Supporting Information:

Cytotoxic protein from the mushroom *Coprinus comatus* possesses a unique mode for glycan binding and specificity

Peilan Zhang[†,‡], Kunhua Li[‖,‡], Guang Yang[†], Changqing Xia[⊥], Jane E. Polston[#], Gengnan Li[‖], Shiwu Li[⊥], Zhao Lin[§], Li-jun Yang[⊥], Steven D. Bruner[‖,*], and Yousong Ding[†,*]

[†]Department of Medicinal Chemistry, Center for Natural Products, Drug Discovery and Development, [‖]Department of Chemistry, [⊥]Department of Pathology, Immunology and Laboratory Medicine, and [#]Department of Plant Pathology, University of Florida, Gainesville, FL, USA

[§]Department of Periodontics, Virginia Commonwealth University, Richmond, VA, USA

Experimental procedures

Molecular cloning, protein expression and purification

A codon-optimized gene encoding the full length Y3 was synthesized (Eurofins Genomics, USA). A construct with the signal peptide removed (SP-free) was amplified via PCR (primers illustrated in **Table S4**). Both genes were cloned into the pPICZαA vector (Thermo Fisher Scientific) (1). The pPICZαA-Y3 construct was used as the template to create Y3 mutants, by site-directed mutagenesis PCR with primers shown in **Table S4**. Plasmids were transformed into *Pichia pastoris* X-33 following manufacturer's instructions (Thermo Fisher Scientific). After selection, multiple single colonies were randomly picked to assess the Y3 production in 10 mL culture. The culture with the highest yield (10 mL) was inoculated into 1 L of YPD (1% yeast extract, 2% peptone, and 2% glucose) to produce Y3 following our previous protocol (2). After three days, the culture supernatant was harvested by centrifugation, filtered and dialyzed twice against 1 L of PBS buffer. The supernatant was further concentrated, and purified by gel filtration chromatography (HiLoad 16/60 SuperDex-75 column, AKTA FPLC System, GE Healthcare) with 150 mM NaCl, and 20 mM Tris-HCl, pH 7.5. Purified Y3 was concentrated to 10 mg·mL^{-1} as determined by Bradford assay (BSA as standard) and the purity of Y3 was determined by SDS-PAGE.

Ellman's test for free thiol determination

Standard Ellman's test was used to detect the free thiol groups in purified Y3 protein. Ten microliter Y3 (100 μM) or L-cysteine (100 μM) was mixed with 990 μL DTNB reagent (100 μM, Thermo Fisher Scientific) and incubated at room temperature for 5 min. The optical absorbance at 412 nm was measured and recorded.

Quantitation of carbohydrate content in Y3 and mutants

Recombinant Y3 and mutants were processed with the phenol-sulfuric acid method for total carbohydrate determination (3, 4). A linear, standard curve was established with D-Glu (0 to 1.6 mg/mL). Protein solutions (50 μL) were added to flat-bottom microplate followed by the rapid mixing with 150 μL of concentrated sulfuric acid. After gentle shaking to achieve homogenized solution, 30 μL of 5% phenol was added and the plate was kept at room temperature for 10 min. Protein free buffer and D-Glu were used as controls. Absorbance was recorded at 490 nm by microplate reader. Carbohydrate content (%) was then calculated using the equation of raw carbohydrate content (mg/mL) / protein concentration (mg/mL) determined by Bradford assay x 100. For the dialysis studies, Y3 (2 mL, 10 mg·mL^{-1}) was dialyzed against 1 L of 0.01 M PBS (pH 7.4) at 4 ºC, which was replaced with fresh buffer every day. The samples were collected daily to

determine the protein concentrations and the carbohydrate contents. All experiments were independently repeated at least three times.

Anti-tobamovirus assay

Purified *Tobacco mild green mosaic virus* (TMGMV) (*Virgaviridae, Tobamovirus*, formerly known as Tobacco mosaic virus strain U2) (GenBank Acc. No. EF469769) (26.5 mg/ml) was obtained from R. Charudattan (BioProdex, Inc.) (5). The ability of recombinant Y3 to inhibit TMGMV infection was evaluated using a range of protein concentrations (0.078, 0.15, 0.31, 0.625, 1.25, 2.5 μM) mixed with a constant concentration of TMGMV (26.5 μg/ml). Leaves (3^{rd}, 4^{th} and 5^{th} leaves from the apex) were detached from approximately 60 8-week old plants of *Nicotiana glutinosa*. Leaves were combined in a bag and then randomly selected for each inoculation. Recombinant Y3 and purified TMGMV were defrosted on ice, diluted with water, mixed to form the final concentrations, and kept on ice until inoculation. Treatments were inoculated in order of Y3 dilution from the most concentrated to the least. The left half of each leaf was inoculated with 50 μl of recombinant Y3 plus TMGMV, and the right half with an equal concentration of TMGMV only. Inoculated leaves were placed in a moist chamber in the dark for 48 h after which time the number of local lesions per each leaf half were recorded. All experiments were repeated 6 times. The percent inhibition was calculated as followed: 1- (No. of local lesions Y3+TMGMV/No. of local lesions from TMGMV alone) x 100%. To examine the effects of time and temperature on the ability of recombinant Y3 to interfere with TMGMV infectivity, a mixture containing recombinant Y3 and purified TMGMV (0.078 μM and 26.5 μg/ml, respectively) were maintained either on ice or at room temperature for 0 min and 20 min before inoculation to the left half of leaves. Purified TMGMV (50 μl of 26.5 μg/ml solution) was inoculated to the right half of leaves 0 and 20 min after incubation. Six leaves were tested per treatment. The percent inhibition was calculated as followed: 1- (No. of local lesions Y3+TMGMV/No. of local lesions from TMGMV alone) x 100%.

Annexin V/7-AAD staining to quantitate cells in early and late apoptosis stages

Apoptosis in Jurkat cells was induced with WT Y3 (0, 0.025, 0.1, 0.25, 1 μM) along with dialyzed Y3 (1 μM) and two Y3 mutants Y3D26A (1 μM) and Y3N122A (1 μM) for 4h or 20h. Cells were then stained with Annexin V-APC and 7-aminoactinomycin D (7-AAD) to evaluate early and late apoptosis by flow cytometry according to manufacturer's protocol (Biolegend). Briefly, 10^6 cells were washed twice with PBS and stained with 5 μL of Annexin-APC (5 $\mu g \cdot ml^{-1}$) and 10 μL of 7-AAD (2.5 $\mu g \cdot mL^{-1}$). After incubation at room temperature for 15 min in the dark, 400 μL of binding buffer was added to each sample. Flow cytometric analysis was carried out using a FACS Fortessa (BD Bioscience) and cells at early and late apoptosis stages were identified using FlowJo program. The total apoptotic cells referred to the combination of early and late ones. Three independent experiments were performed for each set.

