JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

A Multivalent Marine Lectin from *Crenomytilus grayanus* Possesses Anti-cancer Activity through Recognizing Globotriose Gb3

Jiahn-Haur Liao,[†] Chih-Ta Henry Chien,[†,‡] Han-Ying Wu,[§,∥] Kai-Fa Huang,[†] Iren Wang,[†] Meng-Ru Ho,[†] I-Fan Tu,[†] I-Ming Lee,[#] Wei Li,[⊗] Yu-Ling Shih,[†] Chung-Yi Wu,[⊥] Pavel A. Lukyanov,[▽] Shang-Te Danny Hsu,[*,†,§,#] and Shih-Hsiung Wu[*,†,‡,§,#]

[†]Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan

[‡]Department of Chemistry, National Taiwan University, Taipei 106, Taiwan

[§]Institute of Biological Chemistry, Chemical Biology and Molecular Biophysics Program, Taiwan International Graduate Program, Academia Sinica, Taipei 115, Taiwan

[∥]Department of Chemistry, National Tsing Hua University, Hsinchu 30043, Taiwan

[#]Institute of Biochemical Science, National Taiwan University, Taipei 106, Taiwan

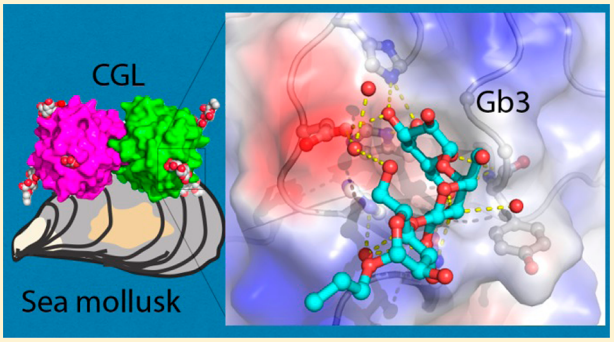
[⊗]Key Laboratory of Aquatic Products Processing and Utilization of Liaoning Province, Dalian Ocean University, Dalian 116023, P.R. China

[⊥]Genomics Research Center, Academia Sinica, Taipei 11529, Taiwan

[▽]G.B. Elyakov Pacific Institute of Bioorganic Chemistry, Far Eastern Branch, Russian Academy of Sciences, Vladivostok 690022, Russian Federation

Supporting Information

ABSTRACT: In this study, we report the structure and function of a lectin from the sea mollusk *Crenomytilus grayanus* collected from the sublittoral zone of Peter the Great Bay of the Sea of Japan. The crystal structure of *C. grayanus* lectin (CGL) was solved to a resolution of 1.08 Å, revealing a β-trefoil fold that dimerizes into a dumbbell-shaped quaternary structure. Analysis of the crystal CGL structures bound to galactose, galactosamine, and globotriose Gb3 indicated that each CGL can bind three ligands through a carbohydrate-binding motif involving an extensive histidine- and water-mediated hydrogen bond network. CGL binding to Gb3 is further enhanced by additional side-chain-mediated hydrogen bonds in each of the three ligand-binding sites. NMR titrations revealed that the three binding sites have distinct microscopic affinities toward galactose and galactosamine. Cell viability assays showed that CGL recognizes Gb3 on the surface of breast cancer cells, leading to cell death. Our findings suggest the use of this lectin in cancer diagnosis and treatment.



1. INTRODUCTION

Lectins are ubiquitous carbohydrate-binding proteins with hemeagglutination activity. Lectins can be found in many animals, plants, and bacteria, and they play important roles in innate immune response by recognizing specific carbohydrate moieties on the surface of potential pathogens.[1] In plants, lectins are known to ward off potential pathogens, while in animals, lectins are also found to aid cell interactions.[2] Lectins recognize carbohydrates through hydrogen bonds, van der Waals forces, hydrophobic interactions, and metal coordination. The carbohydrate binding affinities of individual lectin molecules are generally much weaker—typical dissociation constants (K_d) are in the millimolar (mM) range—than those of protein−protein interactions.[3] Nevertheless, such a weak interaction could be enhanced by several orders of magnitude in biological systems through multivalency, given that carbohydrates are abundantly present on cellular surfaces and thus providing high local concentrations for cross-linking lectins. The diverse biological activities conferred by lectin−carbohydrate binding, as well as the molecular structures and specificities of lectins, allow these proteins to form a large and heterogeneous protein superfamily.[4] Galectins, C-type, I-type, P-type, and pentraxins are well-studied classes of lectins in higher animals.[5]

A number of isolated lectins from sea hydrobionts have been categorized in the C-type lectin family,[6] which features Ca^{2+}-dependent carbohydrate binding capacity. Other marine lectins such as those from the Mytilidae family, however, do not belong to this type of lectin.[7,8] Recently, a novel lectin from the

Received: January 5, 2016
Published: March 24, 2016

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787−4795

sea mollusk *Crenomytilus grayanus* collected from the sublittoral zone of Peter the Great Bay of the Sea of Japan (East Sea) was purified and characterized.[8] The *C. grayanus* lectin (hereafter denoted as CGL) can agglutinate all the types of human erythrocytes, which can be effectively inhibited by *N*-acetyl-D-galactosamine (GalNAc), D-galactose, and D-talose.[8,9] Based on the GalNAc/Gal specificity, CGL is proposed to be a member of the galactose-binding lectin family. However, the amino acid sequence of CGL has no resemblance to galectins or other animal lectins except MytiLec.[7,8] Molecular modeling suggested that CGL has a β-trefoil fold similar to that of the ricin B-like lectin.[9,10] The β-trefoil fold is a prevalent protein fold that is also present in cytokines, agglutinins, and actin-cross-linking proteins. These proteins, however, are evolutionally divergent with a broad range of ligand-binding preferences and distinct biological functions.[11] To investigate CGL in more detail, we determined the crystal structures of CGL in its apo form and in complex with different ligands. We also used solution state NMR spectroscopy to show that CGL exhibits distinct microscopic binding affinities within its three different ligand-binding sites. As CGL was previously reported to display antibacterial and antifungal activities that could be involved in the clearance of bacterial pathogens and inhibition of fungal growth in the sea mussel,[8,12] we demonstrated that CGL, like MytiLec,[7] can recognize globotriose (Gb3; Galα1−4Galβ1−4Glc), which results in Gb3-dependent cytotoxicity in cancer cells.

2. RESULTS

2.1. Overall Structure of CGL. To date, many high-resolution atomic structural models of lectins have been reported,[13] which reveal that many lectins from different organisms are structurally similar despite the lack of sequence similarity. To gain structural insights into CGL, we independently determined the crystal structure of CGL by the sulfur single-wavelength anomalous dispersion phasing method,[14] and the result was essentially identical to the recently reported CGL structure (PDB: 5DUY) determined by molecular-replacement (the positional root mean squared deviation of the backbone Cα atoms between these two structures is 0.2 Å with a Z-score of 34.1).[15] CGL exhibits a typical β-trefoil fold comprising 13 β-strands and 3 α-helixes (Figures 1 and S1). The Ramachandran plot shows that the glutamic acid at position 49 (Glu49) falls in the disallowed region (Figure S2), which may be attributed to its location at the tight turning point of β-hairpin between β-4 and β-5. The primary sequence of CGL contains three 39-residue repeats, which fold into three subdomains that we term as α, β, and γ subdomains. Consequently, CGL exhibits a pseudo-three-fold-symmetric structure with overall dimensions of approximately 31 Å × 32 Å × 35 Å. Furthermore, CGL homodimerizes into a dumbbell-shaped quaternary structure with an extensive dimeric interface that is composed of six hydrogen bonds and several nonbonded contacts to form a 710 Å^2 interface area. In particular, the stacking and intercalation of the aromatic side chains of Phe94, Phe95, and Tyr149 in addition to extensive hydrophobic interactions between Ala91, Met92, Leu147, and Ala150 are important in keeping the stable dimeric assembly (Figure S3). Such a quaternary arrangement is supported by PDBePISA (http://www.ebi.ac.uk/pdbe/pisa/) to be the most stable oligomeric form, where the free energy of assembly dissociation is −1.2 kcal/mol.[16] Although our finding is at variance with previous reports, which indicated that CGL is

