ^{2}FnIIID$_{ChiA1}$ was obtained by expressing a His-tagged protein in *Escherichia coli* BL21(DE3) cells, followed by affinity purification on a Ni^{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/^{2}FnIIID$_{ChiA1}$. Uniform ^{15}N-labeling was achieved by growing the bacteria in M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source; for uniform ^{15}N- and ^{13}C-labeled samples, $^{13}C_6$-glucose was used as the sole carbon source. Typical yields were 2 mg of pure protein per L of bacterial culture. Most NMR experiments were performed with 1.0~1.5 mM ^{2}FnIIID$_{ChiA1}$ samples at 310 K and pH 6.5 (20 mM potassium phosphate, 50 mM KCl, 2mM Pefabloc®) using H_2O/D_2O 9:1 (v/v) as the solvent. Homonuclear 2D NOESY and TOCSY, and 3D HCCH-TOCSY and ^{1}H-^{13}C NOESY-HSQC spectra were obtained with samples dissolved in D_2O.

2.2 NMR Spectroscopy for Structure Calculation

NMR spectra were acquired with a Bruker DMX500, DRX500 or DRX800

spectrometer equipped with a pulse-field-gradient triple resonance probe. Assignment of the ^{1}H, ^{15}N and ^{13}C resonances of $^{2}FnIIID_{ChiA1}$ was obtained using 2D TOCSY, NOESY, ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, 3D ^{1}H-^{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, CCONH, HCCONH, and HCCH-TOCSY spectra (31). The 2D NOESY (150 ms mixing time), 3D ^{1}H-^{15}N NOESY-HSQC (100 ms and 150 ms mixing times), and 3D ^{1}H-^{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 $^{3h}J_{HNCO}$ HNCO experiment was performed (32). The data on DMX 500 and DRX 500 machines were acquired with 512*(HN) x 22*(N) x 54*(C) complex (*) points and 16 scans (CBCACONH); 512*(HN) x 22*(N) x 50*(C), 16 scans (CBCANH); 512*(HN) x 22*(N) x 105*(C), 8 scans (HNCO); 512*(HN) x 22*(N) x 64*(C), 32 scans (CCONH); 512*(HN) x 22*(N) x 40*(H), 32 scans (HCCONH); 512*(H) x 128*(H) x 30*(C), 8 scans (HCCH-TOCSY); 512*(HN) x 30*(N) x 100*(H), and 8 scans (3D ^{1}H-^{15}N TOCSY-HSQC). The data at DRX 800 were acquired with 1024*(HN) x 30*(N) x 35*(C), 32 scans ($^{3h}J_{HNCO}$ HNCO); 1024*(HN) x 32*(N) x 135*(H), 8 scans (3D ^{1}H-^{15}N NOESY-HSQC); 1024*(H) x 50*(C) x

128*(H), and 8 scans (3D ^{1}H-^{13}C NOESY-HSQC). All data were processed with the program NMRPipe (33).

2.3 Resonance Assignments

Backbone sequential assignments (HN, N, Cα, and Cβ) were done mainly using CBCA(CO)NH and CBCANH spectra. First, according to (^{1}HN, ^{15}N) chemical shift of ^{1}H-^{15}N HSQC spectrum, the carbon chemical shifts of inter-residue (Cα(i-1) and Cβ(i-1) from CBCA(CO)NH) and intra-residue (Cα(i) and Cβ(i) from CBCANH) were clustered. Then, in each cluster (i), based on the intra-residue chemical shifts (Cα(i) and Cβ(i)), another cluster (j) which had the same inter-residue chemical shifts (Cα(i-1) and Cβ(i-1)) was searched. All the clusters could be connected into bi-cluster (i-j) sets and thus be assembled entirely. Side-chain assignments (C and H) were acquired mainly using ^{1}H-^{13}C HSQC, C(CO)NH, H(CCO)NH, and HCCH-TOCSY spectra. The C(CO)NH and H(CCO)NH spectra gave the information of side-chain carbons and protons chemical shifts from ^{1}HN and ^{15}N signals of backbone which had already been assigned. This information was combined with the spectra of ^{1}H-^{13}C HSQC and HCCH-TOCSY for the identification of proton and carbon pair. All NMR spectra in this study were analyzed with the program NMRView (34) and in-house programs written

by Tcl/Tk were used to help assignments.

2.4 Structure Calculations

Initially, structure calculation and NOE peak assignment were performed in an iterative and manual manner using the program DYANA (version 1.5) (35). NOE cross-peak intensities were classified as strong, medium or weak, and assigned to restraints of 1.8-3.0, 1.8-4.0 or 1.8-5.0 Å, respectively. On the base of $^{3h}J_{HNCO}$ HNCO spectrum, hydrogen-bonds restraints were applied as 2.5-3.3 Å for N-O pairs and 1.8-2.5 Å for HN-O pairs and used from the initial procedure. Backbone torsion angle restraints were derived from $^{3}J_{HNH\alpha}$ of HMQC-J (31) and TALOS program (36). The backbone ϕ angle restraints used were $-65 \pm 25°$ for $^{3}J_{HNH\alpha} < 4.7$ Hz, $-120 \pm 40°$ for $^{3}J_{HNH\alpha} > 8.5$ Hz and < 9.9 Hz, and $-120 \pm 20°$ for $^{3}J_{HNH\alpha} > 10.0$ Hz. If the predicted TALOS ϕ angle was within the range of $^{3}J_{HNH\alpha}$, the TALOS derived φ angle was used for restraints. The torsion angles $\chi 1$ of Tyr, Phe and Trp were estimated from $^{3}J_{C'C\gamma}$ and $^{3}J_{NC\gamma}$ coupling constants (37). Final refinement employing ambiguous NOE assignments and floating chirality assignments was performed using the program ARIA (version 1.0) and a CNS package as described (38). A total of 150 structures were refined in the last (ninth) iteration, and the quality of 30 lowest energy structures was analyzed using

AQUA and PROCHECK-NMR software (39).

2.5 Sequence Alignments, Homology Modeling, and Structure Comparison

Homologous bacterial sequences of $^{2}FnIIID_{ChiA1}$ were obtained from SMART (smart.embl-heidelberg.de) server (40). In the SMART server, 135 bacterial FnIIIDs from 102 proteins were registered. Among these, I 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 FnIIID sequence in the case of a protein with multiple FnIIIDs. Species redundancies were also considered. For comparison, three FnIIIDs from *Bacillus circulans* were added. Finally 21 FnIIID sequences from 20 carbohydrases from 18 bacterial genera were used in multiple sequence alignments using ClustalW with default parameters (41). CHROMA software was used for analyzing and annotating the results from sequence alignments (42).

The structure of $^{1}FnIIID_{ChiA1}$ was modeled using the $^{2}FnIIID_{ChiA1}$ structure as a template. $^{1}FnIIID_{ChiA1}$ shows 74.4 % amino acid sequence identity with $^{2}FnIIID_{ChiA1}$ and there are no gaps in the sequence alignment. The MODELLER program (43) was used for modeling and the quality of the model was assessed using PROCHECK.

Structure comparisons with other structures of FnIIID-like domain were done using DALI (44) and MOLMOL (45) software.

2.6 Theory of Relaxation Experiment

The magnetic relaxation of nuclei is stimulated by magnetic fields oscillating at the relevant transition frequencies. The relaxation parameters, longitudinal relaxation rate (R_1), transverse relaxation rate (R_2), and $\{^1H\}$-^{15}N NOE (NOE), can be written on the basis of the influences of the appropriate frequencies. For example, the relaxation of a protonated ^{15}N nucleus (e.g., backbone amide in the case of protein) is influenced by HN group motions at the resonant frequencies of the nitrogen (ω_N) and the proton (ω_H), the sum and difference of these frequencies ($\omega_H+\omega_N$, $\omega_H-\omega_N$), and a frequency of zero (corresponding to loss of phase coherence). In the case of ^{15}N relaxation, the two predominant relaxation mechanisms are dipolar-dipolar interaction (DD) between ^{15}N and ^1H, and chemical shift anisotropy (CSA) of ^{15}N. As a consequence of molecular motion, the DD and CSA create oscillating local magnetic fields, and which cause relaxation. Thus relaxation parameters (R_1, R_2, and NOE) which are expressed using DD and CSA termss are linear combinations of spectral density functions of appropriate frequencies as follows (46):

$$R_1 = \frac{d^2}{4}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N)$$

$$R_2 = \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)]$$
$$+ \frac{c^2}{6}[4J(0) + 3J(\omega_N)] + R_{ex}$$

$$NOE = 1 + \frac{d^2}{4R_1}\frac{\gamma_H}{\gamma_N}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]$$

in which $d = \frac{\mu_0 h \gamma_N \gamma_H}{8\pi^2}\left\langle r_{NH}^{-3} \right\rangle$, $c = \frac{\omega_N \Delta\sigma}{\sqrt{3}}$, μ_0 is the permeability of free space; h is Planck's constant; γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N, respectively.

Here the spectral density function, $J(\omega)$, is defined by the Fourier transformation of auto-correlation function, $G(t)$.

$$J(\omega) = \int_{-\infty}^{\infty} G(t)\exp(-\omega\tau)dt$$

To solve these relaxation equations, besides five $J(\omega)$ values, the values of CSA ($\Delta\sigma$), r_{NH}, and R_{ex} must be determined too. To reduce the number of unknowns, the N-H bond length and ^{15}N CSA are usually assumed to be constant with r_{NH}=1.02Å and $\Delta\sigma = -170$ ppm.

The remaining unknowns could be further reduced using Lipari-Szabo "model-free" parameterization of $J(\omega)$ (47, 48). In "model-free", the spectral density function, $J(\omega)$, is defined in terms of (i) the global rotational diffusion of the macromolecule (molecular correlation time, τ_c, typically a few nanoseconds, and possibly the diffusion anisotropy); (ii) the time scale of internal motions faster than the

global tumbling (effective internal correlation time, τ_e, picosecond- nanosecond time scale); and (iii) the degree of restriction (related to amplitude) of these fast internal motions (the square of the order parameter, S^2). These parameters, which can be extracted from the analyses of NMR relaxation experiments, reflect the motional property of each residue.

The spectral density function of model-free treatment varies according to the overall diffusion model of the biomolecule. In the case of isotropic global tumbling, $J(\omega)$ is written as

$$J(\omega) = S^2 \frac{\tau_c}{1+(\omega\tau_c)^2} + (1-S^2)\frac{\tau'}{1+(\omega\tau')^2}$$

where $\frac{1}{\tau'} = \frac{1}{\tau_c} + \frac{1}{\tau_e}$, τ_c is the overall correlation time, τ_e is the effective correlation time for the internal motion, and S^2 is the generalized order parameter. For axially symmetric anisotropic global motion, $J(\omega)$ will be extended as (49-51)

$$J(\omega) = S^2\left[\frac{A_1\tau_1}{1+(\omega\tau_1)^2} + \frac{A_2\tau_2}{1+(\omega\tau_2)^2} + \frac{A_3\tau_3}{1+(\omega\tau_3)^2}\right] + (1-S^2)\frac{\tau'}{1+(\omega\tau')^2}$$

where $A_1 = \frac{(3\cos^2\theta - 1)}{4}$; $\tau_1 = \frac{1}{6D_\perp}$; $A_2 = 3\sin^2\theta\cos^2\theta$; $\tau_2 = \frac{1}{5D_\perp + D_\parallel}$;

$A_3 = \frac{3}{4}\sin^4\theta$; $\tau_3 = \frac{1}{2D_\perp + 4D_\parallel}$; $\frac{1}{\tau'} = 6D + \frac{1}{\tau_e}$; $D = \frac{1}{3}(2D + D_\parallel)$

With spectral density function mentioned above, five model-free models are considered. These included model 1 with S^2 describing the amplitude of spatial

restriction for a backbone amide N-NH vector which can vary from 0 (no spatial restriction) to 1 (complete spatial restriction), and τ_c. Model 2 is with S^2, τ_s, and a correlation time for picosecond time-scale internal motions (τ_e). Model 1 assumes that $\tau_e \rightarrow 0$ and does not contribute to relaxation, whereas model 2 assumes that internal motions are within the 10^{-10}~10^{-12} s time-scale ($0 < \tau_e < \tau_c$) and contribute to relaxation. Model 3 is a modified version of mode 1 to include a parameter describing the contribution of microsecond to millisecond time-scale internal motions to R_2 (R_{ex}, in s^{-1}). Model 4 is a modification of model 2 to include the parameter R_{ex}. Model 5 accounts for internal motions occurring on two time-scales. This model consists of an order parameter for fast picosecond internal motions (S_f^2), an order parameter for nanosecond time-scale internal motions faster than τ_c but slower than τ_c (S_s^2), a correlation time for picosecond internal motions, assumed not to contribute to relaxation ($\tau_e \rightarrow 0$), and a correlation time for nanosecond time-scale internal motions (τ_s) (52).

Model	1	2	3	4	5
Residue specific Parameters	S^2	S^2, τ_e, R_{ex}	S^2, R_{ex}	S^2, τ_e, R_{ex}	S_f^2, S^2, τ_e

The best fit model for the internal motion of residue *i* is determined by simulating the

target function $\chi_i^2 = (\frac{R_{1i}^{obs} - R_{1i}^{calc}}{\sigma_{R_{1i}}})^2 + (\frac{R_{2i}^{obs} - R_{2i}^{calc}}{\sigma_{R_{2i}}})^2 + (\frac{NOE_i^{obs} - NOE_i^{calc}}{\sigma_{NOE_i}})^2$ by means of Monte-Carlo simulations and F-statistical tests as described (52).

2.7 Measurement and Analysis of Relaxation Data

Relaxation experiments were measured using Bruker DMX 500 MHz of ^{1}H frequency with ^{15}N labeled ^{2}FnIIID$_{ChiA1}$ of 0.4 mM at 310 K. R_1 and R_2 measurements were carried out using the pulse sequences reported by Kay et al (53). The R_1 data were obtained with a series of delays: 11.865, 121.865, 251.865 (×2), 451.865, 701.865 (×2), 971.865, and 1301.865 ms. The R_2 data were acquired with a series of delays: 8 (×2), 24, 48, 80 (×2), 128, 192, and 288 ms. All the data were gained with 1024* × 100* complex matrix points and 32 scans. The CPMG sequence was applied with a pulse field strength of 3.4 KHz for ^{15}N. For the {^{1}H}-^{15}N NOE measurement, two 2D spectra were acquired in an interleaved manner. The two data sets of NOE experiment were acquired with 512* × 60* complex points and 64 scans. All the data sets were processed with 45°-shifted sine-square apodization in both dimension using NMRPipe (33). Peak intensities for analysis were extracted using peak picking method of NMRPipe. Uncertainties in peak intensity of R_1 and R_2 datasets were gained from the spectra with duplicated R_1 and R_2 time intervals. Uncertainty in peak intensity of NOE was gained

performing the same experiment twice. Relaxation rates were estimated by fitting the measured peak intensities to a single exponential decay curve, $I_k^{calc} = I_0 \exp(-t_k R_1, R_2)$, using a two-parameter fit of CURVEFIT software (52). For uncertainty in the fitting, Monte-Carlo simulation was employed (54). The model-free analyses for main chain mobility were done by the TENSOR software (55). The ModelFree4.0 software (52) was used for the motional analysis of Trp-580's Nε1-Hε1 vector. In this case, –88.5 ppm was used as the chemical shift anisotropy of Nε1 (56).

3. RESULTS

3.1 Resonance Assignment, Restraints for Structure Calculation, and Structure Determination

Except for the signals from the vector-derived N-terminal His and Met residues, which were not observed in ^{1}H-^{15}N HSQC, all the backbone signals were assigned using CBCA(CO)NH and CBCANH (Figs. 3-1 to 3). Side-chain assignment was achieved using mainly CCONH, HCCONH, and HCCH-TOCSY spectra. Aromatic side-chain chemical shifts of tyrosine, phenylalanine, and tryptophan were assigned from 2D NOESY and TOCSY spectra. A nearly complete assignment of $^{2}FnIIID_{ChiA1}$ was achieved (Table 3-1). This result has been deposited in the BioMagResBank under accession number 5178. To extract distance restraints, NOESY spectra with 100 and 150 ms mixing times were used. No severe spin diffusion effects were found in the spectrum with a 150 ms mixing time. I measured $^{3h}J_{HNCO}$ HNCO spectrum to get hydrogen-bond information directly. By comparing HNCO and $^{3h}J_{HNCO}$ HNCO spectra, 23 inter-hydrogen bonds between the β-strands (Figs. 3-4 and 3-5) were identified. 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 structures gave the global fold. When the calculated structure converged 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 (38).

The structure of the ^{2}FnIIID$_{ChiA1}$ was determined from the distance and torsional angle restraints listed in Table 3-2. An overlay of the final 30 structures shows that the backbone and hydrophobic side chains have been well defined (Figs. 3-6 and 3-7). The average root mean square deviation (RMSD) of these structures (Ala-559 to Thr-644) is 0.431 and 0.714 Å for the backbone and all the heavy-atoms, respectively. There is no distance restraint violation of > 0.3 Å, and no dihedral angle restraint violation of > 5 Å. The detailed statistics for the structures are shown in Table 3-2. The Ramachandran plot of backbone torsion angles are also drawn (Fig. 3-8).

3.2 Sequence Alignments of Bacterial FnIIIDs

Of the 61 FnIIID-containing bacterial carbohydrases which were initially retrieved from SMART database, 21 are from *Streptomyces*, and 16 are from *Bacillus* genera. For a comparison of FnIIIDs across genera, I selected only one FnIIID from

each genus. As the sequences of FnIIIDs from the same enzyme tend to be more similar to each other than to FnIIIDs from other proteins, only one FnIIID sequence per enzyme was selected. If all the bacterial FnIIID sequences were used for alignment, then the apparent similarity in alignment and the proportion of FnIIIDs from Gram-positive bacteria increased. All the information concerning acronyms, enzymatic function, organism name, Gram-sensitivity and order / total number are written in Fig. 3-9 and Table 3-3.

FnIIID from *Yersinia enterocolitica* (O68975) exhibited the lowest sequence identity (32.6%) to $^{2}FnIIID_{ChiA1}$. The lengths of aligned sequences are nearly the same except for those of *Erwinia chrysanthemi* (PEHX_ERWCH) and O68975, which contain ~20 more amino acids in their C'E loops in comparison with the others. A comparison of bacterial FnIIID sequences with animal FnIIID sequences indicates that the C-terminal 20 residues are more conserved in the bacterial sequences.

3.3 Structure Description

The structure of $^{2}FnIIID_{ChiA1}$ is a canonical β-sandwich structure containing two antiparallel β-sheets that are packed face to face (Fig. 3-6 B). One sheet is composed of three β-strands (A, B and E), whereas the other is composed of four

β-strands (C, C′, F and G). Strand G is divided into two sections, G1 and G2, as in animal FnIIIDs. Three loops in the direction of N-terminus (loops BC, C′E and FG) and three in the direction of the C-terminus (loops AB, CC′ and EF) connect the seven β-strands, respectively (Fig. 3-6 B).

The β-sandwich scaffold of ^{2}FnIIID$_{ChiA1}$ is stabilized by an extensive hydrogen-bond network between the β-strands, and by a hydrophobic core formed by inward-facing residues from the β-sheets. It is well known that animal FnIIIDs contain several β-bulge structures on the edge strands A, C′ and G. In ^{2}FnIIID$_{ChiA1}$, three β-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. 3-5, demonstrating its utility.

The residues that are conserved in bacterial FnIIIDs and form the hydrophobic core are well defined in the final structures (Fig. 3-7). 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 FnIIID fold (Figs. 3-7 and 3-14). 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 highly conserved across bacterial and animal FnIIIDs (Figs. 3-7 and 3-9).

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 demonstrated by an artificially engineered FnIIID with altered sequences on the BC and FG loops that binds ubiquitin with high affinity (57). 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. 3-9). The BC loop is stabilized by a hydrogen bond between the main-chain HN group of Ser-583 on the loop and Oη of Tyr-592. In 28 out of 30 final structures, this HN group and the Oη are located within 2.35 Å. This bond is also found in some animal FnIIIDs.

In addition to the loop regions, some exposed residues in the β-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 $^{2}FnIIID_{ChiA1}$ is mostly surrounded by non-charged residues as shown in Fig. 3-10. $^{2}FnIIID_{ChiA1}$ 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 $^{1}FnIIID_{ChiA1}$–$^{2}FnIIID_{ChiA1}$ module, the structure of $^{1}FnIIID_{ChiA1}$ was modeled (Fig. 3-11). As there are no gaps in the sequence alignment and a high sequence identity (74.4%) between $^{2}FnIIID_{ChiA1}$ and $^{1}FnIIID_{ChiA1}$, homology modeling of the $^{1}FnIIID_{ChiA1}$ structure was quite straightforward. The RMSD between the modeled $^{1}FnIIID_{ChiA1}$ and $^{2}FnIIID_{ChiA1}$ is 1.173 Å over the Cα coordinates of Ala-559 to Thr-644 in $^{2}FnIIID_{ChiA1}$. Contacts between the hydrophobic core-forming residues are nearly identical between the two FnIIIDs (Fig. 3-11B). 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. 3-11C. The modeled structure of $^{1}FnIIID_{ChiA1}$ reveals that charged groups are sparsely distributed on the protein surface with no noticeable charged patches, as seen for $^{2}FnIIID_{ChiA1}$. Modeling of the $^{1}FnIIID_{ChiA1}$–$^{2}FnIIID_{ChiA1}$ module using the coordinates of $^{1}FnIIID_{ChiA1}$ and $^{2}FnIIID_{ChiA1}$ 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).

3.4 Low Sequence Complexity

The noteworthy feature of $^{2}FnIIID_{ChiA1}$ 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 proteins with low sequence complexity are more likely to generate relatively extended structures (58), $^{2}FnIIID_{ChiA1}$ adopts a globular and compact fold. The

molecular weight of ^{2}FnIIID$_{ChiA}$ (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 (59). This high content of small side chains probably caused the relatively small difference between backbone and heavy atom RMSDs 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 ^{2}FnIIID$_{ChiA}$ (Fig. 3-12), and thus have little effect on its global fold. Most threonines are not highly conserved in bacterial FnIIIDs; instead, small residues (A, C, S, T, D, N, V, G, and P) occur with a probability greater than 80% at the positions corresponding to threonines in ^{2}FnIIID$_{ChiA}$ (Fig. 3-9). These residues presumably contribute to the solvent-accessible surface of each molecule, and thus may be important functionally.