Cell imaging with FITC-Y3

Y3 was labeled with fluorescein by reacting the purified protein with NHS-fluorescein (Fisher Scientific). Jurkat and HEK293 (control) cells were incubated with Y3-FITC (1 μM) for 1 h, then washed with PBS. Y3-FITC has an absorption maximum at 495 nm and emission maximum at 525 nm. A Zeiss Fluorescent microscope AXIO, AXioCam MRm camera, and AxioVision Rel 4.8.1 software was used to capture imaging pictures.

Western blot analysis of caspase-dependent apoptosis

Jurkat cells were treated with Y3 (0.01 μM, 0.1 μM, 1 μM) or Y3 (1 μM) plus the pan-caspase inhibitor Z-VAD-FMK (20 μM) for 20h. Cells were then collected and washed twice with PBS. An equal amount of protein from the cell lysate of each sample was analyzed by 12% SDS-PAGE. After electrophoresis, proteins were transferred to PVDF membranes, blocked for 1 h with 5% fat-free milk at room temperature, and blotted with the indicated primary antibodies (anti-cleaved

caspase-3, 1:1000; anti-caspase-8, 1:1000; anti-cleaved caspase-9, 1:1000, anti-β-actin, 1:1000, Cell Signaling Technology) with gentle agitation at 4°C overnight. Following a wash step with Tris-buffered saline/Tween-20, the membranes were incubated with HRP-conjugated secondary antibodies at room temperature for 1h and the immune complexes were detected using the standard electrochemiluminescence method.

Hemagglutination assays

Agglutinating activity of Y3 was examined by using human red blood cells, type O^+ and rabbit red blood cells (Innovative research, USA). A serial twofold dilution of the Y3 solution (2 mM starting concentration) in microtiter U-plates (50 μL) was mixed with 50 μL of 2% suspension of red blood cells in 0.01 M PBS (pH 7.4) at room temperature. The results were recorded when the blank was fully sedimented. Experiments were repeated multiple times.

Determination of the carbohydrate binding to Y3

D-Glu, *α*-D-mannose, *β*-D-galactose and lactose were tested with isothermal titration calorimetry (ITC). Purified Y3 was diluted to 100 μM and dialysed against 150 mM NaCl and 20 mM Tris-HCl, pH 7.5 for 16 h. Protein (200 μL) was then placed into a MicroCal iTC200 (Malvern) reaction cell, while different carbohydrates in the same buffer at a concentration of 1-5 mM were injected into the reaction cell over time at 25 °C. Raw data were integrated, normalized and evaluated using ORIGIN 7.0 according to the one-site binding model. Octet RED384 (ForteBio) was used to determine the binding of D-galactose 6-sulfate, *N*-acetyl-D-galactosamine and D-galactosamine to biotinylated Y3 in black 96 wells plates (Nunc™ F96 MicroWell™ plate, Thermo Scientific) following a previous protocol (6). All experiments were performed in triplicate. Data were processed to calculate kinetic and affinity parameters using the ForteBio software.

Crystallization of Y3

Initial Y3 crystallization was carried in a vapor diffusion sitting drop format with a homemade matrix screen designed for glycoproteins (CrystalQuick 96 well, Hampton Research). Small plate clusters of Y3 crystals were identified in a condition containing 20% PEG-4000 and 0.1 M CHES, pH 10.5 in 15 min at 25 °C. Optimization of precipitant and pH, along with microseeding were performed in a hanging drop format (24 well VDX crystallization plate, Hampton Research). Protein (1.8 μL, 10 $mg \cdot mL^{-1}$) plus 2.0 μL of precipitant were balanced against 1 mL of reservoir solution. The resultant plate-shaped single crystals with a size of ~20×20×100 μm were obtained in a final condition that contained 16% PEG-4000, 10% v/v glycerol and 0.1 M CHES, pH 9.5 in 48 h. Crystals of suitable size were harvested and flash frozen in liquid nitrogen with an additional 10% glycerol as cryoprotectant. Heavy atoms derivative crystals for single wavelength derivative (SAD) were obtained through co-crystallization or soaking. Compounds (Heavy Atom Screen Kits, Hampton Research) containing derivative elements (Pt, Au, Hg, I) were selected according to their solubility (pH 9.5~10.5). Y3 native and derivative crystal X-ray diffraction data sets were collected on beamlines 21-ID-G and 21-ID-F of the Life Sciences Collaborative Access Team (LS-CAT) facility at the Advanced Photon Source (APS), Argonne National Laboratory (ANL) with a wavelength of 0.9786 Å at 100 K.

Diffraction data collection and processing

Y3 native and derivative crystal X-ray diffraction data sets were collected on beamlines 21-ID-G and 21-ID-F of the Life Sciences Collaborative Access Team (LS-CAT) facility at the Advanced Photon Source (APS), Argonne National Laboratory (ANL) with a wavelength of 0.9786 Å at 100 K. Data sets were merged using XDS program package (7) and then scaled with AIMLESS from the CCP4 suite (8) to space group $P2_1$, in two diverse forms with different unit cell parameters. Y3 native crystals diffract X-ray to 1.18 Å, while the Pt-derivative data set was truncated at 1.70 Å resolution. Native data sets had large non-origin Patterson peaks as indicated by

PHENIX.XTRIAGE from PHENIX suite, suggesting the presence of translational noncrystallographic symmetry (tNCS).

Atomic structure determination and refinement

The structure of Y3 was determined with Pt-SAD. The number and location of heavy atoms were identified with SHELXD. Initial phase calculation was carried using SHELXE (9) and further refined with MLPHARE, PIRATE (10) and PHASER (11). An interpretable electron-density map was generated with SOLOMON with 20% solvent flattering (12). The model building was initiated with SHELXE, and completed using BUCCANEER (13), ARP/wARP (14) and COOT (15). Final coordinates were refined with REFMAC5 (16) and PHENIX.REFINE (17). Carbohydrates and CHES from the crystallization conditions were added manually with COOT upon careful inspection of the electron-density maps. Derivative ions were added manually into the calculated anomalous difference map, with their occupancy and anisotropic B-factors refined sequentially. The quality of the models was evaluated using a sigma-weighted, simulated annealing composite omit map. Structural illustrations were prepared with PyMOL.

Docking of LDNF into Y3 binding pocket

Docking was conducted using AutoDock 4.2 (18). LDNF was constructed and energy minimized with Spartan '08 (19). Standard algorithms and docking procedures were used for a rigid protein and a flexible ligand in a grid covering the entire protein dimer. The hydrogen bonding/π–π interactions and corresponding estimated free energy of ligand binding (ΔG) were analysed with AutoDockTools 1.5.6. The best docking pose was selected and illustrated.

Mass spectrometry analysis of Y3

For ESI mass spectrometry analysis, protein crystals were picked, washed with the crystallization mother liquor, and dissolved in water. Liquid chromatography (LC) conditions were: 3.0 % - 60 % of 0.1 % TFA-acetonitrile in 0.1 % TFA-water over 40 min with a flow rate of 1.0 $mL \cdot min^{-1}$ (ZORBAX SB-C18, Agilent). For MALDI-TOF-MS, sinapic acid was used as matrix (Protea). To determine the potential bound carbohydrates by LC-MS (Agilent), protein samples were treated with peptide *N*-glycosidase F (PNGase F, New England Biolabs) following manufacturer's protocol.

