A sweet protein monellin as a non-antibody scaffold for synthetic binding proteins

Norihisa Yasui*, Kazuaki Nakamura, Atsuko Yamashita

Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 1-1-1, Tsushima-naka, Kita-ku, Okayama, 700-8530, Japan

*Correspondence to Norihisa Yasui: Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 1-1-1, Tsushima-naka, Kita-ku, Okayama, 700-8530, Japan. E-mail: nyasui@okayama-u.ac.jp

Running title: Monellin scaffold for synthetic binding proteins

Abbreviations: BAS, biotin acceptor sequence; ELISA, enzyme-linked immunosorbent assay; GFPuv, the folding mutant of green fluorescent protein variant; RMSD, root mean square deviations; scMonellin, single-chain monellin; SPR, surface plasmon resonance; SWEEPin, sweet-tasting protein-based synthetic binding protein; TBS, Tris-buffered saline; ySUMO, yeast small ubiquitin-related modifier

Abstract

Synthetic binding proteins that have the ability to bind with molecules can be generated using various protein domains as non-antibody scaffolds. These designer proteins have been used widely in research studies, as their properties overcome the disadvantages of using antibodies. Here, we describe the first application of a phage display to generate synthetic binding proteins using a sweet protein, monellin, as a non-antibody scaffold. Single-chain monellin (scMonellin), in which two polypeptide chains of natural monellin are connected by a short linker, has two loops on one side of the molecule. We constructed phage display libraries of scMonellin, in which the amino acid sequence of the two loops is diversified. To validate the performance of these libraries, we sorted them against the folding mutant of the green fluorescent protein variant (GFPuv) and yeast small ubiquitin-related modifier. We successfully obtained scMonellin variants exhibiting moderate but significant affinities for these target proteins. Crystal structures of one of the GFPuv-binding variants in complex with GFPuv revealed that the two diversified loops were involved in target recognition. scMonellin, therefore, represents a promising non-antibody scaffold in the design and generation of synthetic binding proteins. We termed the scMonellin-derived synthetic binding proteins "SWEEPins."

Keywords: phage display, synthetic binding proteins, non-antibody scaffold, single-chain monellin, combinatorial library

Introduction

Antibodies and their fragments are widely used as diagnostic and research reagents, because of their ability to recognize target molecules (*1-3*). One of the structural features of antibodies that enable them to bind with other molecules is that the diversified loops on the stable immunoglobulin fold are exposed to the solvent. Non-antibody protein domains can also be provided with specific molecular recognition abilities if the domains are equipped with the structural features of antibodies (*4*). It has been demonstrated that protein domains with a non-immunoglobulin fold can be functionalized with novel binding sites by employing directed evolution, in which the combinatorial libraries of protein domains are generated and selected using phage display or other molecular selection techniques. A number of "non-antibody scaffold domains," fibronectin type 3 domain (*5*), lipocalin (*6*), ankyrin repeat protein (*7*), Z domain (*8*), Sso7d protein (*9*), etc., have been reported to generate synthetic binding proteins (*4*). Such synthetic binding proteins are more useful as research reagents than antibodies, because non-antibody scaffolds are generally small in size, monomeric, and easy to express in *Escherichia coli*. These properties overcome the characteristic disadvantages of antibodies, including high molecular weight and the presence of disulfide bonds. In fact, synthetic binding proteins have a wide variety of uses such as altering the specificity of enzymes (*10*), acting as crystallization chaperones in promoting the crystallization of biomacromolecules (*11, 12*), acting as imaging scaffolds to visualize small proteins by cryo-electron microscopy (*13*), and modifying protein-protein interactions in living cells (*14, 15*).

Recently, affimer proteins that were originally called Adhirons (*16*) have been developed for use as synthetic binding proteins (*17-19*). Affimers are composed of a single α-helix and the four anti-parallel β strands in a cystatin-like fold similar to cysteine protease inhibitors, cystatins, and were designed to show high thermal stability (*16*). In terms of

cystatins, two loops on the same side as the N-terminus resides are observed to play a role in the interaction with cysteine proteases to inhibit protease activity (*20-22*), which indicates that the cystatin-like fold is well-suited for interaction with other proteins. In fact, functionally desired affimers have been generated successfully by sorting the phage display library of the designed stable cystatin-like fold scaffold, in which the amino acid sequences of the inserted two loops were diversified (*16-19*).

The sweet protein monellin was originally isolated from the fruit of an African berry *Dioscoreophyllum cumminsii* (*23*). Monellin is composed of two polypeptide chains A and B (*23*), and shows the cystatin-like fold (*24, 25*). Single-chain monellin (scMonellin) proteins have been designed to increase the stability of monellin, in which two polypeptide chains are connected directly (SCM) (*26*) or via a Gly-Phe linker (MNEI) (*27*); these proteins also exhibit the sweetness like natural monellin (*26, 27*). Both types of scMonellin have two loops; one is naturally present in chain A portion, while the other one is artificially introduced between chains A and B (*28-30*). Consequently, scMonellins share structural features with affimer proteins, although, between them, the relative arrangement of the two loops differs slightly, due to variation in the lengths of the β-strands connected by the two loops. Owing to these similarities and differences in the structural features, scMonelins are candidates for a non-antibody scaffold, although this utility has not been demonstrated to date.

Here, we describe the design and generation of synthetic binding proteins using scMonellin as a non-antibody scaffold. We constructed phage display libraries of scMonellin in which the amino acid sequences in the two loops are randomized with the biased composition of the amino acids favorable for protein-protein interactions. We have successfully obtained the synthetic binding proteins targeted to the folding mutant of green fluorescent protein variant (GFPuv) and yeast small ubiquitin-related modifier (ySUMO) by

sorting the libraries. One of the scMonellin variants that showed the affinity for GFPuv was further characterized to reveal the structural basis of the target recognition. The results indicate that scMonellin is a promising protein as a non-antibody scaffold in the design and generation of synthetic binding proteins for various applications.

Materials and methods

Construction of scMonellin library

The chemically synthesized cDNA of scMonellin described by Konno (*31*) in a vector (pIDTAMAP-AMP:scMonellin) was purchased from the Integrated DNA Technologies, Inc. A DNA fragment coding the C-terminal domain of the M13 pIII was amplified by PCR from the wild-type gene III of M13 mp18 (TaKaRa, Accession No. : X02513) using primers 5’-CCGACTCGAGGCTGAAACTGTTGAAAGTTG-3’ (forward) and 5’- CCGGGTACCTTAAGACTCCTTATTACG-3’ and cloned into pBluescript II SK(+) with XhoI and KpnI sites to make pBluescript II SK(+)-pIII. A DNA fragment encoding the signal sequence of DsbA followed by scMonellin was generated by a three-step extension PCR. In the first PCR, pIDTAMAP-AMP:scMonellin was used as a template, and the following primer set was utilized: 5’-CTGGCTTTTTCTGCATCTGCTGCTGGATCCGGCGAATGGGAAATC-3’ (forward) and 5’-GCTGGCTAGCTTACGGCGGCGGCACCGG-3’ (reverse). In the second and third PCR, 5’-CTGGCAGGTCTGGTGCTGGCTTTTTCTGCATCTGC-3’ and 5’-ATACCCATGGATGAAAAAGATCTGGCTGGCTCTGGCAGGTCTGGTGCTG-3’ were used as forward primers, respectively. The resulting DNA fragment was inserted into pET25b using NcoI and NheI sites in order to make pDsbA-scMonellin. To add the segment encoding V5 tag sequence to the 3’- end of DNA encoding DsbA-scMonellin, a three-step extension PCR was carried out. In the first PCR, the PCR fragment was amplified with

5’-TAATACGACTCACTATAGGG-3’ (forward) and 5’-CTTACCGGAGGACGAACTAGTCGGCGGCGGCACCGGGCC-3’ (reverse) from pDsbA-scMonellin. In the second and third PCR, 5’-GAGAGGGTTAGGGATAGGCTTACCGGAGGACGAACTAG-3’ and 5’-TCAGCCTCGAGCGTAGAATCGAGACCGAGGAGAGGGTTAGGGAT AGG-3’ were used as reverse primers, respectively. The third PCR fragment was digested with XbaI and XhoI, and inserted into pBluescript II SK(+)-pIII using the same combination of restriction sites.

