

# SCIENTIFIC REPORTS



OPEN

## The trimeric solution structure and fucose-binding mechanism of the core fucosylation-specific lectin PhoSL

Kazuhiko Yamasaki<sup>1</sup>, Tomoko Yamasaki<sup>1</sup> & Hiroaki Tateno<sup>2</sup>

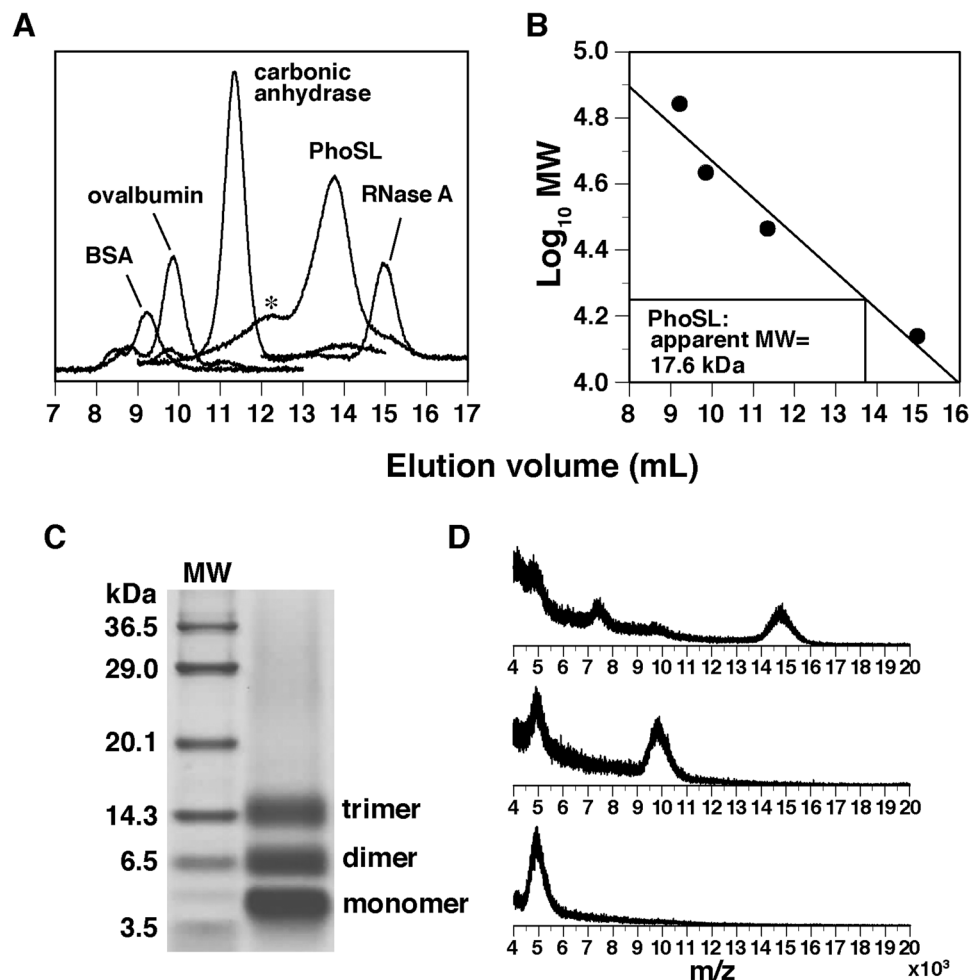
The core  $\alpha$ 1–6 fucosylation-specific lectin from a mushroom *Pholiota squarrosa* (PhoSL) is a potential tool for precise diagnosis of cancers. This lectin consists of only 40 amino acids and can be chemically synthesized. We showed here that a synthesized PhoSL peptide formed a trimer by gel filtration and chemical cross-linking assays, and determined a structure of the PhoSL trimer by NMR. The structure possesses a  $\beta$ -prism motif with a three-fold rotational symmetry, where three antiparallel  $\beta$ -sheets are tightly connected by swapping of  $\beta$ -strands. A triad of Trp residues comprises the structural core, forming NH– $\pi$  electrostatic interactions among the indole rings. NMR analysis with an excess amount of fucose revealed the structural basis for the molecular recognition. Namely, fucose deeply enters a pocket formed at a junction of  $\beta$ -sheet edges, with the methyl group placed at the bottom. It forms a number of hydrophobic and hydrogen-bonding interactions with PhoSL residues. In spite of partial similarities to the structures of other functionally related lectins, the arrangement of the antiparallel  $\beta$ -sheets in the PhoSL trimer is novel as a structural scaffold, and thus defines a novel type of lectin structure.

Fucosylation is a common modification in glycoproteins, and is classified into several types with regard to the modes of linkage, i.e.,  $\alpha$ 1–2,  $\alpha$ 1–3,  $\alpha$ 1–4, and  $\alpha$ 1–6. Among them, the  $\alpha$ 1–6 fucosylation is typically found at the first *N*-acetyl-D-glucosamine (GlcNAc) of the *N*-glycan core (called “core” fucosylation) in the mammalian systems, and is keenly related to cancer development<sup>1</sup>. Most prominently, the level of core fucosylation of *N*-glycan in alpha-fetoprotein (AFP) increases in hepatocellular carcinomas (HCC) and carcinoma metastatic to the liver, but not in benign liver diseases, such as acute viral hepatitis, chronic hepatitis, or liver cirrhosis<sup>2</sup>. This was explained by selective secretion of fucosylated glycoproteins produced in hepatocyte through bile duct, although such bile duct does not exist in tissues of the carcinomas<sup>1</sup>. Therefore, lectins specific to core fucosylation should be particularly useful for the diagnostics of these cancers.

A lectin highly specific to core  $\alpha$ 1–6 fucosylation was recently isolated from a mushroom *Pholiota squarrosa* (PhoSL)<sup>3</sup>. PhoSL binds only to core-fucosylated glycans, but not to the other types of fucosylated glycans, as revealed by frontal affinity chromatography<sup>4</sup>. In addition, this lectin showed stronger affinity than that of a clinically used lectin *Lens culinaris* agglutinin (LCA). Therefore, researchers started to use PhoSL for detecting core fucosylation in diagnosis of cancers and related diseases, i.e., colorectal, prostate, pancreatic, and liver cancers as well as a chronic pancreatitis that is a background of pancreatic ductal adenocarcinoma<sup>5–9</sup>.

The PhoSL molecule purified from the mushroom consists of only 40 residues and can therefore be chemically synthesized<sup>3</sup>. It is likely that PhoSL exists in a trimeric or tetrameric form, as shown by gel filtration, mass spectrometry, and gel electrophoresis analyses<sup>3</sup>. This lectin showed an extreme stability against heat treatment, e.g., 100 °C for 30 min, and also against incubation at pH 2–11. PhoSL is therefore suitable for its use as a medical tool in that it is easy to produce and to keep in the active form. Moreover, it would be easier to incorporate unnatural amino acids by means of the chemical synthesis, in order to produce lectins with novel specificities.

<sup>1</sup>Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, 305-8566, Japan. <sup>2</sup>Biotechnology Research Institute for Drug Discovery, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, 305-8568, Japan. Correspondence and requests for materials should be addressed to K.Y. (email: [k-yamasaki@aist.go.jp](mailto:k-yamasaki@aist.go.jp))



**Figure 1.** Trimeric state of the PhoSL peptide as revealed by gel filtration and chemical cross-linking analyses. Shown are (A) elution profile and (B) relation between elution volume and molecular mass of the proteins in gel filtration analysis, (C) SDS-PAGE gel after the cross-linking reaction with EDC and NHS, and (D) mass spectra of the three respective bands in (C). Asterisk in (A) shows the peak for the presumable hexamer of PhoSL.

In the present study, we determined a novel structure of the PhoSL trimer by NMR, and elucidated interaction with fucose. The results provided a basis for understanding the structural stability and recognition of fucosylated glycans.