Statistical analysis of experimental data

Results were presented as mean ± SE of at least three independent experiments. Statistical significance among multiple groups was analyzed by one-way ANOVA followed by Student's t-test for comparison of the results between two groups using Prism 5 (Graphpad Software, Inc). * and ** indicates $p<0.05$ and $p<0.01$ statistical differences compared to the control, respectively.

References

1. Sambrook J, Russell W (2006) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 3rd Ed.
2. Xie C*, et al.* (2015) Irisin controls growth, intracellular Ca^{2+} signals, and mitochondrial thermogenesis in cardiomyoblasts. *PLOS ONE* 10:e0136816.
3. Masuko T*, et al.* (2005) Carbohydrate analysis by a phenol-sulfuric acid method in microplate format. *Anal Biochem* 339:69-72.
4. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956) Colorimetric method for determination of sugars and related substances. *Anal Chem* 28:350-356.
5. Ferrell J, Charudattan R, Elliott M, Hiebert E (2008) Effects of selected herbicides on the efficacy of tobacco mild green mosaic virus to control tropical soda apple (*Solanum warm*). *Weed Sci* 56:128-132.
6. Yang L, Connaris H, Potter JA, Taylor GL (2015) Structural characterization of the carbohydrate-binding module of NanA sialidase, a pneumococcal virulence factor. *BMC Struct Biol* 15:15.
7. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D* 66:133-144.
8. Winn MD*, et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* 67:235-242.
9. Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66:479-485.
10. Cowtan K (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* 66:470-478.
11. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.
12. Abrahams J, PLeslie AGW (1996) Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallogr D* 52:30-42.
13. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D* 62:1002-1011.
14. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3:1171-1179.
15. Emsley P, Cowtan K (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126-2132.
16. Murshudov GN*, et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D* 67:355-367.
17. Afonine PV*, et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* 68:352-367.
18. Morris GM*, et al.* (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30:2785-2791.
19. Shao Y*, et al.* (2006) Advances in methods and algorithms in a modern quantum chemistry program package. *Phys Chem Chem Phys* 8:3172-3191.

Table S1. Selected cell lines in anticancer activity screening.

Cell line	Types
Jurkat	T cell leukemia
DAN-G	pancreas carcinoma
MIA-PaCa-2	pancreas carcinoma
DU145	prostatic cancer
Hela	cervical cancer
HepG2	liver carcinoma
UM-SCC-1	head and neck squamous carcinoma
HEK 293	non-malignant control

Table S2. Top 20 glycans from glycan microarray screening

Y3 (5 μg/mL)				Y3 (50 μg/mL)			
ID	**Glycan**	**RFU**	**SD**	**ID**	**Glycan**	**RFU**	**SD**
97	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0	16543	1302	97	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0	35336	5997
45	(6S)Galb1-4(6S)Glcb-Sp8	7075	845	22	6S(3S)Galb1-4(6S)GlcNAcb-Sp0	18424	2228
440	(6S)Galb1-3(6S)GlcNAc-Sp0	6718	1217	440	(6S)Galb1-3(6S)GlcNAc-Sp0	16181	864
22	6S(3S)Galb1-4(6S)GlcNAcb-Sp0	5634	774	45	(6S)Galb1-4(6S)Glcb-Sp8	13063	1262
496	Fuca1-2(6S)Galb1-3(6S)GlcNAcb-Sp0	4508	178	295	(6S)Galb1-4(6S)GlcNAcb-Sp0	12862	2289
222	Fuca1-2(6S)Galb1-4(6S)Glcb-Sp0	3871	136	496	Fuca1-2(6S)Galb1-3(6S)GlcNAcb-Sp0	10464	1175
295	(6S)Galb1-4(6S)GlcNAcb-Sp0	3820	628	222	Fuca1-2(6S)Galb1-4(6S)Glcb-Sp0	9781	1200
39	(6S)(4S)Galb1-4GlcNAcb-Sp0	2825	86	155	Galb1-4(6S)Glcb-Sp0	8857	579
237	Neu5Aca2-3Galb1-3(6S)GlcNAc-Sp8	2821	176	505	Galb1-3(6S)GlcNAcb-Sp8	8805	1208
34	(3S)Galb1-4(6S)GlcNAcb-Sp0	2801	205	495	Fuca1-2Galb1-3(6S)GlcNAcb-Sp0	6445	580
35	(3S)Galb1-4(6S)GlcNAcb-Sp8	2637	103	26	(3S)Galb1-4(6S)Glcb-Sp0	6377	283
505	Galb1-3(6S)GlcNAcb-Sp8	2471	613	237	Neu5Aca2-3Galb1-3(6S)GlcNAc-Sp8	5929	532
495	Fuca1-2Galb1-3(6S)GlcNAcb-Sp0	2450	579	39	(6S)(4S)Galb1-4GlcNAcb-Sp0	5616	909
23	6S(3S)Galb1-4GlcNAcb-Sp0	2125	132	34	(3S)Galb1-4(6S)GlcNAcb-Sp0	5538	877
498	GalNAcb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	2032	386	498	GalNAcb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	5351	940
27	(3S)Galb1-4(6S)Glcb-Sp8	1792	431	23	6S(3S)Galb1-4GlcNAcb-Sp0	5251	1201
506	(6S)(4S)GalNAcb1-4GlcNAc-Sp8	1679	95	506	(6S)(4S)GalNAcb1-4GlcNAc-Sp8	4972	293
288	Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp0	1343	143	27	(3S)Galb1-4(6S)Glcb-Sp8	4918	440
47	(6S)GlcNAcb-Sp8	1325	169	375	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	4180	285
42	(6S)Galb1-4Glcb-Sp0	1302	452	40	(4S)Galb1-4GlcNAcb-Sp8	3946	452

Table S3 X-ray data collection, processing and structure refinement

Data Collection	**Y3-Native**	**Y3-Derivative**
Resolution range (Å)	41.8 - 1.18 (1.22 - 1.18)*	41.3 - 1.70 (1.76 - 1.70)*
Space group	$P2_1$	$P2_1$
a b c (Å), β (°)	53.3 56.1 62.7, 92.7	41.1 55.2 41.0, 99.5
Total No. of reflections	567305 (38773)	148850 (11787)
Unique reflections	113409 (9627)	20217 (1790)
Multiplicity	5.0 (4.0)	7.4 (6.6)
Completeness	93% (80%)	99% (87%)
$\langle I/\sigma(I)\rangle$	22.98 (4.44)	14.57 (3.70)
Wilson B factor (Å^2)	8.1	14.0
R_{merge}†	0.041 (0.29)	0.093 (0.44)
R_{meas}†	0.046 (0.34)	0.099 (0.48)
$CC_{1/2}$†	1 (0.91)	1 (0.93)
Refinement	**Y3-Native**	**Y3-Derivative**
R_{work}†	0.172 (0.205)	0.159 (0.187)
R_{free}†	0.189 (0.240)	0.194 (0.248)
No. of non-H atoms	4189	1950
Protein	3420	1751
Ligand	52	34
Protein residues	448	224
RMS Bonds (Å)	0.006	0.019
RMS Angles (°)	1.05	1.54
Ramachandran		
Favored (%)	97	95
Outliers (%)	0	0
Clashscore	1.3	2.9
Average B factors (Å^2)	11.9	18.0
Protein	10.0	16.9
Ligand	14.9	30.5
Water	20.7	27.2