Randomization was carried out using oligonucleotides containing degenerated nucleotide sequences. Large-scale site-directed mutagenesis was performed following a published method (*32*), based on Kunkel mutagenesis using the mixture of oligonucleotides coding a biased amino acid composition that included Tyr (30%) Ser (15%), Gly (10%), Trp (5%), Phe (5%) and 2.5% of each of the other amino acids except for Cys, which was excluded (Japan Bio Services Co., LTD., Saitama, JAPAN) (*33*). The sequence of the oligonucleotides used for the construction of libraries is listed in Table S1. The Kunkel reaction product was amplified by electrotransforming *E. coli* SS320 (Lucigen) carrying the pCDFDuet1-based vector in which the lacI gene is mutated into lacIq (pCDFDuet1-lacIq vector). The library of phagemid vectors was purified and treated with EcoRI and MulI. The DNA treated with restriction enzymes were used in the electroporation of TG1 cells (Lucigen) carrying the pCDFDuet1-lacIq vector (TG1/lacIq). The cells were transferred into 2 L of 2×YT, and then, the helper phages were added to the culture. The cells were incubated at 37°C for 30 min with shaking at 100 rpm. Hyperphage (Progen Biotechnik) (*34*) was used to generate loop library A whereas M13KO7 was used for loop library B. Isopropyl-β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.1 mM, and cells were cultivated at 37°C, overnight. The culture was centrifuged at 5,000 × *g* at

4°C for 15 min, and the supernatant was transferred to a tube. A fifth volume of the solution consisting of 20% (w/v) PEG 8000, 2.5 M NaCl was added to the supernatant and mixed. The mixture was kept on ice for 1 h and centrifuged at 12,000 × *g* at 4°C for 20 min. The phage was suspended in 10 mM Tris-HCl, 1 mM EDTA, pH 7.5. The phage solution was then mixed with a final concentration of 50% (v/v) glycerol and stored at −30°C until use.

Preparation of the biotinylated target proteins for library sorting

A pET25-base expression vector pHFT-GFPuv-BAS was constructed. This vector encodes GFPuv with a segment composed of a decahistidine (His10), FLAG tag and a TEV cleavage site at the N-terminus and the biotin acceptor sequence (BAS) (*35*) at the C-terminus. The DNA encoding GFPuv followed by the BAS was generated by four-step extension PCR. The resulting DNA fragment was inserted into pHFT-GFPuv (*36*) using BamHI and NheI sites to make pHFT-GFPuv-BAS. To prepare the purified GFPuv-BAS protein, the *Escherichia coli* BL21 (DE3) pLysS strain was transformed with pHFT-GFPuv-BAS. The transformant was cultivated in 1 L of LB medium containing 100 μg/mL of carbenicillin and 34 μg/mL of chloramphenicol until the OD_{600} reached ~1.5. Protein expression was induced by adding 0.1 mM IPTG. Cells were supplemented with 50 μM biotin and cultivated for ~16 h at 20°C. Cells were harvested by centrifugation, washed with 20 mM Tris-HCl, pH 8 and stored at −30°C until use. Cells were resuspended in 20 mM Tris-HCl, pH 8.0 and lysed by sonication. After removing the cell debris by centrifugation, the supernatant containing the HFT-GFPuv-BAS was then collected and applied to a Ni-NTA agarose column (QIAGEN). After washing the column with 50 mM imidazole, 300 mM NaCl, and 20 mM Tris-HCl, pH 8.0, proteins were eluted with 250 mM imidazole, 300 mM NaCl, and 20 mM Tris-HCl, pH 8.0. HFT-GFPuv-BAS was treated with His-tagged TEV protease (E:S = 1:6.5) to release the His10-FLAG tag at the N-terminus by incubation at

20°C, overnight. Following the dialysis against 20 mM Tris-HCl, 300 mM NaCl, pH 8.0, the TEV protease and the tag segment were then removed using a second Ni-NTA agarose column.

To prepare the biotinylated ySUMO protein, a pET25-base expression vector pHBAS-WK-ySUMO was constructed to express ySUMO with the BAS. This plasmid encodes the ySUMO (Ser3−Gly98) with the His_{10}-BAS-Trp-Lys segment at the N-terminus. Expression and purification were carried out as for the HFT-GFPuv-BAS protein, without TEV protease treatment.

Sorting of the phage display libraries

In the first round selection, 250 μL of streptavidin-coated magnetic beads was mixed with 500 μL each of the target proteins at ~2 μM (GFPuv-BAS and HBAS-ySUMO) at 4°C for 1 h with rotation. After washing the beads with 20 mM Tris-HCl, 150 mM NaCl, 0.05% (w/v) Tween 20, pH 7.5 (TBS-T), to remove the unbound biotinylated proteins, the beads was treated with 500 μL of 5 μM biotin at 4°C for 5 min. The library phage particles were then mixed with target immobilized beads in 0.5 mL of TBS-T containing 0.5% BSA and 1 μg/mL streptavidin (Nacalai tesque) at 4°C for 1 h with rotation. After washing with 1 mL of TBS-T five times, 3 mL of TG1/lacIq cells was directly infected with the phage/beads mixture by incubation at 37°C for 30 min. After incubation, the infected cells were transferred into 30 mL of 2×YT containing 100 μg/mL of carbenicillin, 100 μg/mL of spectinomycin, 0.1 mM IPTG, and 1.4×10^9 cfu/mL Hyperphage and cultivated at 37°C overnight. Amplified phages were purified by PEG precipitation. In the second through fourth round selections, 40 μL of magnetic beads was used. Aliquots of 500 μL each of the target proteins at 1 μM in the second round selection, 0.5 μM in the third round selection and 0.3 μM in the fourth round selection, respectively, were used for immobilization on the

magnetic beads. Phages bound to the target protein-immobilized beads were eluted with 100 μL of 0.1 M Glycine-HCl, pH 2.5, and neutralized with 20 μL of 2 M Tris-HCl, pH 8.0. A 60 μL of the neutralized eluted phage particles was used to infect 0.5 mL of the log-phase TG1/lacIq cells. Infected cells were then transferred into 2.5 mL of 2×YT containing 100 μg/mL of carbenicillin and 1.4×10^9 cfu/mL Hyperphage and cultivated to amplify the phage particles. At the final round selection, the phage-infected cells were spread on an LB plate containing 100 μg/mL of carbenicillin and 100 μg/mL of spectinomycin to prepare the individual clones.

Phage enzyme-linked immunosorbent assay (ELISA)

Individual TG1/lacIq colonies were grown in ~1 ml of 2×TY with 100 μg/ml of carbenicillin and 100 μg/ml of spectinomycin in a 96-deep well plate at 37°C for 2 h. Hyperphage and 0.1 mM IPTG were added and incubated at 37°C with shaking overnight. Wells of a 96 well plate (F96 Maxisorp nunc-immuno plate, Nunc, cat no. 442404) were coated with 100 μL/well of 5 μg/mL of NeutrAvidin (Thermo Fisher Scientific) in 20 mM Tris-HCl, 150 mM NaCl, pH 7.5 (TBS), by incubation at room temperature for 1 h. After discarding the streptavidin solution, 100 μL/well of 0.5 μM biotinylated proteins (GFPuv-BAS or HBAS-ySUMO) in TBS was added to the wells and incubated at room temperature for 1 h. For direct coating of the antibody, 100 μL/well of 1 μg/mL anti-V5 IgG (FUJIFILM Wako Pure Chemicals) diluted in TBS was added to the wells. For the control well, the same volume of TBS was added. After discarding the protein solution, 130 μL/well of 0.5% BSA in TBS was added to the wells and incubated at room temperature for 1 h. After removing the BSA solution, 50 μL of phage solution from the cell culture was added to the wells and incubated at room temperature for 1 h. After discarding the supernatant, the wells were washed with 200 μL/well of 20 mM Tris-HCl, 150 mM NaCl, 0.05% (w/v) Tween-20,

pH 7.5 (TBS-T), five times, followed by incubation with 100 μL/well of anti-M13 IgG-HRP (GE Healthcare) in TBS-T containing 0.1% BSA (1:2,500). After washing with 200 μL/well of TBS-T five times, 100 μL/well of ABTS solution (Roche) was added and incubated at room temperature for ~10 min. The absorbance at 405 nm was measured on a plate reader, Varioskan Flash (Thermo Scientific).