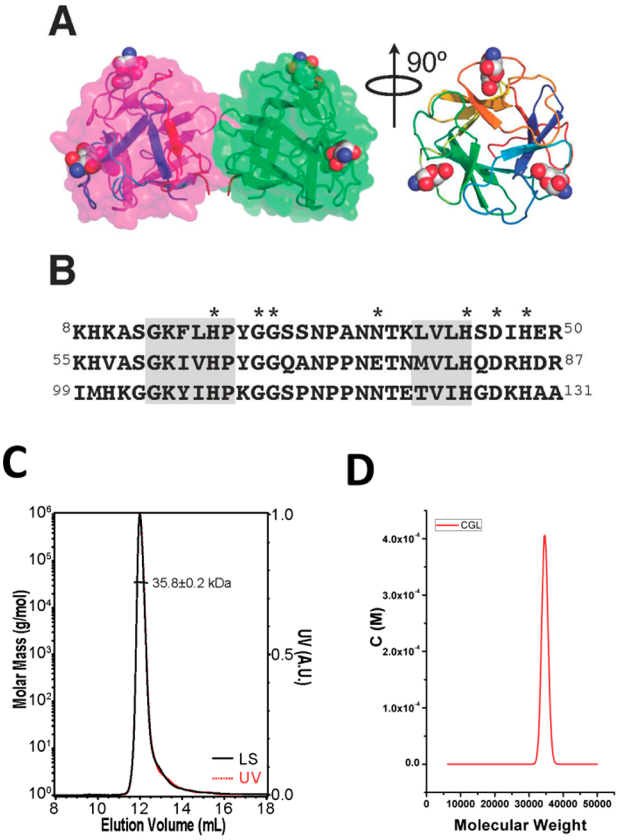


Figure 1. (A) Crystal structure of CGL, showing a homodimer with each protomer having three carbohydrate-binding sites. The structures of galactosamine are shown in spheres with the structure of CGL shown in cartoon representation and color-ramped from blue to red for the N- to C-termini. (B) Sequence alignment of the pseudo-three-fold-symmetric carbohydrate-binding motifs. Residues that adopt β-stranded secondary structures are shed in gray while those that are in contact with carbohydrate are indicated with asterisks on the top of the alignment. (C) SEC-MALS of CGL. The horizontal thin line segments represent the calculated molar mass of the corresponding peak (left ordinate axis). The normalized UV absorbance at 280 nm (right ordinate axis) and intensity of light scattering (LS) are shown in dashed red line and solid black line, respectively. (D) SV-AUC of CGL. The single peak distribution corresponds to a dimeric CGL with a molecular weight of 34 kDa.

primarily monomeric both in the crystalline and solution states,[9,15,17] we independently confirmed the dimeric state of CGL in solution state by SEC-MALS, which indicated that the single, monodispersed (polydispersity = 1.000) elution peak corresponds to a molecular weight (MW) of 35.8 ± 0.2 kDa that is consistent with the expected value for a dimeric CGL (36 kDa) (Figure 1C). The SV-AUC analysis also confirmed the dimeric state of CGL in solution (Figure 1D).

2.2. Structural Homologues of CGL. Given the prevalence of β-trefoil fold in protein structures, we carried out a structural homology search for CGL by the DALI server and identified several structural homologues.[18] Molecular phylogenetic analysis of CGL and these homologues shows that CGL and several lectins from the *Mytilus* genus belong to the same clade (Figure 2). These structural homologues of CGL from the *Mytilus* genus are the only ones that exhibit sequence homology higher than 60% with CGL. In particular, CGL and MytiLec (PDB: 3WMU) share 88% sequence identity with a Z-score value equal to 32.9 according to

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787−4795

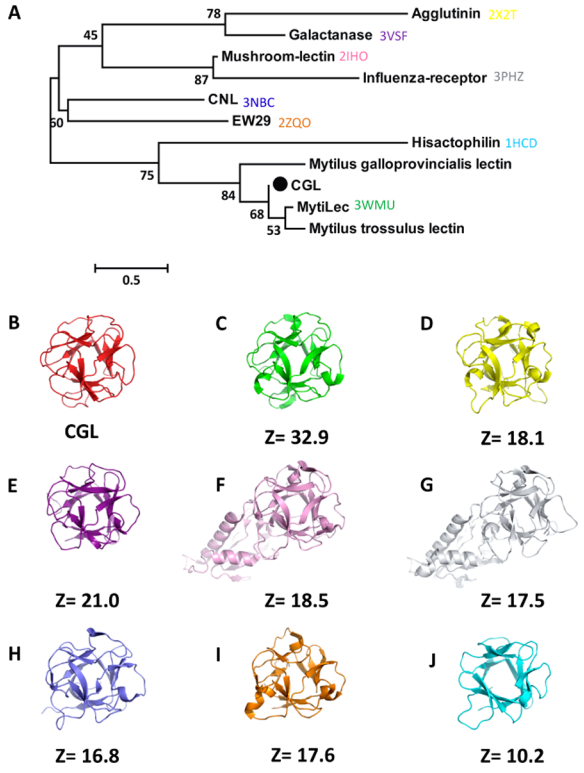


Figure 2. Molecular phylogenetic analysis of CGL. (A) Molecules with trefoil structure are analyzed by the Maximum Likelihood method. CGL falls in the same clade with hisactophilin. The tree is drawn with branch lengths measured in the number of substitutions per site set as the scale. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The trefoil domain structures of these proteins (with PDB codes in parentheses) are shown as follow: (B) CGL, (C) MytiLec (3WMU), (D) agglutinin (2X2T), (E) galactanase (3VSF), (F) mushroom lectin (2IHO), (G) influenza receptor (3PHZ), (H) CNL (3NBC), (I) EW29 (2ZQO), and (J) hisactophilin (1HCD). The corresponding Z-score is shown under each structure.

DALI (Figures 2). The substrate binding modes are highly conserved between CGL and MytiLec with subtle differences in the hydrogen bond network that may be associated with difference in substrate specificity and affinity (Figure S4). The carbohydrate-binding module-13 (CBM13) domain of Ct1,3Gal43A (PDB: 3VSF), which is an exo-β-1,3-galactosidase, has a Z-score value of 21.3;[19] *S. sclerotiorum* agglutinin (SSA, PDB: 2X2T) from a plant pathogenic fungus has a Z-score of 18.1;[20] and galactose-binding lectin, EW29 (PDB: 2ZQO), from the earthworm *Lumbricus terrestris* has a Z-score of 17.5.[21] These three structural homologues exhibit low sequence identity of less than 15% with CGL. Furthermore, their quaternary structures are also different from that of CGL (Figure S5). Finally, an unrelated, engineered trefoil protein (PDB: 3PG0) has a Z-score of 21.2.[11] These four trefoil proteins differ from CGL in the additional β-sheet that is only present at the C-terminus of CGL (β-13; cf. Figure S1). The additional β-sheet at the C-terminus of CGL contains several positive charges on the surface of the CGL dimer, which may be implicated in the previously suggested DNA-binding function (Figure S6).