*Values for the outer shell are given in parentheses

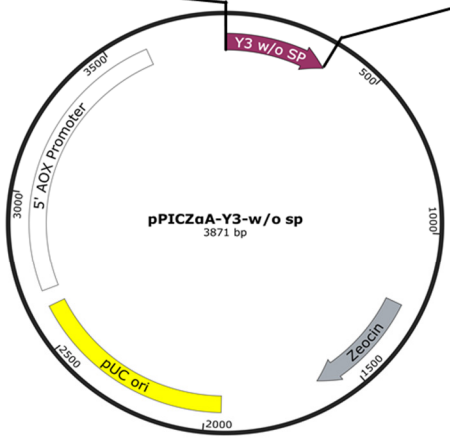
† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| \,/\, \sum_{hkl}\sum_i I_i(hkl)$; $R_{meas} = (\sum_{hkl}\sqrt{n/(n-1)}\sum_{i=1}^{n}|I_i(hkl) - \langle I(hkl)\rangle|) \,/\, \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith intensity measurement of a reflection, $\langle I(hkl)\rangle$ is the average intensity value of that reflection and the summation is over all measurements. $CC_{1/2} = \sum(x - \langle x\rangle)\,(y - \langle y\rangle) / \sqrt{\sum(x - \langle x\rangle)^2 \sum(y - \langle y\rangle)^2}$. *R-factor* $= \sum_{hkl}\big||F_{obs}| - |F_{calc}|\big| \,/\, \sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are measured and calculated structure factors, respectively. R_{free} calculated from 5% of the reflections that were selected randomly and omitted during refinement.

Table S4. Primers used in this work

Primer name	Sequence
Y3FXh	5’-GTATCTCTCGAGAAAAGACAAGATCCTTTG-3’
Y3RNt	5’-TTTTCCTTTTGCGGCCGCTTAAAAATCAGTGG-3’
D26AFw	5’-TGCTATGCCAACTTTGGGAATCGTGATGTTGCAGCA-3’
D26ARv	5’-TGCTGCAACATCACGATTCCCAAAGTTGGCATAGCA-3’
N122AFw	5’-GATCCGGCTGATGGAGACTGTTCCACTGATTTTTAA-3’
N122ARv	5’-TTAAAAATCAGTGGAACAGTCTCCATCAGCCGGATC-3’

A

CAA GAT CCT TTG TCT TGC TAT GAC AAC TTT GGG AAT CGT
GAT GTT GCA GCA TGT GCT AGA TTC ATT GAC GAC TTT TGC
GAT ACC TTG ACA CCA AAC ATT TAC CGA CCA AGA GAT AAC
GGA CAG AGA TGT TAC GTC GTC AAT GGC CAT AAA TGC GAC
TTT ACC GTG TTC AAC ACC AAC AAT GGT GGT TCT CCC ATA
AGA GCT TCA ACT CCT AAC TGT AAG ACT GTT CTT AGA GCT
GCA GCT AAT CGT TGT CCA ACA GGT GGA AGA GGC AAG ATC
AAT CCT AGT GCT CCA TTC CTG TTT GCC ATT GAT CCG AAT
GAT GGA GAC TGT TCC ACT GAT TTT TAA



B

QDPLSCYDNFGNRDVAACARFIDDFCDTLTPNIYRPRDNGQRCYV
VNGHKCDFTVFNTNNGGSPIRASTPNCKTVLRAAANRCPTGGRG
KINPSAPFLFAIDPNDGDCSTDF-

Figure S1. (A) Map of synthesized nucleotide sequence of SP-free Y3. (B) Protein sequence of mature Y3.

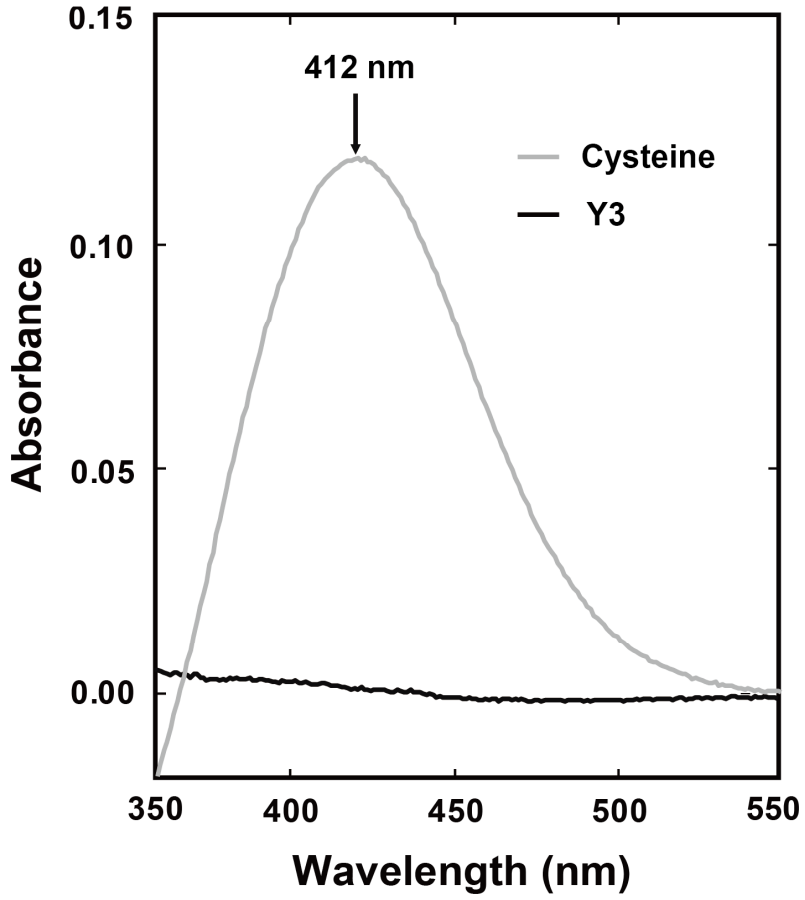


Figure S2. UV-vis spectral analysis of the Ellman's tests found no free thiol group in Y3 protein (black, 100 μM). Cysteine (grey, 100 μM) was used as a positive control.