## Results and Discussion

**Trimerization of the PhoSL peptide.** While the PhoSL molecule purified from the mushroom has been suggested to form a trimer or a tetramer<sup>3</sup>, we determined the oligomeric state of the chemically synthesized peptide. By gel filtration analysis with marker proteins, the molecular weight of PhoSL was estimated to be 17.6 kDa (Fig. 1A,B). This value is slightly less than four times the size of the PhoSL monomer (4.5 kDa).

Next, a cross-linking reaction was carried out by using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and *N*-hydroxysuccinimide (NHS). These reagents covalently link amino groups in Lys residues or at the *N*-terminus and carboxyl groups in Asp/Glu residues or at the *C*-terminus, when pairs of the linkable groups are located in close proximity. After the reaction, three bands were observed in sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (Fig. 1C). These corresponded to monomer, dimer, and trimer, respectively, as determined by mass spectrometry (Fig. 1D). The monomer and dimer are likely to be observed after incomplete reaction or linking of pairs within the same polypeptide that blocks the inter-chain linking. In contrast, the trimer band should represent the oligomeric state of PhoSL, because a band for the tetramer is not observed at all. In this cross-linking approach, presumable dimerization of this trimer, i.e., hexamerization, cannot be excluded; the linkable pairs may not exist around the interface between the two trimers. In the gel filtration profile, we indeed observed a minor peak for a presumable hexamer (~27 kDa), although it is now clear that the main peak is for the trimer (Fig. 1A). Thus, the oligomeric state of the PhoSL peptide was determined by combination of the two methods.

**Solution structure of PhoSL trimer.** We used the chemically synthesized peptide without isotope labeling for the NMR analyses. In the preliminary analyses, the protein dissolved in phosphate buffer (pH 6.0), which

showed significant precipitation (data not shown). Instead, the protein solution was clear when dissolved in Tris buffer (pH 7.5). Therefore, we used this buffer for all the NMR experiments in the present study.

The NMR spectra of PhoSL contained a single set of resonances, which implies a structural symmetry of the PhoSL trimer. Backbone Nuclear Overhauser effect (NOE) connections revealed that PhoSL possessed a secondary structure rich in  $\beta$ -strands (Fig. 2A,B). Namely, we observed intense sequential  $H_{\alpha}H_N$  NOEs, implying close proximity between these protons, which are characteristic of the extended peptide conformation of  $\beta$ -strands<sup>10</sup>. In contrast, we did not observe intense sequential  $H_NH_N$  NOEs and medium-range  $H_{\alpha}H_N$  NOEs ( $|i-j| = 3$  or  $4$ ), which are characteristics of helices<sup>10</sup>. The backbone  $H_{\alpha}H_{\alpha}$  NOEs and  $H_NH_N$  NOEs were observed between the  $\beta$ -strands, revealing the arrangements of the strands in  $\beta$ -sheets (Fig. 2A).

Within the  $\beta$ -sheets, however, we cannot distinguish between intermolecular and intramolecular NOE connections of the  $\beta$ -strands, due to the symmetry. Therefore, in the structure calculation, ambiguities with regard to this distinction were introduced (see Materials and Methods section; Table 1). Following these trials in calculations, the connections between  $\beta$ -strands were determined as those shown in Fig. 2C. Therefore in the final calculation, a part of NOEs in the interfaces between the strands were fixed with regard to intermolecular or intramolecular ones.

The resulting structure satisfied the structural constraints, i.e., NOE distance constraints and torsional angle constraints, as we observed low root mean square deviations (RMSDs) from these constraints (Table 1). The structure also possessed ideal stereochemical properties, with small van der Waals energy, low RMSDs from ideal geometries, and favorable distributions in the Ramachandran plot. Consequently, it showed good convergence, with low RMSD from the mean structure (Table 1, Fig. 3A).

This solution structure revealed a  $\beta$ -prism scaffold with a clear three-fold rotational symmetry (Fig. 3B). It consists of three  $\beta$ -sheets each containing essentially the four  $\beta$ -strands. The N-terminal and C-terminal halves of  $\beta$ -strand 1, with a kink between Thr6 and Lys7, from different monomers, respectively, form contacts within a short range, from which we may say that the  $\beta$ -sheets expand to five-stranded (Fig. 2C). Also, contacts between  $\beta$ -strands 3 and 4 are intermolecular. Thus, the trimer is tightly connected by swapping of  $\beta$ -strands among the three  $\beta$ -sheets.

At the symmetric center, the indole rings of the three Trp32 contact one another and form the major part of the structural core (Fig. 3C). The contacts are achieved through the edge of a ring plane located over another ring plane, defining relative angle of 120 degrees. The indole  $H_{\epsilon 1}$  proton is closest to the ring plane of another Trp with a distance of  $\sim 3$  Å. It appears, therefore, that there are the  $NH-\pi$  electrostatic interactions, which was considered to be a kind of hydrogen bonds, between the rings<sup>11</sup>. We should point out that in these mutual interactions polarization of the N-H bond itself enhances the density of the  $\pi$ -electrons of the acceptor ring, through electron delocalization. Thus, these interactions should significantly contribute to stabilizing the formation of the structural core. This geometry is consistent with an extremely upfield shifted resonance of the  $H_{\epsilon 1}$  proton (4.69 ppm) and NOEs with other ring protons (Fig. S1A in the Supplementary information); a strong ring-current effect<sup>10</sup> may have overcome the downfield shifting by hydrogen bonding.

Cys10 and Cys17 are located on  $\beta$ -strands 1 and 2, respectively, which are adjacent to each other (Fig. 3D). Because of the reductive condition during the NMR measurements, we observed  $H_{\gamma}$  protons of these Cys residues and NOEs with other residues (Fig. S1B). Considering their relative positions, however, we can expect a disulfide bridge under an oxidative condition, which is likely to contribute to the structural stability.

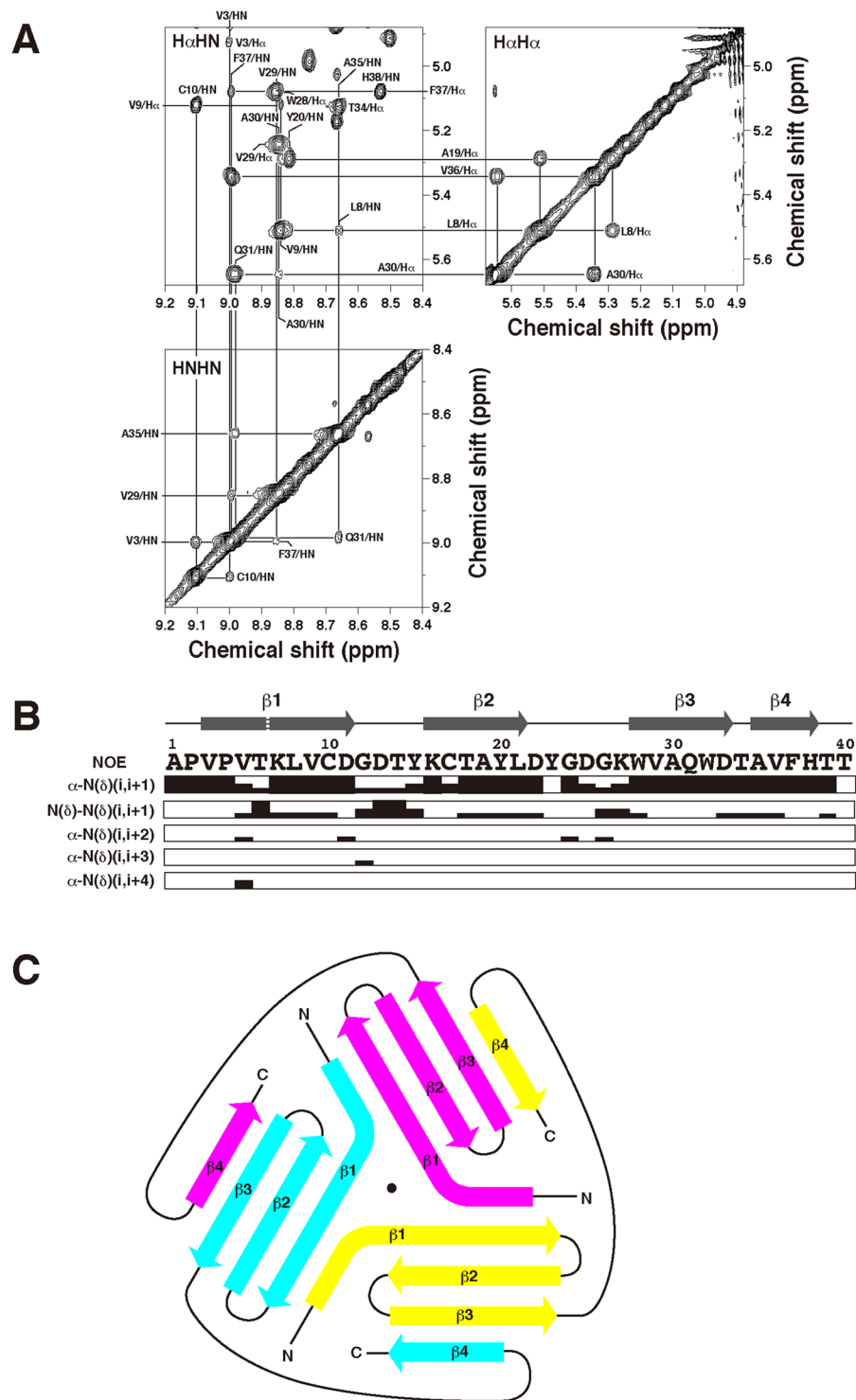
**Innate PhoSL structure.** Although the PhoSL peptide isolated from the mushroom is 40 amino acids in length<sup>3</sup>, the relevant gene cloned from the mushroom codes for a polypeptide of 180 amino acids<sup>12</sup> (Fig. S2A). This polypeptide contains three repeats of PhoSL-like sequence, the first of which corresponds to that used in the present study. It is predictable that, at least immediately after the synthesis, the PhoSL of a single polypeptide forms a pseudotrimer. Two homology models of the pseudotrimer were produced on the basis of the present PhoSL trimer structure by SWISS-MODEL system<sup>13</sup> (Fig. S2B). The two are different in the relative positions of the three repeats in the trimeric structure; either is geometrically allowed by the substantial length of the linkers between the repeats.