2.3. Structural Basis of Ligand Recognition of CGL. CGL-mediated hemagglutination was previously reported to be inhibited by several monosaccharides, including GalNAc, galactose, and talose.[8] We confirmed by biolayer interferometry that CGL binds to galactose and galactosamine with K_d values of 51 and 57 μM, respectively (Figures S7). Nevertheless, CGL exhibits no appreciable binding to glucose, mannose, xylose, or arabinose. We subsequently solved the crystal structures of CGL in complex with galactose and galactosamine to a resolution of 1.56 and 1.70 Å, respectively, by soaking the ligands with apo-form crystals. The galactosamine-bound CGL structure reveals three carbohydrate-binding sites on the surface of each CGL monomer (Figure S8). Site 1 consists of His16, Gly19, Asp35, His37, and Asn119; Site 2 consists of His64, Gly67, Asp83, His85, and Asn27; Site 3 consists of His108, Gly111, Asp127, His129, and Glu75. Superposition of the three carbohydrate-binding sites within the same protomer indicates that all three sites utilize the same amino acid compositions, except for a glutamate-to-asparagine replacement at Site 3, Glu75, to bind to the ligands (Figures 3 and S9).

While our crystallographic finding confirmed the predictions based on homology modeling[17] that the conserved His, Gly, Asp, Asn, and Tyr side chains form a ligand-binding cleft for carbohydrate recognition, our structures also revealed the involvement of several water molecules in mediating ligand formation of galactosamines, which make significant contributions to the binding. Similar binding poses are also found in the galactose-bound CGL structure (Figure 3B). Sequence alignment of the α, β, and γ subdomains of CGL reveals a highly conserved carbohydrate-binding motif, which consists of two histidine residues, a glycine, and an aspartic acid that are conserved (Figure 1). Interestingly, the asparagine and the glutamic acid can be aligned in the same position in the crystal structure. In the crystal structure, galactose is bound to CGL as an α-anomer, except for Site 3 of chain A, which is recognized as a β-anomer. The anomeric hydroxyl group of the β-anomer forms a hydrogen bond with the carbonyl group of Gly111. Electron density map indicates both anomers of galactose may be present in the crystal. Galactose C2-OH does not interact with any residue of CGL. The C3-OH of galactose is recognized by the side chains of Asp35 and His37 at Site 1, Asp83 and His85 at Site 2, and Asp127 and His129 at Site 3. The C4-OH of galactose is further recognized by His16 and His37 at Site 1, His64 and His85 at Site 2, and His108 and His129 at Site 3. The C6-OH is recognized by Gly19 and Asn119 at Site 1 and Gly67 and Asn27 at Site 2, while in Site 3, the C6-OH is recognized by Gly111 and a glutamate, Glu75, instead of an asparagine. Likewise, galactosamine would interact with CGL in the same way (Figure 3).

MytiLec from *Mytilus galloprovincialis* was reported to recognize globotriose Gb3, which is abundantly present on the cancer cell surface.[7] The sequence and structural homologies between CGL and MytiLec led us to investigate the binding ability of Gb3 to CGL. Indeed, biolayer inferometry shows that CGL binds to the Gb3 allyl with a K_d of 14 μM (Figure S10). We subsequently determined the crystal structure of CGL in complex with Gb3 at 1.64 Å resolution (Table 1), which showed that CGL binds to the Gb3 allyl primarily through the terminal galactose moiety (Gal1, see Figure 3) in the same way as galactose binding. In addition to the residues involved in the interaction with the Gal1 of the Gb3 allyl, Sites 1 and 2 have additional interactions, involving the amides of the carboxamide moieties of Asn27 and Asn119, with the second galactose moiety (Gal2) of the Gb3 allyl; the side-chain hydroxyl group of Tyr18 at Site 1 and that of Tyr66 at Site 2 also make an additional water-mediated hydrogen bond with

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787−4795

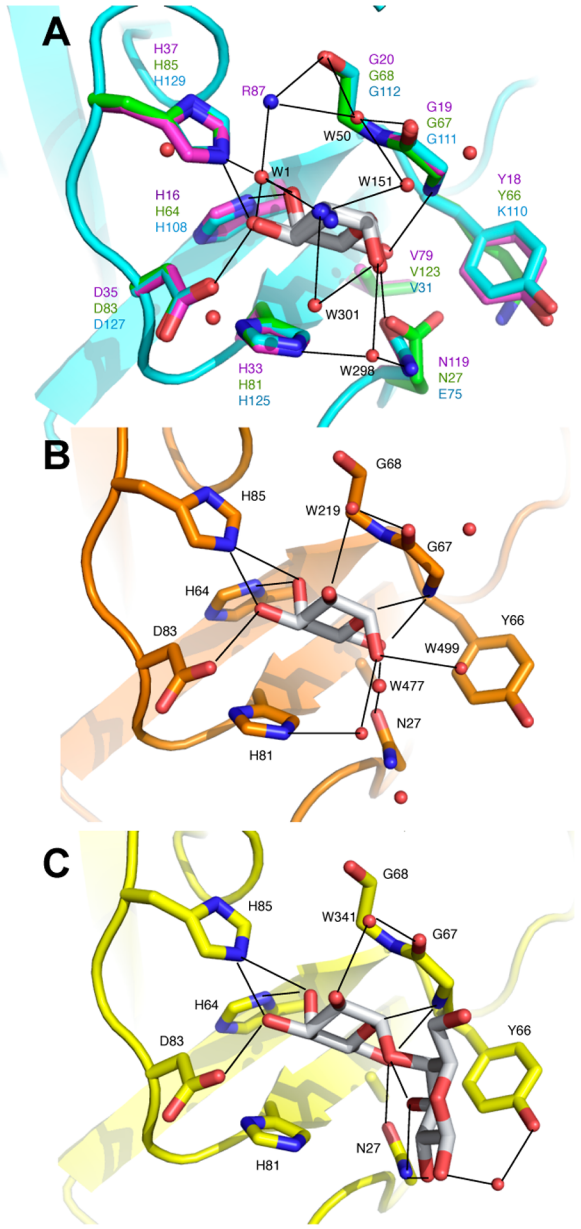


Figure 3. Carbohydrate-binding sites of CGL. (A) Superposition of the three carbohydrate-binding sites within the same protomer. The residues that are involved in galactosamine interactions are shown in stick representation with their respective sequence identities indicated in magenta, green, and cyan for the three different binding motifs, respectively. Solid lines indicate the hydrogen bonds that are involved in galactosamine interactions for the second binding motif (green). (B,C) Representative views of the crystal structures of CGL in complex with galactose (B) and Gb3 (C).

the Gal2 of the Gb3 allyl, which is absence at Site 3 due to the tyrosine-to-lysine replacement (Lys110, see Figure 3). These interactions may be accountable for the better affinity with Gb3 than with galactose alone; this may also account for the partial structure of the Gb3 allyl at Site 3 due to the lack of additional interactions with Lys110. Compared to the previously reported structure of Shiga-like Toxin I in complex with Gb3, which utilizes a Trp side chain to make hydrophobic interactions with the Gal1 of Gb3,[22] CGL displays a unique Gb3 recognition motif among all structurally characterized lectins that may be shared with MytiLec.[23]

