The structure of *Physarum polycephalum* hemagglutinin I suggests a minimal carbohydrate recognition domain of legume lectin fold

Takahide Kouno[1,*], Nobuhisa Watanabe[2,3,4], Naoki Sakai[4], Takashi Nakamura[1], Yuko Nabeshima[1], Masashi Morita[5], Mineyuki Mizuguchi[1], Tomoyasu Aizawa[6], Makoto Demura[4,6], Tsuneo Imanaka[5], Isao Tanaka[4,6], & Keiichi Kawano[6].

[1]Department of Structural Biology, Graduate School of Medicine and Pharmaceutical Sciences, University of Toyama, Toyama 930-0194, Japan

[2]Synchrotron Radiation Research Center, Nagoya University, Nagoya 464-8603, Japan

[3]Department of Biotechnology, Graduate School of Engineering, Nagoya University, Nagoya 464-8603, Japan

[4]Faculty of Advanced Life Science, Hokkaido University, Sapporo 060-0810, Japan

[5]Department of Biological Chemistry, Graduate School of Medicine and Pharmaceutical Sciences, University of Toyama, 2630 Sugitani, Toyama 930-0194, Japan

[6]Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

*Corresponding author; Department of Structural Biology, Graduate School of Medicine and Pharmaceutical Sciences, University of Toyama, Toyama 930-0194, Japan, E-mail: konox005@umn.edu, Telephone: +81-76-434-7595, Fax: +81-76-434-5061

Abstract

Physarum polycephalum hemagglutinin I is a 104-residue protein that is secreted to extracellular space. The crystal structure of hemagglutinin I has a β-sandwich fold found among lectin structures, such as legume lectins and galectins. Interestingly, the β-sandwich of hemagglutinin I lacks a jelly roll motif and is essentially composed of two simple up-and-down β-sheets. This up-and-down β-sheet motif is well conserved in other lectins derived from animals, plants, bacteria, and viruses. It is more noteworthy that the up-and-down β-sheet motif includes many residues that make contact with the target carbohydrates. Our NMR data demonstrate that hemagglutinin I lacking a jelly roll motif also binds to its target glycopeptide. Taken together, the up-and-down β-sheet motif provides a fundamental scaffold for the binding of legume lectin-like proteins to the target carbohydrates, and the structure of hemagglutinin I suggests a minimal carbohydrate recognition domain.

Keywords

Carbohydrate recognition domain, jelly roll motif, lectin, β-sandwich fold, up-and-down β-sheet.

Introduction

Lectins are a large family of carbohydrate-binding proteins found widely in viruses, bacteria, plants, and animals.[1-3] They recognize a specific carbohydrate structure via their carbohydrate recognition domains (CRDs), which are involved in a variety of biological events, such as cell adhesion, differentiation, and host/pathogen recognition. A plant legume lectin has been well known as a model of a CRD-carbohydrate complex that explains how the lectin structure recognizes a sugar chain.[2,4] The legume lectin fold is characterized by a β-sandwich structure consisting of three β-sheets, called "top", "front", and "back" sheets (Figure 1). Two β-sheets, a curved front and a flat back sheet, face each other to form a sandwich, and a relatively small top sheet covers one side of the sandwich. A sugar chain is commonly captured by loop regions between the front and top sheets and/or the surface of the front β-sheet.[4] The topology of β-strands in the CRD structure is conserved among many plant and animal lectins, whereas the loops connecting those β-strands have different lengths. This diversity of loops is likely to determine the specificity of CRD-carbohydrate interaction.

Galectins are animal lectins with a binding affinity for β-galactosides. They have a CRD structure similar to the legume lectin fold, except that they lack the top β-sheet.[4,5] The sugar binding motif of galectin is found on the surface of the front β-sheet and loops connecting the β-strands. The entire β-sandwich structure of galectins bends toward the front β-sheet so that it can hold target carbohydrates. In addition to galectin-carbohydrate interactions, a member of the galectin family, galectin-1, is also involved in protein-protein interactions, such as Ras.[5,6] The protein-protein interactions of galectin-1 occur in cells independently of its lectin activity, whereas the binding of galectin-1 to β-galactosides, i.e., its lectin activity, is found in extracellular spaces.[5]

We identified a 104-residue lectin, called hemagglutinin I (HA1), from a slime mold,

Physarum polycephalum.[7,8] *Physarum* species have a unique life cycle including unicellular amoeba, multinucleated plasmodia, and fruiting bodies, and so has been used as a model system for studies of the eukaryotic mitotic cycle, cell differentiation, and motility.[9]

HA1 is initially found in the intracellular soluble fraction, and is secreted to extracellular space after the plasmodia reach a stationary phase.[8] Although the physiological function of the secreted HA1 remains to be established, HA1 recognizes cell wall polysaccharides of *E. coli*.[10] As the name implies, HA1 agglutinates rabbit erythrocytes, which is inhibited in the presence of glycoproteins, orosomucoid, fetuin, and thyroglobulin.[8] Whereas periodate oxdation of the glycoproteins disrupts the binding of HA1 completely.[10] Therefore, HA1 is likely to recognize sugar chain moieties of glycoproteins. In addition, a binding assay using thyroglobulin indicated that the HA1-carbohydrate interaction requires residues T49 to Y55 of HA1.[11]

In this study, we determined the crystal structure of HA1 and measured nuclear magnetic resonance (NMR) spectra with or without an N-glycosylated peptide from fetuin. Like galectins and plant legume lectins, HA1 adopts a β-sandwich structure. However, the number of β-strands in HA1 is smaller than that of any other lectin with a β-sandwich structure. The NMR data identified glycopeptide-binding sites on the HA1 β-sandwich structure, and suggested that the HA1 tetramer binds to four glycoprotein molecules simultaneously. Our data can explain the hemagglutinating activity of HA1, and propose a minimal CRD fold required for binding to carbohydrates.

Results

Structure of HA1—The crystal structure of HA1 is well-defined at 1.82 Å resolution with an R-free factor of 19.4% (Table 1). The HA1 structure was determined as a tetramer

containing chains 1-4 (Figure 2A). The backbones of chains 1 and 2 are almost completely superimposed on those of chains 4 and 3, respectively. Whereas, chains 1 and 2 (chains 3 and 4) have a different conformation over residues P21-S25. Except for inconsistency in the local conformation, all four chains share a common secondary structure element. The structure of HA1 monomer includes two β-sheets composed of four (β1-β7-β8-β9) and five strands (β2-β3-β4-β5-β6) (Figure 2B). Only the β1 and β7 strands are arranged in a parallel orientation, whereas the others form antiparallel β-sheets. The two β-sheets of HA1 form a sandwich and are highly twisted. Chains 1 and 3 (chains 2 and 4) assemble at an angle to form a concave between them (Figure 2C). Residues G23 and S25 of chain 1 make intermolecular hydrogen bonds with I7 and D9 in strand β1 of chain 3, resulting in the formation of a pseudo-continuous β-sheet (Figure 2E). In contrast, residues G23 and S25 of chain 3 are distant from the neighboring chain 1 to form the concave face (Figure 2C). This asymmetrical assembly of chains 1 and 3 is reflected in the local structural divergence over residues P21-S25, as described above. On the other hand, chains 1 and 2 (chains 3 and 4) are arranged in an antiparallel manner through their β-sandwich surfaces (Figure 2D). The side chains of hydrophobic residues, including Y20 (β2) and Y31 (β3), are exposed on the surface of the HA1 monomer structure (Figure 2F) and form the intermolecular interface. These hydrophobic interactions are likely to participate in the formation of a symmetric homodimer between chains 1 and 2 (chains 3 and 4).