Presumably by proteolytic cleavage of the linkers in the mushroom cells, PhoSL becomes a heterogenic trimer of 40 amino acids polypeptides. It is noticeable that the sequence of PhoSL determined by the N-terminal sequence analysis does not exactly match any of the three repeats in the polypeptide (Fig. S2A). On the other hand, respective amino acids of the former match those of either one of the repeats, except for position 27. This suggests a possibility that the 40-amino acid PhoSL peptide from the mushroom is a mixture of those corresponding to the three repeats in the 180-amino acid polypeptide.

It should be noted that the residues that stabilize the structure, as described above, and those that involved in the fucose binding (see below) are conserved among the three repeats (a conservative substitution of Tyr by Phe at position 23 is observed in repeat 2; Fig. S2A). Therefore, the innate PhoSL or the heterogenic trimer probably existing in the cell should be stable and fully functional.

Recently, a core fucose-specific lectin from a bacterium *Streptomyces rapamycinicus* (SL2-1) was identified, which was highly homologous to PhoSL<sup>14</sup>. Interestingly, SL2-1 is  $\sim 180$  amino acids in length and has three repeats of PhoSL-homologous regions. Therefore, it is likely that the tertiary structure of SL2-1 should be very similar to that of the innate PhoSL protein shown in Fig. S2B. PhoSL, SL2-1, and related proteins from several bacteria and fungi cited in ref.<sup>14</sup>, including another core fucose-specific lectin from *Rhizopus stolonifer*<sup>15</sup>, should comprise a novel family of lectins, which are specific to core fucosylation.

**NMR analysis of binding of fucose.** To understand the mechanism of molecular recognition, we examined binding of L-fucose by NMR (Figs 4A and S3A). Upon titration of fucose, chemical shift perturbations, i.e., changes in positions of some PhoSL peaks, were observed. These gradual changes indicate that binding and



**Figure 2.** Secondary structure of PhoSL. (A) Tiles of a NOESY spectrum (100 ms mixing time) showing intensive sequential  $H_{\alpha}H_N$  NOEs, which are characteristics of  $\beta$ -strands, and  $H_{\alpha}H_{\alpha}$  and  $H_NH_N$  NOEs, which reveal connection of  $\beta$ -strands in  $\beta$ -sheets. The spectrum was recorded at 308 K and frequency of 900 MHz. (B) Profile of sequential and medium-range NOEs. Heights of the bars distinguish the large, medium, and small intensities, corresponding to distance constraints of  $\leq 3.5$  Å,  $\leq 4.5$  Å, and  $\leq 6.0$  Å, respectively. Regions of the  $\beta$ -strands appearing in the structural ensemble, as shown by Procheck-NMR<sup>29</sup>, are indicated above the sequence. A kink between Thr6 and Lys7 in  $\beta$ -strand 1 is shown by a white broken line. (C) A schematic representation of arrangements of the strands in  $\beta$ -sheets. The three monomers are distinguished by different colors. The dot indicates the center of rotational symmetry.

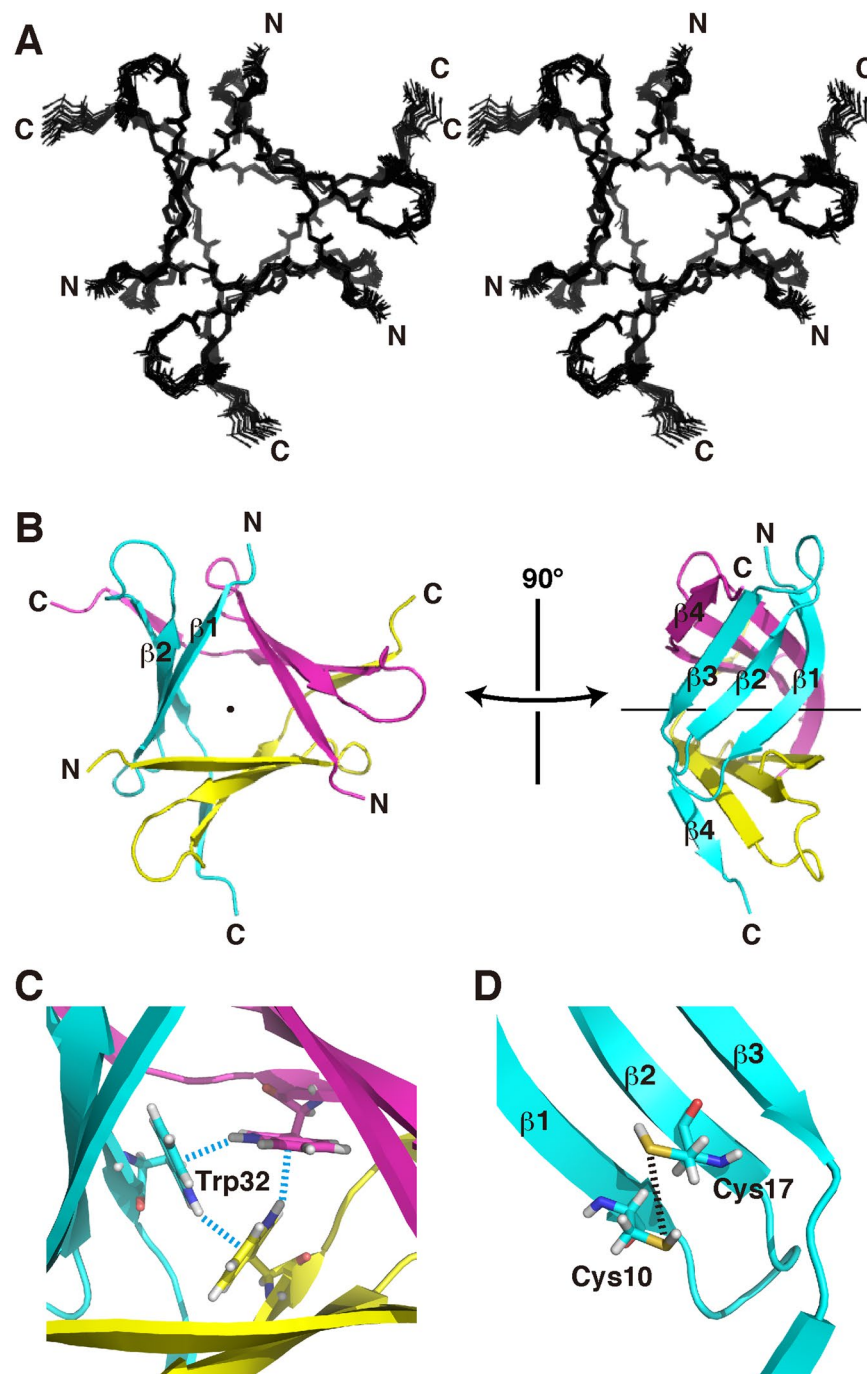
Structural constraints (per trimer)		
Sequential NOE constraints <sup>a</sup>	606	
Intramolecular		606
Medium-range NOE constraints ( $2 \leq  i-j  \leq 4$ ) <sup>a</sup>	378	
Intermolecular		45
Ambiguous		333
Long-range NOE constraints ( $ i-j  > 4$ ) <sup>a</sup>	1107	
Intramolecular		567
Intermolecular		321
Ambiguous		219
Hydrogen bond constraints	120	
Torsion angle constraints	81	
$\chi^1$		75
$\chi^2$		6
Total	2292	
Characteristics		
	Ensemble of 20 structures	Minimized mean structure
RMSD from constraints		
NOEs (Å)	0.0019 ± 0.0008	0.001
Torsion angles (degrees)	0.000 ± 0.000	0
Noncrystallographic symmetry (kcal/mol)	0.00 ± 0.00	0.003
Van der Waals energy (kcal/mol) <sup>b</sup>	6.5 ± 1.0	5.7
RMSD from the ideal geometry		
Bond lengths (Å)	0.0008 ± 0.00004	0.0008
Bond angles (degrees)	0.227 ± 0.003	0.224
Improper angles (degrees)	0.157 ± 0.004	0.153
Average RMSD from the unminimized mean structure (Å)		
Backbone N, C <sub>α</sub> , and C'	0.35 ± 0.06	
All nonproton atoms	0.68 ± 0.09	
Ramachandran plot		
Most favored region (%)	87.6	90.9
Additionally allowed region (%)	12.4	9.1
Generously allowed region (%)	0	0
Disallowed region (%)	0	0