2.4. NMR Analysis of CGL–Ligand Interactions. To further characterize the structure and dynamics of CGL in the context of ligand recognition, we recorded a suite of standard triple resonance NMR experiments to obtain near complete (92%, 1416 out of 1533) overall NMR assignments for CGL (Figure S11), including aromatic side-chain resonances. The chemical shift assignments were deposited in the BioMagResBank (http://www.bmrb.wisc.edu/) under the accession number 25371. Secondary structure prediction using TALOS +[24] and CamCoil[25] showed that CGL contains 13 β-strands and three α-helices that are consistent with the crystal structure of CGL, and that the backbone is highly ordered except for the loops connecting β2 and β3, β6 and β7, and β10 and β11 (Figure S12). ^{15}N spin relaxation dynamics analysis of CGL showed that the backbone amide ^{15}N R_1 and R_2 rates are relatively homogeneous across the sequence. The average R_2 rate of CGL is 20.5 ± 0.2 s^{-1}, and that of the residues in the loop region is 20.4 ± 0.3 s^{-1}, indicating that the loop regions are also highly ordered (Figure S12). Additionally, the rotational correlation time τc derived from the R_1/R_2 ratio (0.057 ± 0.011) is 13.1 ns, which equates to an apparent MW of 22 kDa, according to the correlation between MW and τ_c of proteins studied under the same temperature and external magnetic fields.[26] While the estimated MW is significantly smaller than the expected value for a dimeric CGL (36 kDa), this underestimation may be attributed to the highly asymmetric dumbbell shape of the dimeric assembly leading to rotational diffusion anisotropy.[27,28]

Using the assigned backbone amide ^{15}N–^{1}H correlations, we carried out ligand titration experiments to monitor the carbohydrate binding events. In line with the crystallographic structural findings, the addition of 50-fold galactose and galactosamine (5 mM ligands versus 0.1 mM CGL) resulted in significant chemical shift perturbations in the residues involved in ligand binding, i.e., residues 35–40, 81–90, 124–127 (Figure 4). Furthermore, the extents to which the chemical shifts were affected were very similar for the galactose- and galactosamine-bound forms of CGL, indicating that the overall binding mode and binding site of both carbohydrates were the same. Structural mapping of the observed chemical shift perturbations (significantly perturbed residues are defined as those that exhibit backbone amide chemical shift perturbations of more than one standard deviation) showed localized chemical shift perturbations due to ligand binding (Figure 5). Specifically, Ser22, Ile36, His37, Glu38 were located in Site 1; Ile54, His81, Arg84, His85, Asp86, and Leu89 were located in Site 2; and Asn74, Lys110, Ser113, Ile124, His125, Gly126, and Asp127 were located in Site 3. These results were consistent with the crystallographic findings.

Close examination of the chemical shift perturbation as a function of ligand concentration reveals very different behaviors for the histidines involved in ligand binding. We therefore fit the individual titration curves and obtained their corresponding K_d values, which correspond to the microscopic binding processes at different ligand-binding sites (Figure 5). For Site 2, there were two histidine residues involved in ligand binding, and indeed, their titration curves yielded similar K_d values. We therefore global-fit their titration curves to obtain an apparent K_d value for galactose. The results showed that the binding affinities of the three binding sites were significantly different: the corresponding K_d value for Site 1 was 178 ± 67 μM; Site 2, 1288 ± 259 μM; Site 3, 815 ± 57 μM. Similar analysis was applied to the galactosamine titration analysis (Figure S13).

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787–4795

Table 1. Data Collection and Refinement Statistics[a]

	native CGL	CGL-galactose	CGL-galactosamine	CGL-Gb3 allyl
		Data Collection		
resolution (Å)	30−1.08 (1.12−1.08)	30−1.56 (1.62−1.56)	30−1.70 (1.76−1.70)	30−1.64 (1.70−1.64)
unit cell dimensions, *a*, *b*, *c* (Å)	52.84, 71.06, 93.97	53.26, 72.54, 94.62	53.02, 71.48, 94.07	53.04, 72.46, 94.36
total observations	959 699	680 005	462 570	884 145
unique reflections	149 358 (14 837)	51 874 (4367)	38 277 (3604)	44 506 (3980)
multiplicity	6.4 (5.9)	13.1 (8.6)	12.1 (11.1)	19.9 (10.8)
completeness (%)	98.3 (99.2)	98.1 (83.5)	95.4 (91.2)	99.1 (90.6)
$I/\sigma(I)$	23.9 (2.2)	63.9 (17.9)	40.5 (6.5)	44.1 (4.2)
R_{merge} (%)	7.7 (84.7)	3.7 (11.0)	6.2 (25.3)	11.1 (44.0)
		Refinement		
resolution (Å)	30−1.08	30−1.56	30−1.70	30−1.64
reflections [$>0\sigma(F)$], working/test	134 578/7345	46 541/2635	33 804/1875	39 094/2090
R factor/R_{free}	0.114/0.131	0.105/0.139	0.123/0.152	0.117/0.167
rmsd bond lengths (Å)/angles (deg)	0.008/1.467	0.008/1.396	0.013/1.553	0.014/1.783
average *B* factor (Å^2)/no. of atoms				
protein	8.9/2410	10.4/2396	10.3/2396	11.6/2404
sugar		12.9/72	10.8/72	31.9/157
glycerol/malate	16.1/48	38.8/12	35.2/6	30.1/9
water	31.4/555	30.2/515	31.2/410	30.5/423
Ramachandran plot, residues in %				
most favored regions	88.4	88.8	88.0	88.4
additionally allowed regions	10.8	10.4	11.2	10.8
generously allowed regions	0	0	0	0
disallowed regions	0.8	0.8	0.8	0.8
PDB code	5F8S	5F8W	5F8Y	5F90

[a]These crystals are in space group $P2_12_12_1$, and one asymmetric unit comprises two CGL molecules. Values in parentheses correspond to the highest resolution shell.

The binding affinities of galactosamine in Sites 1 and 3 were better than that for galactose: K_d values were 32 and 302 μM, respectively. In contrast, Site 2 bound to galactosamine with a weaker affinity than that for galactose (K_d = 5135 μM). While CGL makes the same number of direct hydrogen bonds with galactose at different sites, the number of water-mediated hydrogen bonds vary between different sites (Figure S14). As such, the dissociation constants for galactose at different sites exhibit a linear correlation with the number of water-mediated hydrogen bonds in a semilogarithms plot (Figure S15A). In the case of galactosamine, however, such a correlation is lost (Figure S15B). Such a difference may be attributed to the differences in surface charge distributions. The predominantly negatively charged Site 3 surface—in part due to the glutamate-to-asparagine replacement at position 75, Glu75—compared to Sites 1 and 2 may account for the reduction in galactosamine binding affinity due to electrostatic attraction with the positively charged amine group of galactosamine (Figures S8 and S14). Collectively, these data suggest that the micro-environments of these three carbohydrate-binding sites are different, and multivalency may play an important role in the biological functions of CGL in the context of cellular communication and recognition.

2.5. CGL Reduces Cancer Cell Viability. Some lectins have exhibited anti-cancer activity, with the ability to halt tumor cell growth through apoptotis induction.[29] For example, MytiLec can reduce Raji cell viability, which may be attributed to the ability of MytiLec to bind Gb3, which is abundantly distributed on the surface of Raji cells7. Having established the Gb3-binding capacity of CGL, we next tested whether CGL could reduce cell viability of the breast cancer cell line, MCF7, which also possesses abundant Gb3 on the cell surface.[30,31] Indeed, CGL treatment resulted in the detachment of MCF7 cells (Figure 6A,B).[32] Three-dimensional fluorescence imaging reconstruction with fluorescently labeled CGL subsequently revealed CGL to be distributed mostly on the cell surface of the MCF7 cells (Figure 6). Interestingly, we found that Gb3 may be unevenly distributed on the cell surface of MCF7. Using the anti-cancer drug, sorafenib, as a control in an MTT assay, we confirmed the dose-dependent anti-cancer activity of CGL, which was as high as 33% cell death with 200 μg/mL CGL (Figure 6F).