HA1 tetramer formation in solution—The crystal structure of HA1 was determined as a tetramer. To explore the tetramer formation of HA1 in solution, dynamic light scattering (DLS) experiments were carried out with a range of concentrations (5 μM to 1 mM) of HA1. At concentrations less than 30 μM, HA1 showed a relatively broad distribution of

hydrodynamic radii (Figure 3). When the concentration was increased, the fraction of shorter radius was decreased, and HA1 showed a distribution with a sharper peak at 2.7 nm at higher concentrations. The distribution profile at 1 mM concentration is superimposed on that at 100 μM, suggesting that HA1 assembly reaches a plateau. Additionally, we measured the hydrodynamic radius of HA1 at a concentration of 1 mM by pulse field gradient (PFG) NMR.[12] The hydrodynamic radius of HA1 was obtained from a comparison of signal decay rates between HA1 and a standard substance, dioxan, and estimated to be 2.9 nm. This value is consistent with the result from DLS analysis. Wilkins *et al*. reported a relationship between hydrodynamic radii (R_h) of native folded proteins and the number of residues (N); $R_h = (4.75 \pm 1.11)N^{0.29 \pm 0.02}$.[12] According to this equation, the hydrodynamic radii of monomer, dimer, and tetramer of HA1 are expected to be 1.8, 2.2, and 2.7 nm, respectively. In addition, the similar values are estimated with HYDROPRO software[13] on the basis of the coordinate data of the HA1 structure; the hydrodynamic radii of monomer, dimers (chains 1-2 and 1-3), and tetramer of HA1 are expected to be 1.9, 2.3-2.5, and 3.2 nm, respectively. Considering that HA1 shows the same distribution of hydrodynamic radii above 100 μM, and that the hydrodynamic radius of HA1 measured by PFG-NMR is closest to the predicted value for the tetramer, the major part of HA1 is most likely to form a tetramer in solution at concentrations above 100 μM.

NMR experiments—We measured a set of NMR spectra using stable isotope-labeled HA1s and assigned almost all NMR signals on the HSQC spectrum. The resonance assignment revealed that the HSQC spectrum lacks signals of the N-terminal segment, i.e., residues 1-10 and 15-33 of HA1. In the crystal structure of the HA1 tetramer, these residues are located in the interface between chains 1 and 3 (2 and 4), and so they are likely to experience heterogeneous folding and/or dynamics in solution, resulting in severe signal

broadening. In addition, preliminary surface plasmon resonance experiments indicated that an HA1 mutant lacking the first 32 residues is able to bind to a fetuin-derived glycopeptide like full-length HA1 (Figure S1 in Supplemental Information). The missing segment on the HSQC spectrum is not essential for the binding of HA1 to the target carbohydrates.

When the fetuin-derived glycopeptide was added to a solution of HA1, HSQC signals of HA1 were broadened without significant chemical shift perturbations (Figure 4A). The half-widths of some signals were significantly increased more than 1.5-fold. Especially, the signals derived from residues E44, S53, W54, Y55, I56, Y70, E71, G72, and G73 disappeared due to signal broadening when the ratio of HA1 to the glycopeptide reached 1:1 (Figure 4B). The attenuation of NMR signals occurred in a glycopeptide-dependent manner, which means that HA1 interacts with the glycopeptide through a specific region. However, the binding of the glycopeptide does not change the overall HA1 structure itself because HSQC signals of HA1 appear at the same position before and after the addition of the glycopeptide.

The HA1-glycopeptide interface can be identified when the residues showing significant broadening are plotted on the HA1 crystal structure (Figure 4C and 4D). They are located in loops, not on the flat surface, of the β-sandwich structure of HA1, resulting in the formation of a surface patch around loops β4/β5 and β6/β7. This patch includes all residues that disappeared in the presence of an equimolar amount of the glycopeptide, which strongly supports the hypothesis that HA1 binds to the glycopeptide through this surface patch. Interestingly, the signal broadening of residues V34, L98, T99, and L103 implies that the HA1-glycopeptide interface extends to a neighboring monomer. Chains 1 and 2 (3 and 4) appear to form a continuous binding patch on one side of the dimer structure (Figure 4C). The same situation occurs on the opposite side because chains 1

and 2 (3 and 4) are arranged in an antiparallel manner. Therefore, the HA1 dimer composed of chains 1 and 2 possesses two binding sites located at the upper and lower edges of the β-sandwich structure. On the other hand, the binding patch around loops β4/β5 and β6/β7 of chain 1 is far from residues V34, L98, T99, and L103 of chain 3, and so chains 1 and 3 (2 and 4) are unlikely to form a continuous binding surface (Figure 4D). DLS and PFG-NMR indicate that HA1 forms a tetramer in solution. The HA1 tetramer includes two dimers, meaning that the tetramer has four binding patches. The crystal structure of HA1 shows that all of these four binding patches are located on the surface (Figure 5), and so HA1 appears to be able to capture four target carbohydrates simultaneously.

Tryptophan fluorescence of HA1—The tryptophan fluorescence of HA1 was measured with or without the fetuin-derived glycopeptide (Figure S2). In the absence of the glycopeptide, the emission profile gave a peak at 345 nm. When the glycopeptide was added into the HA1 solution, the peak maximum shifted significantly toward a shorter wavelength, and reached a plateau when equimolar glycopeptide was added. HA1 possesses two tryptophan residues, W3 and W54. In the crystal structure of HA1, residue W3 is located in the interior of the HA1 monomer structure, whereas W54 is included in the loop β4/β5 and the side chain of W54 is exposed to solvent (Figure 5). Considering that the NMR experiment using HA1 demonstrated almost no change in the overall structure of HA1 with or without the glycopeptide, a shift of the tryptophan emission peak indicates a change in the solvent accessibility of residue W54. Therefore, when the glycopeptide binds to HA1, the side chain of W54 is most likely to be overlaid with the glycopeptide molecule and lose solvent accessibility.

Discussion

The HA1 structure, determined using X-ray crystallography, includes a β-sandwich fold composed of five-stranded and four-stranded β-sheets. To investigate the structural similarity between HA1 and other proteins, we used the DALI database.[14] As expected, the structural homology search picked some carbohydrate-binding proteins out of the Protein Data Bank; for example, galectin-1[15] and virus capsid protein VP4.[16] In addition, we compared the HA1 structure with a set of carbohydrate-binding module (CBM) families in the CAZy database.[17] We found out that the structure of CBM family 40, such as the CBM of bacterial sialidase NanJ, has a β-sandwich structure similar to HA1.[18] As shown in Figure 6, the secondary structure elements of HA1 are compared to those of galectin-1, a peanut legume lectin,[19] CBM of NanJ, and VP4, which are derived from animals, plants, bacteria, and viruses, respectively. The comparison of the β-strand arrangement among these carbohydrate-binding proteins reveals that their structures are commonly divided into two kinds of β-sandwiches: a jelly roll motif (labeled using Arabic numbers in Figure 6) and a complex of simple up-and-down β-sheets (labeled using Roman numbers). The up-and-down β-sheets is well conserved among the proteins (Figure 6) that are composed of four-stranded and three-stranded β-sheets. Only VP4 has additional β-strands II' and III' between β-strands II and III. In contrast, the region of a jelly roll motif is likely to have a relatively higher diversity than the up-and-down β-sheet region. Peanut legume lectin has some additional strands 2', 2'', and 2''', whereas galectin-1 lacks strand 1. Interestingly, HA1 has only two strands in the region corresponding to the jelly roll motif, and appears to lack the jelly roll motif (Figure 6A). Therefore, we can conclude that the HA1 structure is essentially based on a β-sandwich composed of two up-and-down β-sheets and has a simpler architecture than the other carbohydrate-binding proteins.