**Table 1.** Structural statistics for the NMR structure of PhoSL trimer. <sup>a</sup>Each NOE corresponds to three constraints for the symmetric trimer. <sup>b</sup>Values calculated with the repulsive nonbonded energy function in the CNS software package.

release of fucose from the protein are fast enough to average the peaks in the free and bound states. Therefore, changes in the chemical shifts reflect those in the relative population. Theoretical fitting for the concentration dependencies yielded dissociation constants ( $K_D$ ) of 5.8–6.2 mM ( $5.9 \pm 0.2$  mM; Fig. 4A). Note that this affinity is weaker than that between PhoSL and N-glycans ( $\sim 3 \mu\text{M}$ )<sup>3</sup> by  $\sim 2000$ -fold. Nonetheless, this difference in  $K_D$ s corresponds to that in free energy changes of 4.5 kcal/mol, for which one or two hydrogen bonds may compensate. We suggest, therefore, that there should be such favorable contacts between PhoSL and the other parts of the N-glycans, e.g., GlcNAc moieties, and that the above  $K_D$  value for the fucose monosaccharide is reasonable. This is indeed likely, because an NMR analysis on another core  $\alpha 1-6$  fucosylation-specific lectin from an alga *Bryothamnion triquetrum* showed that the 2nd GlcNAc of the N-glycan core is also involved in the binding, in addition to fucose<sup>16</sup>.

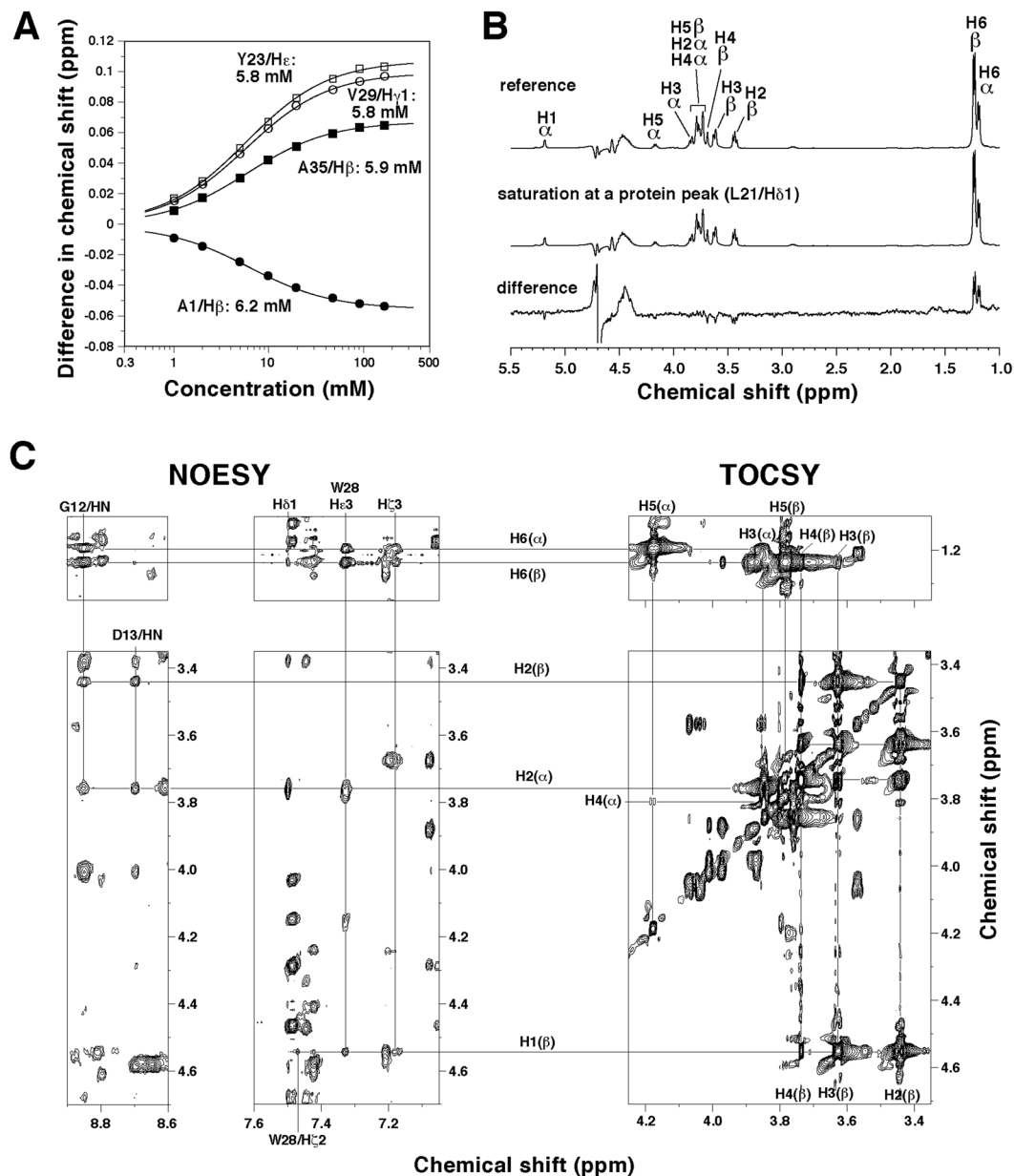
L-fucose is a mixture of the two anomers, i.e.,  $\alpha$ - and  $\beta$ -forms, where the former is biologically relevant. Populations of the  $\alpha$  and  $\beta$  anomers were  $\sim 30\%$  and  $\sim 70\%$ , respectively, as estimated from the NMR peak intensities of H6 methyl protons (Fig. 4B, upper spectrum). Saturation transfer difference (STD) experiment, where saturation of resonances of protein was propagated to those of fucose in rapid equilibrium of association/dissociation, indicated that both the two anomers bound to PhoSL (Fig. 4B). The populations of the anomers bound to PhoSL changed to  $\sim 36\%$  and  $\sim 64\%$ , as seen in the difference spectra, where the  $\alpha$  anomer is likely to be slightly preferable. Thus, we should note that the  $K_D$  value for fucose is an average value for the two anomers in such populations. The STD experiment also showed that H6 methyl group was more involved in the protein binding than the other protons; the only positive peaks in the difference spectrum were those for the H6 protons.