3.5 Structural Comparison

The program DALI was used to search protein databases for structures similar to that of ^{2}FnIIID$_{ChiA1}$ (44). It retrieved 131 structures with Z scores greater than 2.0 from the Protein Data Bank. All of the structures 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 ^{2}FnIIID$_{ChiA1}$ with Z scores larger than 10.0 and a Cα atom RMSD smaller than 2.0 Å: integrin β-4 subunit fragment (PDB code: 1QG3, Z score: 11.9, RMSD: 1.8 (over 83 residues)) (8), tenascin fragment (1QR4, 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 (1BPV, 10.6, 1.9 (84)) (10) (Fig. 3-13). Despite the high similarity of global fold, the surface electrostatic potentials and loop conformations of these proteins are different from those of ^{2}FnIIID$_{ChiA1}$, and also differ between one another. In particular, the loop structures of ^{2}FnIIID$_{ChiA}$ are unique. The ^{2}FnIIID$_{ChiA1}$ loops, except for the BC loop, are very short in comparison with those of other FnIIIDs. The conformation of the BC loop seems to be unique, as no other FnIIID derived by DALI exhibits a similar conformation (Fig. 3-14). Under "Discussion", I compare these structures with ^{2}FnIIID$_{ChiA1}$ from an evolutionary point of view.

3.6 Relaxation Data of ^{2}FnIIID$_{ChiA1}$

Except overlapped (Ala-582, Asn-586, Val-587, Val-594, Leu-614, Lys-627, Ser-641, and Lys-643) and proline (Pro-560 and Pro–563) residues, R_1, R_2, and *NOE*

values of 76 residues were acquired. Estimated R_1, R_2, and *NOE* values are drawn in Figs. 3-15 and 3-16. The mean values (10% trimmed) are 2.43 (± 0.11) s^{-1} (R_1), 5.50 (± 0.32) s^{-1} (R_2), and 0.65 (±0.04) ({^{1}H}-^{15}N *NOE*) (Table 3-4, Figs. 3-15 and 3-16).

To get appropriate motional model (isotropic, axially symmetric, or anisostropic diffusion), the structure which was determined in this study and R_2/R_1 relaxation parameters were used. The residues for which R_2/R_1 ratios may be affected by fast internal motion or R_{ex} contribution were excluded (50, 51). The excluded residues were those with:

(i) {^{1}H}-^{15}N NOE < 0.6, residues with fast internal mobility

(ii) $[\frac{< R_2 > - R_{2i}}{< R_2 >} - \frac{< R_1 > - R_{1i}}{< R_1 >}] > 1.5 \times SD$, residues with significant R_{ex}

As a result, among the initial 76 datasets, 59 datasets were selected. The 500 cycles Monte-Carlo simulation of χ^2 and F-test selected axially symmetric diffusion of prolate as appropriate rotational model. Isotropic diffusion was rejected for its larger value of $\chi^2_{exp} = 78.2$ than that of simulation ($\chi^2_{0.05} = 76$) and an anisotropic diffusion was not selected as there was no statistically significant improvement in F-test (F_{exp}=0.62 < $F_{0.1}$=4.0). In the case of axially symmetric diffusion, two models of prolate and oblate which have two orthogonal minima, can be exist. Oblate diffusion was also rejected for its larger χ^2 value (χ^2_{exp}=72.4 > $\chi^2_{0.05}$=71.2) (Fig. 3-17A).

The asymmetric rotational diffusion ratio ($D_{\parallel}/D_{\perp}$) of the prolate tensor is 1.33 ± 0.05. The relative ratio of the principle components of the inertia tensor which is calculated from the structure is 1.00:0.98:0.36. The $D_{//}$ ($D_{\parallel}$) component of the prolate tensor of ^{2}FnIIID$_{ChiA1}$ makes an angle of 10° with $I_{//}$ component of the inertia tensors (Fig. 3-17B). The overall rotational correlation time ($\tau_c = \frac{1}{6D_{iso}} = \frac{2}{D_{\parallel} + 2D_{\perp}}$) is 4.1 ns and it is very similar to 4.12 ns of ubiquitin at 300 K (50). The molecular weights of ^{2}FnIIID$_{ChiA1}$ (including His-557 and Met-558) and ubiquitin are 8.8 and 8.5 kDa respectively and thus similar each other. It can imply that the ^{2}FnIIID$_{ChiA1}$ has larger hydrodynamic radius than ubiquitin, considering the difference of the temperatures at which the data were gained. The relative moment of the inertia tensor in ubiquitin is 1:0.90:0.64.

Next, assuming asymmetric rotational diffusion tensor, model-free analyses of residue-specific internal mobility was done (Table 3-5 and Fig. 3-18). Of the 76 residue-specific datasets subjected to the analysis, 57 were best fit by model 1, 4 by model 2, 6 by model 3, 8 by model 4, and 1 by model 1.

The mean value of the generalized order parameters is 0.80 (± 0.06). Regions of secondary structure show higher order parameters than those of loop regions (Fig. 3-18A) and 1-S^2 are in general agreement with the residue specific RMSD of the final

30 ensemble structures (Fig. 3-19). As the value of 1-S^2 represents the degree to which orientational information about the bond vector is lost due to local internal motion, rather than due to the roational diffusion of the whole molecule (60), it can mean that the relative disorder of the overlaid structures in the BC and FG loops is the result of their flexibility.

Several residues that required τ_e or R_{ex} terms to describe their relaxation parameters were found. However, there was no residues which revealed statistically meaningful values of τ_e (greater than 100 ps) or R_{ex} (greater than 1 Hz) except N-terminal residue Ala-559 (Fig. 3-18).

Table 3-1. Chemial shift assignments of $^{2}FnIIID_{ChiA1}$

Name	Number	Assigmnets
MET	558	174.576(C) 55.478(CA) 33.442(CB) 31.989(CG) 2.013(HB2) 2.553(HG2)
ALA	559	124.849(N) 50.507(CA) 18.670(CB) 8.146(HN) 4.403(HA) 1.312(HB)
PRO	560	175.861(C) 62.377(CA) 31.893(CB) 50.275(CD) 27.340(CG) 4.635(HA) 2.208/1.742(HB2,HB3) 3.635/3.152(HD2,HD3) 1.800/1.711(HG2,HG3)
THR	561	114.485(N) 173.310(C) 62.229(CA) 69.644(CB) 8.500(HN) 4.219(HA) 1.409(HG2)
ALA	562	127.059(N) 51.029(CA) 18.360(CB) 8.523(HN) 4.695(HA) 1.293(HB)
PRO	563	175.240(C) 62.729(CA) 32.949(CB) 50.456(CD) 27.800(CG) 4.736(HA) 2.483/1.665(HB2,HB3) 4.211/3.623(HD2,HD3) 1.965/1.679(HG2,HG3)
THR	564	107.252(N) 173.346(C) 59.273(CA) 72.948(CB) 22.194(CG2) 7.871(HN) 4.822(HA) 4.246(HB) 1.255(HG2)
ASN	565	114.986(N) 174.520(C) 54.089(CA) 37.500(CB) 8.949(HN) 4.249(HA) 3.000/2.902(HB2,HB3)
LEU	566	121.118(N) 177.092(C) 56.693(CA) 40.767(CB) 27.120(CD1) 27.300(CD2) 27.280(CG) 9.214(HN) 5.023(HA) 1.931/1.098(HB2,HB3) 1.186/1.006(HD1,HD2) 1.511(HG)
ALA	567	128.645(N) 175.864(C) 51.489(CA) 22.176(CB) 9.306(HN) 4.665(HA) 1.362(HB)
SER	568	113.350(N) 175.662(C) 56.328(CA) 64.252(CB) 8.720(HN) 5.413(HA) 3.716/3.716(HB2,HB3)
THR	569	117.446(N) 174.414(C) 61.658(CA) 69.091(CB) 22.376(CG2) 8.628(HN) 4.439(HA) 4.233(HB) 1.080(HG2)
ALA	570	123.301(N) 175.379(C) 52.822(CA) 20.773(CB) 7.835(HN) 4.489(HA) 1.375(HB)
GLN	571	117.530(N) 175.102(C) 56.118(CA) 31.993(CB) 33.682(CG) 8.364(HN) 5.070(HA) 2.211/2.211(HB2,HB3) 2.371/2.371(HG2,HG3)
THR	572	115.061(N) 174.936(C) 60.124(CA) 71.248(CB) 20.616(CG2) 8.904(HN) 4.903(HA) 4.740(HB) 1.206(HG2)
THR	573	109.743(N) 174.721(C) 64.813(CA) 69.302(CB) 19.287(CG2) 8.186(HN) 4.343(HA) 4.454(HB) 1.441(HG2)
SER	574	109.889(N) 173.337(C) 57.310(CA) 65.655(CB) 7.952(HN) 4.721(HA) 4.227/3.779(HB2,HB3)
SER	575	117.005(N) 171.465(C) 57.027(CA) 67.446(CB) 7.602(HN) 5.579(HA) 3.668/3.668(HB2,HB3)
ILE	576	118.912(N) 173.615(C) 59.484(CA) 42.793(CB) 14.863(CD1) 28.234(CG1) 17.661(CG2) 8.543(HN) 4.742(HA) 1.495(HB) 0.800(HD1) 1.027/1.027(HG12,HG13) 0.660(HG2)
THR	577	123.202(N) 173.246(C) 62.219(CA) 70.424(CB) 21.505(CG2) 8.789(HN) 5.119(HA) 3.847(HB) 1.031(HG2)
LEU	578	128.196(N) 175.857(C) 52.892(CA) 46.230(CB) 22.519(CD1) 26.027(CD2) 26.794(CG) 9.284(HN) 5.451(HA) 1.790/1.712(HB2,HB3) 0.549/0.340(HD1,HD2) 1.305(HG)
SER	579	114.858(N) 172.401(C) 57.161(CA) 66.651(CB) 9.063(HN) 5.045(HA) 3.664(HB2) 3.608(HB3)
TRP	580	119.703(N) 173.847(C) 56.679(CA) 30.521(CB) 7.717(HN) 5.088(HA) 3.067/3.340(HB2,HB3) 6.415(HD1) 5.802(HE1) 6.879(HE3) 6.471(HH2) 7.052(HZ2) 6.444(HZ3) 122.890(NE1) 119.703(N)
THR	581	115.990(N) 173.999(C) 62.373(CA) 69.703(CB) 21.995(CG2) 9.046(HN) 4.413(HA) 4.224(HB) 1.322(HG2)
ALA	582	127.402(N) 178.727(C) 52.315(CA) 19.877(CB) 8.562(HN) 4.336(HA) 1.513(HB)
SER	583	116.837(N) 174.916(C) 59.764(CA) 64.323(CB) 8.366(HN) 4.704(HA) 3.845/3.906(HB2,HB3)
THR	584	116.414(N) 174.471(C) 62.710(CA) 69.792(CB) 21.486(CG2) 8.291(HN) 4.337(HA) 1.318(HG2)
ASP	585	120.860(N) 175.677(C) 54.294(CA) 41.461(CB) 8.212(HN) 4.715(HA) 2.747/2.685(HB2,HB3)
ASN	586	118.863(N) 175.817(C) 54.416(CA) 38.793(CB) 8.209(HN) 4.654(HA) 2.912/2.912(HB2,HB3)
VAL	587	119.490(N) 176.676(C) 63.382(CA) 32.736(CB) 20.905(CG1) 8.161(HN) 4.158(HA) 2.165(HB) 0.993(HG1)
GLY	588	110.701(N) 173.848(C) 45.669(CA) 8.571(HN) 4.050/3.972(HA2,HA3)
VAL	589	122.168(N) 177.360(C) 63.271(CA) 32.555(CB) 21.344(CG1) 23.006(CG2) 7.910(HN) 4.187(HA) 2.086(HB) 0.842/0.940(HG1,HG2)
THR	590	119.122(N) 175.106(C) 61.698(CA) 68.807(CB) 22.245(CG2) 8.965(HN) 4.449(HA) 4.231(HB) 1.123(HG2)
GLY	591	109.304(N) 169.200(C) 45.248(CA) 7.382(HN) 4.082/4.082(HA2,HA3)
TYR	592	115.420(N) 174.394(C) 56.782(CA) 41.873(CB) 9.188(HN) 5.360(HA) 2.891/2.575(HB2,HB3) 7.060(HD1) 6.905(HE1)
ASP	593	123.641(N) 174.429(C) 53.055(CA) 43.613(CB) 9.056(HN) 5.149(HA) 2.504/2.231(HB2,HB3)
VAL	594	122.866(N) 174.754(C) 61.133(CA) 32.568(CB) 21.083(CG1) 20.600(CG2) 8.733(HN) 4.451(HA) 1.928(HB) 0.787/0.672(HG1,HG2)
TYR	595	128.126(N) 174.812(C) 58.291(CA) 40.059(CB) 9.631(HN) 4.867(HA) 2.971/2.780(HB2,HB3) 6.704(HD1) 6.670(HE1)
ASN	596	119.635(N) 176.782(C) 50.769(CA) 39.140(CB) 8.872(HN) 4.781(HA) 3.140/2.425(HB2,HB3)
GLY	597	114.858(N) 175.950(C) 47.146(CA) 9.063(HN) 4.230/3.623(HA2,HA3)
THR	598	118.042(N) 174.070(C) 61.868(CA) 69.091(CB) 21.594(CG2) 8.817(HN) 4.326(HA) 4.560(HB) 1.220(HG2)
ALA	599	126.475(N) 176.256(C) 51.209(CA) 20.984(CB) 8.114(HN) 4.681(HA) 1.515(HB)
LEU	600	125.934(N) 176.152(C) 56.258(CA) 41.460(CB) 23.649(CD1) 25.417(CD2) 27.458(CG) 8.593(HN) 3.436(HA) 1.685/1.180(HB2,HB3) 0.432/0.652(HD1,HD2) 1.034(HG)

(Table 3-1. continued)

Name	Numer	Assignments
ALA	601	132.833(N) 177.618(C) 53.453(CA) 20.358(CB) 9.680(HN) 4.560(HA) 1.430(HB)
THR	602	108.912(N) 170.130(C) 60.886(CA) 67.548(CB) 18.683(CG2) 7.610(HN) 4.534(HA) 4.409(HB) 0.970(HG2)
THR	603	121.108(N) 174.223(C) 61.447(CA) 71.826(CB) 20.625(CG2) 7.822(HN) 5.730(HA) 3.799(HB) 1.240(HG2)
VAL	604	118.484(N) 175.912(C) 59.169(CA) 36.277(CB) 18.717(CG1) 22.832(CG2) 9.277(HN) 5.226(HA) 2.744(HB) 1.334/1.221(HG1.HG2)
THR	605	109.736(N) 175.717(C) 62.569(CA) 69.652(CB) 22.070(CG2) 8.648(HN) 4.818(HA) 4.517(HB) 1.435(HG2)
GLY	606	111.282(N) 171.098(C) 44.196(CA) 7.649(HN) 4.582/3.885(HA2.HA3)
THR	607	102.715(N) 172.141(C) 59.203(CA) 66.637(CB) 20.934(CG2) 7.344(HN) 3.273(HA) 4.003(HB) 0.801(HG2)
THR	608	109.005(N) 172.413(C) 59.092(CA) 71.667(CB) 19.998(CG2) 6.445(HN) 4.562(HA) 3.784(HB) 0.886(HG2)
ALA	609	123.654(N) 175.454(C) 51.840(CA) 23.438(CB) 8.521(HN) 4.746(HA) 1.445(HB)
THR	610	118.838(N) 173.100(C) 62.289(CA) 70.424(CB) 21.109(CG2) 8.829(HN) 5.103(HA) 3.877(HB) 1.000(HG2)
ILE	611	130.374(N) 174.514(C) 59.273(CA) 37.323(CB) 12.025(CD1) 27.442(CG1) 18.550(CG2) 9.107(HN) 4.290(HA) 1.777(HB) 0.559(HD1) 0.948(HG1.HG2)
SER	612	121.663(N) 173.713(C) 56.889(CA) 65.374(CB) 8.481(HN) 5.104(HA) 3.849/3.805(HB2.HB3)
GLY	613	108.065(N) 175.538(C) 46.230(CA) 8.491(HN) 4.019/3.863(HA2.HA3)
LEU	614	119.490(N) 176.145(C) 54.327(CA) 41.278(CB) 25.860(CD1) 20.262(CD2) 8.161(HN) 4.191(HA) 1.426/1.248(HB2.HB3) 0.388/0.172(HD1.HD2)
ALA	615	125.070(N) 177.626(C) 51.419(CA) 20.142(CB) 8.273(HN) 4.492(HA) 1.508(HB)
ALA	616	123.577(N) 178.504(C) 52.108(CA) 20.328(CB) 8.309(HN) 4.501(HA) 1.478(HB)
ASP	617	121.632(N) 175.095(C) 55.747(CA) 39.950(CB) 8.439(HN) 4.469(HA) 2.956/2.523(HB2.HB3)
THR	618	114.240(N) 171.381(C) 62.219(CA) 71.896(CB) 19.540(CG2) 8.111(HN) 4.426(HA) 3.673(HB) 0.647(HG2)
SER	619	119.086(N) 173.059(C) 57.801(CA) 64.533(CB) 8.045(HN) 5.162(HA) 3.714/3.617(HB2.HB3)
TYR	620	124.399(N) 175.193(C) 57.679(CA) 43.861(CB) 9.016(HN) 4.692(HA) 2.440/2.440(HB2.HB3) 7.147(HD1) 6.944(HE1)
THR	621	117.594(N) 173.359(C) 61.587(CA) 71.195(CB) 21.976(CG2) 8.558(HN) 5.193(HA) 3.844(HB) 1.109(HG2)
PHE	622	125.936(N) 175.246(C) 56.658(CA) 43.113(CB) 9.293(HN) 6.072(HA) 2.946/2.683(HB2.HB3) 7.142(HD1) 6.955(HE1) 6.670(HZ)
THR	623	108.107(N) 172.492(C) 60.465(CA) 74.842(CB) 23.617(CG2) 8.399(HN) 4.323(HA) 3.918(HB) 1.122(HG2)
VAL	624	117.158(N) 174.621(C) 59.984(CA) 35.413(CB) 21.634(CG1) 8.356(HN) 5.339(HA) 1.061(HB) -0.030/-0.030(HG1.HG2)
LYS	625	122.245(N) 175.188(C) 55.337(CA) 37.032(CB) 29.300(CD) 42.555(CE) 26.511(CG) 8.893(HN) 4.560(HA) 2.011/1.659(HB2.HB3) 1.858/1.519(HD2.HD3) 2.573/2.500(HE2.HE3) 1.228/1.099(HG2.HG3)
ALA	626	124.580(N) 175.677(C) 50.943(CA) 22.492(CB) 8.949(HN) 4.746(HA) 1.466(HB)
LYS	627	118.863(N) 175.107(C) 54.281(CA) 36.046(CB) 29.776(CD) 41.469(CE) 23.300(CG) 8.209(HN) 5.480(HA) 1.591/1.503(HB2.HB3) 2.164/2.048(HD2.HD3) 2.804/2.665(HE2.HE3) 1.300/1.195(HG2.HG3)
ASP	628	120.425(N) 178.313(C) 51.770(CA) 43.284(CB) 8.084(HN) 5.286(HA) 3.345/2.429(HB2.HB3)
ALA	629	120.174(N) 178.327(C) 54.224(CA) 18.389(CB) 8.370(HN) 4.132(HA) 1.447(HB)
ALA	630	120.231(N) 177.558(C) 51.736(CA) 19.151(CB) 7.900(HN) 4.437(HA) 1.417(HB)
GLY	631	106.794(N) 174.885(C) 45.388(CA) 8.015(HN) 4.266/3.541(HA2.HA3)
ASN	632	120.990(N) 174.580(C) 53.853(CA) 39.003(CB) 8.944(HN) 4.744(HA) 3.069/2.558(HB2.HB3)
VAL	633	116.648(N) 176.381(C) 59.694(CA) 34.869(CB) 22.144(CG1) 20.304(CG2) 8.213(HN) 5.337(HA) 2.086(HB) 1.110/0.997(HG1.HG2)
SER	634	119.619(N) 174.489(C) 58.149(CA) 67.342(CB) 9.068(HN) 4.440(HA) 4.241/3.598(HB2.HB3)
ALA	635	119.449(N) 177.643(C) 52.190(CA) 19.160(CB) 8.451(HN) 4.408(HA) 1.520(HB)
ALA	636	122.249(N) 178.807(C) 52.714(CA) 19.537(CB) 8.518(HN) 4.244(HA) 1.398(HB)
SER	637	113.238(N) 172.713(C) 58.502(CA) 66.216(CB) 8.316(HN) 4.572(HA) 4.106/3.745(HB2.HB3)
ASN	638	113.881(N) 173.716(C) 55.276(CA) 37.955(CB) 8.269(HN) 4.508(HA) 2.860(HB2)
ALA	639	124.751(N) 178.767(C) 50.690(CA) 20.487(CB) 8.543(HN) 5.238(HA) 1.224(HB)
VAL	640	120.908(N) 174.281(C) 60.886(CA) 34.869(CB) 21.803(CG1) 8.850(HN) 4.606(HA) 1.925(HB) 0.949(HG1.HG2)
SER	641	122.866(N) 174.106(C) 57.089(CA) 64.504(CB) 8.733(HN) 5.747(HA) 3.867/3.764(HB2.HB3)
VAL	642	124.490(N) 174.832(C) 60.634(CA) 36.407(CB) 21.526(CG1) 20.984(CG2) 9.186(HN) 4.591(HA) 2.210(HB) 1.077/1.102(HG1.HG2)
LYS	643	127.059(N) 176.688(C) 54.844(CA) 34.638(CB) 29.370(CD) 42.176(CE) 24.526(CG) 8.523(HN) 5.366(HA) 1.911/1.668(HB2.HB3) 2.912/2.912(HE2.HE3) 1.348/1.635(HG2.HG3)
THR	644	123.642(N) 62.589(CA) 70.615(CB) 23.331(CG2) 8.912(HN) 4.253(HA) 4.852(HB) 1.280(HG2)

Table 3-2. Structural statistics[a] for $^{2}FnIIID_{ChiA1}$

Restraints for structure calculation	
NOE distance restraints	
Unambiguous	1382
Ambiguous	239
Hydrogen-bond restraints	23x2
Angle restraints	
HMQC-J(ϕ)/TALOS(φ)/χ1	58/55/4
Mean energies (kcal/mol)	
$E_{Lennard\text{-}Jones}$[b]	-131.215 ± 27.2
E_{elec}	-271.614 ± 6.5
Deviation from restraints[c]	
NOE	0.017 ± 5.7 x 10^{-3} Å
Dihedral angle	0.19 ± 3.9 x 10^{-2} °
Hydrogen bond	0.099 ± 3.0 x 10^{-4} Å
Deviation from ideal geometry	
Bonds	0.0011 ± 5.7 x 10^{-5} Å
Angles	0.25 ± 3.9 x 10^{-2} °
Improper	0.16 ± 3.0 x 10^{-2} Å
Ramachandran analysis by PROCHECK (residues Ala-559 to Thr-644) (Fig. 6)	
Residues in most favored regions	74.0 %
Residues in additional allowed regions	23.6 %
Residues in generously allowed regions	2.4 %
Residues in disallowed regions	0.0 %
Coordinate precision (residues Ala-559–Thr-644)	
RMSD of backbone atoms	0.431 Å
RMSD of heavy atoms	0.743 Å

[a]These statistics comprise the ensemble of the final 30 structures calculated with ARIA1.0.