NMR experiments using HA1 and fetuin-derived glycopeptide identified the binding

patches on the HA1 structure (Figure 4C, 4D, and 5). They are located at the lower and upper edges, not on the flat surface, of the β-sandwich structure, suggesting that HA1 binds to a fetuin-derived glycopeptide through loop regions. Specifically, a relatively long loop β4/β5 (II/III) provides a large surface area and is likely to play a major role in the interaction. In addition, this loop includes residue W54. The tryptophan quenching experiment showed that the solvent accessibility of this residue changed dramatically when a fetuin-derived glycopeptide was added, suggesting that the loop β4/β5 (II/III) of HA1 binds directly to the glycopeptide. This idea is consistent with previous data showing that residues T49 to Y55 of HA1 are essential for HA1-glycoprotein interaction.[11] In general, some aromatic residues, especially tryptophans and tyrosines, of CBMs form a hydrophobic surface patch for binding to the pyranose rings of sugar chains on the target carbohydrates.[20-23] In addition, hydroxyl groups of carbohydrates form hydrogen bonds with polar residues of CBMs. These interactions are also found in other lectin/carbohydrate complexes such as galectin-1/galactoside[15] and VP4/sialoside.[16] HA1 has aromatic amino acid residues W54 and Y55 and polar residues S43, E44, S53, E71, and T99 on the surface of the binding patch (Figure 5). These residues will help HA1 to capture the target glycoproteins.

Furthermore, it is noteworthy that the carbohydrate-binding surface of HA1 expands to the neighboring monomer. The crystal structure of HA1 tetramer indicates that the assembly of monomers is heterogeneous, i.e., chains 1 and 2 (3 and 4) interact with each other through the flat surface of the β-sandwich structure, whereas chains 1 and 3 (2 and 4) associate through edge-to-edge interaction of the β-sandwich (Figure 2). In the dimer composed of chains 1 and 2, the C-terminus of chain 2 is close to the loop β4/β5 (II/III) of chain 1 (Figure 2C and 2D). Thus, HA1 multimerization can explain formation of the continuous binding patch. DLS and PFG-NMR analyses indicated that HA1 forms a

tetramer at a concentration at which the NMR titration was carried out. When HA1 concentration was decreased to less than 100 μM, a fraction of shorter hydrodynamic radius was increased, implying that HA1 is in equilibria between the monomer/dimer and tetramer at lower concentrations. HA1 dimer in solution has two potential combinations, i.e., a combination of chains 1 and 2 or chains 1 and 3. In the crystal structure, the interface area between chains 1 and 3 (2 and 4) is estimated to be ~310 $Å^2$, which is much less than that between chains 1 and 2 (3 and 4) (~850 $Å^2$). Additionally, PISA analysis[22] indicated that only the dimer composed of chains 1 and 2 (3 and 4) is potentially biologically relevant. These features suggest that HA1 dimer is composed of chains 1 and 2, and still achieves formation of the extended binding surface even at lower concentrations. This idea will explain a preliminary result that HA1 binds to a glycoprotein with a relatively high affinity; the apparent association constant value is 4.5×10^7 M^{-1}.

As described above, HA1 binds to a glycopeptide through the loop regions, especially loop β4/β5 (II/III), of the β-sandwich fold composed of up-and-down β-sheets. When the binding site of HA1 is compared with the CRDs of the other carbohydrate-binding proteins, we see that the region of the up-and-down β-sandwich includes the majority of residues that make direct contact with a ligand sugar chain (Figure 6). In addition, those residues are predominantly found on the "front" sheet region of the β-sandwich except for VP4. VP4 includes an additional β-sheet II'/III' covering the front sheet surface and has unique binding sites.[16] In the structures of peanut legume lectin and CBM of NanJ, some residues in the jelly roll motif also participate in binding to carbohydrates, which is likely to increase the affinity and specificity. As a whole, the up-and-down β-sandwich, particularly the front β-sheet region, appears to provide a scaffold for the binding of carbohydrate-binding proteins to their ligands. In other words, the up-and-down β-sandwich is a minimal CRD in legume lectin fold. This is consistent with the finding that an HA1 lacking the first 32

residues still maintains the binding to glycoprotein fetuin (Figure S1).

In general, lectins form multimers. For example, peanut legume lectin and galectin-1 form a homotetramer and a homodimer, respectively.[25,26] These self-multimerization properties are expected to cause their agglutination. The DLS and PFG-NMR data indicated that HA1 also forms dimer and tetramer in solution. It is likely that HA1 dimer is composed of chains 1 and 2 via hydrophobic interactions, including the side chains of Y20 (β2) and Y31 (β3), i.e., HA1 dimerizes through the "front" sheet surface of the β-sandwich structure. In the cases of galectin-1 and CBM of NanJ, the front β-sheet surface provides a binding interface with sugar chains, whereas HA1 binds to glycopeptides through the edge of the β-sandwich (Figure 6). This situation in HA1 is similar to that in VP4 in which the front sheet surface of VP4 makes contact with an additional β-sheet II'/III', not carbohydrates.[16] Instead, the ligand-binding site of VP4 is located at the edge of the β-sandwich structure. HA1 binds to carbohydrates through the loop regions, even if the β-sheet surface is occupied by the neighboring monomer due to the dimerization. Moreover, tetramerization is also unlikely to disrupt the binding of HA1 to the target carbohydrates because the binding patches of HA1 are still located on the surface in the tetramer (Figure 5). These features suggest that the dimerization and tetramerization enable HA1 to act as a linker protein between ligands possessing sugar chains, such as glycoproteins and hemocytes, to form hemagglutinations. Thus, our data demonstrate that HA1 has a rational structure to explain its carbohydrate-binding and hemagglutinating activities, and propose a minimal CRD structure, i.e., a β-sandwich structure composed of two up-and-down β-sheets.

Materials and Methods

Preparation of recombinant HA1—The recombinant HA1 was expressed in *E. coli* BL21

(DE3) pLysS (Novagen) and purified as described previously.[7] To prepare the ^{15}N-, $^{13}C/^{15}N$-, and $^{2}H/^{13}C/^{15}N$-labeled proteins, the cells were cultivated in M9 media including $^{15}NH_4Cl$, [^{13}C]-glucose, and/or $^{2}H_2O$. After purification using a DEAE-sepharose Fast Flow column (GE Healthcare) and a thyroglobulin-agarose column (Sigma Chemical), the HA1 solution was dialyzed against 20 mM sodium acetate (pH 5.5). The purity of HA1 was confirmed by SDS-PAGE and mass spectrometry.

Preparation of an N-linked glycopeptide—An N-linked glycopeptide was prepared according to the protocol described by Rice et al.[27] The target glycopeptide derived from fetuin (Sigma Chemical) was finally purified by reverse-phase high-pressure liquid chromatography using a C-18 column. To identify the fraction containing the target glycopeptide, 100 μl of each fraction was assayed by the phenol-sulfuric acid assay.[28] The purified glycopeptide was sequenced by Edman degradation (PPSQ-21 protein sequencer, Shimadzu).