Construction of expression vectors

The genes for scMonellin variants were cloned in a pET25-based expression vector, pHFT (*36*). The pHFT vector expresses a cloned gene product with a decahistidine His10, a FLAG tag, and a TEV cleavage site fused to the N-terminus. The DNA fragments encoding the scMonellin variants were amplified and subcloned into the pHFT treated with BamHI and NheI. The DNA encoding ySUMO (Ser3−Gly98) was subcloned into the same vector using BamHI and NheI sites. All constructs were verified by DNA sequencing.

Protein expression and purification

BL21 (DE3) cells were transformed with the expression vectors. Protein expression was induced using autoinduction media for 22~24 h at 30°C (*37*). Proteins were purified with Ni-affinity chromatography. The N-terminal tag was cleaved by TEV protease, and the cleaved protein was purified by Ni-affinity chromatography. For surface plasmon resonance measurement, the tag-cleaved GFPuv (*36*) and ySUMO were further purified on an ENrich Q 5 × 50 anion-exchange column (C.V.: 0.98 mL, Bio-Rad) to remove the residual tagged species. The column was equilibrated with 20 mM Tris-HCl, pH 8.0 and elution was performed with a linear gradient from 0 to 0.5 M NaCl over a 20-column volume at a flow rate of 1 ml/min for GFPuv purification. For ySUMO purification, the proteins were eluted with a linear gradient from 0 to 1 M NaCl over a 20-column volume.

Size exclusion chromatographic analysis

The purified scMonellin variants were subjected to size exclusion chromatography on an ENrich SEC 70 10 × 300 column equilibrated with 20 mM Tris-HCl,150 mM NaCl, pH 7.5 at a flow rate of 1 ml/min with NGC Quest 10 Plus (Bio-Rad).

Differential scanning fluorimetry

The thermal stability for scMonellin and its variants were assessed by protein thermal shift assay using the Protein Thermal Shift kit (Applied Biosystems). The purified protein samples were dialyzed against 20 mM HEPES-Na, 150 mM NaCl, pH 7.5. The dialyzed protein (~1 μg) and Protein Thermal Shift Dye were mixed in the dialysis buffer to prepare 20 μL of the protein melt reaction. For the measurement of scMonellin WT, ~5 μg of the purified sample was used because of low signal when measured using 1 μg of the protein sample. Fluorescent intensity was measured by the StepOne Real-Time PCR System (Applied Biosystems). The mixtures were denatured by raising the temperature from 25°C to 99°C at a rate of 0.022°C/sec. The apparent thermal denaturation temperatures (T_m) were estimated by the two-state Boltzmann model using Protein Thermal Shift Software 1.3 (Applied Biosystems).

Surface plasmon resonance measurement

Surface plasmon resonance analysis was carried out using a Biacore 2000 instrument (GE Healthcare) at a constant temperature of 20°C. His-tagged scMonellin variants were immobilized on Ni-NTA sensor chip. His-tagged ySUMO protein was immobilized on the surface of the reference cell at approximately the same level as that of scMonellin variants on the main surface to prevent non-specific interaction of the analytes with the surface.

Sensorgrams were collected after infusing various concentrations of analyte proteins in 10 mM Tris-HCl, 150 mM NaCl, 50 μM EDTA, 0.005% (w/v) Tween-20, pH 7.5 at a flow rate of 30 μL/min. The surface was regenerated by a pulse infusion of 10 mM Tris-HCl, 150 mM NaCl, 350 mM EDTA, 0.005% (w/v) Tween 20, pH 7.5 after each run. The obtained sensorgrams were processed with BIAevaluation software. The double-referenced sensorgrams were obtained by subtracting the response from the reference cell and subsequently subtracting the sensorgram of buffer (i.e., zero concentration of analyte) injection. Values for the dissociation constants (K_D) were estimated from plots of equilibrium response values against analyte (GFPuv or ySUMO) concentrations by fitting the 1:1 binding model using Igor Pro software (WaveMetrics) with the following equation:

$$R_{eq}(C) = \frac{R_{max} \times C}{K_D + C}$$

where $R_{eq}(C)$ is the response at equilibrium observed at the analyte concentration, C and R_{max} is the difference in the R_{eq} in the absence and presence of saturating concentrations of the analyte proteins.

Crystallization, data collection and structural determination

The separately purified GFP-40 and GFPuv were mixed at a 1:1 molar ratio and concentrated to ~20 mg/mL. The concentrated sample was subjected to crystallization screening via the sitting-drop vapor diffusion method using a Crystal Screen kit (Hampton Research). Crystals of form I of the GFP-40/GFPuv complex were grown at 20°C in hanging drops with a reservoir solution containing 15% (w/v) PEG 4000, 200 mM $MgCl_2$, 0.1 M Tris-HCl, pH 8.5. Crystals of form II of the GFP-40/GFPuv complex were also grown at 20°C in hanging drops with a reservoir solution containing 5.5% (w/v) PEG 8000, 5% (v/v) ethylene glycol, 50 mM $[Co(NH_3)_6]Cl_3$, 0.1 M HEPES-Na, pH 7.5.

Prior to data collection, the crystals were soaked in a reservoir solution with added 20% (v/v) ethylene glycol and then flash-frozen in liquid nitrogen. The diffraction data sets used for the structural determination were collected at a wavelength of 1.0000 Å on a SPring-8 BL41XU using an EIGER X 16M (DECTRIS) detector. Diffraction data were processed using the HKL-2000 program package (*38*). Initial phases were determined via molecular replacement with Phaser (*39*) in the CCP4 program suite. The orientations and positions of GFPuv and GFP-40 were determined by using the structure of GFP (PDB ID: 1B9C) and the structure of scMonellin (PDB ID: 2O9U), of which loops 1 and 2 were omitted as the search models, respectively. Clear solutions were obtained for both the GFPuv and GFP-40 molecules. Crystal form I contained two complexes of GFPuv and GFP-40 in the asymmetric unit, whereas crystal form II contained one GFP-40/GFPuv complex in the asymmetric unit. The resulting models were improved by iterative cycles of manual model correction with COOT (*40*) and refinement with Phenix.refine (*41*). A summary of the data collection and refinement statistics is shown in Table 1.

For the structural analysis, the binding interface was analyzed with CONTACT in the CCP4 program suite (*42*) and 'Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (*43*). The structure superposition was performed with GESAMT in the CCP4 program suite (*44*). All figures of the protein structures were prepared with PyMOL (The PyMOL Molecular Graphics System, Version 2.2 Schrödinger, LLC.).

Results

Library design and construction

The single chain monellins (scMonellin) SCM (*26*) and MNEI (*27*) are composed of 94 and 96 amino acid residues, respectively, and they show the structures and activities that

316 are similar to those of natural monellin (*28, 30*). These engineered proteins consist of a
317 five-strand anti-parallel β-sheet and an α-helix on the concave side of the β-sheet; they also
318 have two loops (L_{23} and L_{45}) on the same side of the molecule as the N-terminus (*25, 30*). In
319 the following, we use the terms loop 1 and loop 2, instead of L_{23} and L_{45}, respectively (Fig.
320 1B). scMonellin, not natural monellin, was chosen as a scaffold because a single polypeptide
321 form is more suited to be displayed on the phage surface, and it also allowed for simultaneous
322 randomization of the two loops simultaneously.