We then analyzed the NOESY spectrum of PhoSL with an excess amount of fucose (100 mM), and compared the chemical shifts of all protons with those of free PhoSL (Fig. S3B). Because chemical shift perturbations are related to changes in microscopic environment, we may predict the residues involved in the fucose binding. The results suggested that regions around the junctions of  $\beta$ -sheet edges were involved.



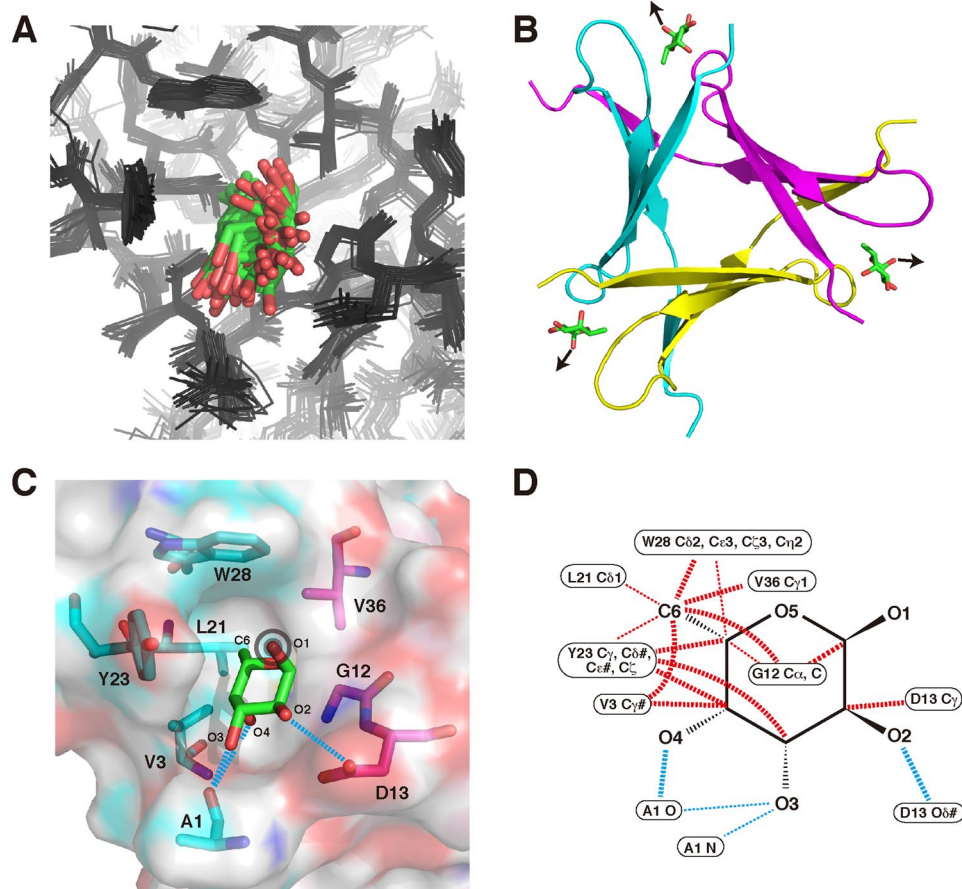
**Figure 3.** The solution structure of the PhoSL trimer. (A) The ensemble of the 20 selected structures in a stereo view. (B) Ribbon diagrams of the minimized mean structure where the orientation in the left panel is the same as that in (A). The three monomers are distinguished by different colors. The axis of the rotational symmetry is shown by a dot in left or a horizontal line in right. (C) A triad of Trp32 in the structural core. Blue broken lines indicate the NH- $\pi$  electrostatic interactions. (D) Two Cys residues in the vicinity. They should form a disulfide bridge under an oxidized condition, as shown by a broken line. These figures were produced by PyMOL ver. 1.8 (Schrodinger, LLC, New York, USA).

**Mechanism of molecular recognition.** In the same NOESY spectrum, we identified several NOE cross-peaks between PhoSL and fucose, although the molecules are in an equilibrium state, with rapid binding and release (Fig. 4C). We separately assigned the cross-peaks for the two anomers of fucose, and calculated the respective structures in a similar manner for the free PhoSL (Figs 5 and S4; Table S1). With these NOEs, the structure around the bound fucose did not strongly converge (Figs 5A and S4A). Nonetheless, the binding site and orientation became clear in the structures. The following descriptions on the binding mechanism are for  $\alpha$ -fucose, although they are essentially the same for  $\beta$ -fucose, unless otherwise stated.



**Figure 4.** NMR analyses of binding of fucose. (A) Changes in chemical shifts for selected resonances of PhoSL upon titration of fucose (see also Fig. S3A). By theoretical fitting of data,  $K_D$  values were obtained, as indicated. (B) STD experiment for the interaction of fucose with PhoSL. The solution contained 50  $\mu$ M PhoSL peptide and 5 mM L-fucose. (C) Two-dimensional NMR spectra of PhoSL–fucose mixture solution, containing 0.3 mM PhoSL peptide and 100 mM fucose. (Left four panels) A NOESY spectrum containing NOE cross peaks between PhoSL and the two anomers of fucose. (Right two panels) Those of a TOCSY spectrum showing intramolecular connectivity in the two fucose anomers (right). The mixing times for NOESY and TOCSY were 100 ms and 50 ms, respectively. The NMR assignments for the two anomers appear in literature<sup>30</sup>, and were confirmed here by the 2D NMR. The spectra were recorded at 308 K and at a frequency of 500 MHz in (B) or 900 MHz in (C).

The fucose-binding sites were located indeed where the chemical shift perturbation suggested, i.e., junctions of the  $\beta$ -sheet edges (Fig. 5B). At each site, a pocket was formed by several residues from two PhoSL chains (Fig. 5C). Fucose molecule deeply entered the pocket, with the C6 methyl group placed at the bottom, which was consistent with the STD result. Between fucose and the residues that comprise the pocket, a number of hydrophobic and hydrogen-bonding interactions are formed (Fig. 5C,D). Namely, the C6 methyl group and other carbon atoms contact aliphatic side chains of Val3, Leu21, and Val36, aromatic side chains of Tyr23 and Trp28, backbone carbon of Gly12, and aliphatic moiety of Asp13. Of noticeable is the contacts to Y23, which have a CH– $\pi$  attractive effect<sup>17</sup>, other than the hydrophobic effect. Also, hydrogen bonds are formed between the O<sub>2</sub>, O<sub>3</sub>, or O<sub>4</sub> hydroxyl groups and nitrogen or oxygen atoms of Ala1, and Asp13 (Gly12 is also involved in hydrogen bonds for  $\beta$ -fucose; Fig. S4C,D). Because fucose in the pocket is not strongly fixed in the structural ensemble, some of these contacts