Crystallization and data collection—Crystals of HA1 were grown by the hanging drop vapor diffusion method over a reservoir of 10-25% (v/v) PEG3350, 150 mM NaCl, 0.1 M MES pH 6-7. Drops were formed by mixing 1 μl of HA1 solution (18.0 mg/ml in 20 mM Na-acetate pH 5.5) and 1 μl of reservoir solution, with equilibration over the reservoir at 293 K. Iodinated crystal for the SAD phasing was grown under the same conditions except NaI was used as a substitute for NaCl. The diffraction data were collected using an in-house X-ray source (Rigaku FR-E SuperBright with a Cu/Cr dual target) and a Rigaku R-AXIS VII imaging-plate detector, and the SAD data were collected using the capillary-top crystal mounting method.[29] The collected intensities were indexed, integrated, corrected for absorption, and scaled using HKL2000.[30]

Structure determination of HA1—Initial phases were determined at 2.68 Å resolution by the SAD method using NaI crystal data. Iodine positions were determined using SELEXC and SHELXD[31] using the HKL2MAP interface,[32] and SOLVE was used to refine the iodine sites and to calculate the initial SAD phases.[33] Further phase improvement and auto-model building were carried out using RESOLVE[34] and REFMAC5[35] using the RESOLVE_BUILD script. With this partial structure model, the initial phases of the 1.82 Å resolution NaCl crystal data were determined using the molecular replacement program MOLREP.[36] The structure of HA1 was then built by ARP/wARP,[37] manually fit with Coot[38] and refined by REFMAC5. The crystallographic parameter and refinement statistics are summarized in Table 1.

NMR spectroscopy—The samples for NMR experiments contained 1 mM HA1 (^{15}N-, ^{13}C/^{15}N-, or ^{2}H/^{13}C/^{15}N-labeled proteins), 20 mM sodium phosphate (pH 5.5), and 10% D_2O. To achieve the backbone resonance assignments, the following TROSY-type experiments were performed at 313 K on a Bruker DMX500 spectrometer: HNCA, HN(CO)CA, HNCO, HN(CA)CO, HN(CA)CB, and HN(CO)CACB.[39] The ^{1}H chemical shifts were directly referenced to the resonance of 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt, while the ^{13}C and ^{15}N chemical shifts were indirectly referenced with the absolute frequency ratio Ξ (^{13}C/^{1}H) = 0.251449530 and Ξ (^{15}N/^{1}H) = 0.101329118.[40] To identify the carbohydrate-binding sites of HA1, ^{1}H-^{15}N HSQC spectra were acquired using ^{15}N-labeled HA1 with or without a fetuin-derived glycopeptide at a 1:1 molar ratio. To perform PFG-NMR experiments, HA1 was dissolved in D_2O buffer containing 20 mM sodium phosphate (pH 5.5) and 0.05% dioxan at a final concentration of 1 mM. PFG-NMR measurements were performed on a Bruker Avance 500 spectrometer with a

reported pulse sequence.[12] 20 spectra were acquired with the strength of the diffusion gradient varying between 1% and 100% of the maximum power level, and the signal decay was fitted to a single exponential function. The hydrodynamic radius of HA1 (R_h^{HA1}) was estimated from a relationship, $R_h^{HA1} = d^{dioxan}/d^{HA1}$ (R_h^{dioxan}) where d^{dioxan} and d^{HA1} are the decay rates of NMR signal obtained from dioxan and HA1, respectively, and R_h^{dioxan} is hydrodynamic radius of dioxan, 2.12 Å.[12] All NMR data were processed using NMRPipe[41] and analyzed using NMRView.[42]

DLS measurements—The DLS experiments were carried out on a DynaPro instrument (Protein Solutions). HA1 solution was prepared in 20 mM sodium phosphate (pH 5.5) at concentrations ranging from 5 μM to 1 mM and centrifuged at 20,000g immediately before the measurements. The scattering intensities at 90˚ were collected at 25˚C using 10 μl of HA1 solution. A single measurement represents an average of 10 independent scans of 5 seconds. Data collection and processing were followed by the regularized procedures with Dynamics software (Protein Solutions).

Accession number

Atomic coordinates of HA1 were deposited in the Protein Data Bank with the accession code 3A5P.

Acknowledgments

This work was supported by grants from the National Project on Protein Structural and Functional Analyses (Japan). We thank Drs. K.-M. Chen, J. K. Lee, and L. A. Randles for their suggestions and editorial helps.

References

1. Sharon, N. (2008). Lectins: past, present and future. Biochem. Soc. Trans. 36, 1457-1460.

2. Sharon, N. (2007). Lectins: carbohydrate-specific reagents and biological recognition molecules. J. Biol. Chem. 282, 2753-2764.

3. Drickamer, K. & Taylor, M. E. (2003). Identification of lectins from genomic sequence data. Methods Enzymol. 362, 560-567.

4. Chandra, N. R., Prabu, M. M., Suguna, K. & Vijayan, M. (2001). Structural similarity and functional diversity in proteins containing the legume lectin fold. Protein Eng. 14, 857-866.

5. Camby, I., Mercier, M. L., Lefranc, F. & Kiss, R. (2006). Galectin-1: a small protein with major functions. Glycobiology 16, 137R-157R.

6. Paz, A., Haklai, R., Elad-Sfadia, G., Ballan, E. & Kloog, Y. (2001). Galectin-1 binds oncogenic H-Ras to mediate Ras membrane anchorage and cell transformation. Oncogene 20, 7486-7493.

7. Morita, M. (1998). Molecular analysis of Physarum haemagglutinin I: lack of a signal sequence, sulphur amino acids and post-translational modifications. Microbiology 144, 1077-1084.

8. Morita, M. & Nishi, A. (1992). Isolation and partial characterization of haemagglutinin from plasmodia of Physarum polycephalum. J. Gen. Microbiol. 138, 619-625.

9. Schreckenbach, T. & Werenskiold, A. K. (1986). Gene expression during plasmodial differentiation. In The Molecular Biology of Physarum polycephalum (Dove, W. F., Dee, J., Hatano, S., Haugli, F. B. & Wohlfarth-Bottermann, K. E., eds), pp. 131-150, Plenum Press, New York.

10. Morita, M., Minami, E., Zheng, L. & Nishi, A. (1995). Physarum polycephalum

haemagglutinins: effect of nutrition on synthesis, and their possible role in nature. Microbiology 141, 2315-2322.

11. Morita, M., Iwado, Y. & Okamura, S. (1998). Identification of a carbohydrate-binding site in Physarum haemagglutinin I. Biochem. Mol. Biol. Int. 46, 233-240.

12. Willkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. & Smith, L. J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry 38, 16424-16431.

13. de la Torre, J. G., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys. J. 78, 719-730.

14. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. Bioinformatics 24, 2780-2781.

15. Bourne, Y., Bolgiano, B., Liao, D. I., Strecker, G., Cantau, P., Herzberg, O., Feizi, T. & Cambillau, C. (1994). Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides. Nat. Struct. Biol. 1, 863-870.

16. Dormitzer, P. R., Sun, Z. Y., Wagner, G. & Harrison, S. (2002). The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO J. 21, 885-897.

17. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37, D233-D238.

18. Boraston, A. B., Ficko-Blean, E. & Healey, M. (2007). Carbohydrate recognition by a large sialidase toxin from Clostridium perfringens. Biochemistry 46, 11352-11360.

19. Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1996). Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex. J. Mol. Biol. 259, 281-296.

20. Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. Cell. Mol. Life Sci. 63, 2954-2967.

21. Boraston, A. B., Bolan, D. N., Gilbert, H. J. & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382, 769-781.

22. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabó, L., McKie, V. A., Gilbert, H. J. & Davies, G. J. (2002). Promiscuity in ligand-binding: The three-dimensional structure of a Piromyces carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. Proc. Natl. Acad. Sci. U.S.A. 99, 14077-14082.

23. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J. & Williamson, M. P. (2000). The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. J. Biol. Chem. 275, 41137-41142.

24. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.

25. Cho, M. & Cummings, R. D. (1995). Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. I. Physical and chemical characterization. J. Biol. Chem. 270, 5198-5206.

26. Lotan, R., Skutelsky, E., Danon, D. & Sharon, N. (1975). The purification, composition, and specificity of the anti-T lectin from peanut (Arachis hypogaea). J. Biol. Chem. 250, 8518-8523.