3. DISCUSSION

Reagents that can induce tumor cell death comprise one of the key strategies for cancer chemotherapy. A vast number of lectins have been shown to halt the growth of tumor cells through apoptotis induction,[2] which they effect through a variety of pathways. For example, *Bauhinia forficata* lectin induces cell death and inhibits integrin-mediated adhesion of MCF7 human breast cancer cells.[33] *Sclerotium rolfsii* lectin strongly inhibits proliferation and induces apoptosis in MCF-7 and ZR-75 human breast cancer cells while weakly inhibiting proliferation of non-tumorigenic MCF-10A and HMEC human breast cells.[34] In addition to anti-cancer activities, lectins also possess other biological activities. For instance, extracellular galectin-3 can induce programmed cell death of human T-cells through CD7 and CD29 binding, which activates an apoptosis signal to the mitochondria.[35]

Herein, we report the structure and function of a novel galactose-binding lectin, CGL, which adopts a β-trefoil fold (Figure 1). The Pfam protein family database (http://pfam.

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787−4795

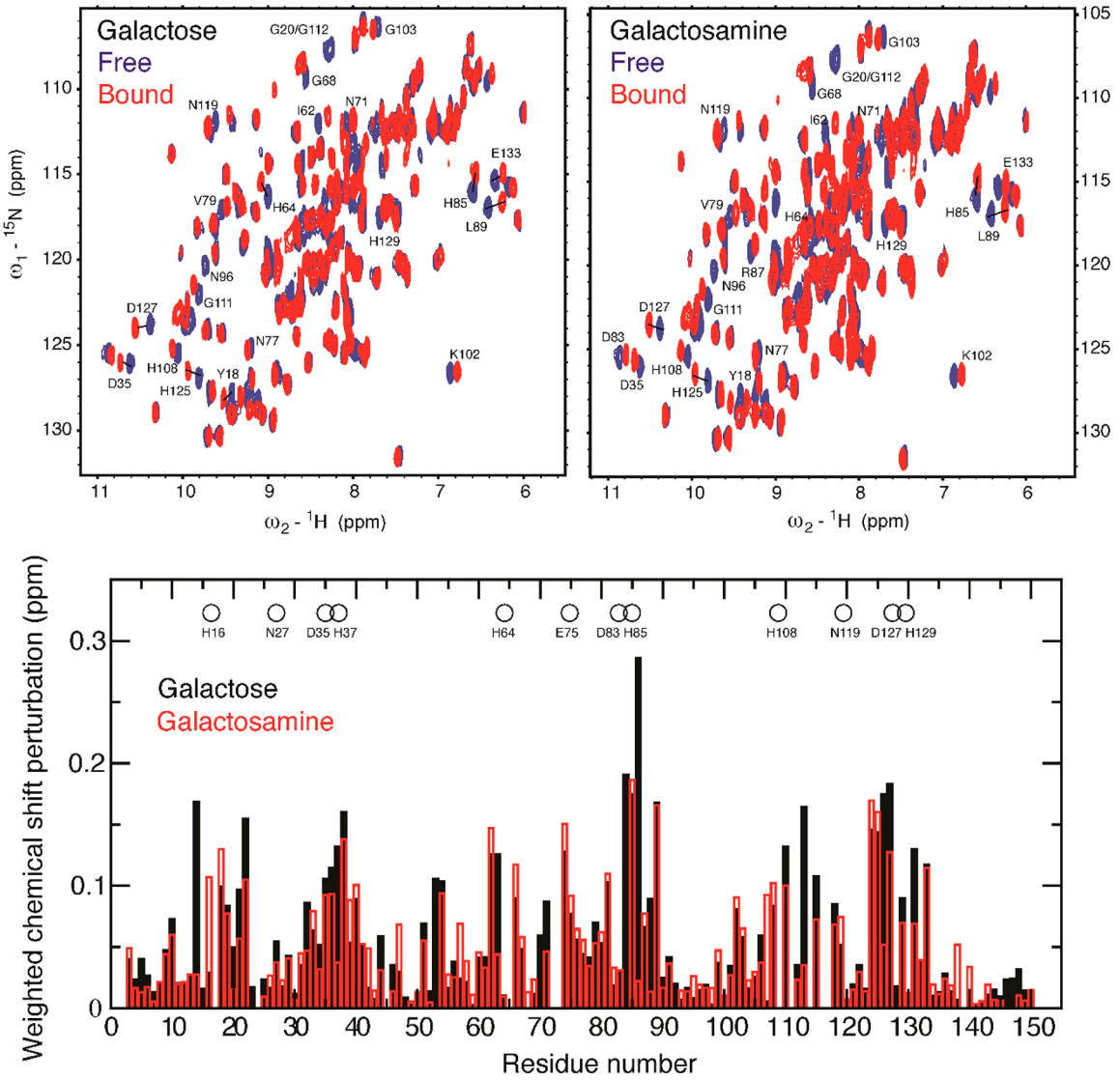


Figure 4. Chemical shift perturbation of CGL upon galactose and galactosamine binding. (A) $^{15}N-^{1}H$ HSQC spectrum overlay of free (blue) and galactose-bound (red) CGL. (B) $^{15}N-^{1}H$ HSQC spectrum overlay of free (blue) and galactosamine-bound (red) CGL. (C) Comparison of weighted chemical shift perturbation of galactose binding (black) and galactosamine binding (red) as a function of residue number.

sanger.ac.uk/) contains 15 subfamilies within the β-trefoil superfamily; these include ricin B lectin, agglutinin, Kunitz lequme, fibroblast growth factor (FGF), interleukin-1, fascin, and *Clostridium botulinum* hemagglutinin.[36] While CGL is structurally similar to hisactophilin (PDB: 1HCD), a member of the fascin family[37] (Z-score = 10.1; Figure 2), CGL shows little sequence identity with hisactophilin (9.3%). Functionally, CGL is a lectin but hisactophilin is an actin-binding protein. Meanwhile, many β-trefoil proteins indeed exhibit carbohydrate-binding ability. For example, Ct1,3Gal43A (PDB: 3VSF) is an exo-β-1,3-galactosidase that consists of a GH43 domain, a CBM13 domain, and a dockerin domain, and possesses the unusual ability to hydrolyze β-1,3-galactan.[19] Furthermore, the carbohydrate-binding module-13 (CBM13) domain utilizes an aromatic residue (Tyr431), charged residues (Asp419, Glu420, and Asp416), and a polar residue (Asn438) to bind galactose. Finally, while CBM13 is structurally similar to CGL, it utilizes amino acids to bind galactose (Figure 2). These natural variants of β-trefoil proteins exemplify the functional diversity of proteins with the same fold. ThreeFoil is an engineered β-trefoil protein that displays multivalent carbohydrate-binding capacity—its substrates include galactose—with remarkable thermal stability and solubility.[11] It is an example of a de novo designed protein that contains repeating modules for better folding stabilities and functionalities, as has been exploited in numerous natural occurring β-trefoil proteins.