323 The phagemid vector for displaying scMonellin on the M13 phage was designed to
324 have the signal sequence of DsbA at the N-terminal and V5 tag for detection at the
325 C-terminal, connected to the full-length of pIII protein of the M13 phage according to a
326 previous study (*33*) (Fig. 1A). Substitution of the cysteine residue with serine (C41S) was
327 introduced to avoid the intermolecular disulfide formation. Amino acid residues in loops 1
328 and 2 of scMonellin were then diversified to generate the combinatorial libraries. EcoRI and
329 MluI sites were introduced into loops 1 and 2, respectively, to remove the parent sequence
330 during library construction (Fig. 1A).

331 We designed and generated two different combinatorial libraries using scMonellin as
332 a non-antibody scaffold. One library, named "loop library A," was constructed using
333 scMonellin, in which the lengths of loop 1 and loop 2 were fixed at seven and five residues,
334 respectively. The length of loop 1 was the same as that of scMonellin MNEI, in which two
335 polypeptide chains are connected via a Gly-Phe linker. We did not diversify the Tyr residue
336 at the beginning of loop 2 (boxed in Fig. 1A), with the hope that this Tyr would contribute to
337 the interaction with targets, because Tyr residues are suitable in making a binding interface
338 (*45*). In the other library, "loop library B," the lengths of the two loops were varied. The
339 lengths of loop 1 and loop 2 were five to 10 and five or six residues, respectively (Fig. 1C).
340 We limited the variation in the length of loop 2 in loop library B, because this loop seems to

form β-turn, according to the crystal structure of scMonellin MNEI (PDB ID: 2O9U) (*30*). In both libraries, the two loops were diversified with highly biased amino acid residue mixtures, as employed in a previous study on obtaining synthetic binding proteins using monobody libraries (*33*). Both loop library A and loop library B were constructed in the phage-display format with estimated numbers of independent sequences of 2.0×10^9 and 7.0×10^{10}, respectively.

Library sorting

To examine the performance of the libraries, we sorted them against two target proteins, GFPuv (*46*) and ySUMO. GFP has been widely used for applications in the life sciences (*47-49*), while ySUMO is known as a protein tag that is efficient in enhancing protein expression and solubility (*50, 51*). These proteins have also been targeted with non-antibody scaffold libraries in previous studies due to their usefulness (*16, 52-56*). We have chosen these proteins as model targets because we can compare the properties of the scMonellin variants with those of reported binders that have been derived from other non-antibody scaffolds.

We first sorted loop library A against GFPuv and ySUMO by phage display. After four rounds of library selection for each target, the phage clones that showed affinity for GFPuv or ySUMO were identified by enzyme-linked immunosorbent assay (ELISA). In total, 22 of the 23 clones for GFPuv gave rise to ELISA signals (Fig. 2A), and DNA sequencing analysis of 20 of these clones revealed 6 different scMonellin variants (Fig. 3A). On the other hand, all of 22 clones for ySUMO tested have exhibited the binding signals in phage ELISA (Fig. 2B). DNA sequencing analysis revealed that all of the clones shared the same amino acid sequence (Fig. 3A).

We also sorted another library, loop library B, against the same targets in almost the same way for loop library A. After four rounds of library selection, 19 of the 22 clones obtained by selection against GFPuv gave rise to ELISA signals, whereas 11 of the 22 clones obtained by selection against ySUMO exhibited ELISA signals (Figs. 2C and 2D). DNA sequencing analysis of these clones identified 5 and 10 distinct variants of scMonellin, which exhibited binding affinities for GFPuv and ySUMO, respectively (Fig. 4A).

Characterization of selected scMonellin variants exhibiting affinities to the targets

In order to characterize selected scMonellin variants, these were expressed in *E.coli* and purified using affinity chromatography. We initially tested the interaction between the purified scMonellin variants and their target proteins by size-exclusion chromatography (Fig. S1). After this preliminary test, several scMonellin variants were selected for further characterization of the target binding by surface plasmon resonance (SPR) measurement.

Among the scMonellin variants against GFPuv from loop library A, the target binding of GFP-40 was analyzed by SPR measurement. GFP-40 was observed to bind to GFPuv with fast binding and dissociation rates (Fig. 3B, *right*). The dissociation constant at the equilibrium state (K_D) was estimated to be approximately 24 μM (Fig. 3B, *right*). On the other hand, the dissociation constant could not be estimated for the wild-type of scMonellin (scMonellin WT) (Fig. 3B, *left*). These observations indicated that GFP-40 acquired the ability to bind to GFPuv protein when the scMonellin scaffold was made to contain the appropriate amino acid sequences in the loops. The only ySUMO-targeted scMonellin variant selected from loop library A, SUMO-31, interacted with ySUMO with fast binding and dissociation rates, although the dissociation rate was slower than that observed in the interaction between GFP-40 and GFPuv (Fig. 3C). The K_D value of the interaction between SUMO-31 and ySUMO was estimated to be ~3.5 μM (Fig. 3C).

Next, we characterized the purified protein samples of scMonellin variants derived from loop library B. Among the five scMonellin variants that provided the binding signals for GFPuv in the phage ELISA, SPR measurements of three variants named GFP-kz02, GFP-kz06 and GFP-kz09 were carried out to further characterize their interactions with GFPuv. All three variants were observed to bind to GFPuv with fast binding and dissociation rates, as observed for the GFP-40 variant derived from loop library A (Fig. 4B). The K_D values for GFP-kz02, GFP-kz06, and GFP-Kz09, estimated at the equilibrium state, were 4.6 μM, 3.4 μM, and 12 μM, respectively (Fig. 4B). These values were two to seven times lower than that of GFP-40, which suggested that loop library B is more efficient than loop library A in obtaining the synthetic binding proteins with higher affinities. The scMonellin variants targeted ySUMO, SUMO-kz03 and SUMO-kz11, showed the sensorgrams indicating that interactions occurred with fast binding and dissociation rates (Fig. 4C). The equilibrium K_D values of SUMO-kz03 and SUMO-kz11 for ySUMO binding were estimated to be 0.9 μM and 1 μM, respectively (Fig. 4C), which were three to four times lower than that estimated for SUMO-31. This observation suggested that loop library B again outperformed loop library A in the efficiency of obtaining higher-affinity binders.

We next characterized the solution behavior of the purified protein samples of the scMonellin variants as well as the wild type using size-exclusion chromatography. All of the scMonellin variants tested here, along with the wild type, were eluted predominantly as single peaks (Fig. 5). The relative molecular mass of scMonellin WT was estimated to be ~10.2 kDa, which is comparable with the predicted molecular mass of scMonellin variants (11.2 kDa). Several variants such as GFP-kz09, GFP-kz06, SUMO-31, and GFP-40 were eluted at the volumes corresponding to relative molecular masses smaller than expected, suggesting that these variants interacted with the resin of the column during chromatography.

We further investigated the thermal stability of the scMonellin variants along with wild-type by differential scanning fluorimetry. Apparent thermal denaturation temperature (T_m) for scMonellin WT was estimated to be 74.2°C (Table 1 and Fig. S2), which is comparable with that of scMonellin MNEI (74.2°C), as investigated by circular dichroism spectroscopy (*57*). The T_m values for all tested scMonellin variants were estimated to be lower than that of scMonellin WT (Table 1 and Fig. S2). These variants exhibited a monophasic transition in fluorescence melt curve like scMonellin WT (Fig. S2B). These results indicated that the scMonellin scaffold was robust to the alteration in lengths of and introduction of mutations into the two loops in terms of the solution behavior.

As described above, the use of scMonellin as a non-antibody scaffold enabled us to generate synthetic binding proteins with the ability to bind to the model target proteins, GFPuv and ySUMO. We named the synthetic binding proteins based on the scMonellin scaffold "SWEEPins; sweet-tasting protein-based synthetic binding proteins."