27. Rice, K. G., Rao, N. B. & Lee, Y. C. (1990). Large-scale preparation and characterization of N-linked glycopeptides from bovine fetuin. Anal. Biochem. 184, 249-258.

28. Fox, J. D. & Robyt, J. F. (1991). Miniaturization of three carbohydrate analyses using a microsample plate reader. Anal. Biochem. 195, 93-96.

29. Kitago, Y., Watanabe, N. & Tanaka, I. (2005). Structure determination of a novel protein by sulphur SAD using chromium radiation in combination with a new crystal mounting method. Acta Cryst. D61, 1013-1021.

30. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307-326.

31. Sheldrick, G. (2008). A short history of SHELX. Acta Cryst. A64, 112-122.

32. Pape, T. & Schneider, T. R. (2004). HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. J. Appl. Cryst. 37, 843-844.

33. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Cryst. D55, 849-861.

34. Terwilliger, T. C. (2003). Automated main-chain model building by template matching and iterative fragment extension. Acta Cryst. D59, 38-44.

35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst. D53, 240-255.

36. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. J. Appl. Cryst. 30, 1022-1025.

37. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. 6, 458-463.

38. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Cryst. D60, 2126-2132.

39. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. (1999). TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. J. Am. Chem. Soc. 121, 844 - 848.

40. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995). ^{1}H, ^{13}C and ^{15}N chemical shift referencing in

biomolecular NMR. J. Biomol. NMR 6, 135-140.

41. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277-293.

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

Figure 1: Complex structure of peanut legume lectin and its ligand. The legume lectin structure is composed of three β-sheets: front (green), back (blue), and top (red) sheets. The lectin and its ligand are shown with a ribbon and ball-and-stick model, respectively. The coordinate data were obtained from the Protein Data Bank (accession code 2PEL).

Figure 2: Crystal structure of *Physarum polycephalum* HA1. (A) The backbone trace of HA1 tetramer. Chains 1 and 4 (red) (2 and 3; blue) share the same structure, whereas chains 1 and 2 adopt a different conformation over residues P21-S25. Arrows represent directions of view of panels C and D. (B) Stereoview of the ribbon representation of HA1 monomer (chain 1). (C) The ribbon representation of HA1 dimer composed of chains 1 and 3. Residues I7-D9, and G23-S25 are depicted with a ball-and-stick scheme. A rectangle represents a region of close-up view in panel E. (D) The ribbon representation of the HA1 dimer composed of chains 1 and 2. Residues Y20 and Y31 are depicted with a ball-and-stick scheme. A rectangle represents a region of close-up view in panel F. (E) Close-up view of the interface between chains 1 and 3. The broken lines represent intermolecular hydrogen bonds. (F) Close-up view of the interface between chains 1 and 2.

Figure 3: Hydrodynamic radii of HA1. DLS experiments were carried out using HA1 at concentrations ranging from 5 μM to 1 mM. The percentage of number of different species was plotted against the hydrodynamic radius.

Figure 4: NMR interaction assay between HA1 and fetuin-derived glycopeptide. (A) A selected region of ^{1}H-^{15}N HSQC spectra of HA1 with (red) or without (blue) an equimolar glycopeptide. (B) A linewidth ratio of the HSQC signals with and without an equimolar

glycopeptide. Residues with ratios above 1.5 are presented by yellow-colored bars. Arrows colored in red represent the signals that completely disappear in the presence of glycopeptides. The secondary structure element of HA1 is depicted in the upper part of the panel. (C, D) The residues that produce signal broadening and disappearing in the presence of the glycopeptide are presented with a space-filling model colored in yellow/green and red/blue, respectively. Red/yellow-colored residues belong to chain 1, whereas blue/green-colored residues belong to chain 2 (C) or chain 3 (D).

Figure 5: Solvent-accessible surface of HA1 tetramer. The surfaces generated by residues that produce signal broadening and disappearing in the presence of the glycopeptide are colored in green. The side chain of W54 is presented with a ball-and-stick model. The orientation of the molecule is identical to that in figure 2A.

Figure 6: Comparison of secondary structure elements among HA1 (A), peanut legume lectin (B; PDB code 2PEL), galectin-1 (C; 1SLB), CBM of NanJ (D; 2V73), and VP4 (E; 1KQR). α-Helices and β-strands are represented by shaded rectangles and cyan arrows, respectively. The β-strands that belong to a jelly roll motif and an up-and-down β-sheet motif (the region in yellow) are labeled with Arabic and Roman numbers, respectively. Red ovals roughly represent positions of the residues that make contact with the target carbohydrates.

Table 1 Crystallographic data collection and refinement statistics.

Data set	Iodine derivative	Native
X-ray source	Cr	Cu
Wavelength (Å)	2.29	1.54
Space group	$P2_1$	$P2_12_12_1$
Cell constants	*a* = 59.3 Å	*a* = 64.9 Å
	b = 97.7 Å	*b* = 78.0 Å
	c = 67.1 Å	*c* = 86.1 Å
	β = 103.5 º	
Resolution (Å)	2.68 (2.78 - 2.68)	1.82 (1.89 - 1.82)
Unique reflections	20,963	40,006
Redundancy	7.2 (6.7)	4.6 (4.5)
Completeness (%)	99.9 (99.1)	99.6 (100.0)
R_{merge} (%)	9.9 (18.5)	3.0 (16.1)
I/ σ(I)	25.9 (11.8)	49.4 (11.0)
Mosaicity (º)	0.60	0.94
Refinement range (Å)		30.37 - 1.82
Number of reflections		39,795
Completeness (%)		99.6
$R_{work}/R_{free\ (for\ 5\%)}$		15.4 / 19.4
No of waters		858
RMS bond length (Å)		0.013
RMS bond angles (º)		6.234
Average B-factor ($Å^2$)		23.2
Ramachandran plot (%)		
favoured		98.5
allowed		1.5
outlier		0.0

Supplemental Data

Figure S1: Surface plasmon resonance profile obtained from interactions between a full-length HA1 and fetuin-derived glycopeptide (A) and between an HA1 mutant lacking the first 32 residues and fetuin-derived glycopeptide (B). HA1-glycopeptide interaction was detected using a BIAcore J biosensor (GE Healthcare). A fetuin-derived glycopeptide was immobilized on a CM5 chip using an amine-coupling reagent (GE Healthcare). After equilibrating with 20 mM sodium acetate buffer (pH 5.5), 180 μl of HA1 solutions at different concentrations (125 nM to 2 μM) were injected at 60 μl/min over immobilized glycopeptide and control flow cells. The dissociation of HA1 from fetuin-derived glycopeptide was observed with a constant flow (60 μl/min) of sodium acetate buffer for 3 min. After each injection, the biosensor surface was regenerated using 60 μl of 10 mM glycine-HCl buffer (pH 2.0). The expression plasmid of the N-terminal truncated HA1 was constructed by a PCR-based method, and the protein was purified in the same way as the original full-length HA1.

Figure S2: Tryptophan fluorescence profiles of HA1 with (broken line) or without (solid line) an equimolar fetuin-derived glycopeptide. Tryptophan fluorescence was measured on an F-4500 fluorescence spectrometer (Hitachi) with a 2-mm quartz cuvette. An HA1 sample was excited at 295 nm, and the fluorescence emission spectrum was recorded from 310 to 420 nm at 25 ˚C. The sample includes 20 μM HA1 and 20 mM sodium acetate

buffer (pH 5.5). HA1 solution was titrated with a fetuin-derived glycopeptide (0-50 μM).