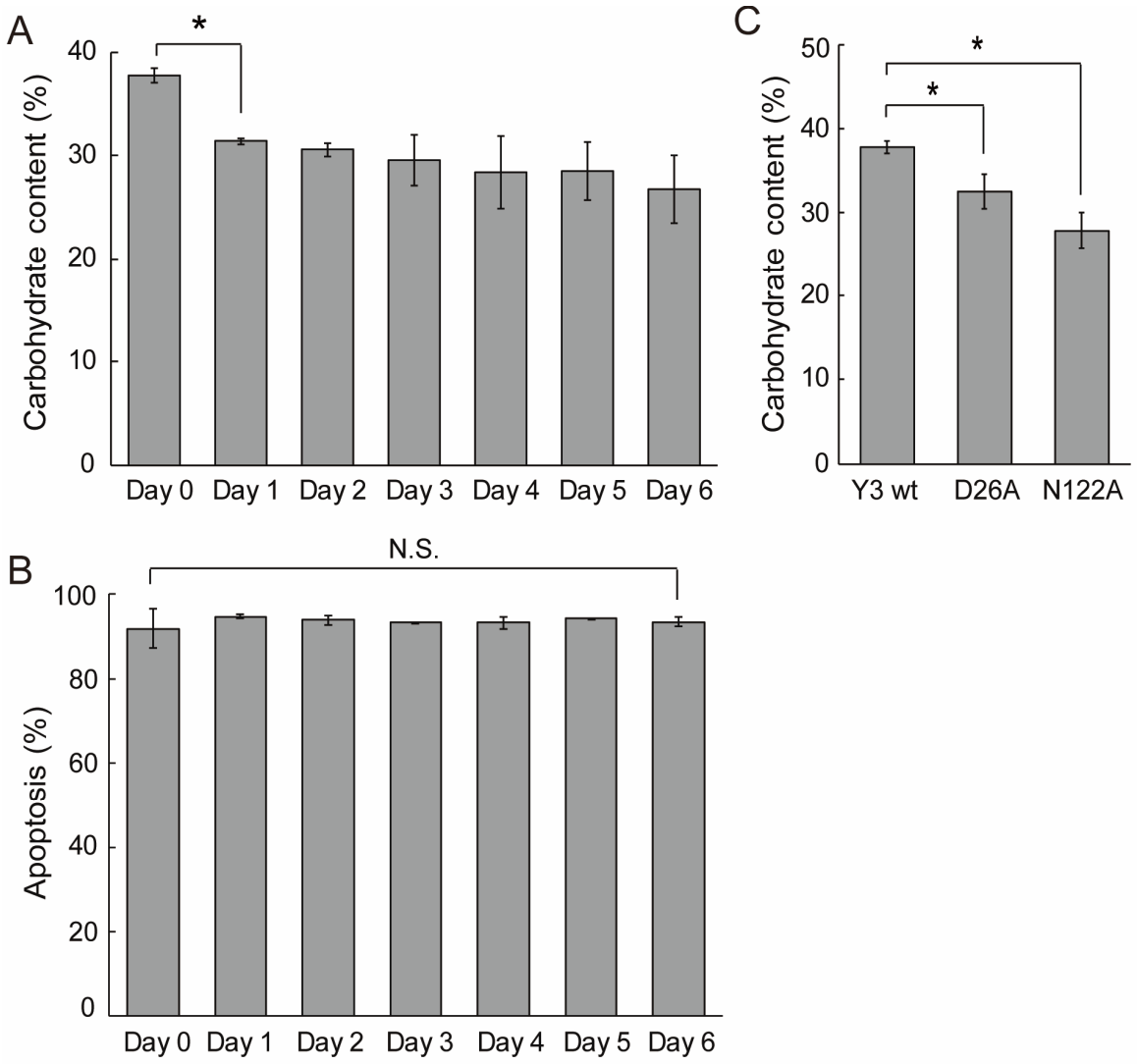


Figure S3. (A) Carbohydrate contents of Y3 (day 0) and dialyzed Y3 samples (day 1 to 6) were determined by phenol-sulfuric acid analysis. D-Glucose was used to establish a linear standard curve. Dialysis led to a statistically significant decrease of carbohydrate content in Y3. (B) Y3 and dialyzed Y3 samples (1 μM) showed no significant difference in inducing the total apoptotic Jurkat cells after 20 h. The apoptotic cells were quantitated by annexin V/7-AAD staining analysis. (C) D26A and N122A mutants (10 $mg \cdot mL^{-1}$) contained significantly less amount of carbohydrate contents in comparison with of Y3 (10 $mg \cdot mL^{-1}$), as determined by phenol-sulfuric acid analysis. Data are shown as the means ± SD (n = 3). Significant differences between the untreated Y3 and dialyzed Y3 or mutants are shown (*: $p<0.05$).

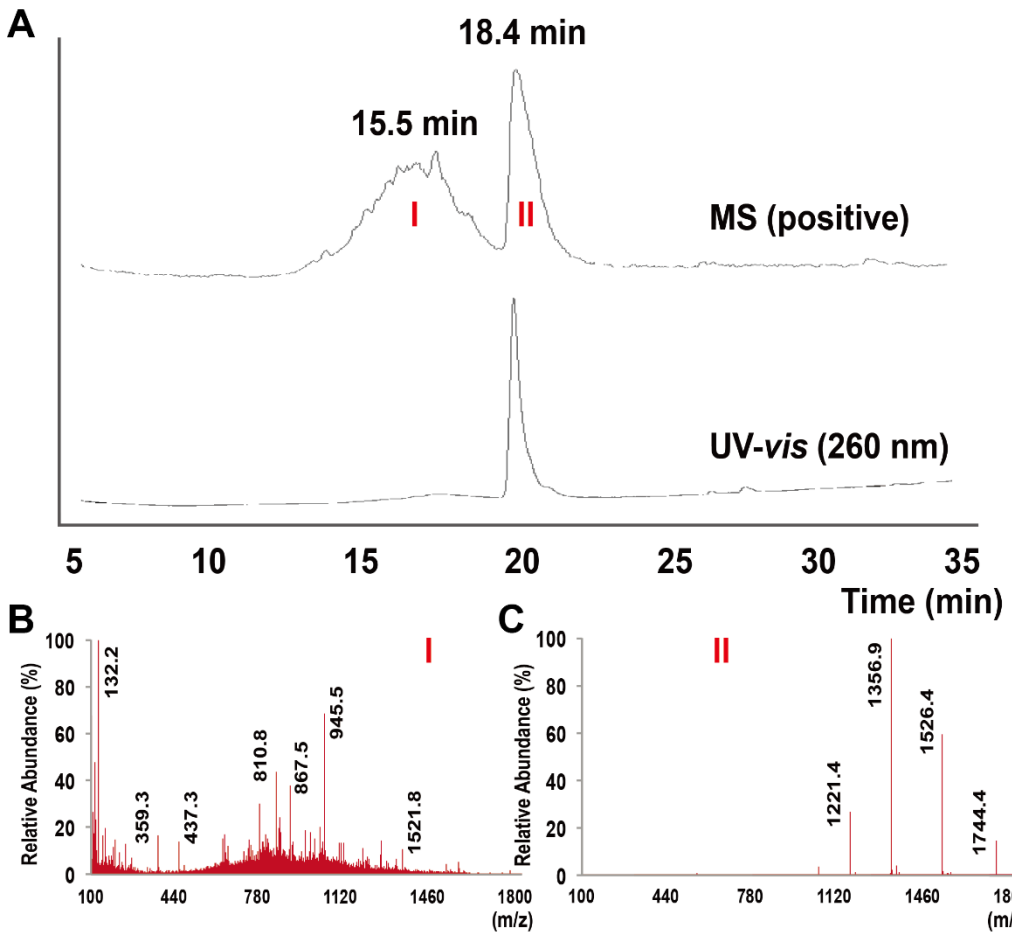


Figure S4. LC-ESI-MS analysis of recombinant Y3 from *P. pastoris* expression system. (A) LC and MS traces of Y3 sample. (B) ESI-MS spectrum of peak I. (C) ESI-MS spectrum of peak II.