Structural analysis of the SWEEPin-GFPuv complex

To reveal the structural basis for the target recognition by the scMonellin variant, the crystal structures of the SWEEPin GFP-40/GFPuv complex were determined. We obtained diffraction quality crystals of the GFP-40/GFPuv complex under two conditions. Data sets for both types of crystal that had different space groups, $P2_1$ (crystal form I) and $P2_12_12_1$ (crystal form II), were successfully collected. The asymmetric units of crystal form I and form II contained two and one complex(es), respectively. The crystal structures were determined by the molecular replacement method using the structures of GFPuv [Protein Data Bank (PDB) ID: 1B9C] and scMonellin (PDB ID: 2O9U) as search models. The GFP-40/GFPuv complex structures of crystal form I and form II were then refined at 1.7 Å and 2.0 Å resolutions, respectively.

The overall structures of the complexes and conformations of two loops of GFP-40 were similar (Figs. S3A and S3B) when the three complexes were superimposed (root mean square deviations in the range of 0.397–1.53 Å; Fig. S3C), although differences in orientation of the body of GFP-40 among three complexes were observed. The structures of chains A (GFP-40) and B (GFPuv) of crystal form I are described below as a representative of the GFP-40/GFPuv complex, unless otherwise stated.

The overall structure of the GFP-40/GFPuv complex has revealed that GFP-40 is bound to the base of the β-can fold of GFPuv, opposite to the side on which the N and C termini are located, using loops 1 and 2, as expected (Fig. 6A). Superposition between scMonellin (PDBID: 2O9U) and SWEEPin GFP-40 in the complex (RMSD of 0.747 for the Cα atoms of 90 aligned residues) revealed that their overall structures were similar (Fig. S3D). Two differences in backbone structures were found. The backbone structure of the segment composed of Arg41-Pro42-Ser43 on strand 2 in SWEEPin GFP-40 was different from the corresponding region of scMonellin, probably due to the substitution of Cys with Ser to avoid the intermolecular disulfide formation (Fig. S3D). Another minor difference was found in the position of Ile57 at the beginning of strand 3, shortening strand 3 by one residue (Fig. S3D). On the other hand, the GFPuv in the complex showed a similar structure to GFPuv alone (PDB ID: 1B9C, chain A) with an RMSD of 0.495 Å for the Cα atoms of 224 aligned residues, which indicated that no major conformational changes occur upon GFP-40 binding.

The total solvent-accessible surface area buried in the interface between SWEEPin GFP-40 and GFPuv in the complex was 1285 $Å^2$, which is comparable to the standard physiological protein-protein interfaces (1600 ± 400 $Å^2$) (*58*). The amino acid residues outside the loop 1 and loop 2 did not appear to largely contribute to the binding interface, even though Phe13 on an α-helix of SWEEPin GFP-40 was located within 4 Å of Pro211 and

Asn212 of GFPuv. The side chains of the three amino acid residues (Asn52, Arg54, and Tyr50) in loop 1 and two amino acid residues (Gln82 and Tyr84) in loop 2 were involved in the interaction with the GFPuv molecule, mainly by hydrogen bonding (Fig. 6B). In particular, Asn52 in loop 1 appeared to form a hydrogen bond with the side chain of Glu142 residue in GFPuv (Fig. 6B). The side chain of Arg54 seemed to form a salt bridge with the side chain of Glu142 in GFPuv. On the other hand, Gln82 and Tyr84 residues in loop 2 formed hydrogen bonds with the side chains of Glu172 and Arg215 residues in GFPuv, respectively (Fig. 6B). Analysis of the binding interface revealed that Ala51, Ser53, and Gly55 residues in loop 1 and Pro85 residue in loop 2 of SWEEPin GFP-40 were not located within 4 Å of any atom of GFPuv, which indicated that these residues do not contribute largely into making the binding interface. The epitopes for GFP-40 did not contain the mutation sites specific for GFPuv (i.e., F99S, M153T, and V163A), which implied that GFP-40 is a pan-binder for the GFP variants and thus is not specific to the GFPuv variant.

In addition to the direct interactions between amino acid residues in the two loops of GFP-40 and GFPuv, water-mediated hydrogen bonding networks were found at the binding interface (Fig. 6C). The water molecules found at the binding interface may also contribute to stabilizing the GFP-40/GFPuv complex.

Comparison of the GFP-binding mode of GFP-40 and other GFP-binders

Many kinds of synthetic binding proteins targeted to GFP and its variants have been reported. Furthermore, structural information on the binding sites on the GFPs is known for several of these binders including αRep (*52*), DARPin (*53*), and nanobodies (*54, 55*). We compared the SWEEPin GFP-40 with five other synthetic binding proteins in terms of their binding sites on the GFP molecule. Most GFP binders mainly recognize the side of the β-can fold of GFP, unlike the SWEEPin GFP-40 (Fig. 7A). α-Rep (PDB ID: 4XL5) and DARPin

(PDB ID: 5MA5) are observed to wrap around the β-can fold, making a large interface with GFP (Fig. 7A, *upper*). The three kinds of nanobodies (PDB IDs: 3K1K, 3G9A, and 6LR7), which were focused on in this study, had different binding sites on the GFP and did not share the main binding site with GFP-40 (Fig. 7A, *lower*). Among the GFP binders investigated, αRep shared limited epitopes (Lys52, Gly138, His139, Lys140, Tyr143, Glu172, Lys209, Pro211, Glu213, Asp216, and His217) on the base of the β-can fold with GFP-40 (Fig. 7B), simply because this GFP-binding protein had a particularly large binding interface with the GFP variant, EGFP (Fig.7A, *upper middle*). The structural inspection performed here revealed that the binding site of GFP-40 on GFPuv seems unique among the GFP binders, although available structural information on GFP binders remains limited.

Discussion

In this study, we have described the use of the sweet protein scMonellin as a non-antibody scaffold in generating the synthetic binding proteins that target proteins of interest. Phage display libraries with diversified amino acid sequences of two loops within scMonellin were constructed, and the synthetic binding proteins targeted for GFPuv or ySUMO were successfully obtained by sorting these scMonellin loop libraries. The biggest problem is the lower affinity of the scMonellin-based binders for the target proteins, even though the scaffold is structurally similar to affimers, the successful synthetic binding proteins. One possible explanation for this is that the lengths of the randomized loops were too short to obtain the scMonellin variants with high affinities for the target proteins. The lengths of loop 2 in the scMonellin libraries constructed in this study are five or six residues, whereas both loops of the affimers are ten residues long (*16*). Extending the length of loop 2 of the scMonellin scaffold needs to be addressed in the future. Another possible explanation is that the concentrations of the target proteins during library selection were too high, which

514 allows the weak affinity binders to be preferentially enriched. A single weak affinity binder
515 for ySUMO was obtained from library A, implying this possibility. The sorting condition will
516 need optimizing in the future to obtain binders that show high affinity.

517 Despite the relatively low affinity for GFPuv, it is noteworthy that the scMonellin
518 variant GFP-40 has an extremely rare binding site when compared with other non-antibody
519 GFPs binders (Fig. 7). This variant binds to the base of the β-can fold of GFPuv using the
520 variable loops forming the convex paratope. In contrast, other non-antibody GFP binders
521 interact with the side of the β-can fold of GFPs using flat or concaved paratopes (Fig. 7A).
522 Especially, GFP-binding nanobodies utilize their framework regions, in addition to the
523 variable loops, in the recognition of GFPs, resulting in forming the relatively flat paratopes
524 (*55*). These differences in the shape of the binders' paratope may explain why GFP-40 has a
525 unique epitope on GFPuv.

526 Some nanobodies, such as enhancer and minimizer, can modulate the fluorescence
527 properties of GFPs, which is useful to many applications in living cells (*55, 59*). The effect of
528 these nanobodies in altering GFP properties depends on their binding sites on the GFP
529 molecules. Therefore, how GFP-40 affects the fluorescence properties of GFP should be
530 investigated. We did not undertake fluorescence measurements, because the concentrations of
531 the protein samples were too high for these measurements to be carried out. This was due to
532 the low affinity of the scMonellin variant GFP-40. An affinity maturation procedure to
533 increase the affinity of GFP-40 for GFP will be required before fluorescence measurements
534 can be taken.