[b]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.

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

Table 3-3. Bacterial FnIIIDs used for multiple sequences alignment

ID	Organism	Enzyme	Gram
2Fn3	*Bacillus circulans*	chitinase	+
1Fn3	*Bacillus circulans*	chitinase	+
CHID_BACCI	*Bacillus circulans*	chitinase	+
Q48494	*Kurthia zopfii*	chitinase	+
Q9ZNH8	*Acidovorax*	PHBdepolymerase	–
Q9LBN6	*Leptothrix*	PHBdepolymerase	–
O24719	*Comamonas testosteroni*	PHBdepolymerase	–
PHB_ALCFA	*Alcaligenes faecalis*	PHBdepolymerase	–
Q52155	*Burkholderia pickettii*	PHBdepolymerase	–
Q9WXD3	*Serratia marcescens*	chitinase	–
Q9I1H5	*Pseudomonas aeruginosa*	chitinase	–
O30678	*Xanthomonas maltophilia*	chitinase	–
GUXB_CELFI	*Cellulomonas fimi*	exoglucanase	+
GUN4_THEFU	*Thermomonospora fusca*	endoglucanase	+
Q9L036	*Streptomyces coelicolor*	alpha-amylase	+
Q46348	*Cytophaga*	collangenase	–
Q9AJB4	*Ruminococcus albus*	xylanase	+
Q9CE95	*Lactococcus lactis*	chitinase	+
APU_THETU	*Thermoanaerobacter thermosulfurogenes*	amylopullulanase	+
PEHX_ERWCH	*Erwinia chrysanthemi*	exo-poly-α-D-galacturonosidase	–
O68975	*Yersinia enterocolitica*	exo-poly-galacturonase	–

Table 3-4. R_1, R_2, $\{^1H\}$-^{15}N NOE, R_2/R_1 data of $^2FnIIID_{ChiA1}$

Name	No	R1(s^{-1})	dR1	R2(s^{-1})	dR2	NOE	dNOE	R2/R1	dR2/R1
ALA	559	1.554	0.051	4.159	0.165	-0.099	0.127	2.677	0.106
PRO	560	N/A							
THR	561	2.384	0.072	6.169	0.203	0.572	0.062	2.588	0.085
ALA	562	2.511	0.049	5.579	0.125	0.626	0.043	2.221	0.050
PRO	563	N/A							
THR	564	2.487	0.101	5.201	0.242	0.707	0.059	2.091	0.097
ASN	565	2.544	0.071	5.520	0.176	0.647	0.050	2.170	0.069
LEU	566	2.557	0.056	5.738	0.141	0.680	0.042	2.244	0.055
ALA	567	2.612	0.077	5.768	0.196	0.678	0.055	2.208	0.075
SER	568	2.429	0.053	4.849	0.131	0.713	0.042	1.996	0.054
THR	569	2.520	0.082	5.590	0.215	0.662	0.052	2.219	0.085
ALA	570	2.347	0.051	6.117	0.155	0.581	0.040	2.607	0.066
GLN	571	2.317	0.054	5.949	0.268	0.794	0.048	2.568	0.116
THR	572	2.426	0.065	5.924	0.202	0.689	0.050	2.442	0.083
THR	573	2.135	0.111	4.709	0.282	0.671	0.109	2.206	0.132
SER	574	2.378	0.184	4.749	0.439	0.637	0.118	1.997	0.184
SER	575	2.505	0.065	6.052	0.179	0.769	0.041	2.416	0.071
ILE	576	2.522	0.069	5.897	0.184	0.666	0.044	2.338	0.073
THR	577	2.473	0.076	6.454	0.228	0.716	0.045	2.610	0.092
LEU	578	2.463	0.081	5.578	0.222	0.613	0.059	2.265	0.090
SER	579	2.548	0.063	5.620	0.169	0.612	0.045	2.206	0.066
TRP	580	2.628	0.062	5.947	0.177	0.647	0.040	2.263	0.067
THR	581	2.562	0.079	5.840	0.226	0.712	0.054	2.279	0.088
ALA	582	N/A							
SER	583	2.339	0.068	5.499	0.468	0.706	0.063	2.351	0.200
THR	584	2.288	0.099	5.397	0.297	0.626	0.092	2.359	0.130
ASP	585	2.244	0.061	5.301	0.191	0.550	0.053	2.362	0.085
ASN	586	N/A							
VAL	587	N/A							
GLY	588	2.171	0.075	4.609	0.196	0.548	0.089	2.123	0.090
VAL	589	2.240	0.069	5.801	0.192	0.570	0.064	2.590	0.086
THR	590	2.518	0.078	5.882	0.199	0.594	0.049	2.336	0.079
GLY	591	2.410	0.045	5.755	0.125	0.641	0.040	2.388	0.052
TYR	592	2.484	0.075	5.546	0.205	0.684	0.048	2.233	0.082
ASP	593	2.502	0.083	5.711	0.234	0.641	0.051	2.283	0.093
VAL	594	N/A							
TYR	595	2.533	0.085	5.593	0.220	0.660	0.050	2.208	0.087
ASN	596	2.418	0.082	5.777	0.242	0.667	0.050	2.389	0.100
GLY	597	2.521	0.056	5.296	0.148	0.644	0.042	2.101	0.059
THR	598	2.152	0.218	5.936	0.668	0.655	0.178	2.758	0.310
ALA	599	2.495	0.058	5.803	0.176	0.729	0.042	2.326	0.070
LEU	600	2.302	0.037	5.029	0.102	0.652	0.035	2.185	0.044
ALA	601	2.380	0.069	5.614	0.205	0.681	0.048	2.358	0.086
THR	602	2.388	0.035	5.239	0.100	0.699	0.032	2.194	0.042
THR	603	2.276	0.056	5.500	0.169	0.666	0.039	2.417	0.074
VAL	604	2.508	0.064	5.845	0.181	0.668	0.042	2.331	0.072
THR	605	2.374	0.104	5.705	0.292	0.761	0.064	2.403	0.123

(Table 3-4. continued)

Name	No	R1(s^{-1})	dR1	R2(s^{-1})	dR2	NOE	dNOE	R2/R1	dR2/R1
GLY	606	1.917	0.036	4.804	0.113	0.564	0.033	2.506	0.059
THR	607	2.218	0.057	5.146	0.168	0.662	0.043	2.320	0.076
THR	608	2.467	0.051	5.607	0.138	0.702	0.037	2.273	0.056
ALA	609	2.523	0.058	5.605	0.153	0.620	0.040	2.221	0.061
THR	610	2.282	0.061	5.363	0.171	0.682	0.046	2.350	0.075
ILE	611	2.520	0.069	5.071	0.183	0.690	0.050	2.012	0.073
SER	612	2.268	0.057	5.892	0.179	0.682	0.041	2.598	0.079
GLY	613	2.313	0.056	4.857	0.151	0.700	0.048	2.100	0.065
LEU	614	N/A							
ALA	615	2.333	0.052	5.537	0.161	0.689	0.045	2.373	0.069
ALA	616	2.305	0.035	5.334	0.099	0.721	0.036	2.315	0.043
ASP	617	2.373	0.045	5.158	0.121	0.664	0.039	2.174	0.051
THR	618	2.383	0.061	5.280	0.166	0.617	0.045	2.216	0.070
SER	619	2.331	0.051	5.160	0.129	0.629	0.042	2.214	0.055
TYR	620	2.400	0.072	5.004	0.183	0.621	0.050	2.085	0.076
THR	621	2.345	0.066	5.322	0.176	0.614	0.043	2.270	0.075
PHE	622	2.412	0.078	5.261	0.209	0.638	0.053	2.181	0.087
THR	623	2.601	0.072	5.444	0.190	0.692	0.046	2.093	0.073
VAL	624	2.676	0.087	5.550	0.538	0.649	0.054	2.074	0.201
LYS	625	2.619	0.091	5.690	0.246	0.645	0.057	2.172	0.094
ALA	626	2.565	0.088	5.648	0.245	0.654	0.055	2.202	0.095
LYS	627	N/A							
ASP	628	2.482	0.084	6.102	0.244	0.626	0.059	2.458	0.098
ALA	629	2.396	0.055	5.348	0.150	0.705	0.053	2.232	0.063
ALA	630	2.194	0.044	4.944	0.133	0.606	0.038	2.253	0.061
GLY	631	2.280	0.056	4.981	0.162	0.657	0.045	2.184	0.071
ASN	632	2.464	0.066	5.773	0.178	0.621	0.047	2.343	0.072
VAL	633	2.386	0.085	5.771	0.237	0.554	0.066	2.418	0.099
SER	634	2.579	0.078	5.836	0.202	0.667	0.059	2.263	0.078
ALA	635	2.518	0.057	6.417	0.166	0.721	0.044	2.548	0.066
ALA	636	2.550	0.049	5.333	0.132	0.655	0.042	2.091	0.052
SER	637	2.608	0.054	5.380	0.148	0.661	0.043	2.063	0.057
ASN	638	2.544	0.047	5.295	0.134	0.658	0.042	2.082	0.053
ALA	639	2.542	0.055	5.404	0.155	0.629	0.052	2.125	0.061
VAL	640	2.543	0.068	5.354	0.173	0.630	0.046	2.106	0.068
SER	641	N/A							
VAL	642	2.454	0.065	5.000	0.168	0.641	0.044	2.038	0.068
LYS	643	N/A							
THR	644	2.487	0.061	5.359	0.154	0.613	0.041	2.155	0.062
NE1	580	2.175	0.034	5.827	0.116	0.673	0.032	2.679	0.053

Table 3-5. Model free data of $^{2}FnIIID_{ChiA1}$

Name	No.	Model	S^2	dS^2	τ_e(s)	$d\tau_e$	R_{ex}(Hz)	dR_{ex}	Sf^2	dSf^2
ALA	559	4	0.419	0.019	9.927e-11	1.628e-11	1.046	0.216		
PRO	560	N/A								
THR	561	3	0.789	0.024			0.984	0.246		
ALA	562	1	0.834	0.012						
PRO	563	N/A								
THR	564	1	0.807	0.024						
ASN	565	1	0.840	0.018						
LEU	566	1	0.856	0.014						
ALA	567	1	0.870	0.021						
SER	568	1	0.776	0.014						
THR	569	1	0.838	0.019						
ALA	570	4	0.771	0.017	6.772e-11	1.863e-11	0.593	0.202		
GLN	571	3	0.777	0.019			0.646	0.304		
THR	572	1	0.832	0.017						
THR	573	1	0.709	0.026						
SER	574	1	0.755	0.044						
SER	575	1	0.857	0.017						
ILE	576	1	0.856	0.019						
THR	577	3	0.827	0.023			0.847	0.288		
LEU	578	1	0.828	0.020						
SER	579	1	0.846	0.015						
TRP	580	1	0.882	0.018						
THR	581	1	0.863	0.024						
ALA	582	N/A								
SER	583	1	0.779	0.022						
THR	584	1	0.779	0.027						
ASP	585	2	0.737	0.018	6.266e-11	1.987e-11				
ASN	586	N/A								
VAL	587	N/A								
GLY	588	1	0.707	0.018						
VAL	589	4	0.718	0.024	4.931e-11	2.379e-11	0.975	0.252		
THR	590	4	0.812	0.026	7.020e-11	2.988e-11	0.438	0.236		
GLY	591	4	0.790	0.015	3.847e-11	1.661e-11	0.383	0.165		
TYR	592	1	0.824	0.021						
ASP	593	1	0.842	0.020						
VAL	594	N/A								
TYR	595	1	0.844	0.021						
ASN	596	1	0.828	0.022						
GLY	597	1	0.822	0.015						
THR	598	1	0.775	0.056						
ALA	599	1	0.841	0.013						
LEU	600	1	0.760	0.010						
ALA	601	1	0.808	0.019						
THR	602	1	0.793	0.009						
THR	603	1	0.778	0.015						
VAL	604	1	0.847	0.017						
THR	605	1	0.810	0.024						

(Table 3-5. continued)

Name	No.	Model	S^2	dS^2	τ_e(s)	$d\tau_e$	R_{ex}(Hz)	dR_{ex}	Sf^2	dSf^2
GLY	606	4	0.626	0.013	3.610e-11	7.359e-12	0.337	0.133		
THR	607	1	0.749	0.015						
THR	608	1	0.829	0.012						
ALA	609	2	0.819	0.017	6.555e-11	2.740e-11				
THR	610	1	0.773	0.015						
ILE	611	1	0.807	0.016						
SER	612	3	0.765	0.020			0.600	0.228		
GLY	613	1	0.751	0.014						
LEU	614	N/A								
ALA	615	1	0.793	0.013						
ALA	616	1	0.777	0.010						
ASP	617	1	0.785	0.012						
THR	618	1	0.793	0.016						
SER	619	1	0.776	0.013						
TYR	620	1	0.782	0.019						
THR	621	4	0.756	0.022	3.937e-11	1.917e-11	0.336	0.187		
PHE	622	1	0.799	0.021						
THR	623	1	0.849	0.017						
VAL	624	1	0.882	0.024						
LYS	625	1	0.863	0.024						
ALA	626	1	0.853	0.022						
LYS	627	N/A								
ASP	628	3	0.821	0.028			0.724	0.303		
ALA	629	1	0.801	0.015						
ALA	630	2	0.714	0.011	4.017e-11	1.192e-11				
GLY	631	1	0.751	0.014						
ASN	632	1	0.838	0.019						
VAL	633	4	0.764	0.031	7.211e-11	2.982e-11	0.583	0.290		
SER	634	1	0.862	0.020						
ALA	635	3	0.836	0.017			0.863	0.193		
ALA	636	1	0.830	0.011						
SER	637	1	0.845	0.014						
ASN	638	1	0.828	0.014						
ALA	639	1	0.833	0.015						
VAL	640	1	0.830	0.015						
SER	641	N/A								
VAL	642	5	0.784	0.089	3.173e-09	1.335e-09			8.277e-01	2.420e-02
LYS	643	N/A								
THR	644	2	0.803	0.016	5.559e-11	2.331e-11				
NE1	580	3	0.842	0.013			0.673	0.140		

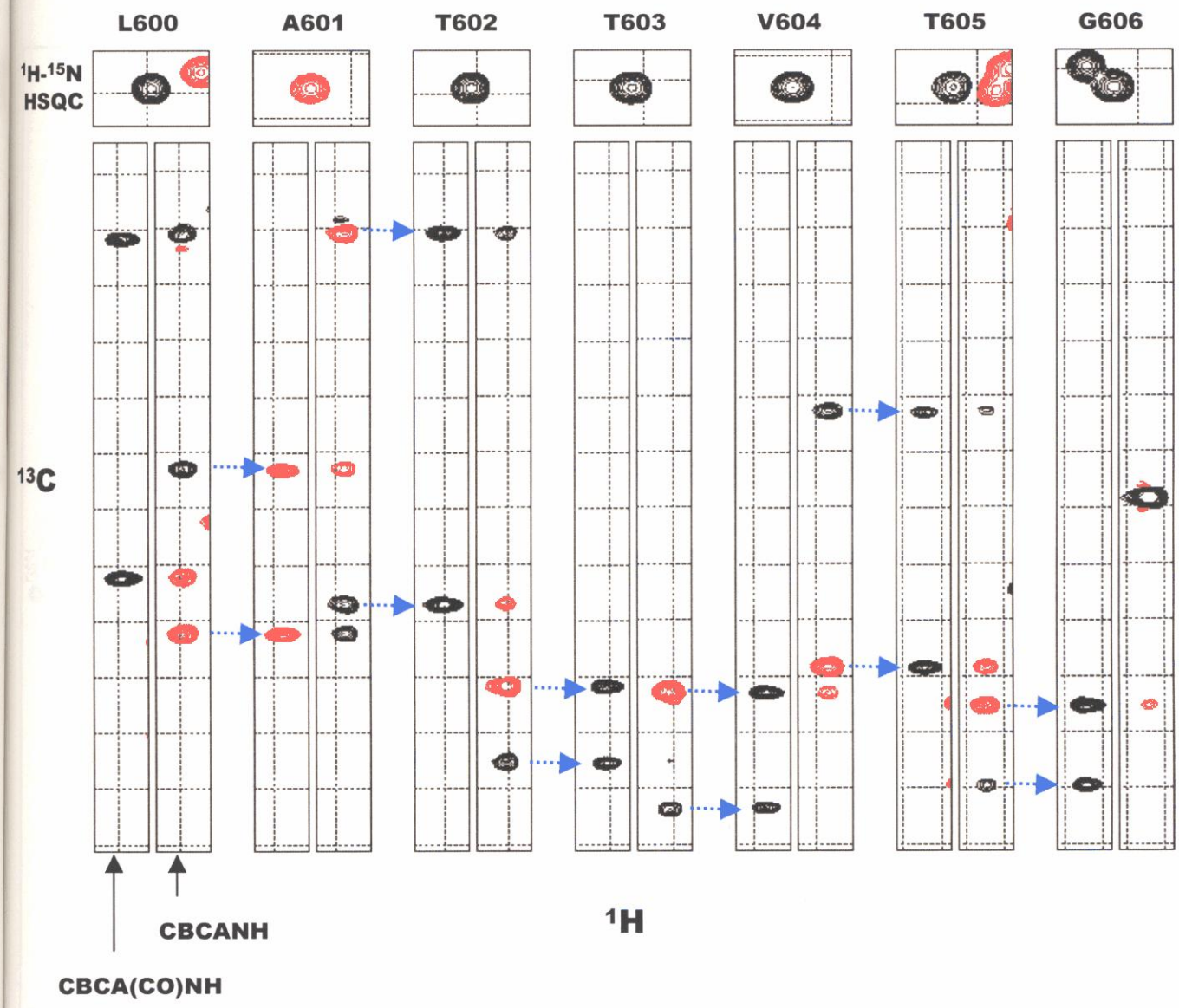


Fig. 3-1. Strip plots from Leu-600 to Gly-606

Corresponding regions of ^{1}H-^{15}N HSQC, CBCA(CO)NH, and CBCANH are selected. Inter-residue connectivities are drawn with blue arrows.

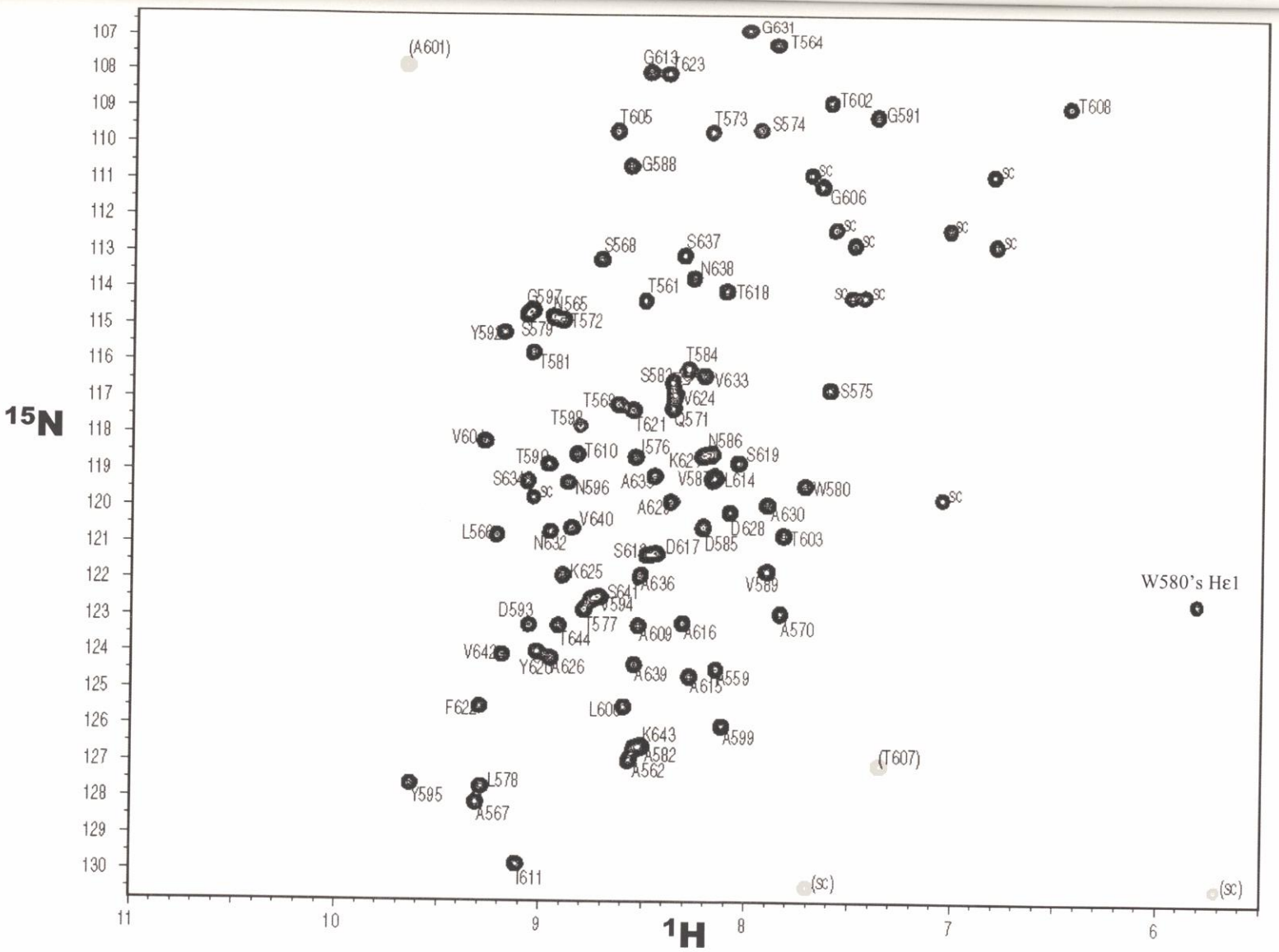


Fig. 3-2. The ^{1}H-^{15}N HSQC spectrum of ^{2}FnIIID$_{ChiA1}$.
The assignments of the backbone amide groups are labeled. Labels in parentheses indicate negative peaks resulted from aliasing. The 'sc' labels indicate side-chain peaks from asparagine or glutamine residues. The Hε1 atom of Trp-580 side chain resonates at 5.80 ppm.