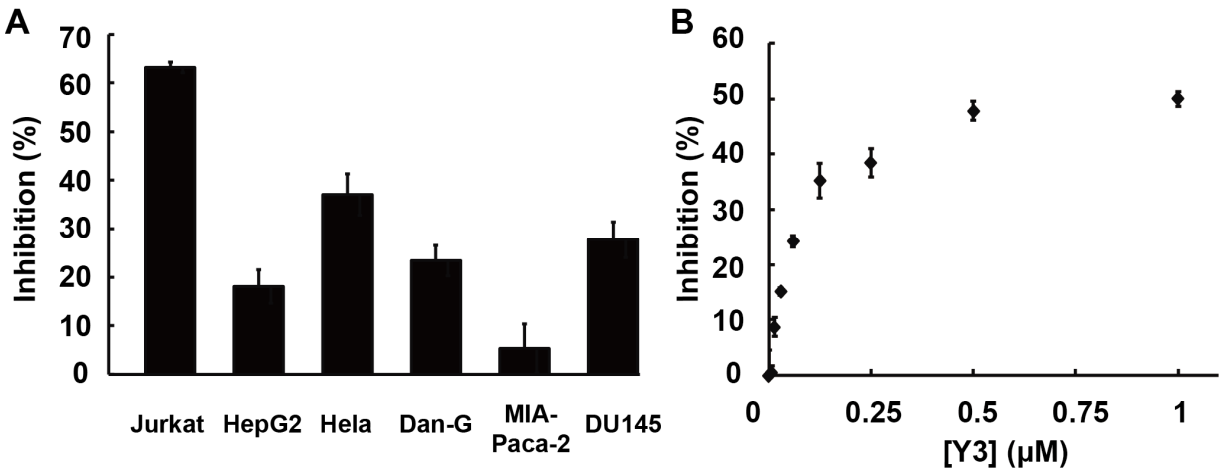


Figure S5. (A) Anticancer activity screening of Y3 at 10 μM by MTT assay. (B) Dose-dependent inhibition of Jurkat cells by recombinant Y3.

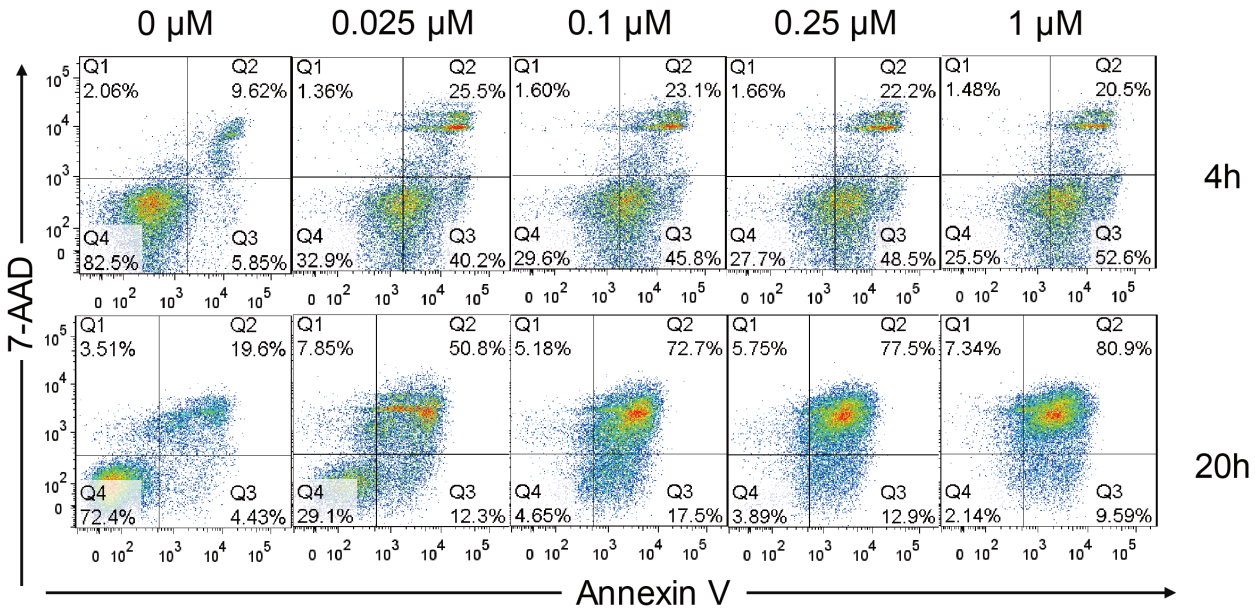


Figure S6. Flow cytometry analysis of cell apoptosis induced by the treatment of Y3 (0, 0.025, 0.1, 0.25, and 1 μM) for 4 h and 20 h. The appearance of apoptosis cells was detected by flow cytometry using Annexin V/7-AAD staining.

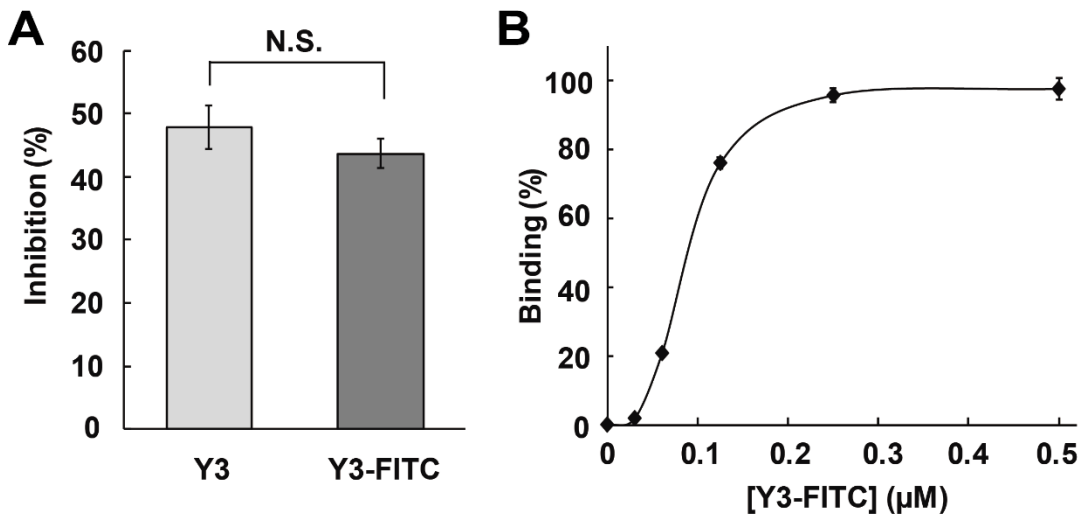


Figure S7. (A) Comparable anti-Jurkat activity of Y3 and Y3-FITC at 0.5 μM in MTT assay. Y3-FITC retained activity at similar level of Y3. (B) Dose-dependent binding of Y3-FITC to Jurkat cells. Jurkat cells were incubated with serial concentration of Y3-FITC. After mild washing, the binding was quantitated by flow cytometry.