**Figure 5.** Mechanism for recognition of  $\alpha$ -fucose. **(A)** Ensemble of 20 structures of PhoSL (black wire)– $\alpha$ -fucose (stick; green: carbon; red: oxygen) complex obtained by calculations using NOEs between PhoSL and fucose (see Fig. 4C; Table S1). Non-proton atoms in fucose and surrounding residues (Ala1, Val3, Gly12, Asp13, Leu21, Tyr23, Trp28, and Val36) are aligned. **(B)** The minimized mean structure of PhoSL (ribbon)– $\alpha$ -fucose (stick) complex, where the PhoSL monomers are shown in different colors. Arrows indicate presumable directions of the  $\alpha$ 1–6 linkage to the GlcNAc of the N-glycan. **(C)** A binding pocket for  $\alpha$ -fucose seen in the surface of the minimized mean structure. The residues that comprise the pocket are indicated in stick, as colored for the protein chains. Fucose–protein hydrogen bonds (distances  $\leq 3.5$  Å between N or O atoms) are shown by blue broken lines. As in (B), a sign on the O<sub>1</sub> atom indicates the direction of  $\alpha$ 1–6 linkage to GlcNAc, pointing toward this side of the plane. **(D)** Schematic representations of contacts between PhoSL and  $\alpha$ -fucose. Hydrophobic contacts (distances  $\leq 4.5$  Å between C atoms: red) and hydrogen bonds (defined above: blue) are indicated by broken lines. Bold lines indicate the contacts that are observed in 50% or more of the 60 pockets in the structure ensemble, while thin lines indicate those that are observed in 20% or more. The figures in (A–C) were produced by PyMOL.

are formed in a partial fraction; this may be more fixed when GlcNAc or N-glycan is attached. It should be noted that the O<sub>1</sub> atom of  $\alpha$ -fucose points outward from the pocket and is open for connection to GlcNAc (Fig. 5B,C).

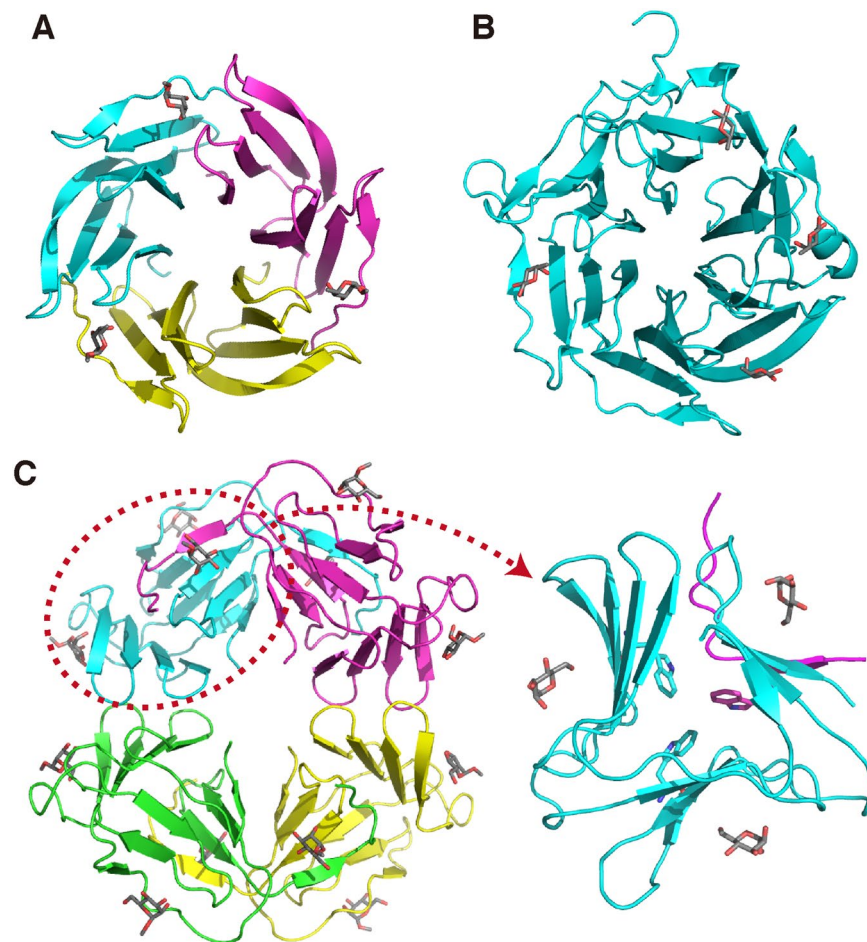
In the present study, we identified the binding site of fucose and a number of intermolecular interactions, which provided the structural basis for the molecular recognition. To reveal the mechanism of specificity for the  $\alpha$ 1–6 linkage, however, it will be necessary to analyze the structure of complex with a sugar containing GlcNAc. There may be favorable contacts between GlcNAc and PhoSL, or machinery for discrimination against the other types of linkages.

**Sharing four-stranded antiparallel  $\beta$ -sheets with other lectins.** A fucose-binding lectin from a bacterium *Ralstonia solanacearum* (RSL) forms a symmetric trimer as well (Fig. 6A)<sup>18</sup>. A monomer consists of a tandem of two four-stranded antiparallel  $\beta$ -sheets, resulting in a six-bladed  $\beta$ -propeller structure with a pseudo-six-fold symmetry for the trimer.

A very similar six-bladed  $\beta$ -propeller structure is observed for another fucose-binding lectin from a fungus *Aleuria aurantia* (AAL) (Fig. 6B)<sup>19</sup>. The latter, however, is a monomeric polypeptide possessing six repeats of four-stranded  $\beta$ -sheets.

A mannose-specific lectin (or agglutinin) from a higher plant *Galanthus nivalis* (GNA) forms a tetramer (Fig. 6C)<sup>20</sup>. Each monomer consists of three four-stranded antiparallel  $\beta$ -sheets. When two of the three  $\beta$ -sheets





**Figure 6.** Comparison of related lectin structures. (A) The crystal structure of RSL trimer in complex with methylfucose (PDB ID: 2BT9)<sup>18</sup>. (B) The crystal structure of AAL, consisting of a single polypeptide, in complex with fucose (PDB ID: 1OFZ)<sup>19</sup>. (C) The crystal structure of GNA tetramer in complex with methylmannose (PDB ID: 1MSA)<sup>20</sup>. In the right panel, a monomeric part of GNA with a swapping of a  $\beta$ -strand (circled in the left panel) is focused. In these figures, different polypeptide chains are distinguished by colors, and bound sugars are shown in stick representations. In the right panel of (C), a triad of Trp residues in the core is indicated also by stick.

in each monomer are considered, the structure of the tetramer appears to be an eight-bladed  $\beta$ -propeller structure partly similar to those of RSL and AAL. On the other hand, the three  $\beta$ -sheets in the monomeric part are arranged in a symmetric manner and comprise a pseudotrimeric  $\beta$ -prism structure similar to that of PhoSL (Fig. 6C, right panel). It also contains a triad of Trp residues at this pseudosymmetric center.

Nonetheless, the above resemblance of GNA monomer to PhoSL trimer is superficial. Namely,  $\beta$ -prism of GNA is formed essentially by a single polypeptide, with swapping of one  $\beta$ -strand with another monomer. Also, the  $\beta$ -strands in the front layer direct clockwise, while the corresponding  $\beta$ -strands of PhoSL direct counterclockwise (Figs 2B and 6C); this is also true when the structures are looked from the opposite direction. In addition the three Trp rings of GNA contact among one another through simple hydrophobic forces, but not through the NH- $\pi$  electrostatic forces. Furthermore, the binding pockets for sugar are formed over the  $\beta$ -sheet planes.

As above, RSL, AAL, and GNA commonly contain four-stranded antiparallel  $\beta$ -sheets as units. By rearrangements in the polypeptide level, they form  $\beta$ -propellers or similar structures. PhoSL is similar to them only in that it also contains similar antiparallel  $\beta$ -sheet as the structural unit. Indeed, when we used the Dali program<sup>21</sup> to search the data base for similar structures, only those possessing antiparallel  $\beta$ -sheets of similar length, including several  $\beta$ -propeller structures, were identified. This was also essentially the same when we artificially treated the PhoSL trimer as a single polypeptide before submitting to Dali. Therefore, the arrangement of the three  $\beta$ -sheets in PhoSL defines a novel type of structural scaffold.