535 The consensus sequences, Y-X-N and Q-X-(Y/W)-P, were found in loop 1 and loop
536 2, respectively, when comparing the amino acid sequences of GFP-40 and the variants
537 GFP-kz02 and GFP-kz06 (Figs. 3A and 4A). In the crystal structure of the GFP-40/GFPuv
538 complex, the Pro residue (Pro85 in the case of GFP-40) in the consensus sequence in loop 2

539 did not make a direct contact with GFPuv, suggesting that this conserved Pro residue plays a
540 role in forming the specific main chain conformation. Thus, GFP-kz02 and GFP-kz06 may
541 share the binding site on GFPuv with GFP-40, although this possibility will need to be tested
542 through competitive binding experiments and/or structural determination. An advantage of
543 the SWEEPins scaffold in terms of affinity maturation is the feasibility of simultaneous
544 engineering of two loops. Affinity maturation of GFP-40 is likely to be achievable by
545 generating a library in which the conserved residues that are described above are fixed and
546 the residues at other positions in two loops are simultaneously randomized.

547 We chose ySUMO as another model target protein to demonstrate the efficiency in
548 sorting the phage display libraries of scMonellin. In the biological context, ySUMO (SMT3)
549 is covalently attached to other proteins, and regulates the function of these modified proteins
550 through interactions with proteins containing the SUMO-interacting motif (SIM) (*60-62*). We
551 generated 11 variants of scMonellin targeted to ySUMO in this study. In the loops of these
552 ySUMO binders, acidic residues were found frequently, especially in loop 1 (Figs. 3A and
553 4A). The highly negative region within loop 1 of the ySUMO-binding scMonellin variants
554 might interact with the basic residues in the vicinity of the hydrophobic cleft comprising the
555 SIM-binding site on ySUMO through long-range electrostatic interactions (*63, 64*).

556 Non-antibody scaffold proteins, monobodies and affimers, that bind to ySUMO have
557 previously been generated and well characterized (*16, 56*). In the case of the ySUMO-binding
558 monobody (ySMB-1), Tyr residues in FG loop of the fibronectin type III domain contribute
559 to making the binding interface. The crystal structure of the ySMB-1/ySUMO complex (PDB
560 ID: 3QHT) revealed that FG loop of ySMB-1 forms a β-hairpin and docks in the hydrophobic
561 region of the SIM binding site, which indicates that this interacting loop mimics the binding
562 mode of SIMs (*56*). On the other hand, 22 distinct affimer proteins that bind to ySUMOs
563 have been generated (*16*). Tiede *et al*. have pointed out that some of them have the sequences

similar to SIMs in their loops, like the monobody ySMB-1. For example, the IDLT sequence in loop 1, and the consensus sequence (W/F/Y)(E/D)$_{2–4}$(W/F/Y) in two loops are found among the ySUMO-binding affimers (*16*). Other motifs, PX$_{1–3}$(N/Q)(W/F/Y) or G(L/I), were also identified in loop 2, in addition to the SIM-related motifs. Despite the similarities in the structural features of the scaffold and two loops randomized in the libraries, our ySUMO-binding scMonellin variants did not contain the consensus sequences identified in the affimers. Therefore, the scMonellin scaffold may be useful in generating the synthetic binding proteins equipped with molecular properties distinct from the affimer proteins.

The scMonellin variants are more unstable than the wild-type of scMonellin, as judged by the thermal denaturation profiles (Table S1 and Fig. S2B). The scMonellin scaffold will have to be stabilized for use in various applications. There is a wealth of information on the folding properties of scMonellin, which is its advantage over affimers in terms of simplicity in stabilization. For example, the structure-guided design of stabilized mutants of scMonellin has been reported (*65*). The mutation sites of these stabilized scMonellin proteins are located outside loop 1 and loop 2. Introducing the stabilizing mutations will provide us with an alternative design of the scMonellin scaffold to, for example, vary the lengths of loop 2.

The scMonellin loop libraries constructed in this study will be useful in engineering monellin proteins, for example, in enhancing the sweetness. The sites on monellin protein that determine sweetness and are involved in the receptor binding have been explored mainly by structure-guided mutagenesis analysis (*66-69*). Although almost all amino acid residues reported to be involved in the sweetness characteristic are located on the convex side of monellin, it has recently reported that the amino acid residues in loop 1 (Arg53) and loop 2 (Arg82) of scMonellin MNEI are important for exhibiting sweetness (*70*). This finding indicates that the two loops 1 and 2 may also affect receptor-binding property and sweetness

589 of scMonellin. Thus, the design and generation of the scMonellin mutants, focusing on loop 1
590 and loop 2, offer promising ways of making the artificial sweet proteins, which can be tested
591 using the libraries generated in this study.
592

593 ## References

594 (1) Conroy, P.J., Law, R.H., Caradoc-Davies, T.T., and Whisstock, J.C. (2017)
595 Antibodies: From novel repertoires to defining and refining the structure of
596 biologically important targets. *Methods*. **116**, 12-22

597 (2) Weisser, N.E., and Hall, J.C. (2009) Applications of single-chain variable fragment
598 antibodies in therapeutics and diagnostics. *Biotechnol Adv*. **27**, 502-520

599 (3) Hudson, P.J., and Souriau, C. (2003) Engineered antibodies. *Nature Medicine*. **9**,
600 129-134

601 (4) Sidhu, S.S., and Koide, S. (2007) Phage display for engineering and analyzing protein
602 interaction interfaces. *Curr Opin Struct Biol*. **17**, 481-487

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

605 (6) Schlehuber, S., Beste, G., and Skerra, A. (2000) A novel type of receptor protein,
606 based on the lipocalin scaffold, with specificity for digoxigenin. *J Mol Biol*. **297**,
607 1105-1120

608 (7) Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., and Pluckthun, A. (2003)
609 Designing repeat proteins: well-expressed, soluble and stable proteins from
610 combinatorial libraries of consensus ankyrin repeat proteins. *J Mol Biol*. **332**, 489-503

(8) Nord, K., Nilsson, J., Nilsson, B., Uhlen, M., and Nygren, P.A. (1995) A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Eng*. **8**, 601-608

(9) Zhao, N., Schmitt, M.A., and Fisk, J.D. (2016) Phage display selection of tight specific binding variants from a hyperthermostable Sso7d scaffold protein library. *Febs j*. **283**, 1351-1367

(10) Tanaka, S., Takahashi, T., Koide, A., Ishihara, S., Koikeda, S., and Koide, S. (2015) Monobody-mediated alteration of enzyme specificity. *Nat Chem Biol*. **11**, 762-764

(11) Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O., and Grutter, M.G. (2007) Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol*. **5**, e7

(12) Stockbridge, R.B., Kolmakova-Partensky, L., Shane, T., Koide, A., Koide, S., Miller, C., and Newstead, S. (2015) Crystal structures of a double-barrelled fluoride ion channel. *Nature*. **525**, 548-551

(13) Liu, Y., Huynh, D.T., and Yeates, T.O. (2019) A 3.8 A resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat Commun*. **10**, 1864

(14) Yasui, N., Findlay, G.M., Gish, G.D., Hsiung, M.S., Huang, J., Tucholska, M., Taylor, L., Smith, L., Boldridge, W.C., Koide, A., Pawson, T., and Koide, S. (2014) Directed network wiring identifies a key protein interaction in embryonic stem cell differentiation. *Mol Cell*. **54**, 1034-1041

(15) Sha, F., Gencer, E.B., Georgeon, S., Koide, A., Yasui, N., Koide, S., and Hantschel, O. (2013) Dissection of the BCR-ABL signaling network using highly specific monobody inhibitors to the SHP2 SH2 domains. *Proc Natl Acad Sci U S A*. **110**, 14924-14929