Figure 1

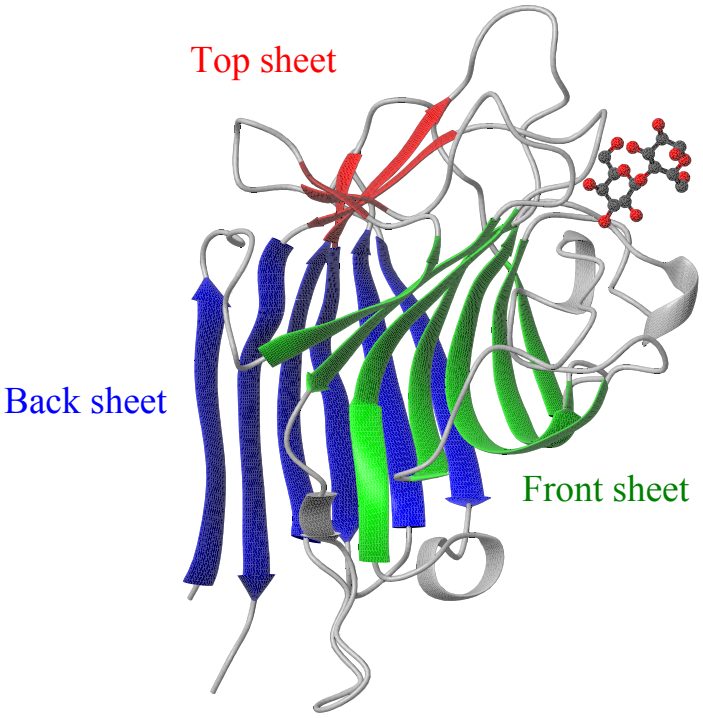


Figure 2

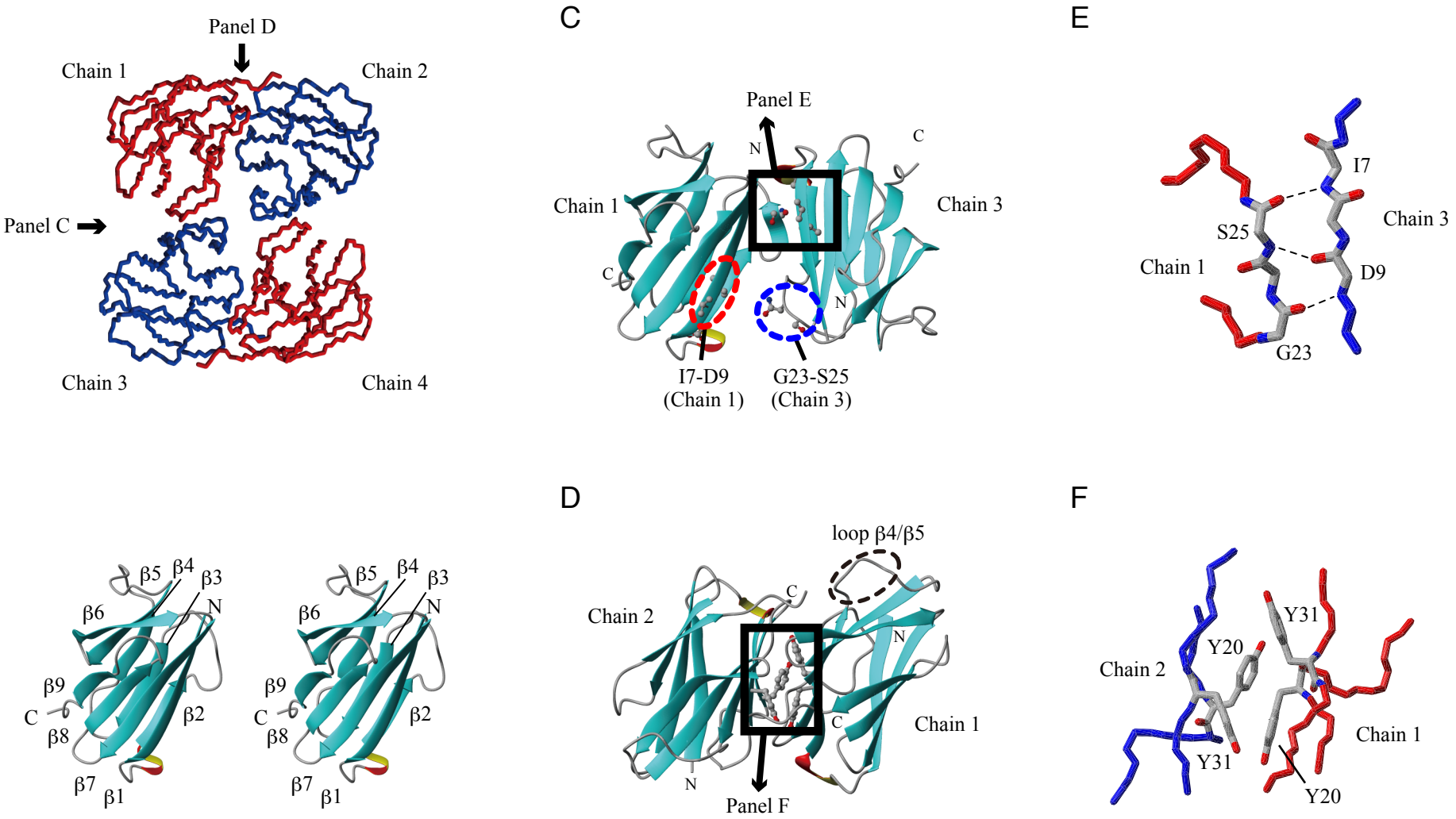

A
Panel D
Chain 1
Chain 2
Panel C
Chain 3
Chain 4
C
Panel E
N
C
Chain 1
Chain 3
C
N
I7-D9
(Chain 1)
G23-S25
(Chain 3)
E
I7
Chain 3
S25
Chain 1
D9
G23
B
β5
β4
β3
N
β6
β9
C
β2
β8
β7
β1
D
loop β4/β5
Chain 2
C
N
Chain 1
C
N
Panel F
F
Y31
Y20
Chain 2
Chain 1
Y31
Y20

Figure 3

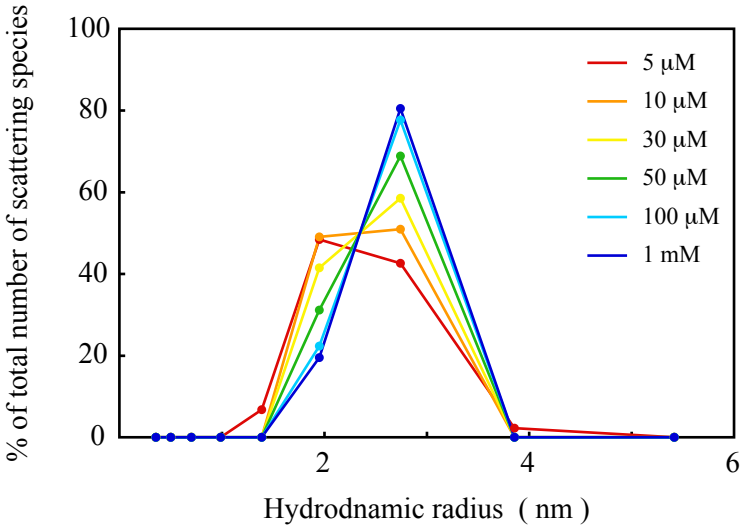

% of total number of scattering species
100
80
60
40
20
0
2
4
6
Hydrodnamic radius (nm)
5 μM
10 μM
30 μM
50 μM
100 μM
1 mM