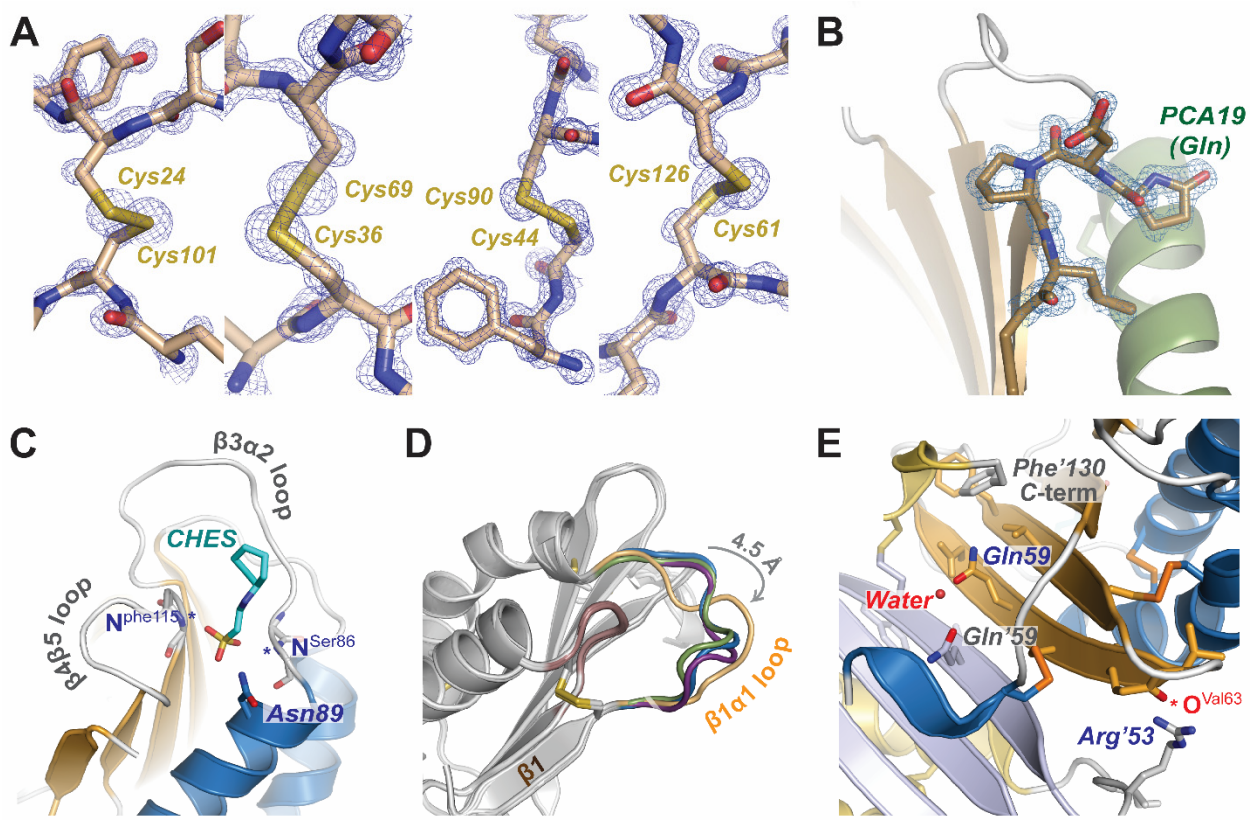


Figure S8. High-resolution crystal structure of Y3. (A) Four intramolecular disulfide bridges stabilize Y3 monomers. Calculated 2*mFo-DFc* map is displayed at 2σ contour level. (B) The *N*-terminal Met was cleaved during Y3 expression. PCA19 was formed through Glu19 cyclization. (C) A bound CHES molecule was well resolved near β4β5 and β3α2 loops. (D) Y3 monomers mainly differed in the β1α1 loop region. The orange loop (chain D) shifted ~ 4.5 Å comparing with the loops of other monomers. (E) Y3 protein formed a strong dimer through a β-β interaction Gln59 and neighbouring Arg'53 also stabilized the dimeric interface.

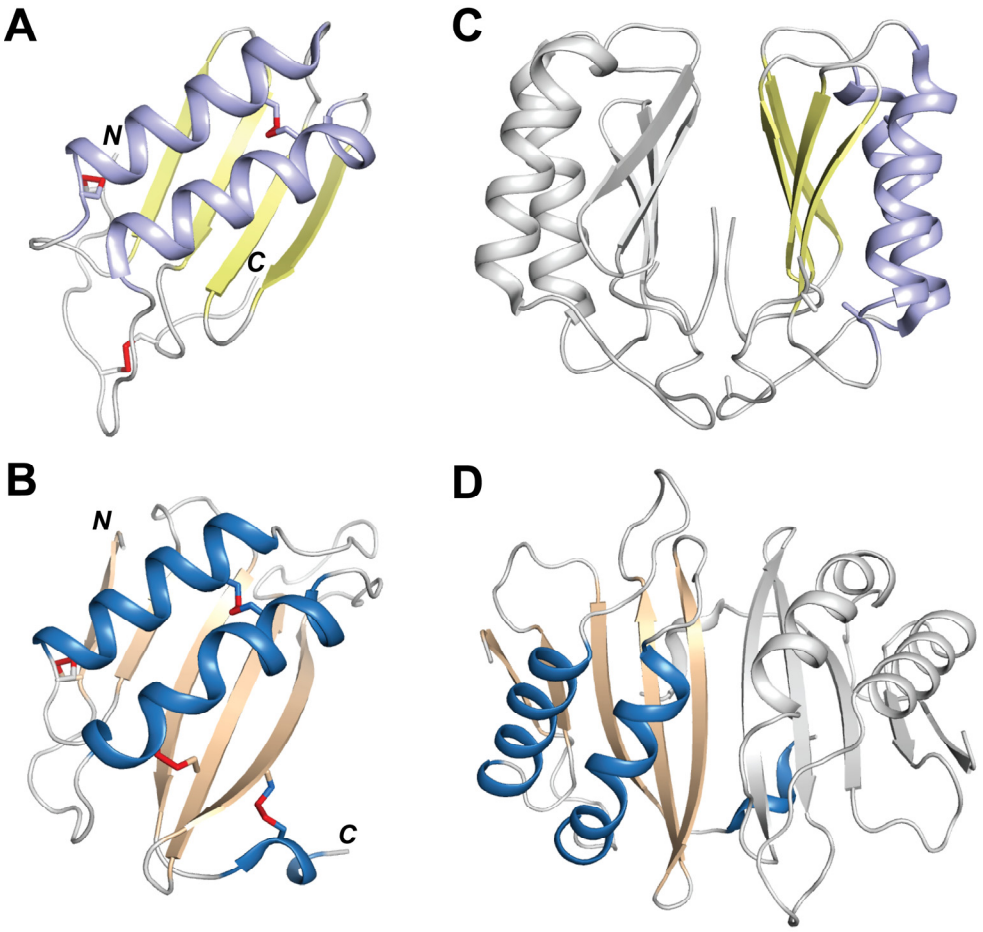


Figure S9. (A-B) LDL (pdb entry 4NDV, A) monomer shares considerable structural similarity with Y3 (RMSD of 2.4 Å, B). The key differences of these two monomers lie in the α3-helix present in Y3, the number of β sheets, along with the number and locations of disulfide bridges. (C-D) LDL (C) dimerizes in a different fashion as compared with Y3 (D).