## Materials and Methods

**Sample preparation.** A PhoSL peptide without N-terminal acetylation or C-terminal amidation (APVPVTKLVCDGDTYKCTAYLDYGDGK~~WVAQWDTAVFHTT~~) was chemically synthesized by the standard solid-phase method and purified to >95% by HPLC (BEX Co., Ltd., Tokyo, Japan). The above sequence is derived from the gene relevant to the lectin described in a published patent application<sup>12</sup> and is different from

that in the literature<sup>3</sup>, APVPVTKLVCDGDTYKCTAYLDFEGDGRWVAQWDTN<sup>V</sup>FHTG, at four positions (23rd, 27th, 35th, and 40th amino acids from the N-terminus, as underlined; see Fig. S2A). It was shown, however, that these alterations do not influence the specificity to core-fucosylated glycans<sup>12</sup>. In this patent, it was mistakenly described that this lectin was derived from a mushroom *P. terrestris*. However, it was later corrected to be derived from *P. squarrosa* in literature<sup>3</sup>. The concentration of the peptide in solutions was estimated from absorbance at 280 nm ( $A_{280}$ ), where molar absorption coefficient was calculated from the amino acid sequences<sup>22</sup>. The molar concentrations for PhoSL are shown as a monomer, unless otherwise stated.

**Gel filtration analysis.** The PhoSL peptide (96  $\mu$ M: 0.43 mg mL<sup>-1</sup>) and molecular marker proteins (0.5 mg mL<sup>-1</sup>), i.e., bovine serum albumin (69.3 kDa; Sigma-Aldrich), ovalbumin (42.9 kDa; Sigma-Aldrich), carbonic anhydrase (29.0 kDa; Sigma-Aldrich), and ribonuclease A (13.7 kDa; Sigma-Aldrich), were dissolved in 20 mM Tris-HCl (pH 7.5), 100 mM NaCl. The solutions of 200  $\mu$ L each were analyzed separately on a Superdex 75 10/300 column (GE Healthcare).

**Cross-linking analysis.** PhoSL (0.2 mM) was dissolved in 20 mM 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid (HEPES) (pH 7.5) (DOJIN, Kumamoto, Japan), 50 mM EDC (Peptide Institute, Inc., Osaka, Japan), 50 mM NHS (Wako, Osaka, Japan), and kept at room temperature for 1 h, allowing the linking reaction. This was stopped by adding sample solution for SDS-PAGE [50 mM Tris-HCl (pH 6.8), 2% SDS, 6%  $\beta$ -mercaptoethanol, 10% glycerol, and 0.02% bromophenol blue: final concentrations] and heating at 90 °C for 5 min. After SDS-PAGE, gel moieties containing bands were cut out. The gels were incubated in 100 mM NH<sub>4</sub>HCO<sub>3</sub> for 10 min and then crushed with a homogenization pestle. After purified with ZipTip C4 (Millipore) according to the protocol provided by the vendor, proteins were analyzed by mass spectrometry on Shimadzu Axima TOF<sup>2</sup>.

**NMR measurements.** For structure determination of PhoSL trimer, solution consisted of 0.3 mM PhoSL peptide, 20 mM d<sub>11</sub>-Tris-HCl (pH 7.5) (Isotec Inc.), 1 mM d<sub>10</sub>-dithiothreitol (DTT) (Isotec Inc.), 0.1 mM sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), and 5% D<sub>2</sub>O. In PhoSL-fucose titration analysis, a solution containing 50  $\mu$ M PhoSL peptide, 20 mM d<sub>11</sub>-Tris-HCl (pH 7.5), 0.2 mM d<sub>10</sub>-DTT, 0.1 mM DSS, and 5% D<sub>2</sub>O was titrated with a solution of 500 mM or 1 M L-fucose (Tokyo Chemical Industry, Co., Ltd., Tokyo, Japan) in 20 mM d<sub>11</sub>-Tris-HCl (pH 7.5). Solution for STD<sup>23</sup> experiment consisted of 50  $\mu$ M PhoSL peptide, 5 mM L-fucose, 20 mM d<sub>11</sub>-Tris-HCl (pH 7.5), 0.2 mM d<sub>10</sub>-DTT, 0.1 mM DSS, and 5% D<sub>2</sub>O. That for structure determination of PhoSL-fucose complex or analysis of chemical shift perturbation consisted of 0.3 mM PhoSL peptide, 100 mM L-fucose, 20 mM d<sub>11</sub>-Tris-HCl (pH 7.5), 1 mM d<sub>10</sub>-DTT, 0.1 mM DSS, and 5% D<sub>2</sub>O.

NOE spectroscopy (NOESY), total correlation spectroscopy (TOCSY), and double-quantum-filtered correlation spectroscopy (DQF-COSY) spectra<sup>10</sup> for structure determination and chemical shift perturbation were recorded on Bruker Avance III-900 (900.13 MHz for <sup>1</sup>H) spectrometers at 308 K, where mixing times of 100 ms and 50 ms for NOESY and TOCSY, respectively. Fucose-titration and STD experiments were performed on a Bruker Avance III-500 (500.13 MHz for <sup>1</sup>H) spectrometer at 308 K. For STD, a spectrum with irradiation on resonance of Leu21/H<sub>8</sub>1 for 2 sec was subtracted from a reference spectrum with an off-resonance irradiation. NOESY, TOCSY, and simple <sup>1</sup>H spectra of a sample dissolved in 99.96% D<sub>2</sub>O (Isotec Inc.) after lyophilization were recorded at 298 K, where 20 hydrogen bond donors were identified.

**Spectral analysis and structure determination of PhoSL trimer.** The NOESY, TOCSY, and DQF-COSY spectra were analyzed by Felix ver. 2007 (Felix NMR, Inc., San Diego, CA). The backbone and side chain <sup>1</sup>H resonances were assigned on the basis of the sequential NOEs<sup>10</sup>. Chemical shifts were referenced to the internal DSS. By analyzing the spectra, 14 and 5 pairs of H <sub>$\beta$</sub>  and valine H <sub>$\gamma$</sub>  resonances, respectively, were assigned stereospecifically and, at the same time, the relevant rotamers around the  $\chi^1$  angles were determined (methods described in refs<sup>24,25</sup>). Also, two pairs of Leu H <sub>$\delta$</sub>  resonances were spectroscopically assigned, and the relevant  $\chi^2$  rotamers were determined. Similarly, the  $\chi^1$  rotamers of three Thr residues were estimated.

A simulated annealing protocol with random initial velocities<sup>26</sup> and an extended initial structure was carried out by CNS ver. 1.3<sup>27</sup>. The distance constraints derived from the NOESY spectra and torsion angle constraints on the determined  $\chi^1$  and  $\chi^2$  rotamers were imposed as described previously<sup>28</sup>. For structure determination of the symmetric trimer, weak non-crystallographic symmetry (NCS) restraints were applied at 10 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Conditions for the molecular dynamics were: 50,000 K for 2,000 steps in torsion-angle space at high-temperature annealing stage; 50,000 K (initial) for 2,000 steps in torsion-angle space at the first slow cooling stage; and 2,000 K (initial) for 2,000 steps in Cartesian space at the second slow cooling stage.

NOEs were categorized with regard to inter- or intramolecular ones. All the sequential NOEs (NOEs between adjacent residues) were treated as intramolecular ones. For medium-range ( $2 \leq |i-j| \leq 4$ ;  $i$  and  $j$  are residue numbers of the two protons for the NOE) and long-range ( $|i-j| > 4$ ) NOEs, ambiguity in terms of this classification was introduced by means of the sum-averaging function. In order to reduce the search space of the random simulated annealing and increase the acceptance ratio, however, some NOEs from the interstrand interfaces were fixed as either intramolecular or intermolecular ones. This was systematically introduced, as follows.