Figure 4

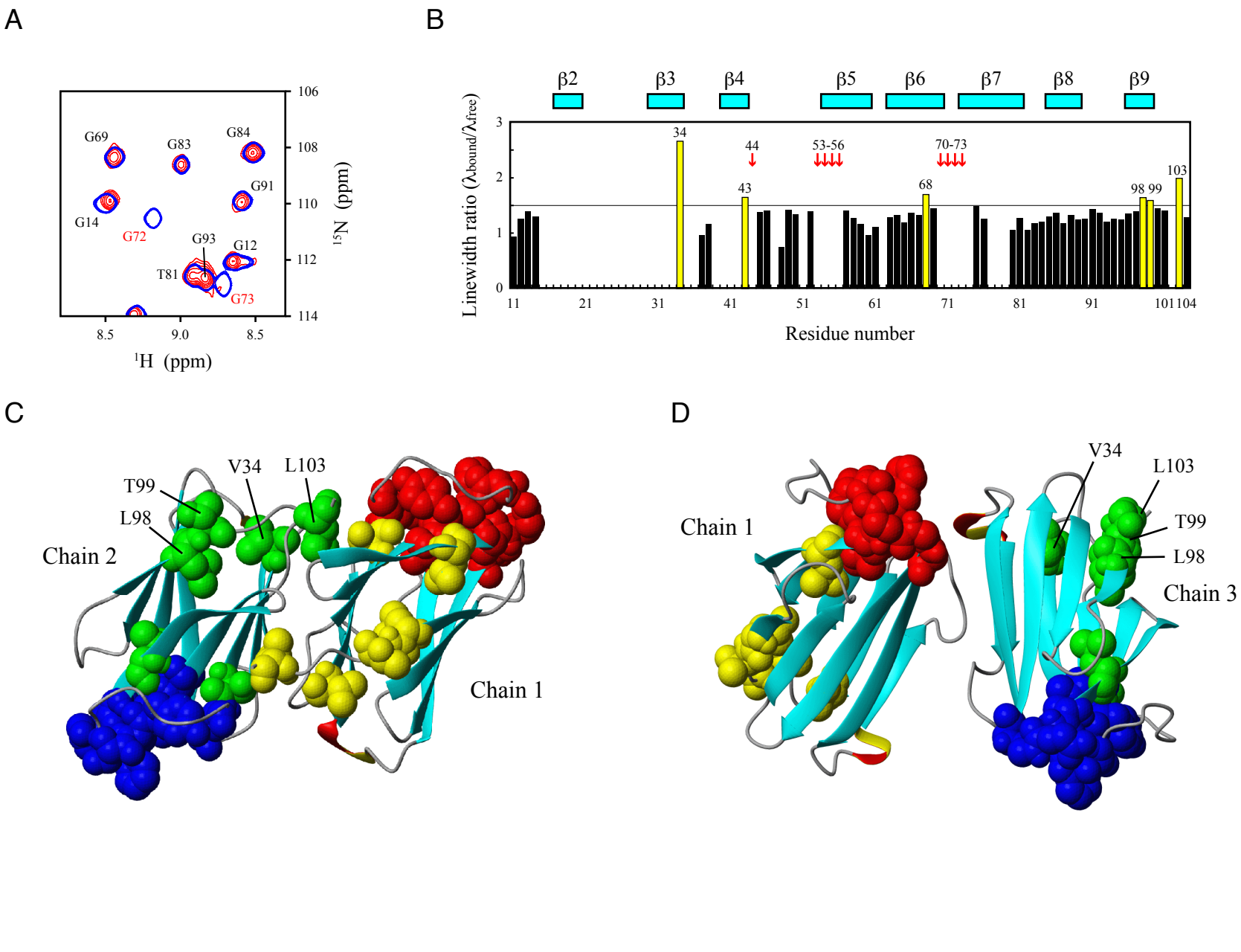

A
G69
G83
G84
G91
G14
G72
G93
G12
T81
G73
106
108
110
112
114
8.5
9.0
8.5
^{15}N (ppm)
^{1}H (ppm)
B
β2
β3
β4
β5
β6
β7
β8
β9
34
44
43
53-56
70-73
68
98 99
103
Linewidth ratio ($\lambda_{bound}/\lambda_{free}$)
0
1
2
3
11
21
31
41
51
61
71
81
91
101 104
Residue number
C
T99
V34
L103
L98
Chain 2
Chain 1
D
V34
L103
Chain 1
T99
L98
Chain 3