(16) Tiede, C., Tang, A.A., Deacon, S.E., Mandal, U., Nettleship, J.E., Owen, R.L., George, S.E., Harrison, D.J., Owens, R.J., Tomlinson, D.C., and McPherson, M.J. (2014) Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. *Protein Eng Des Sel*. **27**, 145-155

(17) Michel, M.A., Swatek, K.N., Hospenthal, M.K., and Komander, D. (2017) Ubiquitin Linkage-Specific Affimers Reveal Insights into K6-Linked Ubiquitin Signaling. *Mol Cell*. **68**, 233-246 e235

(18) Robinson, J.I., Baxter, E.W., Owen, R.L., Thomsen, M., Tomlinson, D.C., Waterhouse, M.P., Win, S.J., Nettleship, J.E., Tiede, C., Foster, R.J., Owens, R.J., Fishwick, C.W.G., Harris, S.A., Goldman, A., McPherson, M.J., and Morgan, A.W. (2018) Affimer proteins inhibit immune complex binding to FcgammaRIIIa with high specificity through competitive and allosteric modes of action. *Proc Natl Acad Sci U S A*. **115**, E72-E81

(19) Tiede, C., Bedford, R., Heseltine, S.J., Smith, G., Wijetunga, I., Ross, R., AlQallaf, D., Roberts, A.P., Balls, A., Curd, A., Hughes, R.E., Martin, H., Needham, S.R., Zanetti-Domingues, L.C., Sadigh, Y., Peacock, T.P., Tang, A.A., Gibson, N., Kyle, H., Platt, G.W., Ingram, N., Taylor, T., Coletta, L.P., Manfield, I., Knowles, M., Bell, S.,

Esteves, F., Maqbool, A., Prasad, R.K., Drinkhill, M., Bon, R.S., Patel, V., Goodchild, S.A., Martin-Fernandez, M., Owens, R.J., Nettleship, J.E., Webb, M.E., Harrison, M., Lippiat, J.D., Ponnambalam, S., Peckham, M., Smith, A., Ferrigno, P.K., Johnson, M., McPherson, M.J., and Tomlinson, D.C. (2017) Affimer proteins are versatile and renewable affinity reagents. *Elife*. **6**,

(20) Machleidt, W., Thiele, U., Laber, B., Assfalg-Machleidt, I., Esterl, A., Wiegand, G., Kos, J., Turk, V., and Bode, W. (1989) Mechanism of inhibition of papain by chicken egg white cystatin. Inhibition constants of N-terminally truncated forms and cyanogen bromide fragments of the inhibitor. *FEBS Lett*. **243**, 234-238

(21) Renko, M., Požgan, U., Majera, D., and Turk, D. (2010) Stefin A displaces the occluding loop of cathepsin B only by as much as required to bind to the active site cleft. *Febs j*. **277**, 4338-4345

(22) Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990) The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *Embo j*. **9**, 1939-1947

(23) Morris, J.A., and Cagan, R.H. (1972) Purification of monellin, the sweet principle of Dioscoreophyllum cumminsii. *Biochim Biophys Acta*. **261**, 114-122

(24) Ogata, C., Hatada, M., Tomlinson, G., Shin, W.C., and Kim, S.H. (1987) Crystal structure of the intensely sweet protein monellin. *Nature*. **328**, 739-742

672 (25) Murzin, A.G. (1993) Sweet-tasting protein monellin is related to the cystatin family of
673 thiol proteinase inhibitors. *J Mol Biol*. **230**, 689-694
674 (26) Kim, S.H., Kang, C.H., Kim, R., Cho, J.M., Lee, Y.B., and Lee, T.K. (1989)
675 Redesigning a sweet protein: increased stability and renaturability. *Protein Eng*. **2**,
676 571-575
677 (27) Tancredi, T., Iijima, H., Saviano, G., Amodeo, P., and Temussi, P.A. (1992)
678 Structural determination of the active site of a sweet protein. A 1H NMR
679 investigation of pMNEI. *FEBS Lett*. **310**, 27-30
680 (28) Somoza, J.R., Jiang, F., Tong, L., Kang, C.H., Cho, J.M., and Kim, S.H. (1993) Two
681 crystal structures of a potently sweet protein. Natural monellin at 2.75 A resolution
682 and single-chain monellin at 1.7 A resolution. *J Mol Biol*. **234**, 390-404
683 (29) Spadaccini, R., Crescenzi, O., Tancredi, T., De Casamassimi, N., Saviano, G.,
684 Scognamiglio, R., Di Donato, A., and Temussi, P.A. (2001) Solution structure of a
685 sweet protein: NMR study of MNEI, a single chain monellin. *J Mol Biol*. **305**,
686 505-514
687 (30) Hobbs, J.R., Munger, S.D., and Conn, G.L. (2007) Monellin (MNEI) at 1.15 A
688 resolution. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. **63**, 162-167
689 (31) Konno, T. (2001) Multistep nucleus formation and a separate subunit contribution of
690 the amyloidgenesis of heat-denatured monellin. *Protein Sci*. **10**, 2093-2101

691 (32) Sidhu, S.S., and Weiss, G.A. (2004) Constructing phage display libraries by
692 oligonucleotide-directed mutagenesis. in *Phage Display–A Practical Approach* (T., C.,
693 and H.B., L., ed.)^eds.), pp. 27–41, Oxford University Press, Oxford, UK
694 (33) Wojcik, J., Hantschel, O., Grebien, F., Kaupe, I., Bennett, K.L., Barkinge, J., Jones,
695 R.B., Koide, A., Superti-Furga, G., and Koide, S. (2010) A potent and highly specific
696 FN3 monobody inhibitor of the Abl SH2 domain. *Nat Struct Mol Biol*. **17**, 519-527
697 (34) Rondot, S., Koch, J., Breitling, F., and Dubel, S. (2001) A helper phage to improve
698 single-chain antibody presentation in phage display. *Nat Biotechnol*. **19**, 75-78
699 (35) Beckett, D., Kovaleva, E., and Schatz, P.J. (1999) A minimal peptide substrate in
700 biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci*. **8**, 921-929
701 (36) Nakatani, T., Yasui, N., Tamura, I., and Yamashita, A. (2019) Specific modification
702 at the C-terminal lysine residue of the green fluorescent protein variant, GFPuv,
703 expressed in Escherichia coli. *Scientific Reports*. **9**, 4722
704 (37) Studier, F.W. (2005) Protein production by auto-induction in high density shaking
705 cultures. *Protein Expr Purif*. **41**, 207-234
706 (38) Otwinowski, Z., and Minor, W. (1997) [20] Processing of X-ray diffraction data
707 collected in oscillation mode. *Methods Enzymol*. **276**, 307-326
708 (39) McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and
709 Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*. **40**, 658-674
710 (40) Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and
711 development of Coot. *Acta Crystallogr D Biol Crystallogr*. **66**, 486-501

(41) Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. **66**, 213-221

(42) Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., and Wilson, K.S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. **67**, 235-242

(43) Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. **372**, 774-797

(44) Krissinel, E. (2012) Enhanced fold recognition using efficient short fragment clustering. *J Mol Biochem*. **1**, 76-85

(45) Koide, S., and Sidhu, S.S. (2009) The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins. *ACS Chem Biol*. **4**, 325-334

(46) Crameri, A., Whitehorn, E.A., Tate, E., and Stemmer, W.P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol*. **14**, 315-319

(47) Tsien, R.Y. (1998) The green fluorescent protein. *Annu Rev Biochem*. **67**, 509-544

(48) Zimmer, M. (2002) Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem Rev*. **102**, 759-781

(49) Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994) Green fluorescent protein as a marker for gene expression. *Science*. **263**, 802-805

(50) Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S.D., and Butt, T.R. (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics*. **5**, 75-86

(51) Marblestone, J.G., Edavettal, S.C., Lim, Y., Lim, P., Zuo, X., and Butt, T.R. (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci*. **15**, 182-189