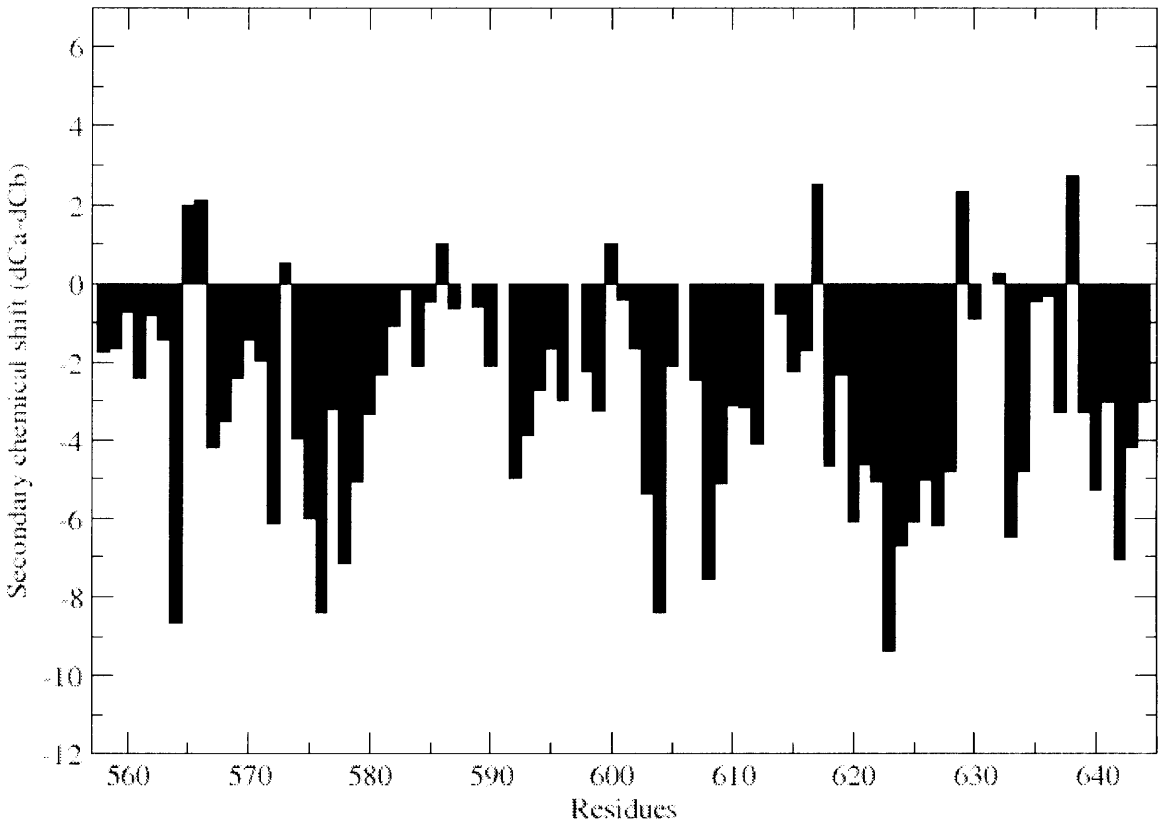


Fig. 3-3. Backbone secondary chemical shift

Based on assigned chemical shifts, secondary chemical shifts, $(C\alpha_{exp}\text{-}C\alpha_{random})\text{-}(C\beta_{exp}\text{-}C\beta_{random})$, were calculated and plotted. The negative values of this plot mean the β-strand structures (79) Using this plot, after finishing assignment, the secondary structures can be predicted readily.

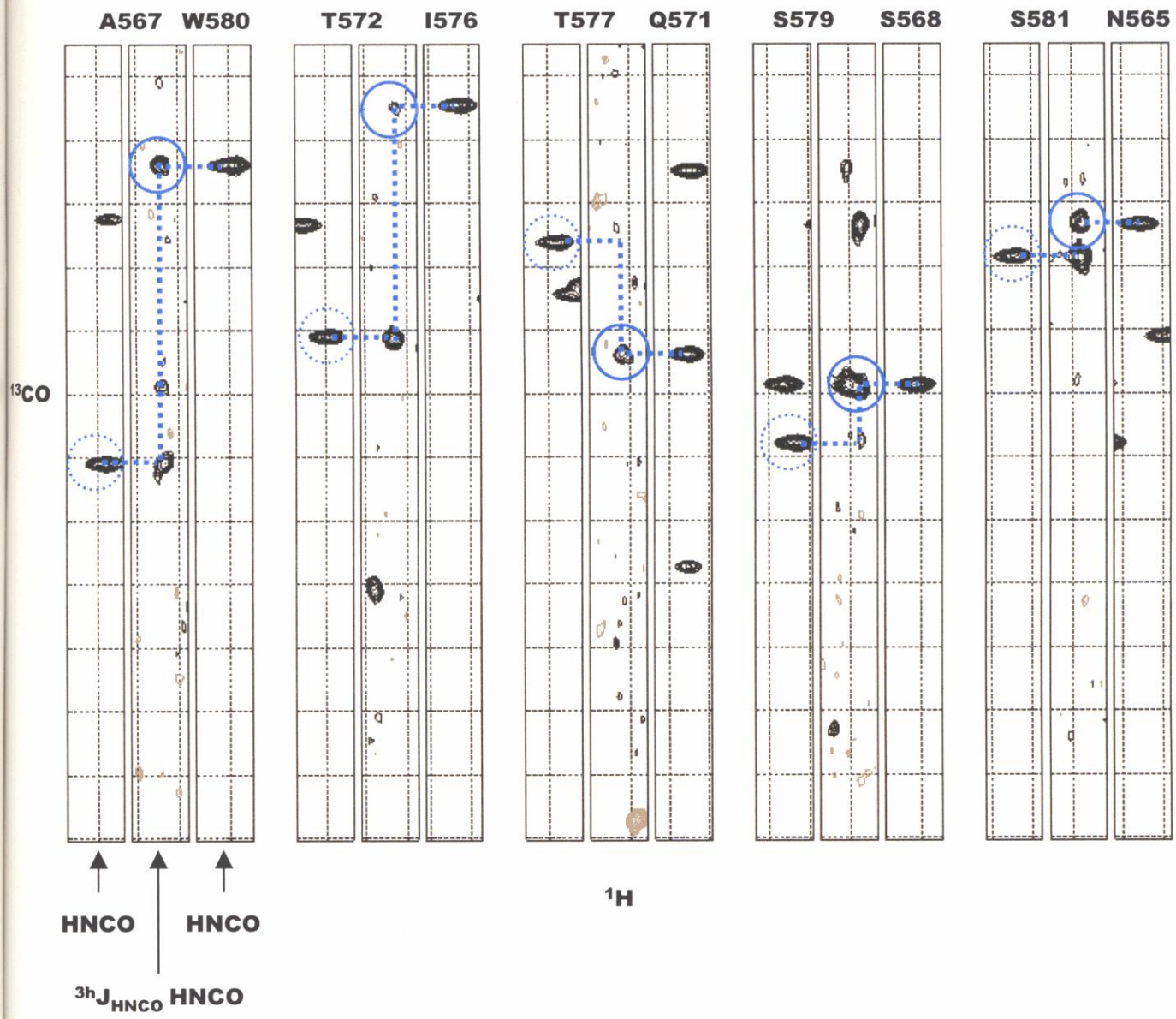


Fig. 3-4. Direct observed hydrogen bonds between β-strands A and B.

Directed detected hydrogen bonds between β strands A and B were drawn using HNCO, $^{3h}J_{HNCO}$ HNCO strip plots. The left one exhibits the strip plot of hydrogen bond donor from HNCO, center strip plot was extracted from $^{3h}J_{HNCO}$ HNCO, and right strip plot comes from HNCO for hydrogen bond acceptor. Hydrogen bond connectivity was drawn by dotted blue line.

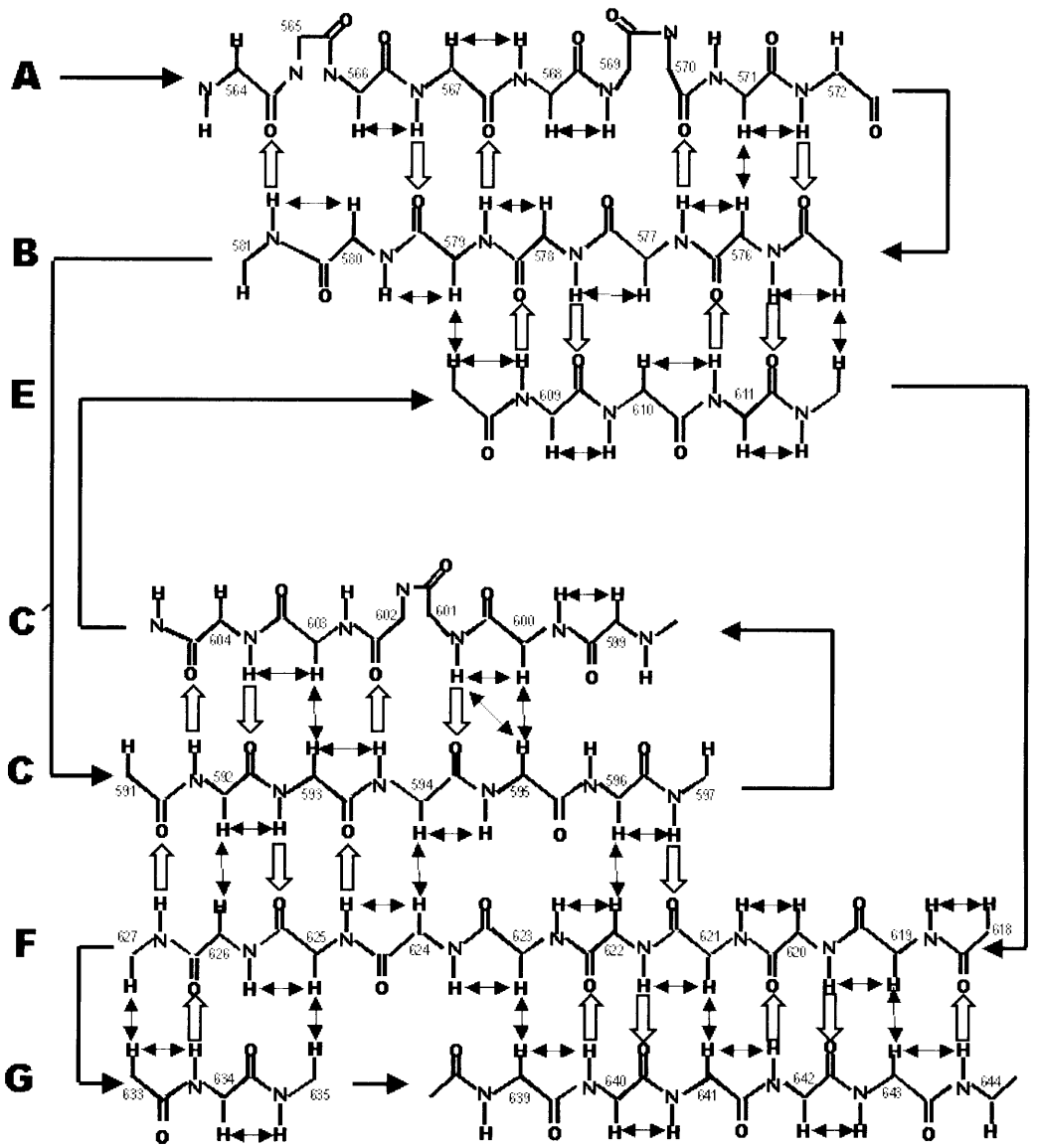


Fig. 3-5. Schematic diagram of the ^{2}FnIIID$_{ChiA1}$ β-sheets.

The hydrogen bonds identified through scalar couplings across hydrogen bonds ($^{3h}J_{HNCO}$) are displayed as broad, open arrows. The observed NOEs are indicated as double-headed thin arrows. The β-strands are labeled on the left. In total, 23 hydrogen bonds were detected. The β-bulge structures on the A (Asn-565/Leu-566 and Thr-569/Ala-570) and C′ (Ala-601/Thr-602) strands are indicated.

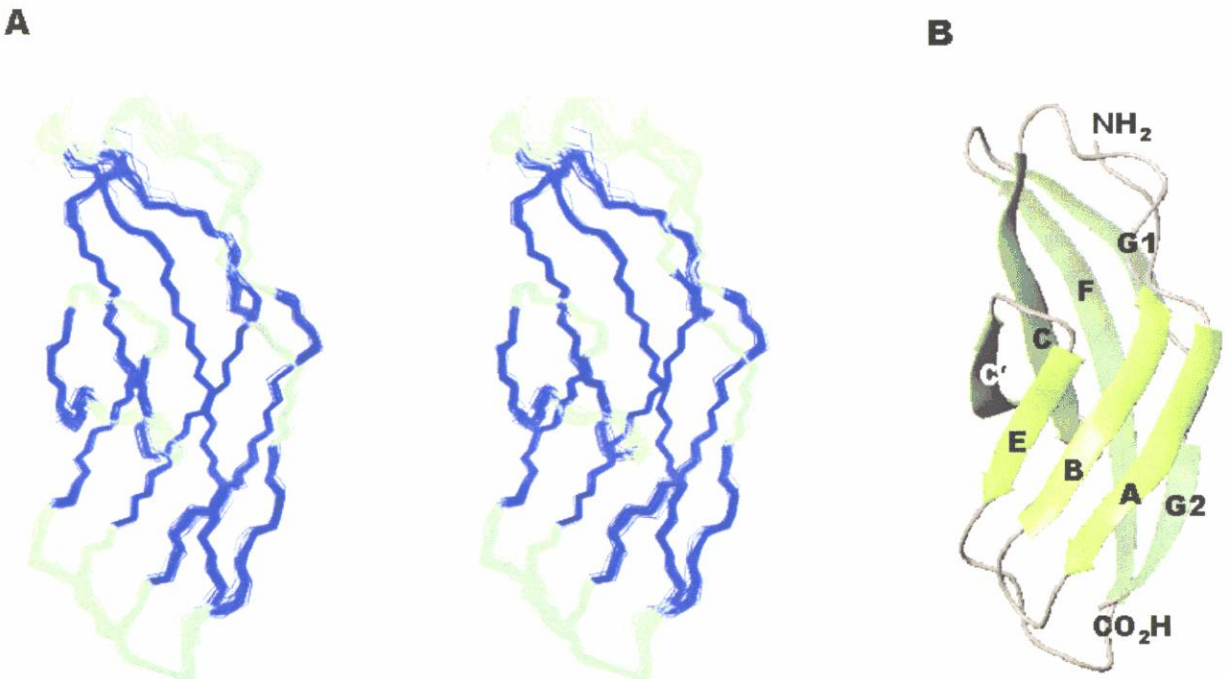


Fig. 3-6. Solution structure of ^{2}FnIIID$_{ChiA1}$.

(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 β-strands (A, B, C, C′, E, F and G (G1 and G2)) are labeled. All structure figures were prepared using the program MOLMOL (45).

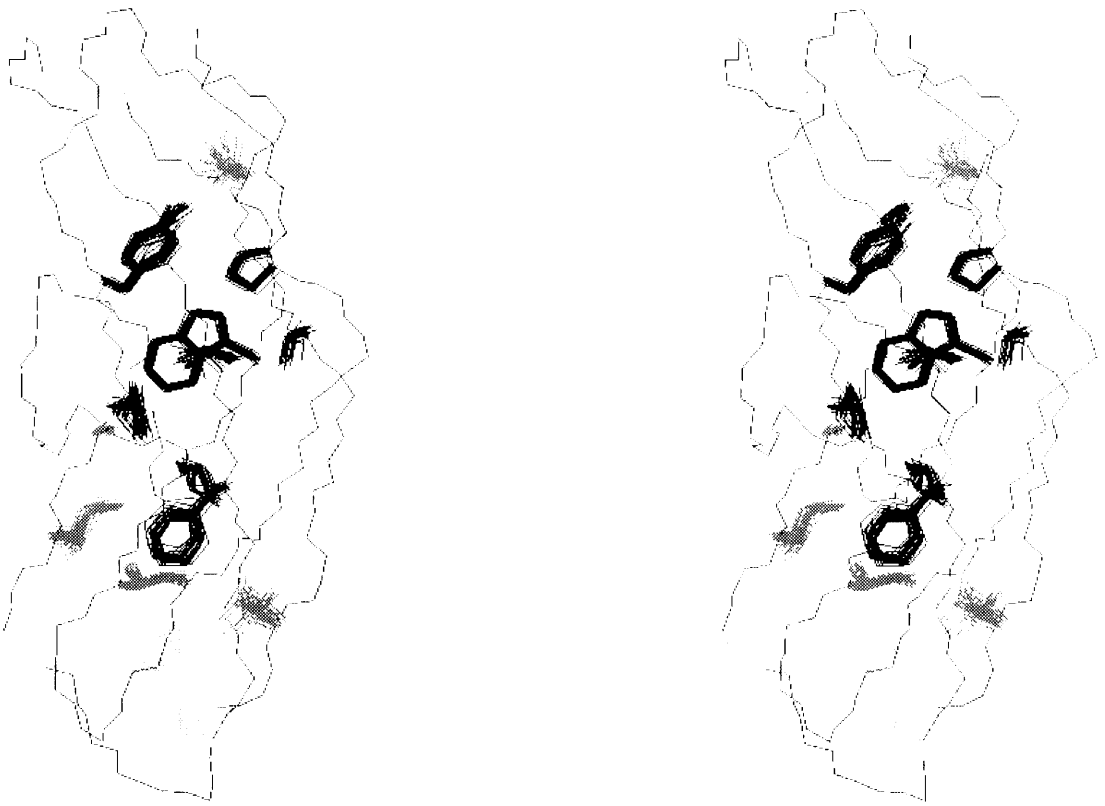


Fig. 3-7. Conserved or hydrophobic-core-forming residues.

Stereo view showing the superposition of conserved residues in bacterial FnIIIDs or hydrophobic-core-forming residues with the backbone trace of the lowest energy structure. The residues which are conserved in sequences alignment and hydrophobic-core-forming (solvent accessibility below 7%) are colored black (Pro-563, Leu-578, Trp-580, Tyr-592, Val-594, Phe-622, Val-624, Ala-626 and Ser-637). The other residues which play roles in hydrophobic-core-forming are colored gray (Ile-576, Ser-583, Ala-609, Ile-611 and Val-642). The other conserved residues are drawn in light gray (Leu-566, Gly-613, Leu-614, Thr-618, Tyr-620, Asp-628, Gly-631, Ser-634 and Thr-644).