Figure 5

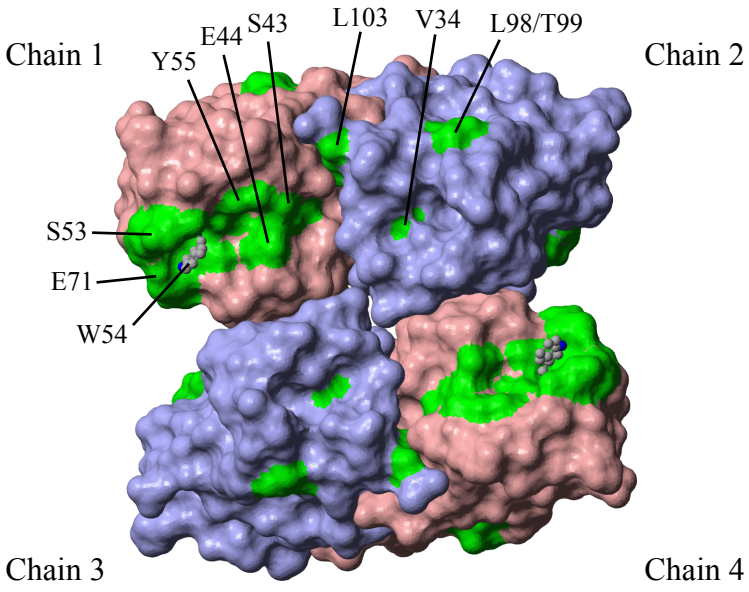

Chain 1
Chain 2
Chain 3
Chain 4
Y55
E44
S43
L103
V34
L98/T99
S53
E71
W54

Figure 6

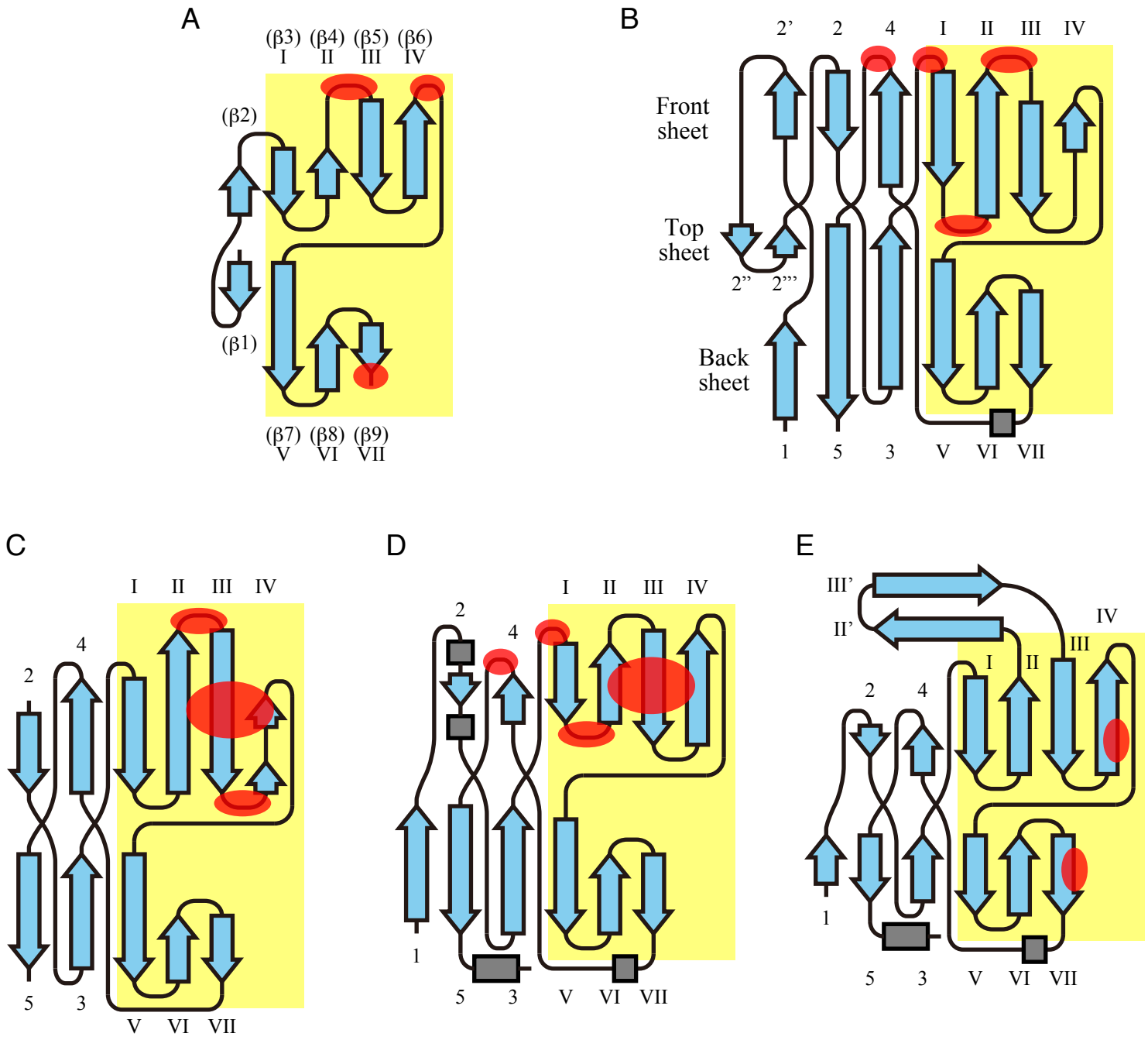
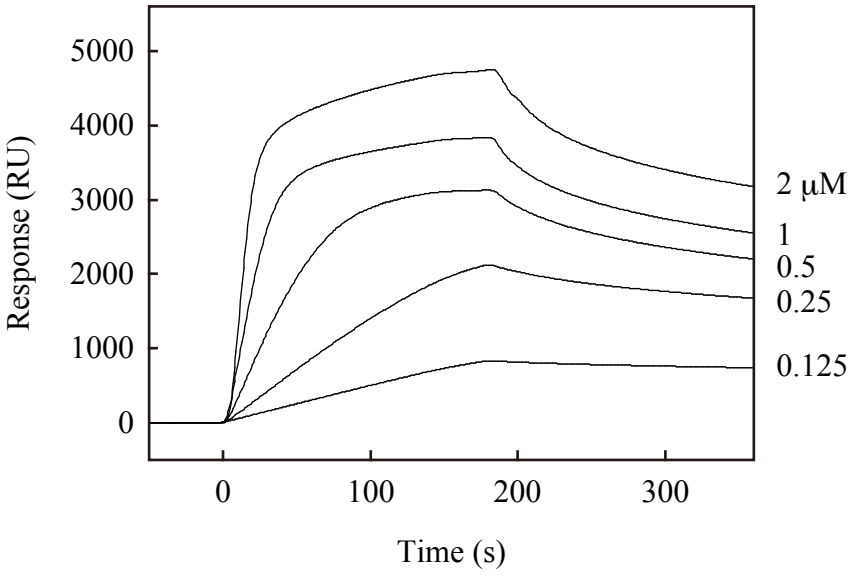

A
(β3) (β4) (β5) (β6)
I II III IV
(β2)
(β1)
(β7) (β8) (β9)
V VI VII
B
2' 2 4 I II III IV
Front sheet
Top sheet
2" 2"'
Back sheet
1 5 3 V VI VII
C
I II III IV
2
4
5 3
V VI VII
D
I II III IV
2
4
1
5 3 V VI VII
E
III'
II'
IV
III
I
II
2 4
1
5 3 V VI VII

Figure S1

A


5000
4000
3000
2000
1000
0
Response (RU)
0
100
200
300
Time (s)
2 μM
1
0.5
0.25
0.125

B

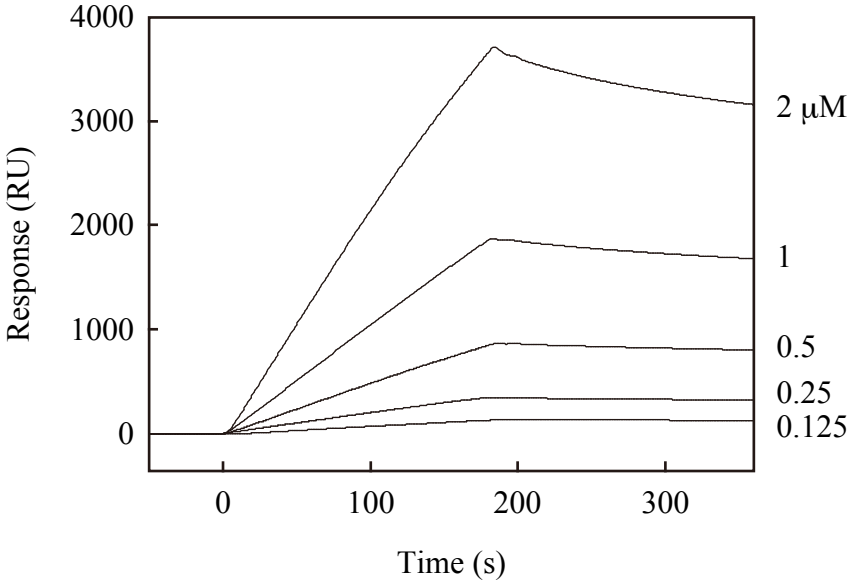
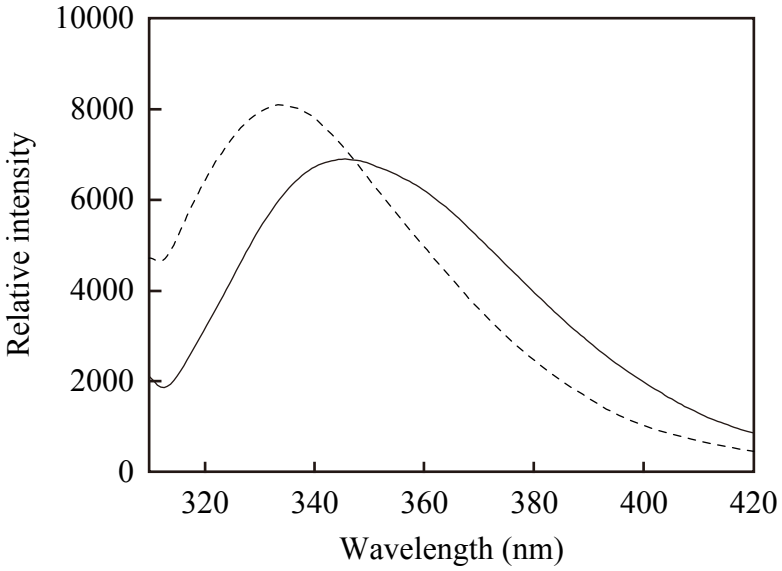

4000
3000
2000
1000
0
Response (RU)
0
100
200
300
Time (s)
2 μM
1
0.5
0.25
0.125

Figure S2


10000
8000
6000
4000
2000
0
Relative intensity
320
340
360
380
400
420
Wavelength (nm)