In contrast to previously identified lectins, whose structure–function relationships have been systematically compared,[38] CGL is void of the commonly used carbohydrate-binding motif—G-X-X-X-Q-X-(W/Y)—and is not classified as a member of the ricin B-like family.[13,39] Canonical β-trefoil fold proteins, e.g., ricin B, may result from the triplication of an ancestral gene encoding a polypeptide chain of about 40 amino acid residues.[40] Likewise, three carbohydrate-binding sites of CGL may be copies of an ancestral gene. Unlike previously reported β-trefoil lectins, however, histidine residues play a central role in the carbohydrate binding of CGL. The presence of an aromatic residue facing the nonpolar face of galactose is a common feature of galactose-binding lectins.[41] In contrast, CGL binds to galactose very differently: the carbohydrate-

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787–4795

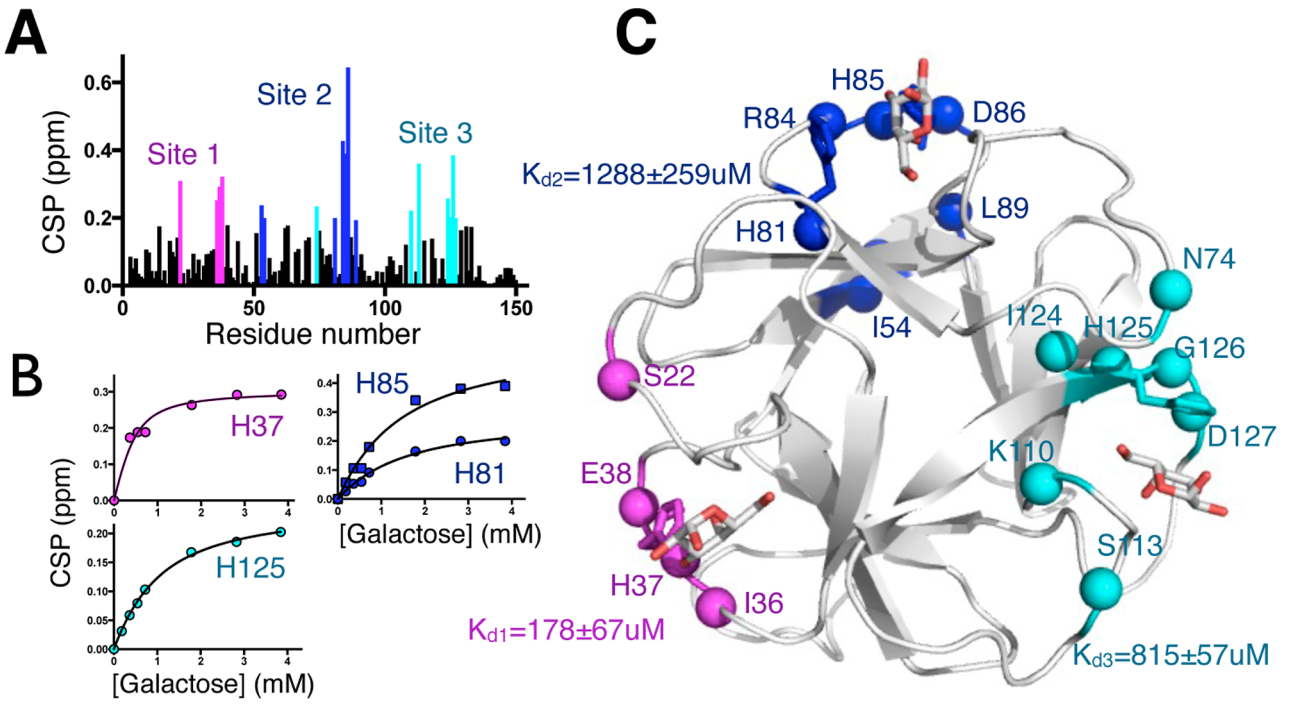


Figure 5. NMR titration analysis of galactose binding to CGL. (A) Weighted chemical shift perturbations (CSPs) between apo- and galactose-bound CGL are plotted as a function of residue number. The residues that shift more than one standard deviation are highlighted by magenta (Site 1), blue (Site 2), and cyan (Site 3). (B) CSPs of H37, H85/H81, and H125 as a function of galactose concentration are used to extract the apparent K_d values corresponding to three ligand-binding sites. (C) Residues that shift more than one standard deviation were mapped onto the crystal structure.

binding sites of CGL are solvent-exposed with a shallow cleft that can accommodate both the α and β anomers of galactose. Such a novel binding motif may be a unique feature that only exists within the *Mytilus* genus.[17,23,38]

An important finding from our structural investigation of CGL is that the dimeric quaternary structure and the triplication of the ligand-binding site within each protomer can give rise to a total of six ligand-binding sites within a CGL dimer. This could confer multivalency that is critical for several biological functions such as hemagglutination.[9,12,42] Our NMR titration analysis reveals that subtle amino acid composition variations within the microenvironments of the ligand-binding sites as the ones observed in the crystal structure (Figure 3) can significantly affect the binding affinity of each binding site (Figures 5 and S14). This also raises the possibility that the three different binding sites may in fact have different substrate specificities, which remains to be established. The previously reported hemagglutinating activity,[8,9,12] is directly related to the potential multivalency of CGL. A recent mutagenesis study based on homology modeling indeed demonstrated that CGL possess three carbohydrate-binding sites with differential mucin-binding activities with Site 3 being the least critical one.[17] Many of the carbohydrate interacting residues proposed by the homology modeling in this study can be now confirmed by our high-resolution crystal structures (Figure 3).

Here we demonstrate a novel function of CGL to reduce cell viability of a breast cancer cell line that may be associated with Gb3 binding, MCF7. Gb3 contains a terminal galactose but other oligomers such as the TF molecule (α-Gal-(1,3)-β-GalNAc) contains a terminal galactose as well. MCF7 expresses TF molecules may also be recognized by CGL. Several lectins have been shown to bind Gb3 resulting changes in cancer cells.[7,43−46] *Silurus asotus* lectin binds to Gb3 on Raji cells that results in the disappearance of the membrane-bound form of HSP70.[43] *Silurus asotus* egg lectin (SAL) down-regulates the membrane transporter protein MRP1 (multidrug resistance associate protein-1) in Burkitt's lymphoma cells through Gb3 (Galα1−4Galβ1−4Glc)-glycosphingolipid. Here, we show that CGL can also bind Gb3 and mediate the cell death of MCF7 cells (Figure 6). A previous report reveals that Gb3 activates c-Src and β-catenin signaling and up-regulates the expression of FGF-2, CD44, and Oct-4, enriching tumorigenesis in breast cancer stem cells, while silencing glucosylceramide synthase expression and disrupting Gb3 synthesis would lead to the death of breast cancer stem cells.[31]

Although the cellular mechanism of how CGL leads to the death of MCF7 remains unknown, our in vitro binding assay, structural analysis, and fluorescence imaging results suggest that its anti-cancer may be associated with Gb3 binding. Indeed, cell-surface-displayed Gb3 has been found in cell lines for breast cancer,[30,31] lung cancer,[47] lymphoma,[48] and colorectal cancer.[49] Lectins have also been utilized in applications beyond their biomedical relevance, such as lectin blotting, lectin histochemistry, and cytochemistry.[50] A very recent report has demonstrated that Mytilec interacts with Gb3 on Burkitt's lymphoma cells thereby triggering apoptosis through multiple pathways.[23]

The carbohydrate−lectin interactions have been exploited in biological detection systems as label-free carbohydrate sensors[51] for glyco-biomarkers.[52] Glyco-biomarkers, particularly the disease-specific glycans, are emerging indicators of specific diseases such as cancer.[52,53] The lectin microarray is one such diagnostic tool developed to recognize glyco-markers.[54] In addition, elevated levels of Gb3 have also been seen in drug-resistant cancers and cell lines.[47,55] In this regard, CGL could be developed into a lectin-based biosensor for cancer diagnostics.