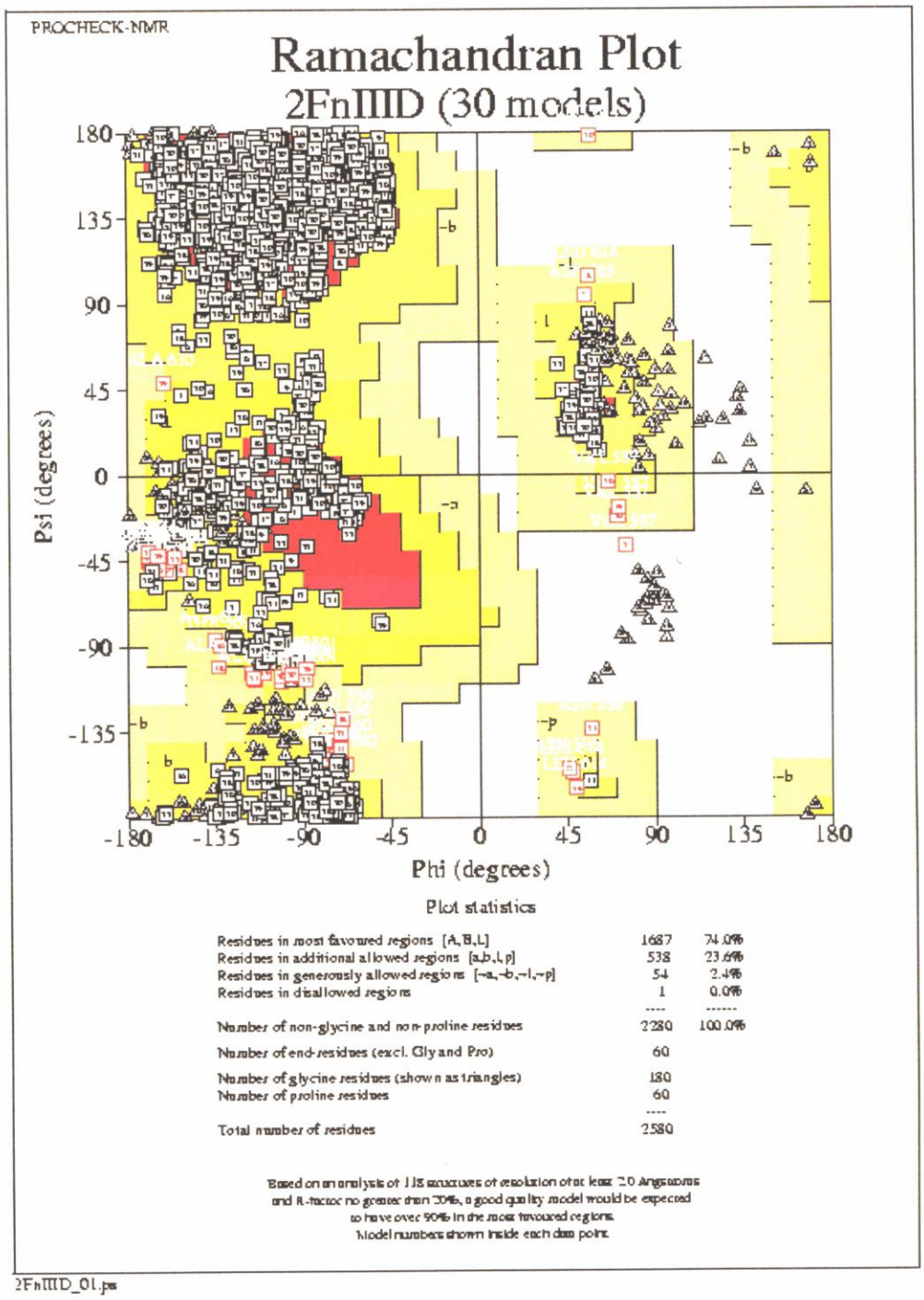


Fig. 3-8. Ramachandran Plot of 30 overlay structures.

All the residues of $^{2}FnIIID_{ChiA1}$ (Ala-559 to Thr-644) were considered. This plot was prepared by PROCHECK-NMR software (39).

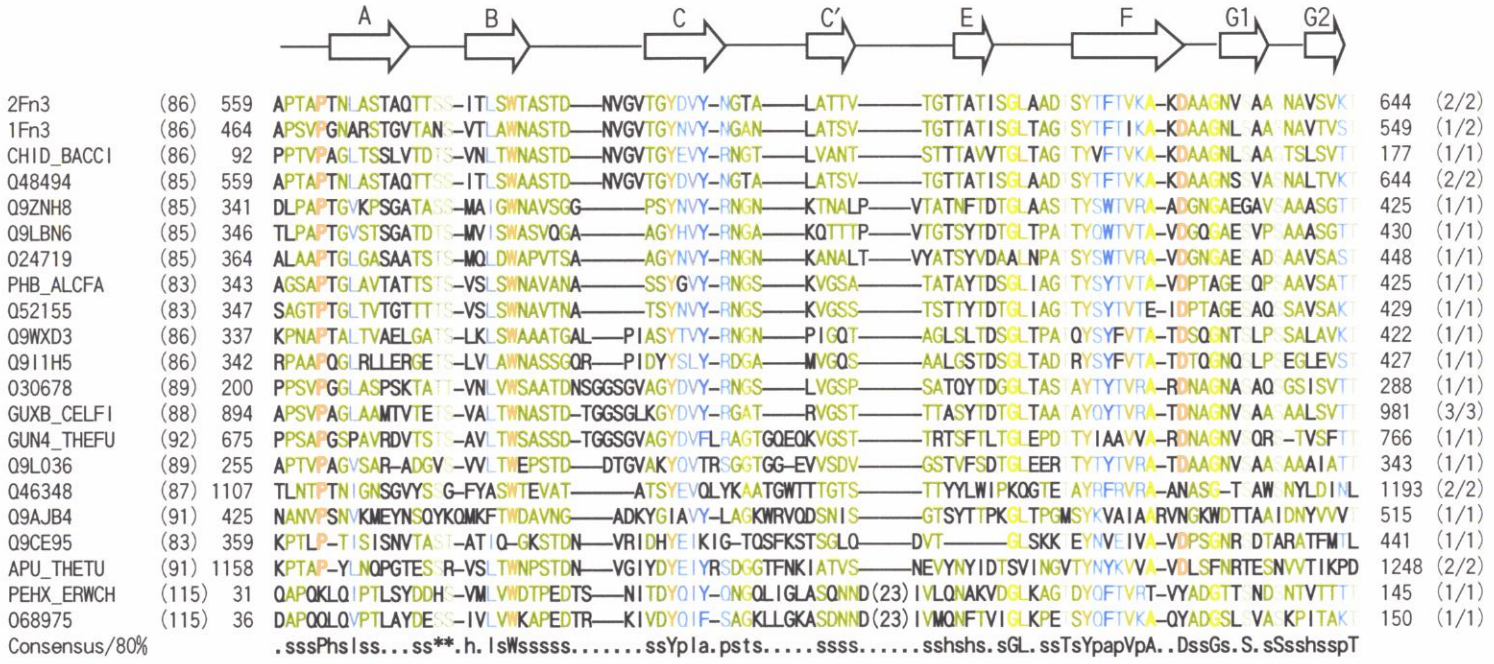


Fig. 3-9. Multiple sequences alignment of bacterial FnIIIDs.

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: + is applied for positively charged residues (H, K and R), - for negatively charged residues (D and E), a for aromatic residues (F, H, W and Y), h for hydrophobic residues (A, C, F, H, I, L, M, V, W and Y), l for aliphatic residues (I, L and V), * for alcohol residues (S and T), p for polar residues (D, E, H, K, N, Q, R, S and T), s for small residues (A, C, S, T, D, N, V, G and P), t for tiny residues (A, G and S) and upper case 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. $^{2}\text{FnIIID}_{\text{ChiA1}}$ and $^{1}\text{FnIIID}_{\text{ChiA1}}$ are abbreviated as 2Fn3 and 1Fn3 respectively. Secondary structure elements of $^{2}\text{FnIIID}_{\text{ChiA1}}$ are indicated above the alignments. Sequence alignment was carried out by ClustalW (41) and annotated using CHROMA (42).

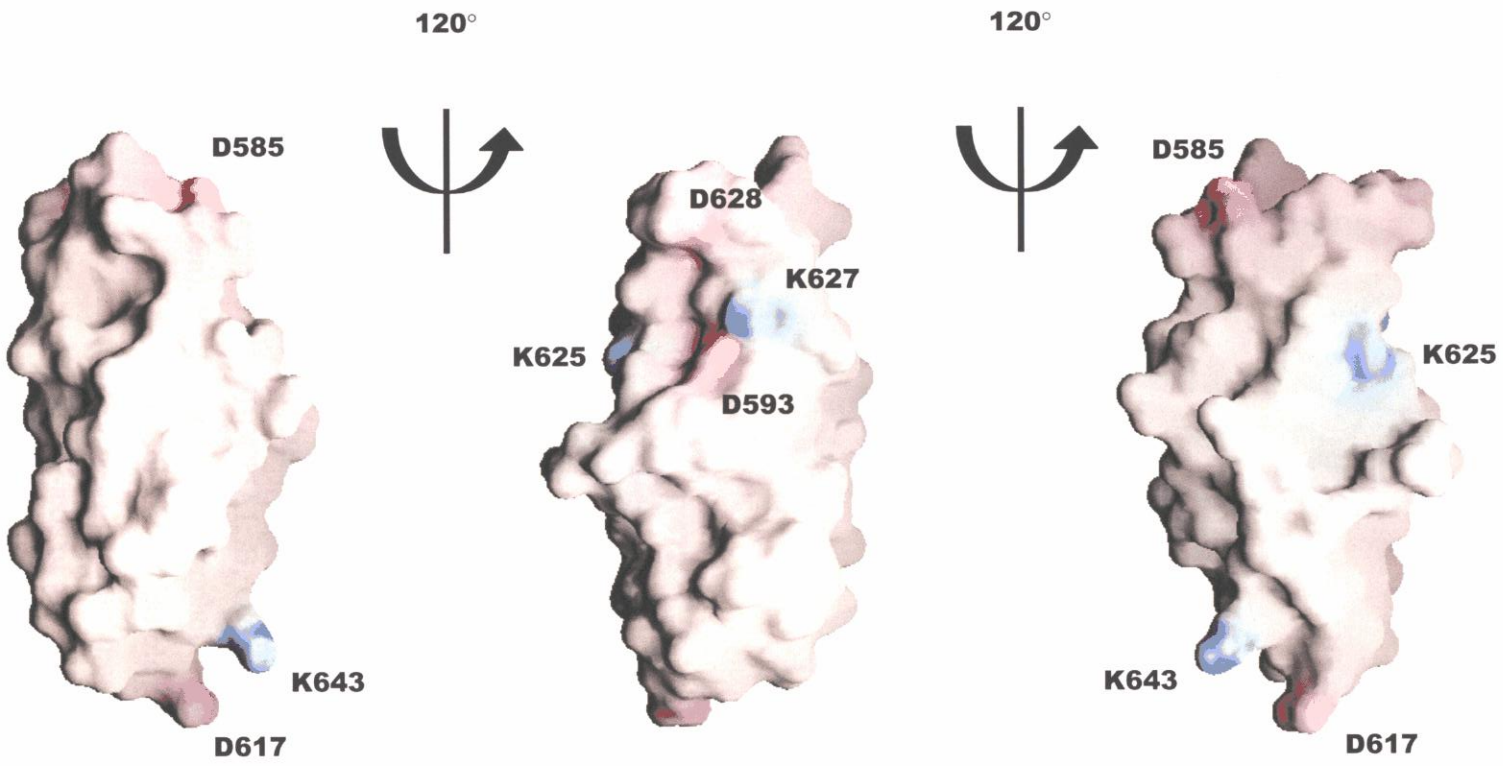


Fig. 3-10. Surface diagram.

Electrostatic potentials on the solvent accessible surface of $^{2}FnIIID_{ChiA1}$. The potentials are colored from red (negative charge) to blue (positive charge). Images were rotated by 120° along *y*-axis.

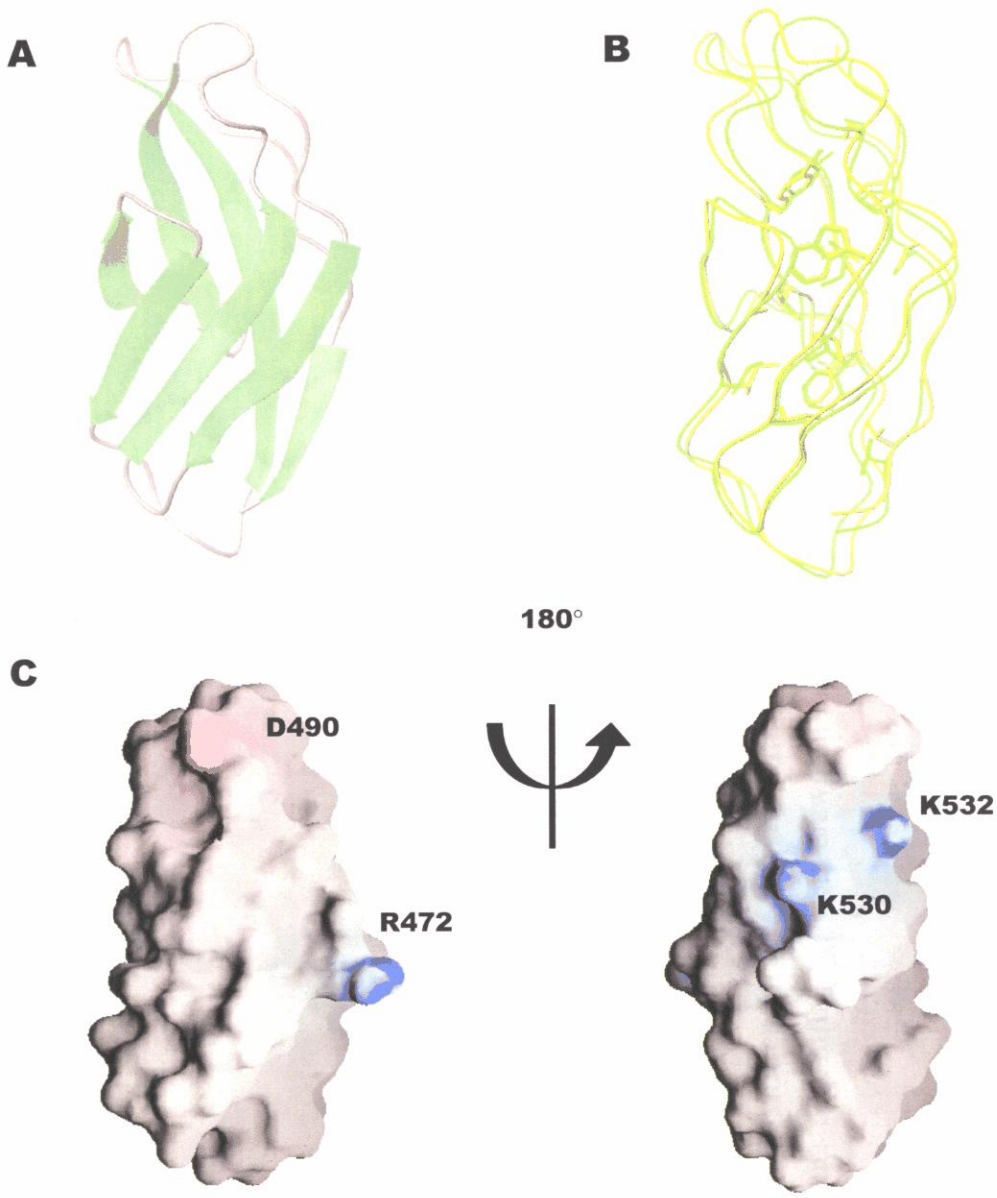


Fig. 3-11. Modeled Structure of $^{1}FnIIID_{ChiA1}$.

A, Ribbon diagram of 1FnIIIDChiA1. B, Overlayed structures of $^{2}FnIIID_{ChiA1}$ (yellow) and $^{1}FnIIID_{ChiA1}$ (green). Backbones and hydrophobic core forming residues were drawn two. C, Electrostatic potentials on the solvent accessible surface of $^{1}FnIIID_{ChiA1.}$ The potentials are colored from red (negative charge) to blue (positive charge). The right-hand image was generated by 180° rotation along *y*-axis of the left-hand molecule.

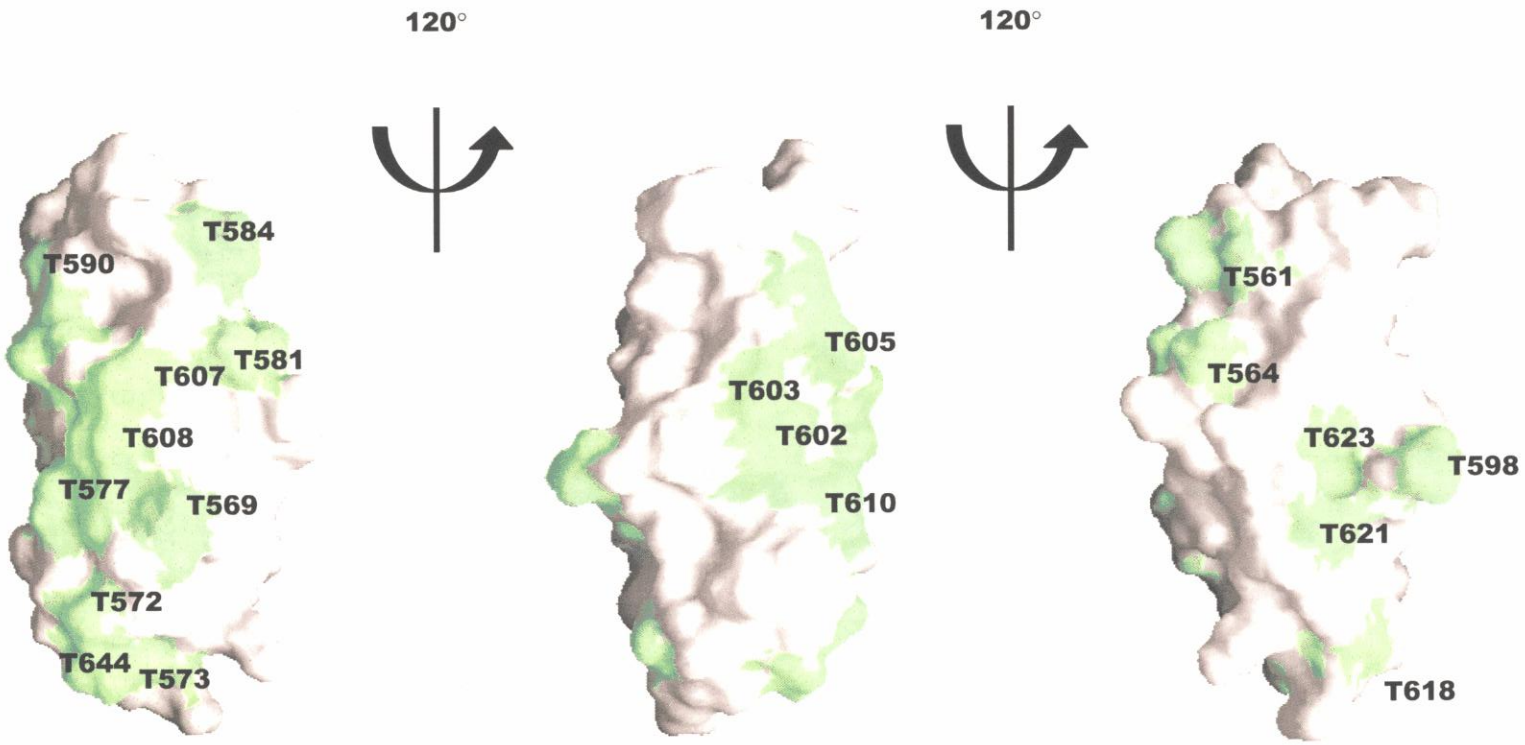


Fig. 3-12. Threonine residues on the Surface diagram.

Twenty threonine residues of $^{2}FnIIID_{ChiA1}$ were drawn on the surface. Images were rotated by 120° along *y*-axis.

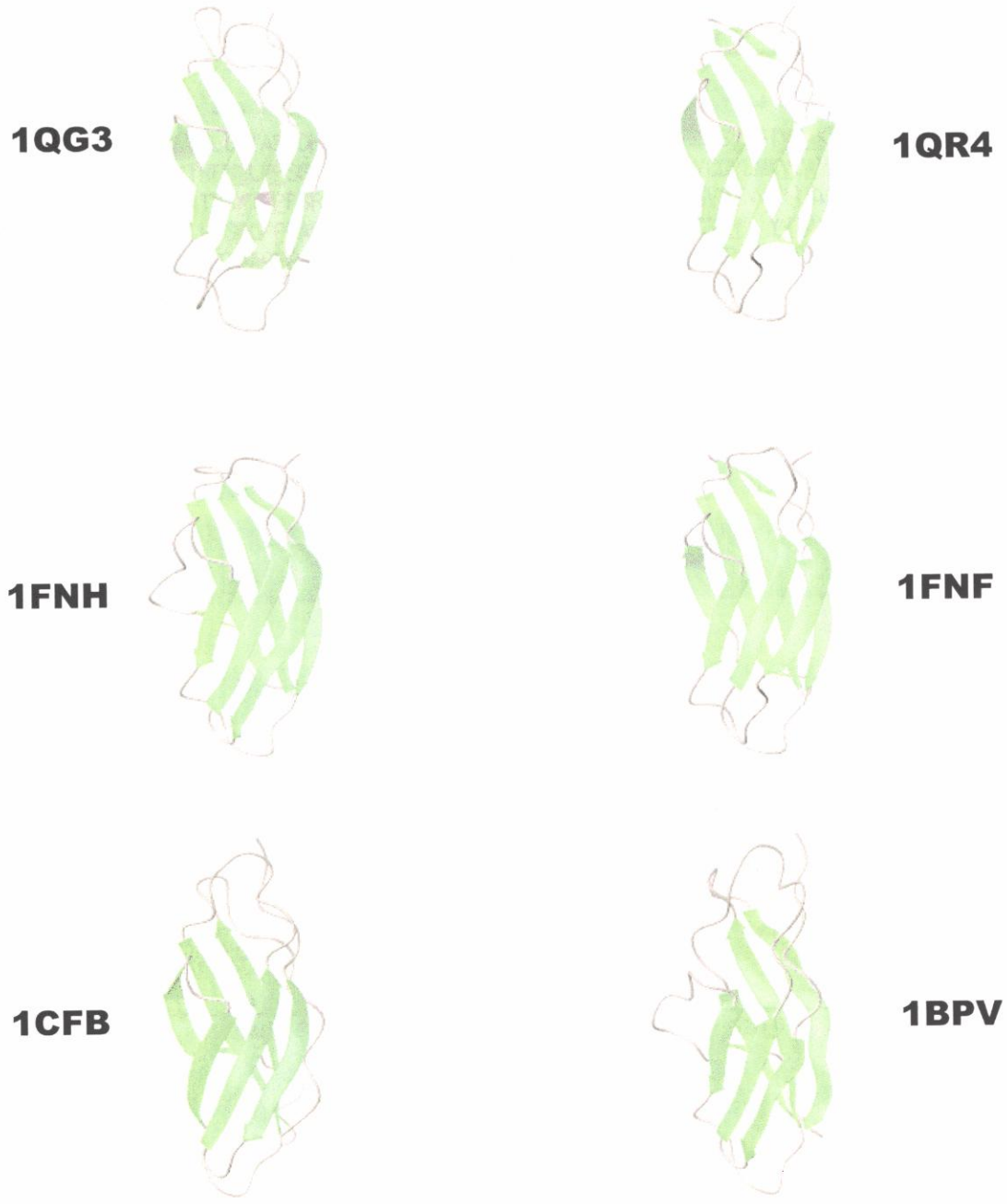


Fig. 3-13. Animal FnIIID structures with high similarity to $^{2}FnIIID_{ChiA1}$.
These structures reveal DALI Z scores of above 10.0 against $^{2}FnIIID_{ChiA1}$. Each structure was labeled with its PDB code: 1QG3 (FnIIID from intergrin β4 subunit), 1QR4 (from tenascin), 1FNH (from Fibronectin), 1FNF (from Fibronectin), 1CFB (from neuroglian), and 1BPV (from titin). For clarity, each of the structures was oriented using the structure alignment with a structure of $^{2}FnIIID_{ChiA1}$. All these structures can be superposed into that of $^{2}FnIIID_{ChiA1}$ within Cα RMSD 2.0.