When NOEs around the interface between  $\beta$ -strands 3 and 4, i.e., those between regions Lys27–Gln31 and Thr34–Thr39, which included typical interstrand backbone NOEs (Val29/HN–Phe37/HN, Ala30/H $\alpha$ –Val36/H $\alpha$ , and Gln31/HN–Ala35/HN in Fig. 2A), were treated as intramolecular ones, we could not obtain any acceptable structures on the basis without a distance violation greater than 0.2 Å, no torsion angle violation greater than 2° from 3000 trials. Alternatively, when these NOEs were treated as intermolecular ones, we obtained eight acceptable structures from 3000 trials. In all of the accepted structures, the strands were arranged as shown in Fig. 2C. Therefore, NOEs around other interfaces between  $\beta$ -strands were fixed accordingly. Namely, the NOEs were

intermolecular for Pro2–Val5 and Val9–Gly12 (the N-terminal and C-terminal halves of  $\beta$ -strand 1); intramolecular for Val3–Val9 and Cys17–Tyr23 ( $\beta$ -strands 1 and 2); and intramolecular for Lys16–Leu21 and Lys27–Thr34 ( $\beta$ -strands 2 and 3). Based on the calculated structure, acceptors of the hydrogen bonds were determined, and the relevant distance constraints were applied. Finally, in 50 trials, 20 structures were selected on the bases without a distance violation greater than 0.2 Å, no torsion angle violation greater than 2°, and the lowest total energies.

The structures of PhoSL– $\alpha$ -fucose and PhoSL– $\beta$ -fucose complexes were calculated respectively by a simulated annealing protocol similar to the above. Intensities of the NOEs between PhoSL and either of the two anomers were calibrated to the relative occupancies seen in the STD experiment (Fig. 4B), as if the binding pocket was fully occupied by a particular anomer. For the initial structure of the calculation, the minimized mean structure obtained by a calculation in advance without the PhoSL–fucose NOEs was used, instead of the extended structure. Conditions for the molecular dynamics were weakened, i.e., 1,000 K for 1,000 steps in torsion-angle space at the annealing stage; 1,000 K (initial) for 1,000 steps in torsion-angle space at the first cooling stage; and 1,000 K (initial) for 1,000 steps in Cartesian space at the second cooling stage. The NCS restraints were applied only to the protein moiety. Calculation was performed for 20 trials, and all of the structures were accepted, without a distance violation greater than 0.2 Å or a torsion angle violation greater than 2°.

**Accession numbers.** The co-ordinates of the determined structure and <sup>1</sup>H chemical shifts for the free state of PhoSL have been deposited to the Protein Data Bank (PDB) and BMRB under accession IDs 5xzk and 36103, respectively.

## References

- Miyoshi, E., Moriwaki, K. & Nakagawa, T. Biological function of fucosylation in cancer biology. *J. Biochem.* **143**, 725–729 (2008).
- Aoyagi, Y. *et al.* The usefulness of simultaneous determinations of glucosaminylation and fucosylation indices of alpha-fetoprotein in the differential diagnosis of neoplastic diseases of the liver. *Cancer* **67**, 2390–2394 (1991).
- Kobayashi, Y. *et al.* A novel core fucose-specific lectin from the mushroom *Pholiota squarrosa*. *J. Biol. Chem.* **287**, 33973–33982 (2012).
- Tateno, H., Nakamura-Tsuruta, S. & Hirabayashi, J. Frontal affinity chromatography: sugar-protein interactions. *Nat. Protoc.* **2**, 2529–2537 (2007).
- Shimomura, M. *et al.* Establishment of a novel lectin-antibody ELISA system to determine core-fucosylated haptoglobin. *Clin. Chim. Acta* **446**, 30–36 (2015).
- Fujita, K. *et al.* Decreased fucosylated PSA as a urinary marker for high Gleason score prostate cancer. *Oncotarget* (2016).
- Llop, E. *et al.* Improvement of Prostate Cancer Diagnosis by Detecting PSA Glycosylation-Specific Changes. *Theranostics* **6**, 1190–1204 (2016).
- Okuyama, M. *et al.* Target-selective photo-degradation of AFP-L3 and selective photo-cytotoxicity against HuH-7 hepatocarcinoma cells using an anthraquinone-PhoSL hybrid. *Chem. Commun. (Camb.)* **52**, 2169–2172 (2016).
- Ueda, M. *et al.* Specific increase in serum core-fucosylated haptoglobin in patients with chronic pancreatitis. *Pancreatology* **16**, 238–243 (2016).
- Wüthrich, K. *NMR of Proteins and Nucleic Acids*. (John Wiley & Sons, Inc., 1986).
- Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M. & Tanabe, K. Origin of the attraction and directionality of the NH/pi interaction: Comparison with OH/pi and CH/pi interactions. *J. Am. Chem. Soc.* **122**, 11450–11458 (2000).
- Kobayashi, Y. *et al.* Jpn. Kokai Tokkyo Koho JP2011–148735 (2011).
- Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–258 (2014).
- Vainauskas, S. *et al.* Profiling of core fucosylated N-glycans using a novel bacterial lectin that specifically recognizes alpha1,6 fucosylated chitobiose. *Sci. Rep.* **6**, 34195 (2016).
- Oda, Y. *et al.* A new fungal lectin recognizing alpha(1-6)-linked fucose in the N-glycan. *J. Biol. Chem.* **278**, 32439–32447 (2003).
- Do Nascimento, A. S. *et al.* Algal lectin binding to core (alpha1-6) fucosylated N-glycans: structural basis for specificity and production of recombinant protein. *Glycobiology* **25**, 607–616 (2015).
- Laughrey, Z. R., Kiehna, S. E., Riemen, A. J. & Waters, M. L. Carbohydrate-pi interactions: what are they worth? *J. Am. Chem. Soc.* **130**, 14625–14633 (2008).
- Kostlanova, N. *et al.* The fucose-binding lectin from *Ralstonia solanacearum*. A new type of beta-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan. *J. Biol. Chem.* **280**, 27839–27849 (2005).
- Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C. & Imberty, A. Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for *Aleuria aurantia* lectin. *J. Biol. Chem.* **278**, 27059–27067 (2003).
- Hester, G., Kaku, H., Goldstein, I. J. & Wright, C. S. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nat. Struct. Biol.* **2**, 472–479 (1995).
- Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–549 (2010).
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption-coefficient of a protein. *Protein Sci.* **4**, 2411–2423 (1995).
- Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **123**, 6108–6117 (2001).
- Zuiderweg, E. R., Scheek, R. M. & Kaptein, R. Two-dimensional 1H-nmr studies on the lac repressor DNA binding domain: further resonance assignments and identification of nuclear Overhauser enhancements. *Biopolymers* **24**, 2257–2277 (1985).
- Hyberts, S. G., Marki, W. & Wagner, G. Stereospecific assignments of side-chain protons and characterization of torsion angles in Eglin c. *Eur. J. Biochem.* **164**, 625–635 (1987).
- Nilges, M., Clore, G. M. & Gronenborn, A. M. Determination of 3-Dimensional Structures of Proteins from Interproton Distance Data by Dynamical Simulated Annealing from a Random Array of Atoms - Circumventing Problems Associated with Folding. *FEBS Lett.* **239**, 129–136 (1988).
- Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* **54**, 905–921 (1998).
- Yamasaki, K. *et al.* A novel zinc-binding motif revealed by solution structures of DNA-binding domains of Arabidopsis SBP-family transcription factors. *J. Mol. Biol.* **337**, 49–63 (2004).
- Laskowski, R. A., Rullmann, J. A. C., Macarthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
- Hansman, G. S. *et al.* Structural basis for norovirus inhibition and fucose mimicry by citrate. *J. Virol.* **86**, 284–292 (2012).

## Acknowledgements

This project was supported by Leading Engine program for Accelerating Drug discovery (LEAD) of AIST. We thank Mr. O. Tani of AIST for his support in the mass spectrometry, and Drs. T. Kubota, A. Shimizu, and Y. Chiba for helpful discussions. We also thank RIKEN Yokohama NMR facility for the access to the two 900-MHz NMR instruments, where we especially appreciate Drs. F. Hayashi and A. Matsugami for supports in the measurements.

## Author Contributions

K.Y. and H.T. designed the study. K.Y. conducted NMR analyses, and wrote the paper. T.Y. conducted gel-filtration and cross-linking analyses. All authors reviewed the results and approved the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-25630-2>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018