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787−4795

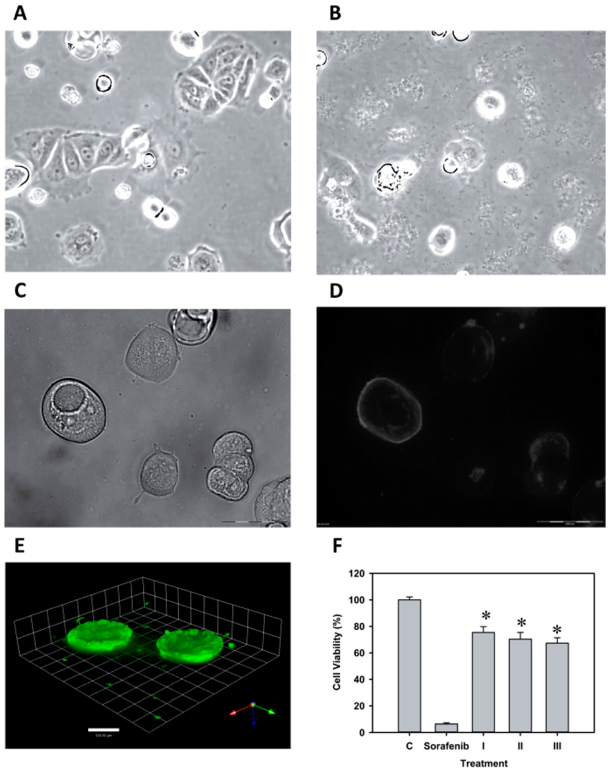


Figure 6. Cell viability assays of MCF7 cell line after addition of CGL. (A) MCF7 cells adhere on the plate without treatment of CGL. (B) MCF7 cells become detached from the plate after treatment of CGL. (C) Light microscope image of MCF7 cells after treatment of FITC-CGL. (D) The image of fluorescence microscopy of previous mentioned MCF7 cells indicates FITC-CGL mainly bound on the membrane of MCF7 cells. (E) The 3D images of fluorescence microscopy showed that MCF7 cells were surrounded by FITC fluorescence. (F) Cell viability assays of MCF7 evaluated using an MTT assay. The CGL protein was added to each well at various concentrations (I, 50; II, 100; and III, 200 μg/mL), and then the cells were incubated for 24 h before MTT assay. Sorafenib (2 μg/mL) was used as positive control. C (control): MCF7 without CGL treatment. *$P < 0.05$ for comparison with control.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b00111.

Experimental details (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*sthsu@gate.sinica.edu.tw

*shwu@gate.sinica.edu.tw

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the use of the instruments in the Biophysics Core Facility, Scientific Instrument Center at Academia Sinica. The NMR spectra were obtained at the Core Facility for Protein Structural Analysis supported by the National Core Facility Program for Biotechnology. We are grateful to the staff at the beamlines in the National Synchrotron Radiation Research Center (NSRRC), Taiwan, and SPring-8, Japan, for assistance with synchrotron data collection. This research was supported by the National Science Council and Ministry of Science and Technology (NSC 100-2113-M-001-022-MY3; NSC 102-2811-M-001-132, MOST 103-2113-M-001-029-MY3, MOST 103-2811-M-001-126, NSC 103-2923-M-001-006-MY3, NSC 100-2113-M-001-031-MY2, NSC 102-2113-M-001-017-MY2) and Academia Sinica, Taipei, Taiwan. S.-T.D.H. is a recipient of the Career Development Award (CDA-00025/2010-C) from the International Human Frontier Science Program.

■ REFERENCES

(1) Varki, A.; Etzler, M. E.; Cummings, R. D.; Esko, J. D. Discovery and Classification of Glycan-Binding Proteins. *Essentials of Glycobiology*, 2nd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009; Part IV, Chapt. 26.

(2) Yau, T.; Dan, X.; Ng, C. C.; Ng, T. B. *Molecules* **2015**, *20*, 3791.

(3) Goldstein, I. J.; Reichert, C. M.; Misaki, A. *Ann. N. Y. Acad. Sci.* **1974**, *234*, 283.

(4) Sharon, N. *J. Biol. Chem.* **2007**, *282*, 2753.

(5) Kilpatrick, D. C. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, *1572*, 401.

(6) Chernikov, O. V.; Molchanova, V. I.; Chikalovets, I. V.; Kondrashina, A. S.; Li, W.; Lukyanov, P. A. *Biochemistry (Moscow)* **2013**, *78*, 760.

(7) Fujii, Y.; Dohmae, N.; Takio, K.; Kawsar, S. M.; Matsumoto, R.; Hasan, I.; Koide, Y.; Kanaly, R. A.; Yasumitsu, H.; Ogawa, Y.; Sugawara, S.; Hosono, M.; Nitta, K.; Hamako, J.; Matsui, T.; Ozeki, Y. *J. Biol. Chem.* **2012**, *287*, 44772.

(8) Belogortseva, N. I.; Molchanova, V. I.; Kurika, A. V.; Skobun, A. S.; Glazkova, V. E. *Comp. Biochem. Physiol., Part C: Pharmacol., Toxicol. Endocrinol.* **1998**, *119*, 45.

(9) Kovalchuk, S. N.; Chikalovets, I. V.; Chernikov, O. V.; Molchanova, V. I.; Li, W.; Rasskazov, V. A.; Lukyanov, P. A. *Fish Shellfish Immunol.* **2013**, *35*, 1320.

(10) Rutenber, E.; Ready, M.; Robertus, J. D. *Nature* **1987**, *326*, 624.

(11) Broom, A.; Doxey, A. C.; Lobsanov, Y. D.; Berthin, L. G.; Rose, D. R.; Howell, P. L.; McConkey, B. J.; Meiering, E. M. *Structure* **2012**, *20*, 161.

(12) Chikalovets, I. V.; Chernikov, O. V.; Pivkin, M. V.; Molchanova, V. I.; Litovchenko, A. P.; Li, W.; Lukyanov, P. A. *Fish Shellfish Immunol.* **2015**, *42*, 503.

(13) Hashimoto, H. *Cell. Mol. Life Sci.* **2006**, *63*, 2954.

(14) Huang, K. F.; Liaw, S. S.; Huang, W. L.; Chia, C. Y.; Lo, Y. C.; Chen, Y. L.; Wang, A. H. *J. Biol. Chem.* **2011**, *286*, 12439.

(15) Jakob, M.; Lubkowski, J.; O'Keefe, B. R.; Wlodawer, A. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2015**, *71*, 1429.

(16) Krissinel, E.; Henrick, K. *J. Mol. Biol.* **2007**, *372*, 774.

(17) Kovalchuk, S. N.; Golotin, V. A.; Balabanova, L. A.; Buinovskaya, N. S.; Likhatskaya, G. N.; Rasskazov, V. A. *Fish Shellfish Immunol.* **2015**, *47*, 565.

(18) Holm, L.; Rosenstrom, P. *Nucleic Acids Res.* **2010**, *38*, W545.

(19) Jiang, D.; Fan, J.; Wang, X.; Zhao, Y.; Huang, B.; Liu, J.; Zhang, X. C. *J. Struct. Biol.* **2012**, *180*, 447.

(20) Sulzenbacher, G.; Roig-Zamboni, V.; Peumans, W. J.; Rouge, P.; Van Damme, E. J.; Bourne, Y. *J. Mol. Biol.* **2010**, *400*, 715.

(21) Suzuki, R.; Fujimoto, Z.; Kuno, A.; Hirabayashi, J.; Kasai, K.; Hasegawa, T. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1895.