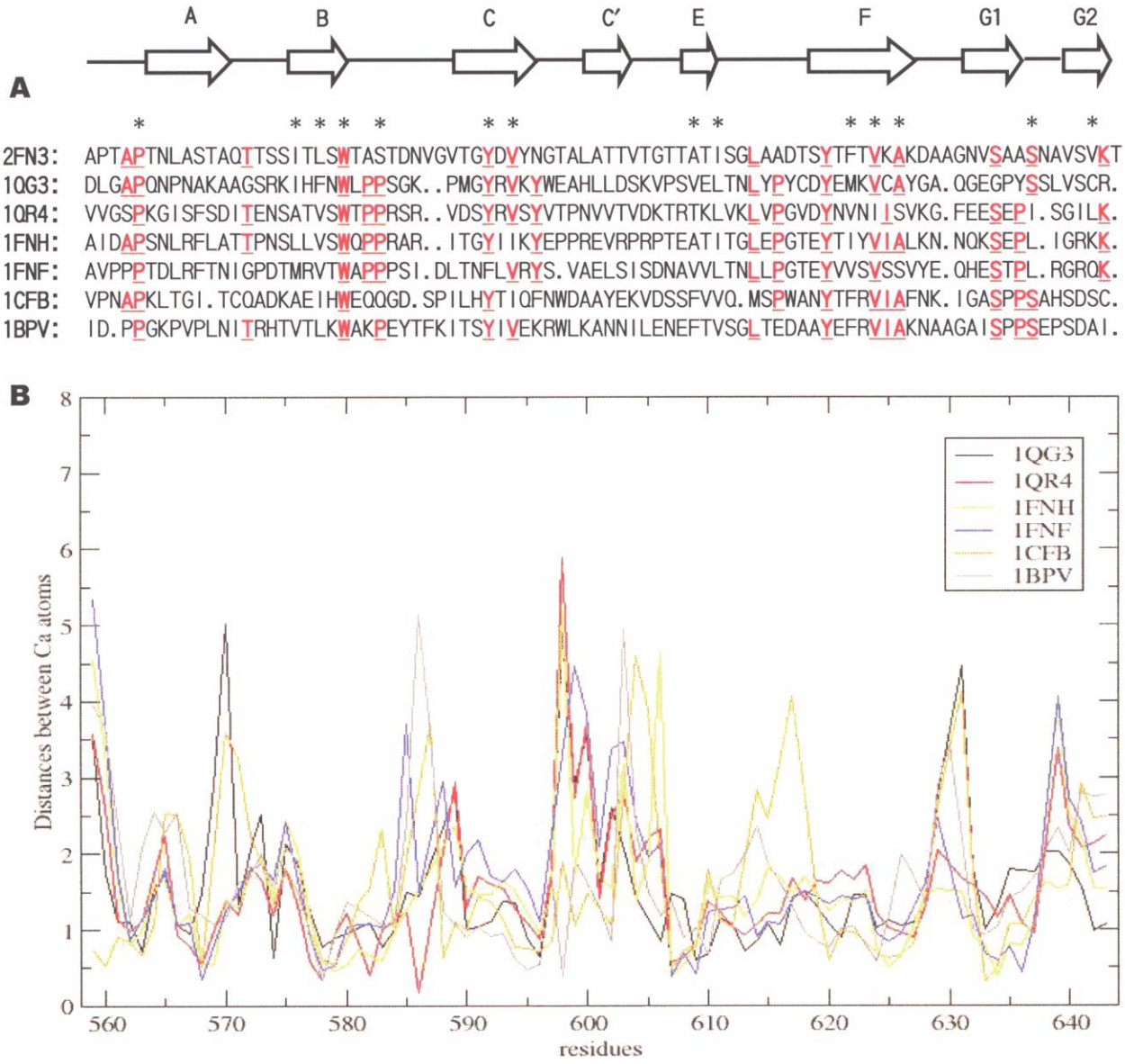


Fig. 3-14. Comparison of $^{2}FnIIID_{ChiA1}$ structure with those of animal FnIIIDs.

A, structure based sequence alignment of animal FnIIIDs.Using DALI program, the sequences of animal FnIIIDs showing high structure similarity ($Z > 10$) to that of $^{2}FnIIID_{ChiA1}$ were aligned. Conserved amino acids beyond 4 sequences were drawn in red and underlined. The hydrophobic core forming residues of $^{2}FnIIID_{ChiA1}$ are labeled with *. The sequence of $^{2}FnIIID_{ChiA1}$ was abbreviated as 2FN3 and other sequences were named after their PDB codes. B, distance diagram of aligned structures. The distances between Cα atoms of $^{2}FnIIID_{ChiA1}$ and animal FnIIIDs in each aligned residue were plotted for structure comparison. For clarity, secondary structure were indicated at the top.

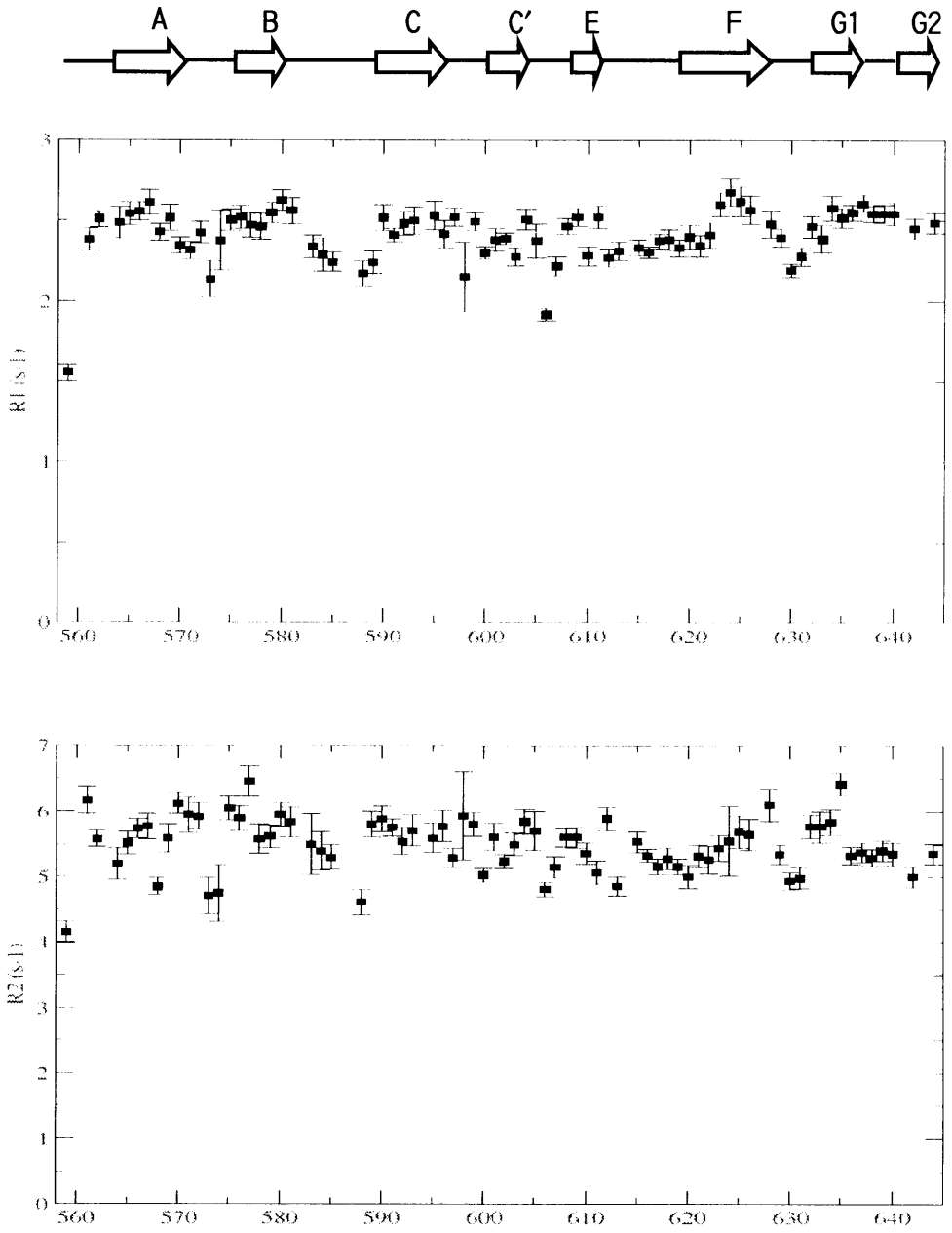


Fig. 3-15. Plots of R_1 and R_2.

All the data were measured at ^{15}N frequency of 50.7 MHz. Among the 86 residues (Ala-559 to Thr-644), except prolines (Pro-560 and Pro–563) and overlapping residues in HSQC (Ala-582, Asn-586, Val-587, Val-594, Leu-614, Lys-627, Ser-641, and Lys-643), the data from 76 residues were acquired. The error bar represent standard deviations in the estimated parameters.

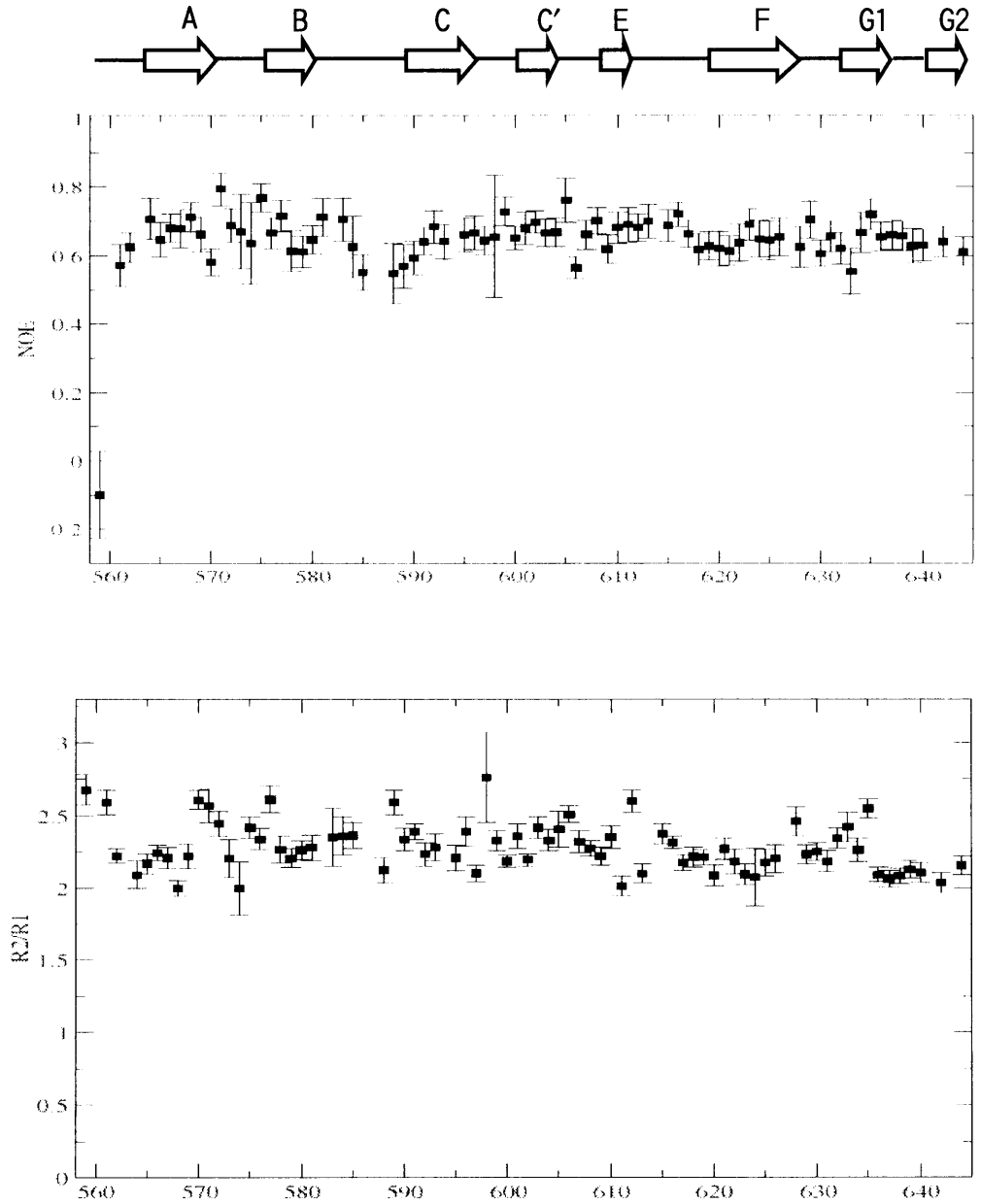


Fig. 3-16. Plots of {^{1}H}-^{15}N NOE and R_2/R_1.

All the data were measured at ^{15}N frequency of 50.7 MHz. Among the 86 residues, the data from 76 residues were acquired. A schematic diagram of secondary structure of ^{2}FnIIID$_{ChiA1}$ was drawn at the top. To acquire overall diffusion tensor, these data were used initially.

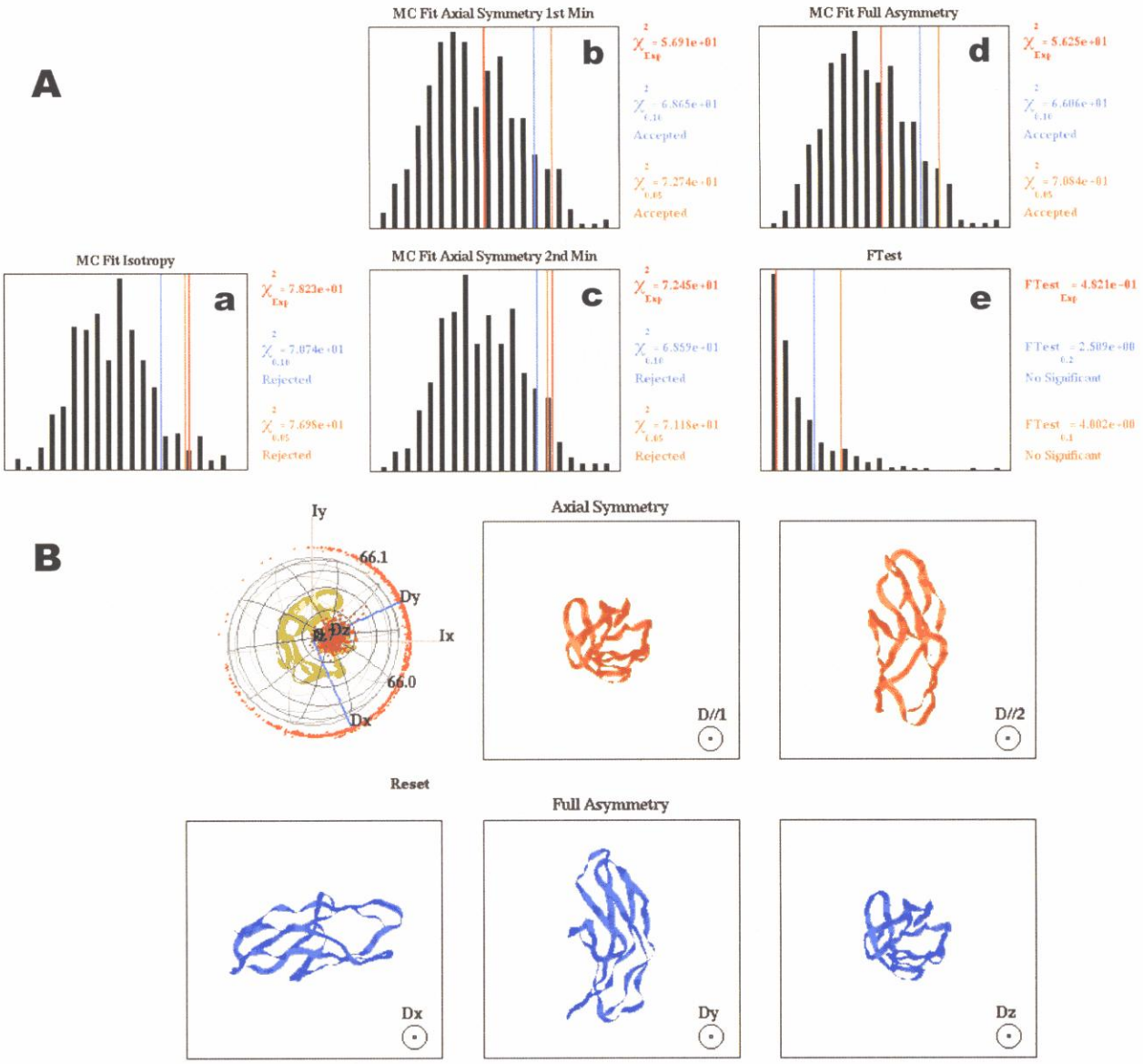


Fig. 3-17. Overall diffusion model of ^{2}FnIIID$_{ChiA1}$.

A, Statistical analysis of χ^2 and F-test. B, The axes of diffusion tensor. (A) The 500 cycles of Monte-Carlo simulation of χ^2 were done in isotropic (a), axial symetric prolate (b) and oblate (c), and anisotropic model (d) respectively. The experimental χ^2 values were written in orange color and compared with $\chi^2_{0.05}$ values. When χ^2_{exp} was greater $\chi^2_{0.05}$, the model was rejected. Axial symmetric prolate and anisotropic model could be selected. However, as F-test indicate there was no significant improvement in anisotropic model (e), finally axial symmetric prolate model was select as diffusion tensor. (B) The upper left panel is the ellipse form of rotational diffusion tensor. The Iz component of inertia tensor makes a angle of 10° with Dz component of diffusion tensor. The two panels of the upper right show axially symmetric tensor. The three lower panels were taken from fully asymmetric tensor's x (Dx), y (Dy), and z (Dz) axes. The prolate (D//1) tensor shows similar orientation to full symmetric tensor (Dz). These figures were captured from the program Tensor.

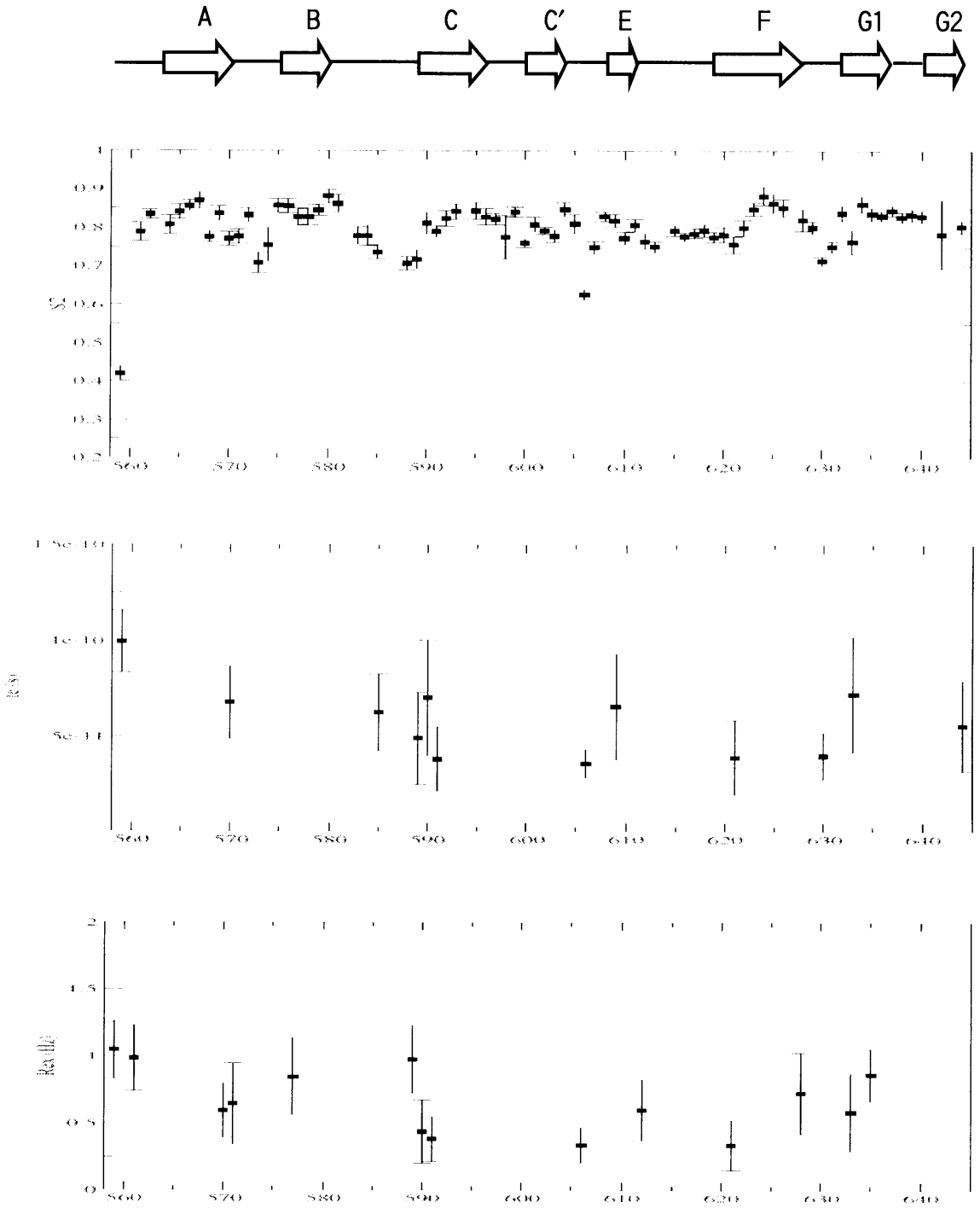


Fig. 3-18. Internal dynamics from model free analysis of relaxation data.