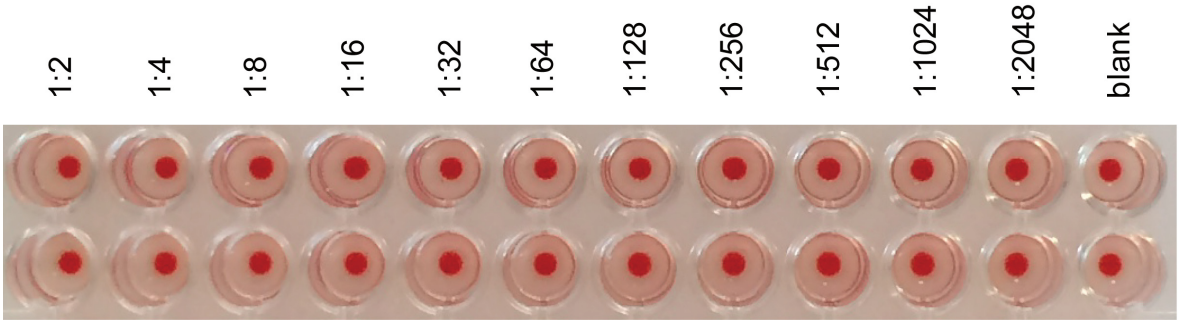


Figure S10. Hemagglutination assay of Y3 using human (upper panel) and rabbit (lower panel) erythrocytes. Y3 solution at 2 mM was serially diluted by two folds (2 to 2048 times).

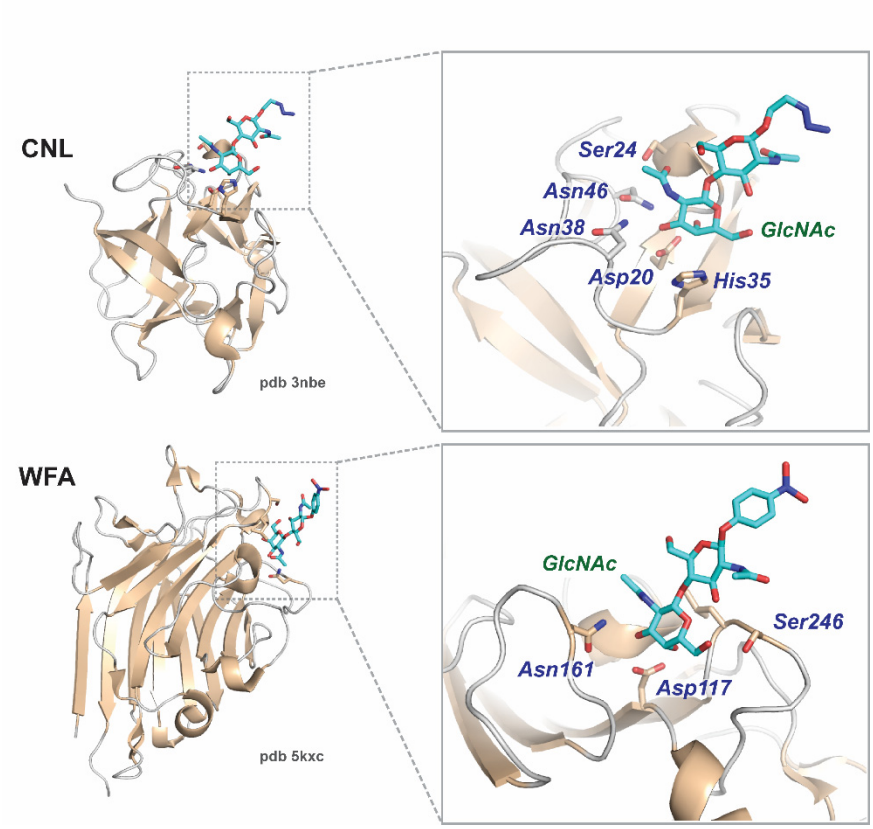


Figure S11. GlcNAc binding in the structures of *Wisteria floribunda* lectin (PDB 5KXC) and *Clitocybe nebularis* ricin B-like lectin (PDB 3NBE).

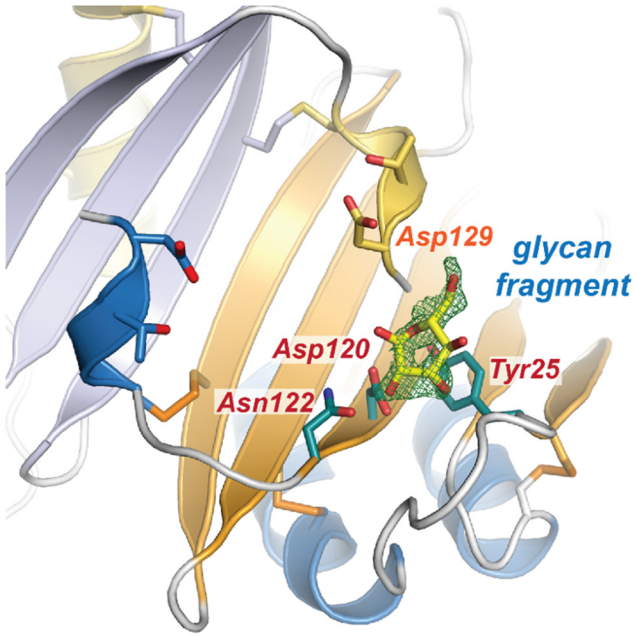


Figure S12. Unassigned electron density near the GlcNAc interacting region in Y3 high-resolution structure.

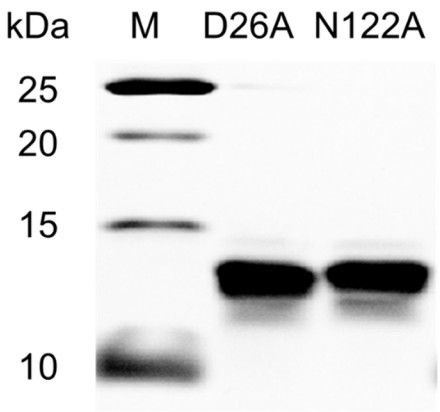


Figure S13. SDS-PAGE analysis of purified Y3D26A and Y3N122A.