(22) Ling, H.; Boodhoo, A.; Hazes, B.; Cummings, M. D.; Armstrong, G. D.; Brunton, J. L.; Read, R. J. *Biochemistry* **1998**, *37*, 1777.

(23) Hasan, I.; Sugawara, S.; Fujii, Y.; Koide, Y.; Terada, D.; Iimura, N.; Fujiwara, T.; Takahashi, K. G.; Kojima, N.; Rajia, S.; Kawsar, S. M.; Kanaly, R. A.; Uchiyama, H.; Hosono, M.; Ogawa, Y.; Fujita, H.; Hamako, J.; Matsui, T.; Ozeki, Y. *Mar. Drugs* **2015**, *13*, 7377.

(24) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213.

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787−4795

(25) De Simone, A.; Cavalli, A.; Hsu, S. T.; Vranken, W.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 16332.
(26) http://www.nmr2.buffalo.edu/nesg.wiki/NMR_determined_Rotational_correlation_time
(27) Tsan, P.; Hus, J. C.; Caffrey, M.; Marion, D.; Blackledge, M. *J. Am. Chem. Soc.* **2000**, *122*, 5603.
(28) Hsu, S. T. D.; Cabrita, L. D.; Fucini, P.; Dobson, C. M.; Christodoulou, J. *J. Mol. Biol.* **2009**, *388*, 865.
(29) Yau, T.; Dan, X.; Ng, C. C.; Ng, T. B. *Molecules* **2015**, *20*, 3791.
(30) Bremer, E. G.; Levery, S. B.; Sonnino, S.; Ghidoni, R.; Canevari, S.; Kannagi, R.; Hakomori, S. *J. Biol. Chem.* **1984**, *259*, 14773.
(31) Gupta, V.; Bhinge, K. N.; Hosain, S. B.; Xiong, K.; Gu, X.; Shi, R.; Ho, M. Y.; Khoo, K. H.; Li, S. C.; Li, Y. T.; Ambudkar, S. V.; Jazwinski, S. M.; Liu, Y. Y. *J. Biol. Chem.* **2012**, *287*, 37195.
(32) Collart-Dutilleul, P. Y.; Panayotov, I.; Secret, E.; Cunin, F.; Gergely, C.; Cuisinier, F.; Martin, M. *Nanoscale Res. Lett.* **2014**, *9*, 564.
(33) Silva, M. C.; de Paula, C. A.; Ferreira, J. G.; Paredes-Gamero, E. J.; Vaz, A. M.; Sampaio, M. U.; Correia, M. T.; Oliva, M. L. *Biochim. Biophys. Acta, Gen. Subj.* **2014**, *1840*, 2262.
(34) Savanur, M. A.; Eligar, S. M.; Pujari, R.; Chen, C.; Mahajan, P.; Borges, A.; Shastry, P.; Ingle, A.; Kalraiya, R. D.; Swamy, B. M.; Rhodes, J. M.; Yu, L. G.; Inamdar, S. R. *PLoS One* **2014**, *9*, e110107.
(35) Nakahara, S.; Oka, N.; Raz, A. *Apoptosis* **2005**, *10*, 267.
(36) Fujimoto, Z. *Biosci., Biotechnol., Biochem.* **2013**, *77*, 1363.
(37) Habazettl, J.; Gondol, D.; Wiltscheck, R.; Otlewski, J.; Schleicher, M.; Holak, T. A. *Nature* **1992**, *359*, 855.
(38) Zurga, S.; Pohleven, J.; Renko, M.; Bleuler-Martinez, S.; Sosnowski, P.; Turk, D.; Kunzler, M.; Kos, J.; Sabotic, J. *FEBS J.* **2014**, *281*, 3489.
(39) Hirabayashi, J.; Dutta, S. K.; Kasai, K. *J. Biol. Chem.* **1998**, *273*, 14450.
(40) Robertus, J. D.; Ready, M. P. *J. Biol. Chem.* **1984**, *259*, 13953.
(41) Sujatha, M. S.; Sasidhar, Y. U.; Balaji, P. V. *Biochemistry* **2005**, *44*, 8554.
(42) Tong, C. Q.; Li, W.; Jin, Q.; Tan, C. Y.; Qu, M.; Chen, W.; Molchanova, V.; Chikalovets, I.; Chernikov, O.; Lukyanov, P. *African J. Biotechnol.* **2011**, *10*, 18532.
(43) Sugawara, S.; Kawano, T.; Omoto, T.; Hosono, M.; Tatsuta, T.; Nitta, K. *Biochim. Biophys. Acta, Gen. Subj.* **2009**, *1790*, 101.
(44) Fujii, Y.; Sugawara, S.; Araki, D.; Kawano, T.; Tatsuta, T.; Takahashi, K.; Kawsar, S. M.; Matsumoto, R.; Kanaly, R. A.; Yasumitsu, H.; Ozeki, Y.; Hosono, M.; Miyagi, T.; Hakomori, S. I.; Takayanagi, M.; Nitta, K. *Protein J.* **2012**, *31*, 15.
(45) Fujii, Y.; Kawsar, S. M.; Matsumoto, R.; Yasumitsu, H.; Ishizaki, N.; Dogasaki, C.; Hosono, M.; Nitta, K.; Hamako, J.; Taei, M.; Ozeki, Y. *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.* **2011**, *158*, 30.
(46) Watanabe, Y.; Tateno, H.; Nakamura-Tsuruta, S.; Kominami, J.; Hirabayashi, J.; Nakamura, O.; Watanabe, T.; Kamiya, H.; Naganuma, T.; Ogawa, T.; Naude, R. J.; Muramoto, K. *Dev. Comp. Immunol.* **2009**, *33*, 187.
(47) Tyler, A.; Johansson, A.; Karlsson, T.; Gudey, S. K.; Brannstrom, T.; Grankvist, K.; Behnam-Motlagh, P. *Exp. Cell Res.* **2015**, *336*, 23.
(48) LaCasse, E. C.; Bray, M. R.; Patterson, B.; Lim, W. M.; Perampalam, S.; Radvanyi, L. G.; Keating, A.; Stewart, A. K.; Buckstein, R.; Sandhu, J. S.; Miller, N.; Banerjee, D.; Singh, D.; Belch, A. R.; Pilarski, L. M.; Gariepy, J. *Blood* **1999**, *94*, 2901.
(49) Holst, S.; Wuhrer, M.; Rombouts, Y. *Adv. Cancer Res.* **2015**, *126*, 203.
(50) Dan, X.; Liu, W.; Ng, T. B. *Med. Res. Rev.* **2016**, *36*, 221.
(51) Zeng, X.; Andrade, C. A.; Oliveira, M. D.; Sun, X. L. *Anal. Bioanal. Chem.* **2012**, *402*, 3161.
(52) Lee, J. H.; Cho, C. H.; Kim, S. H.; Kang, J. G.; Yoo, J. S.; Chang, C. L.; Ko, J. H.; Kim, Y. S. *Mol. Cell. Proteomics* **2015**, *14*, 782.
(53) Tan, B.; Matsuda, A.; Zhang, Y.; Kuno, A.; Narimatsu, H. *Mol. BioSyst.* **2014**, *10*, 201.
(54) Hirabayashi, J.; Yamada, M.; Kuno, A.; Tateno, H. *Chem. Soc. Rev.* **2013**, *42*, 4443.
(55) De Rosa, M. F.; Sillence, D.; Ackerley, C.; Lingwood, C. *J. Biol. Chem.* **2004**, *279*, 7867.

DOI: 10.1021/jacs.6b00111
J. Am. Chem. Soc. 2016, 138, 4787−4795