The S2, τ_e, and R_{ex} from model free analysis were plotted against primary sequence. Of the 76 residues subjected to the analysis, 57 were best fit to model1, 4 by model 2, 6 by model 3, 8 by model, and 1 by model. The Nε1 of Trp-580 which is not shown here is included into model 3 (see Table 3-5).

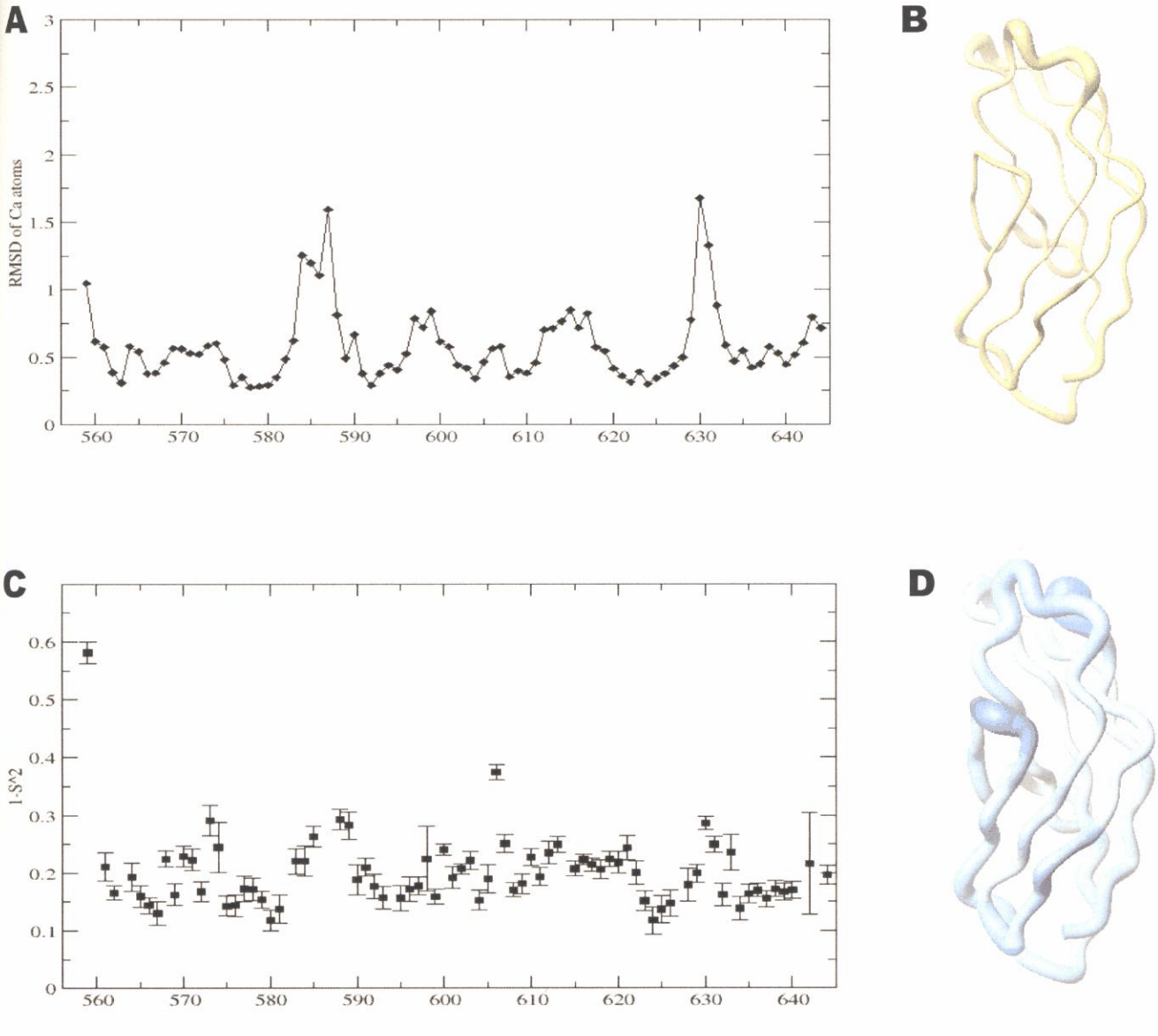


Fig. 3-19. Plots of backbone RMSD difference and 1-S^2.

The RMSDs of backbone Cα atoms in final 30 overlaid were plotted in (A) and drawn as ribbon in (B). The width of ribbon is proportional to RMSD. The values of 1-S^2 from model free analysis were plotted in (C) and drawn as ribbon in (D). The width and color of ribbon get wider and darker according to the value of 1-S^2 from 0.1 to 0.6. These figures indicate the correlation between RMSD and 1-S^2.

4. DISCUSSION

4.1 The residues under evolutional constraints are similar between the bacterial and animal FnIIIDs.

The NMR solution structure of ^{2}FnIIID$_{ChiA1}$ and multiple sequences alignment of bacterial FnIIIDs shows that these domains are surprisingly similar each other in spite of 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 evolutional pressure, such as residues on loops or exposed on β-sheets, are also preserved, as shown in Figs. 3-7 and 3-9. Although the possibility that the threonines on the surface of ^{2}FnIIID$_{ChiA1}$ have specialized functions such as those of antifreeze proteins (61, 62) cannot be excluded, 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 ^{2}FnIIID$_{ChiA}$ compared with in other bacterial FnIIIDs.

So where did the bacterium acquire FnIIID from initially? Owing to the high sequence similarity between FnIIIDs of bacteria and animals, it was suggested that the

bacterial FnIIID was transferred across phyla horizontally, and this hypothesis has been widely accepted through phylogenetic analysis (17, 21). My structural information for $^{2}FnIIID_{ChiA1}$ may increase the evidence in support of this concept. There is good correlation in hydrophobic-core-forming residues between bacterial and animal FnIIIDs and structural details such as β-bulges at the edge strands are also common in both bacterial and animal domains. To the best of my 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 (63, 64). 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 FnIIID from titin proteins, which are intracellular members of the FnIIID-containing family, is most similar to that of $^{2}FnIIID_{ChiA1}$ with an amino acid identity of 34%, the titin FnIIID is

used as an example. Fortunately, the tertiary structure of a titin module from human cardiac muscle ($FnIIID^{1BPV}$; PDB code, 1BPV) has been reported (10). $FnIIID^{1BPV}$, which shares a sequence identity of 32 % with $^{2}FnIIID_{ChiA1}$ and shows the Cα RMSD of 1.9 Å over corresponding 81 residues, reveals a very similar hydrophobic-core packing. Of the 14 residues of $^{2}FnIIID_{ChiA1}$ with a solvent-accessible surface area of less than 7%, 11 residues from $FnIIID^{1BPV}$ (including 8 identical residues: Pro-563, Leu-578, Trp-580, Tyr-592, Val-594, Phe-622, Ala-626 and Ser-637 after $^{2}FnIIID_{ChiA1}$ numbering) have the same hydrophobic feature (Fig. 4-1).

My 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 (HHM) approach, have recently revealed new FnIIIDs from yeast and plant (65, 66), despite low sequence similarities with bacterial FnIIIDs (~15% identities with the sequence of $^{2}FnIIID_{ChiA1}$). Of these, the crystal structure of the FnIIID in purple acid phosphatase from kidney bean has been reported ($FnIIID^{4KBP}$; PDB code, 4KBP) (67). $FnIIID^{4KBP}$ shares 12% sequence identity with $^{2}FnIIID_{ChiA1}$ and the Cα RMSD over 73 corresponding residues between the two domains is 2.3 Å with a Z score of 5.5. Among the $FnIIID^{4KBP}$ residues corresponding to the 14 residues that form the hydrophobic core of $^{2}FnIIID_{ChiA1}$, 8 residues (including 5

identical ones: Pro-563, Trp-580, Val-594, Ile-611 and Val-624 after $^{2}FnIIID_{ChiA1}$ numbering) reveal the same feature (Fig. 4-1).

Besides apparent similarity in their overall structure, NMR parameters indicate that there is striking similarity in the arrangement of the core-forming residues of $^{2}FnIIID_{ChiA1}$ and $FnIIID^{1BPV}$. The chemical shift of $FnIIID^{1BPV}$ is a good indicator for this structural resemblance. Whereas carbon chemical shifts reflect secondary structure, proton chemical shifts are sensitive to tertiary environments. Trp-580 and Tyr-592, which are the central residues of the FnIIID scaffold and highly conserved, make unique side-chain interactions with each other. The indole nitrogen of Trp-580 forms a stacking interaction with the aromatic ring of Tyr-592. This interaction causes an unusual upfield shift of the chemical shift of Trp-580 Hε1: the chemical shift is 5.80 ppm, which is the most upfield-shifted value for a tryptophan Hε1 in the BMRB database (Figs. 3-2 and 4-2). The values of mean and standard deviation of Trp's Hε1 in the database is 10.09 ± 0.75 ppm. Therefore 5.80 ppm is the extremely rare value to occur. Intriguingly, the same proton of $FnIIID^{1BPV}$ exhibits a very similar value of 5.81 ppm (BMRB No. 4295) (10, 68), although the chemical shift is highly sensitive to the relative orientation of the tryptophan and tyrosine residues. This result indicates that the electromagnetic environments, and thus the relative orientation, of the residues in the central

hydrophobic core of $^{2}FnIIID_{ChiA1}$ and $FnIIID^{1BPV}$ are highly similar.

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.

4.2 The structure of $^{2}FnIIID_{ChiA1}$ shows different features from other β-sandwich domains of bacterial carbohydrases.

It has been reported that the FnIIIDs of chitinase A1 from *Bacillus circulans* WL-12 are not directly involved in chitin binding. The truncated form of chitinase A1 from *Bacillus circulans* WL-12, lacking either $^{2}FnIIID_{ChiA1}$, or both $^{1}FnIIID_{ChiA1}$ 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 (27).

A structure resembling that of FnIIID has been found in another bacterial carbohydrase, chitinase A from *Serratia marcescens* (69). This enzyme consists of three domains: an N-terminal domain (ChiN), a catalytic domain, and a small (α+β) domain.

The sequence of the catalytic domain is conserved between this enzyme and chitinase A1 from *Bacillus circulans*, and their tertiary structures are similar (RMSD between corresponding Cα atoms, 1.25 Å), suggesting that they share an identical catalytic mechanism. ChiN adopts a global fold similar to $^{2}FnIIID_{ChiA1}$ but with an additional β-α-β element between the A and B strands that interacts with the other domain (Fig. 4-3). 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 (70). 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 (Figs. 3-14A and 4-3C).

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 *Serratia marcescens* that has three domains including an FnIIID (71). Although its FnIIID 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 *Bacillus 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 FnIIID have been found in bacterial carbohydrases, such as Domain N of α-amylase II (72), C-terminal starch-binding domain of β-amylase (73), cellulose-binding domain (74) and xylan-binding domain (75). The secondary structure topologies of these modules are different from that of FnIIID, although their tertiary structures show limited similarity to that of $^{2}FnIIID_{ChiA1}$ 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 (Fig. 4-4). Only the structure of $^{2}FnIIID_{ChiA1}$ does not exhibit such features.

The high-degree of conservation of the global structures among the bacterial

FnIIID-containing carbohydrases and the lack of known function of the domain and of surface features tempt us to postulate that FnIIID has the role of a spacer in bacterial carbohydrases. To degrade insoluble substrate efficiently, catalytic and binding domains must adopt various relative positions. The fact that most FnIIIDs are located between the catalytic and binding domains may support this hypothesis. Moreover, mechanical elasticity, which is an intrinsic property of FnIIID, may make the FnIIID the best candidate for this type of role.

4.3 The ^{2}FnIIID$_{ChiA1}$ shows less restricted internal motion than other FnIIIDs.

As the generalized order parameters (S^2) are gained in each HN-N vector, regardless of the diffusion tensor selected as overall tumbling motion, it can be used as good criteria for dynamics comparison of homologous proteins.

To data, the relaxation data of two FnIIIDs, the tenth FnIIID from fibronection (FNfn10, PDB code: 1FNA) and the third FnIIID from tenascin (TNfn3, PDB code: 1TEN), were reported (56). The FNfn10 reveals Cα RMSD of 1.7 over corresponding 76 residues of ^{2}FnIIID$_{ChiA1}$ with the Dali Z score of 10.4. Also TNfn3 show high similarity (Z score: 10.9, RMSD: 2.1 over 80 residues). These proteins have longer loops than ^{2}FnIIID$_{ChiA1}$. Among the loops, the RGD motif in FG loop has the affinity

with integrin and thus mediates contacts between extracellular matrix and cell-surface receptors (76).

The values of S^2 in ^{2}FnIIID$_{ChiA1}$ are smaller than those of FNfn10 and TNfn3. The average S^2 in secondary structures of ^{2}FnIIID$_{ChiA1}$ is 0.82 ± 0.03, while those of FNfn10 and TNfn3 are ~ 0.85. In spite of the difference in S^2 at regular structures, the dynamic features of the tryptophan side-chains (Nε1 of Trp) which are completely conserved in tertiary structure (Trp-580 of ^{2}FnIIID$_{ChiA1}$, Trp-22 of FNfn10, and Trp-22 of TNfn3) are nearly same. All the three side-chains can be classified into model 3. Their S^2 and R_{ex} values are 0.84 ± 0.01 and 0.7 ± 0.01 Hz for ^{2}FnIIID$_{ChiA1}$ (Table 3-5), 0.83 ± 0.01 and 1.5 ± 0.2 Hz for FNfn10, and 0.83 ± 0.02 and 0.8 ± 0.01 Hz for TNfn3. Therefore, these side-chains of the three domains experience the same degree of motional restriction. This similarity emphasizes the difference of S^2 in these proteins in regular structures. The smallness of the S^2 in ^{2}FnIIID$_{ChiA1}$ can be identified when comparing with other proteins. The average value of S^2 from Indiana Dynamics Database (IDD, pooh.chem.indiana.edu/IDD.html) is 0.839 ± 0.106 for all the data and 0.85 ± 0.07 for β structures (77) (cf. 0.80 ± 0.06 and 0.82 ± 0.03 respectively in ^{2}FnIIID$_{ChiA1}$). The simplest explanation for the phenomena may be the high content of amino acids of small side-chains in ^{2}FnIIID$_{ChiA1}$. Stone et al. performed statistical

analysis of the IDD and indicated that the amino acids of small side-chains have greater flexibility and lower S^2 (77). This interpretation can be hold in this case, even though whether it has a special role is not known at all. It is notable that FNfn10 has a remarkable core plasticity and mutation to amino acids of small side-chain don't change the free energy much (78). I am trying experiments to know the thermodynamic feature of $^2FnIIID_{ChiA1}$ and to evoke the correlation between high content of light amino acids and the feature, assuming $^2FnIIID_{ChiA1}$ as higshly mutated form into amino acids of small side-chain. The dynamic data of $^2FnIIID_{ChiA1}$ which has been reported in this thesis will contribute to further experiment as basal one.

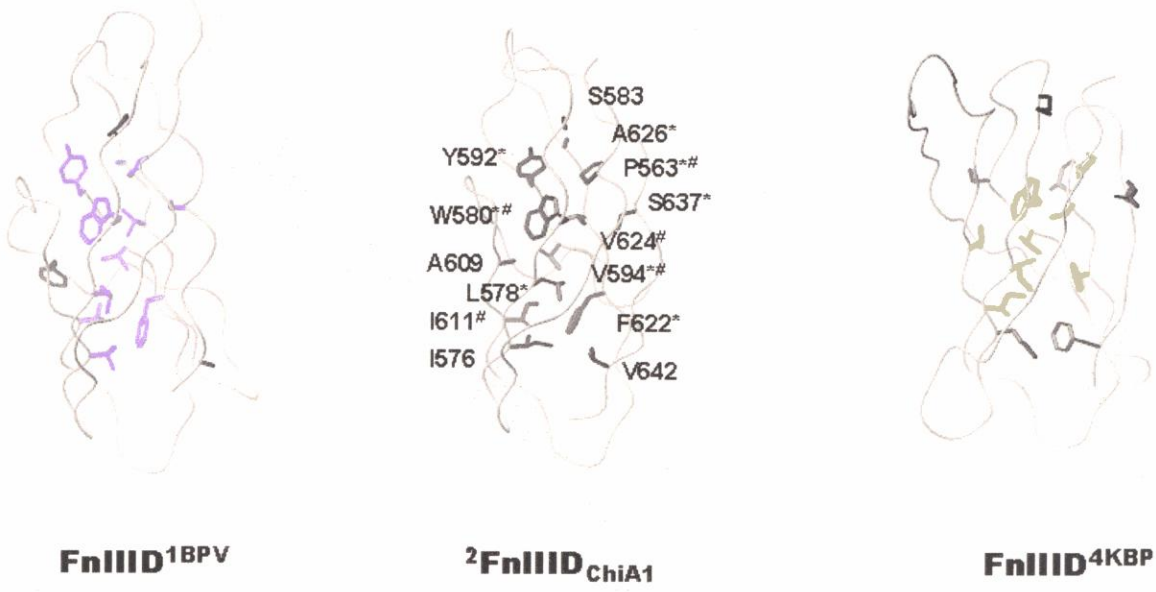


Fig. 4-1. Hydrophobic cores of ^{2}FnIIID$_{ChiA1}$, FnIIID^{1BPV} and FnIIID^{4KBP}.
The sequences of ^{2}FnIIID$_{ChiA1}$, FnIIID^{1BPV} and FnIIID^{4KBP} were aligned on the base of their structures using DALI program (52). The core-forming residues of ^{2}FnIIID$_{ChiA1}$ (Pro-563, Ile-576, Leu-578, Trp-580, Ser-583, Tyr-592, Val-594, Ala-609, Ile-611, Phe-622, Val-624, Ala-626, Ser-637 and Val-642) and their counterpart residues of FnIIID^{1BPV} and FnIIID^{4KBP} are drawn. Among the counterparts, the residues showing similar hydrophobic properties are colored violet (FnIIID^{1BPV}) and green (FnIIID^{4KBP}). The core-forming residues of ^{2}FnIIID$_{ChiA1}$ which are identical in FnIIID^{1BPV} and in FnIIID^{4KBP} are labeled with * and # respectively.

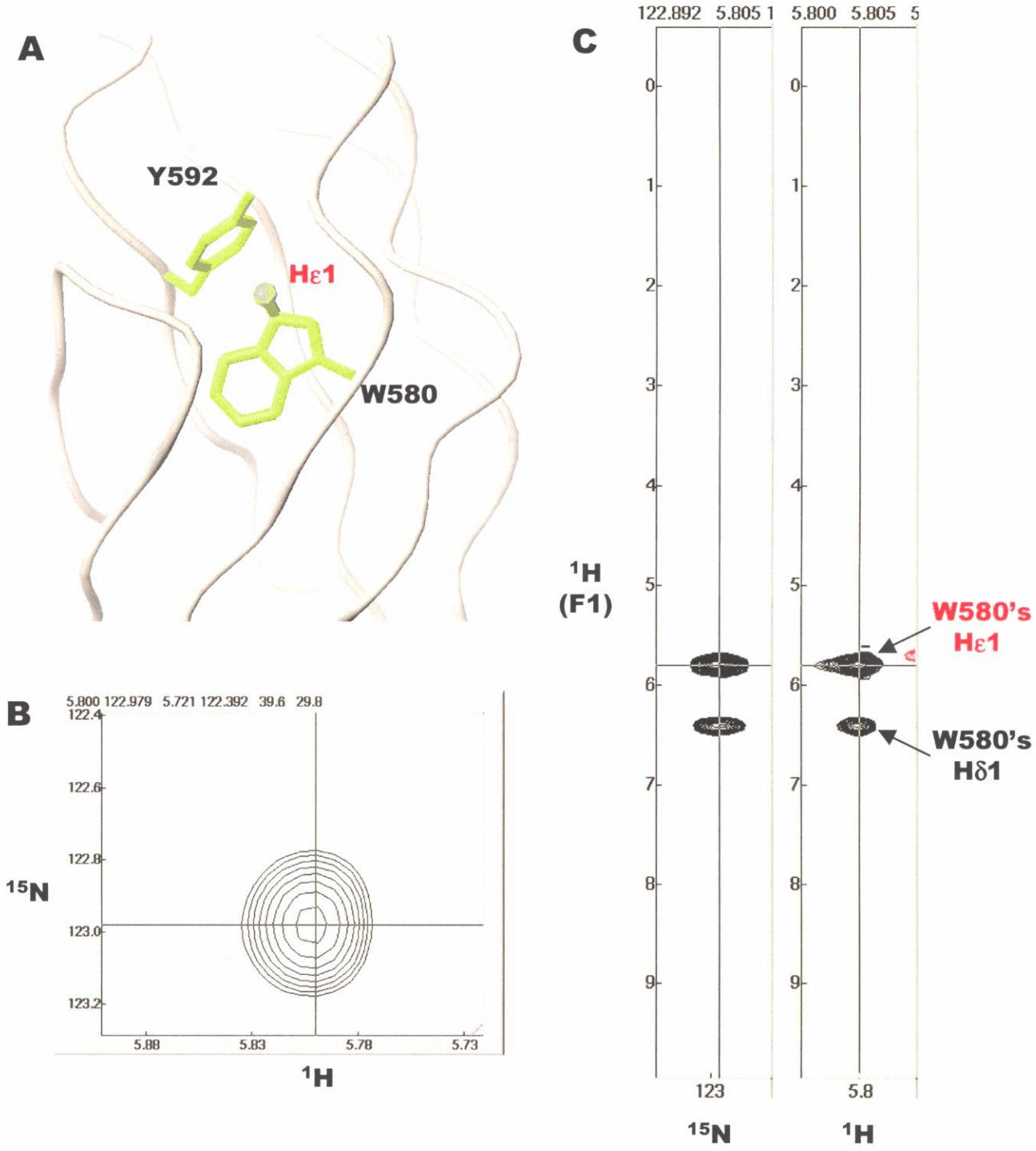


Fig. 4-2. Structure and NMR spectra of Trp-580's Hε1

A, Orientation of Trp-580's Hε1. Side chains of Trp-80 and Tyr-592 are colored green. The Hε1 atom of Trp-580 is drawn with ball and labeled red. B, ^{1}H-^{15}N HSQC spectrum of Trp-580's Hε1. C, 3D ^{15}N TOCSY-HSQC strip plot of Trp-580's Hε1. The Hε1 and Hδ1 peaks from Trp-580 are labeled red and black respectively. The observation of the Hδ1 peak through TOCSY correlation from Hε1 atom supports the assignment of Hε1.

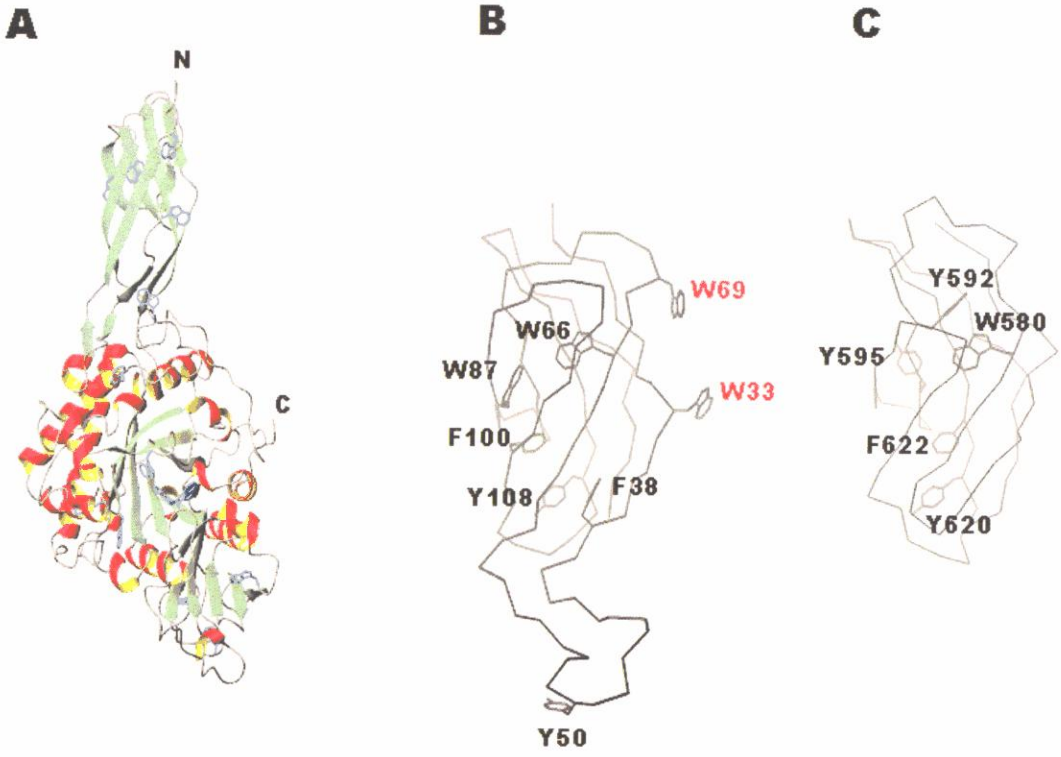


Fig. 4-3. Structural comparison of ChiN and ^{2}FnIIID$_{ChiA1}$.

(A) Ribbon diagram of chitinase A from *Serratia marcescens*. Tryptophan side chains are displayed in blue. (B, C) Cα traces of the N-terminal domain of chitinase A from *Serratia marcescens* (ChiN) (B) and ^{2}FnIIID$_{ChiA1}$ (C). Aromatic side chains are also shown. Tryptophan residues of ChiN, which are involved in the chitin interaction, are labeled in red.

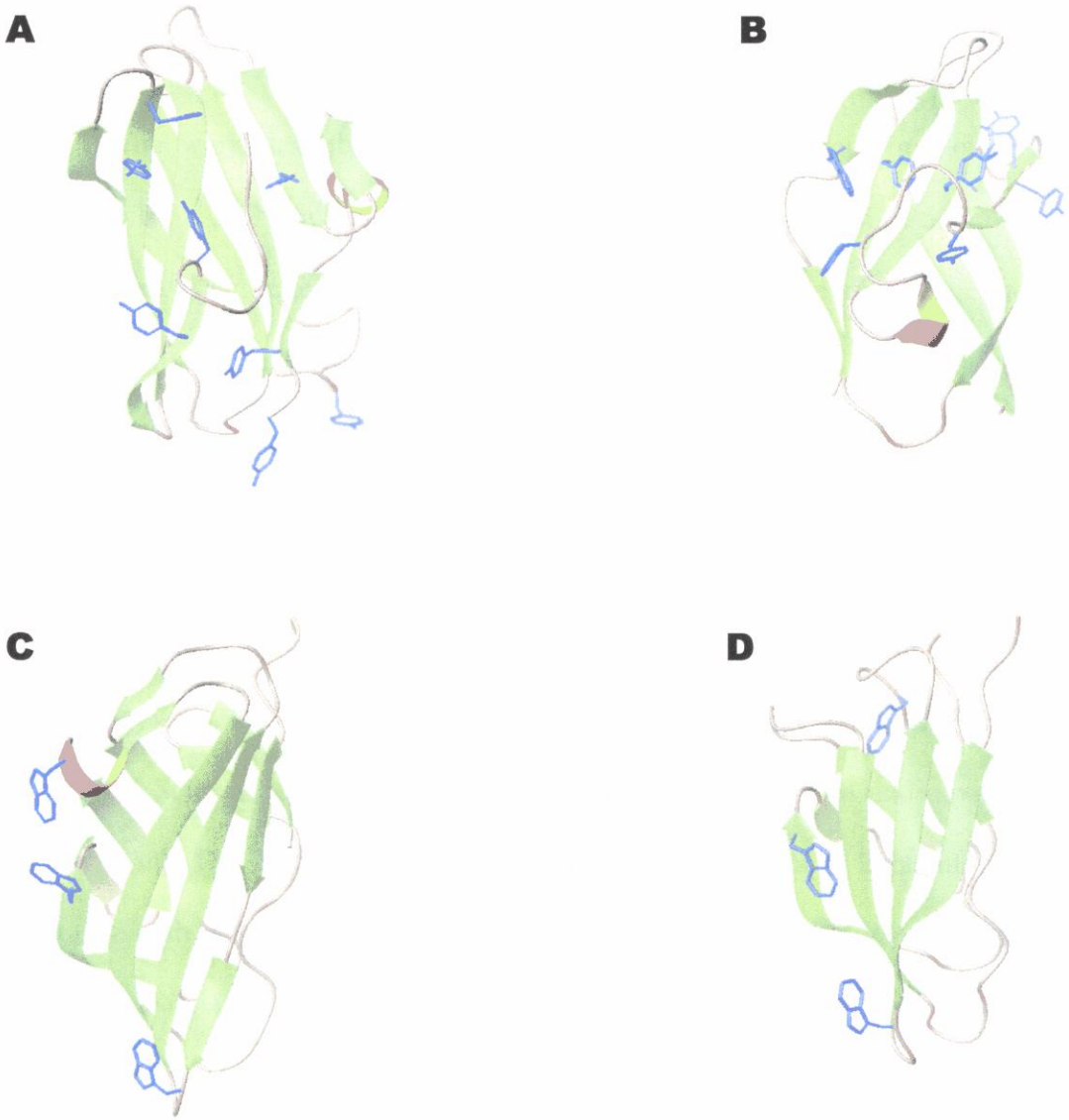


Fig. 4-4. Other bacterial β-sandwich domains.

Ribbon diagrams of Domain N of α-amylase II (A), C-terminal starch binding domain (B), cellulose-binding domain (C), and xylan-binding domain (D). For clarity, only surface-exposed aromatic residues are drawn and colored blue.

5. ACKNOWLEDGEMENTS

The work described in this thesis has been carried out in the Division of Structural Biology, Nara Institute of Science and Technology, for three and a half years. I have much to be grateful to many people who have given me knowledge, new ideas, and friendship during the period.

I would like to acknowledge and thank Prof. M. Shirakawa (Yokohama City University) for giving me a chance to study in Japan and instructing me in the course of this study. I would like to thank Prof. T. Hakoshima (Nara Institute of Science and Technology) for his encouragement. I would like to thank Prof. C. Kojima (Nara Institute of Science and Technology) for his scientific discussion and advice. I would like to thank Dr. T. Ikegami (Nara Institute of Science and Technology) for his assistance with NMR spectroscopy and useful discussion. I would like to thank Prof. T. Watanabe (Niigata University) for his nice collaboration. I would like to thank Dr. M. Mishima (Nara Institute of Science and Technology) for his support. I also express my gratitude to all the colleagues of Division of Structural Biology, Nara Institute of Science and Technology.

Finally, I would like to thank my wife, Yoon-Joo Sur, for her patience, understanding, advice, encouragement, and love.

6. REFERENCES

1. Bork, P., Downing, A.K., Kieffer, B., and Campbell, I.D. (1996) Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys.* **29**, 119-167.

2. de Vos, A.M. , Ultsch, M., and Kossiakoff, A.A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. (1992) *Science* **255**, 306-312.

3. Huber, A. H., Wang, Y. M., Bieber, A. J., and Bjorkman, P. J. (1994) Crystal structure of tandem type III fibronectin domains from *Drosophila neuroglian* at 2.0 Å. *Neuron* **12**, 717-731.

4. Leahy, D.J., Aukhil, I., and Erickson, H.P. (1996) 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* **84**,155-164.

5. Banner, D.W., D'Arcy, A., Chene, C., Winkler, F.K., Guha, A., Konigsberg, W.H., Nemerson, Y., and Kirchhofer, D. (1996) The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* **380**, 41-46.

6. Bisig, D., Weber, P., Vaughan, L., Winterhalter, K., and Piontek, K. (1999) Purification, crystallization and preliminary crystallographic studies of a two

fibronectin type-III domain segment from chicken tenascin encompassing the heparin- and contactin-binding regions. *Acta Crystallogr. Sect.D* **55**, 1069-1073.

7. Sharma, A., Askari, J. A., Humphries, M. J., Jones, E. Y., and Stuart, D. I. (1999) Crystal structure of a heparin- and integrin-binding segment of human fibronectin. *EMBO J.* **18**, 1468-1479.

8. de Pereda, J.M., Wiche, G., and Liddington, R. C. (1999) Crystal structure of a tandem pair of fibronectin type III domains from the cytoplasmic tail of integrin alpha6beta4. *EMBO J.* **18**, 4087-4095.

9. Main, A.L., Harvey, T.S., Baron, M. , Boyd, J., and Campbell, I.D. (1992) The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* **71**, 671-678.

10. Goll, C.M., Pastore, A., and Nilges, M. (1998) The three-dimensional structure of a type I module from titin: a prototype of intracellular fibronectin type III domains. *Structure* **6**, 1291-1302.

11. Copie, V., Tomita, Y., Akiyama, S. K., Aota, S., Yamada, K. M., Venable, R. M., Pastor, R. W., Krueger, S., and Torchia, D. A. (1998) Solution structure and dynamics of linked cell attachment modules of mouse fibronectin containing the RGD and synergy regions: comparison with the human fibronectin crystal structure. *J. Mol. Biol.*

277, 663-682.

12. Fattorusso, R., Pellecchia, M., Viti, F., Neri, P., Neri, D., and Wüthrich, K. (1999) NMR structure of the human oncofoetal fibronectin ED-B domain, a specific marker for angiogenesis. *Structure* **7**, 381-390.

13. Bateman, A., Jouet, M., MacFarlane, J., Du, J.S., Kenwrick, S., and Chothia, C. (1996) Outline structure of the human L1 cell adhesion molecule and the sites where mutations cause neurological disorders. *EMBO. J.* **15**, 6050-6059.

14. Oberhauser, A.F., Marszalek, P.E., Erickson, H.P., and Fernandez, J.M. (1998) The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393**, 181-185.

15. Plaxco, K.W., Spitzfaden, C., Campbell, I.D., and Dobson, C.M. (1996) Rapid refolding of a proline-rich all-beta-sheet fibronectin type III module. *Proc. Natl. Acad. Sci. USA* **93**, 10703-10706.

16. Watanabe, T., Suzuki, K., Oyanagi, W., Ohnishi, K., and Tanaka, H. (1990) Gene cloning of chitinase A1 from *Bacillus circulans* WL-12 revealed its evolutionary relationship to *Serratia* chitinase and to the type III homology units of fibronectin. *J. Biol. Chem.* **265**, 15659-15665.

17. Bork, P., and Doolittle, R.F. (1992) Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA* **89**, 8990-8994.

18. Lawrence, J.G. (1999) Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* **2**, 519-523.

19. Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.

20. Whittam, T.S. (1992) *Curr. Biol.* **2**, 676-678.

21. Little, E., Bork, P., and Doolittle, R.F. (1994) Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. *J. Mol. Evol.* **39**, 631-643.

22. Alam, M. M., Mizutani, T., Isono, M., Nikaidou, N., and Watanabe, T. (1996) *J. Ferment. Bioeng.* **82**, 28-36.

23. Watanabe, T., Oyanagi, W., Suzuki, K., and Tanaka. H. (1990) Chitinase system of *Bacillus circulans* WL-12 and importance of chitinase A1 in chitin degradation. *J. Bacteriol.* **172**, 4017-4022.

24. Matsumoto, T., Nonaka, T., Hashimoto, M., Watanabe, T., and Mitsui, Y. (1999) *Proc. Japan Acad.* **75**, 269-274.

25. Watanabe, T., Ishibashi, A., Ariga, Y., Hashimoto, M., Nikaidou, N., Sugiyama, J., Matsumoto, T., and Nonaka, T. (2001) Trp122 and Trp134 on the surface of the catalytic domain are essential for crystalline chitin hydrolysis by *Bacillus circulans* chitinase A1. *FEBS Lett.* **494**, 74-78.

26. Ikegami, T., Okada, T., Hashimoto, M., Seino, S., Watanabe, T., and Shirakawa, M. (2000) Solution structure of the chitin-binding domain of *Bacillus circulans* WL-12 chitinase A1. *J. Biol. Chem.* **275**, 13654-13661.

27. Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S., and Tanaka, H. (1994) The roles of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation. *J. Bacteriol.* **176**, 4465-4472.

28. Kay, L.E. (1998) Protein dynamics from NMR. *Nat. Struct. Biol.* Suppl. 513-517.

29. Ishima, R., and Torchia, D.A. (2000) Protein dynamics from NMR. *Nat. Struct. Biol.* **9** 740-743.

30. Palmer, A.G. III (2001) NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **30** 129-155.

31. Cavanagh, J., Fairbrother, W.J., Palmer, A.G. III, and Skelton, N.J. (1996) *Protein NMR Spectroscopy*, Academic Proess, San Diego.

32. Cornilescu, G., Hu, J.-S., and Bax, A. (1999) Identification of the Hydrogen Bonding Network in a Protein by Scalar Couplings. *J. Am. Chem. Soc.* **121**, 2949-2950.

33. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J.*

Biomol. NMR **6**, 277-293.

34. Johnson, B.A., and Blevins, R.A. (1994) NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603-614.

35. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283-298.

36. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biolmol. NMR* **13**, 289-302.

37. Hu, J.-S., Grzesiek, S., and Bax, A. (1997) Two-Dimensional NMR Methods for Determining χ_1 Angles of Aromatic Residues in Proteins from Three-Bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ Couplings. *J. Am. Chem. Soc.* **119**, 1803-1804.

38. Linge, J.P., O' Donoghue, S.I., and Nilges, M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods in Enzymology* **339**, 71-90.

39. Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477-486.

40. Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., and Bork, P. (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231-234.

41. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

42. Goodstadt, L., and Ponting, C.P. (2001) CHROMA: consensus-based colouring of multiple alignments for publication. *Bioinformatics* **9**, 845-846.

43. Säli, A., and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815.

44. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

45. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51-55.

46. Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Clarendon, Oxford

47. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theroy and range of

validity. *J. Am. Chem. Soc.* **104**, 4546-4559.

48. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* **104**, 4559-4570.

49. Woessner, D.E. (1962) Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. *J. Chem. Phys.* **37**, 647-654.

50. Tjandra, N., Feller, S.E., Pastor, R.W., and Bax, A. (1995) Rotational Diffusion Anisotropy of Human Ubiquitin from 15N NMR Relaxation. *J. Am. Chem. Soc.* **117**, 12562-12566.

51. Tjandra, N., Wingfield, P., Stahl, S., and Bax, A. (1996) Anisotropic rotational diffusion of perdeuterated HIV protease from 15N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR.* **3**, 273-284.

52. Mandel, A.M., Akke, M., and Palmer, A.G. III (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**, 144-163.

53. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied

by 15N NMR relaxation. *Biochemistry* **33**, 5984-6003.

54. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) *Numerical Recipes in C*, Cambridge University Press, New York.

55. Dosset, P., Hus, J.C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* **16**, 23-28.

56. Carr, P.A., Erickson, H.P., and Palmer, A.G. III (1997) Backbone dynamics of homologous fibronectin type III cell adhesion domains from fibronectin and tenascin. *Structure* **5**, 949-959.

57. Koide, A., Bailey, C.W., Huang, X., and Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol.* **284**, 1141-1151.

58. Wootton, J.C. (1994) Sequences with 'unusual' amino acid compositions. *Curr. Opin. Struct. Biol.* **4**, 413-421.

59. Bairoch, A., and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45-48.

60. LeMaster, D.M. (1999) NMR relaxation order parameter analysis of the dynamics of protein side chains. *J. Am. Chem. Soc.* **121**, 1726-1742.

61. Liou, Y.C., Tocilj, A., Davies, P.L., and Jia, Z. (2000) Mimicry of ice structure by

surface hydroxyls and water of a beta-helix antifreeze protein. *Nature* **406**, 322-324.

62. Graether, S.P., Kuiper, M.J., Gagne, S.M., Walker, V.K., Jia, Z., Sykes, B.D., and Davies, P.L. (2000) Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature* **406**, 325-328.

63. Nei, M. *Molecular Evolutionary Genetics* (1987) Columbia Univ. Press, New York.

64. Li, W.-H. *Molecular Evolution* (1997) Sinauer, Sunderland, MA.

65. Bateman, A., and Chothia, C. (1996) Fibronectin type III domains in yeast detected by a hidden Markov model. *Curr. Biol.* **6**, 1544-1547.

66. Tsyguelnaia, I., and Doolittle, R.F. (1998) Presence of a fibronectin type III domain in a plant protein. *J. Mol. Evol.* **46**, 612-614.

67. Strater, N., Klabunde, T., Tucker, P., Witzel, H., and Krebs, B. (1995) Crystal structure of a purple acid phosphatase containing a dinuclear Fe(III)-Zn(II) active site. *Science* **268**, 1489-1492.

68. Muhle-Goll, C., Nilges, M., and Pastore, A. (1997) 1H and 15N NMR resonance assignments and secondary structure of titin type I domains. *J. Biomol. NMR* **9**, 2-10.

69. Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S., and Vorgias, C. E. (1994) Crystal structure of a bacterial chitinase at 2.3 Å resolution. *Structure* **2**, 1169-1180.

70. Uchiyama, T., Katouno, F., Nikaidou, N., Nonaka, T., Sugiyama, J., and Watanabe, T. (2001) Roles of the exposed aromatic residues in crystalline chitin hydrolysis by chitinase A from *Serratia marcescens* 2170. *J. Biol. Chem.* **276**, 41343-41349.

71. Suzuki, K., Taiyoji, M., Sugawara, N., Nikaidou, N., Henrissat, B., and Watanabe, T. (1999) The third chitinase gene (chiC) of *Serratia marcescens* 2170 and the relationship of its product to other bacterial chitinases. *Biochem. J.* **343**, 587-596.

72. Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonozuka, T., and Sakano, Y. (1999) Crystal structure of *Thermoactinomyces vulgaris* R-47 alpha-amylase II (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. *J. Mol. Biol.* **287**, 907-921.

73. Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R., and Utsumi, S. (1999) Structure of raw starch-digesting *Bacillus cereus* beta-amylase complexed with maltose. *Biochemistry* **38**, 7050-7061.

74. Xu, G.Y., Ong, E., Gilkes, N.R., Kilburn, D.G., Muhandiram, D.R., Harris-Brandts, M., Carver, J.P., Kay, L.E., and Harvey, T.S. (1995) Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry* **34**, 6993-7009.

75. Simpson, P.J., Bolam, D.N., Cooper, A., Ciruela, A., Hazlewood, G.P., Gilbert, H.J.,

and Williamson, M.P. (1999) A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. *Structure Fold. Des.* **7**, 853-864.

76. D'Souza, S.E., Ginsberg, M.H., and Plow, E.F. (1991) Arginyl-glycyl-aspartic acid (RGD): a cell adhesion motif. *Trends Biochem. Sci.* **16**, 246-250.

77. Goodman, J.L., Pagel, M.D., and Stone, M.J. (2000) Relation between protein structure and dynamics from a database of NMR-derived backbone order parameters. *J. Mol. Biol.* **295**, 963-978.

78. Cota, E., Hamill, S.J., Fowler, S.B., and Clarke, J. (2000) Two proteins with the same structure respond very differently to mutation: the role of plasticity in protein stability. *J. Mol. Biol.* **302**, 713-725.

79. Wishart, D.S., and Sykes, B.D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR.* **2**, 171-180.