(52) Chevrel, A., Urvoas, A., Li de la Sierra-Gallay, I., Aumont-Nicaise, M., Moutel, S., Desmadril, M., Perez, F., Gautreau, A., van Tilbeurgh, H., Minard, P., and Valerio-Lepiniec, M. (2015) Specific GFP-binding artificial proteins (alphaRep): a new tool for in vitro to live cell applications. *Biosci Rep*. **35**,

(53) Hansen, S., Stuber, J.C., Ernst, P., Koch, A., Bojar, D., Batyuk, A., and Pluckthun, A. (2017) Design and applications of a clamp for Green Fluorescent Protein with picomolar affinity. *Sci Rep*. **7**, 16292

(54) Zhang, Z., Wang, Y., Ding, Y., and Hattori, M. (2020) Structure-based engineering of anti-GFP nanobody tandems as ultra-high-affinity reagents for purification. *Sci Rep*. **10**, 6239

754 (55) Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M.,
755 Muyldermans, S., Casas-Delucchi, C.S., Cardoso, M.C., Leonhardt, H., Hopfner, K.P.,
756 and Rothbauer, U. (2010) Modulation of protein properties in living cells using
757 nanobodies. *Nat Struct Mol Biol*. **17**, 133-138
758 (56) Gilbreth, R.N., Truong, K., Madu, I., Koide, A., Wojcik, J.B., Li, N.S., Piccirilli, J.A.,
759 Chen, Y., and Koide, S. (2011) Isoform-specific monobody inhibitors of small
760 ubiquitin-related modifiers engineered using structure-guided library design. *Proc*
761 *Natl Acad Sci U S A*. **108**, 7751-7756
762 (57) Zheng, W., Yang, L., Cai, C., Ni, J., and Liu, B. (2018) Expression, purification and
763 characterization of a novel double-sites mutant of the single-chain sweet-tasting
764 protein monellin (MNEI) with both improved sweetness and stability. in *Protein Expr*
765 *Purif* ed.)^eds.), pp. 52-56, Elsevier,
766 (58) Lo Conte, L., Chothia, C., and Janin, J. (1999) The atomic structure of protein-protein
767 recognition sites. *J Mol Biol*. **285**, 2177-2198
768 (59) Schmidthals, K., Helma, J., Zolghadr, K., Rothbauer, U., and Leonhardt, H. (2010)
769 Novel antibody derivatives for proteome and high-content analysis. *Analytical and*
770 *Bioanalytical Chemistry*. **397**, 3203-3208
771 (60) Li, S.J., and Hochstrasser, M. (1999) A new protease required for cell-cycle
772 progression in yeast. in *Nature* ed.)^eds.), pp. 246-251, Nature Publishing Group,
773 (61) Tanaka, K., Nishide, J., Okazaki, K., Kato, H., Niwa, O., Nakagawa, T., Matsuda, H.,
774 Kawamukai, M., and Murakami, Y. (1999) Characterization of a fission yeast

775 SUMO-1 homologue, pmt3p, required for multiple nuclear events, including the
776 control of telomere length and chromosome segregation. in *Molecular and Cellular*
777 *Biology* ed.)^eds.), pp. 8660-8672, American Society for Microbiology Journals,
778 (62) Mossessova, E., and Lima, C.D. (2000) Ulp1-SUMO crystal structure and genetic
779 analysis reveal conserved interactions and a regulatory element essential for cell
780 growth in yeast. in *Mol. Cell* ed.)^eds.), pp. 865-876,
781 (63) Song, J., Zhang, Z., Hu, W., and Chen, Y. (2005) Small ubiquitin-like modifier
782 (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. in
783 *J. Biol. Chem.* ed.)^eds.), pp. 40122-40129, American Society for Biochemistry and
784 Molecular Biology,
785 (64) Sekiyama, N., Ikegami, T., Yamane, T., Ikeguchi, M., Uchimura, Y., Baba, D.,
786 Ariyoshi, M., Tochio, H., Saitoh, H., and Shirakawa, M. (2008) Structure of the small
787 ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing
788 chromatin-associated factor 1 bound to SUMO-3. in *J. Biol. Chem.* ed.)^eds.), pp.
789 35966-35975, American Society for Biochemistry and Molecular Biology,
790 (65) Leone, S., Pica, A., Merlino, A., Sannino, F., Temussi, P.A., and Picone, D. (2016)
791 Sweeter and stronger: enhancing sweetness and stability of the single chain monellin
792 MNEI through molecular design. *Sci Rep.* **6**, 34045
793 (66) Sung, Y.H. (2001) Solution Structure, Backbone Dynamics, and Stability of a Double
794 Mutant Single-chain Monellin. STRUCTURAL ORIGIN OF SWEETNESS. in *J.*
795 *Biol. Chem.* ed.)^eds.), pp. 19624-19630,

(67) Somoza, J.R., Cho, J.M., and Kim, S.H. (1995) The taste-active regions of monellin, a potently sweet protein. *Chem Senses*. **20**, 61-68

(68) Esposito, V., Gallucci, R., Picone, D., Saviano, G., Tancredi, T., and Temussi, P.A. (2006) The Importance of Electrostatic Potential in The Interaction of Sweet Proteins with the Sweet Taste Receptor. in *Journal of Molecular Biology* ed.)^eds.), pp. 448-456,

(69) Templeton, C.M., Ostovar pour, S., Hobbs, J.R., Blanch, E.W., Munger, S.D., and Conn, G.L. (2011) Reduced Sweetness of a Monellin (MNEI) Mutant Results from Increased Protein Flexibility and Disruption of a Distant Poly-(L-Proline) II Helix. in *Chemical Senses* ed.)^eds.), pp. 425-434,

(70) Yang, L., Zhu, K., Yu, H., Zhang, X., and Liu, B. (2019) The Flexible Loop is a New Sweetness Determinant Site of the Sweet-Tasting Protein: Characterization of Novel Sweeter Mutants of the Single-Chain Monellin (MNEI). *Chem Senses*. **44**, 607-614

Figure legends

Figure 1. Design of scMonellin phage display libraries. (A) Amino acid and coding sequences of the DsbA signal sequence-scMinellin-V5 tag segment in the phagemid vector constructed in this study. The secondary structure elements are indicated below the amino acid sequence. The residues that were randomized in loops 1 and 2 are highlighted in magenta. The Tyr residue in loop 2 that was not diversified is boxed. EcoRI and MluI sites introduced into the portions coding loop 1 and loop 2 are indicated. **(B)** The crystal structure of scMonellin (PDB ID: 2O9U) shown in two views rotated 90 ˚ about the y-axis. Two loops diversified in the combinatorial library are shown in magenta. The secondary structure elements are labeled according to Hobbs *et al*. (*30*). According to the shape of the β-sheet, the side on which the α-helix is located and the opposite side are called the concave and convex sides, respectively (*30*). **(C)** Design of the scMonellin libraries. Loop lengths were fixed in loop library A but were varied in loop library B. X refers to a mixture of codons with compositions containing 30% tyrosine, 15% serine, 10% glycine, 5% each of tryptophan, phenylalanine, and 2.5% each of all other residues, except cysteine.

Figure 2. Phage ELISA for screening the phage clones displaying target-binding scMonellin variants. Clones isolated by sorting loop library A against GFPuv **(A)** and ySUMO **(B)**. Clones isolated by sorting loop library B against GFPuv **(C)** and ySUMO **(D)**. Phage clones were incubated in wells coated with Tris-buffered saline (TBS; cyan), NeutrAvidin (NAV; magenta), biotinylated target proteins via NeutrAvidin (GFPuv/NAV or ySUMO/NAV; green) or anti-V5 monoclonal antibody (anti-V5; gray). Bound phages were

833 detected using an HRP-conjugated antibody against the M13 phage. The absorbance of the
834 ABTS product at 405 nm is shown for each clone.
835
836 **Figure 3. The target-binding scMonellin variants isolated by sorting loop library A. (A)**
837 Loop sequences of the target-binding scMonellin variants derived from loop library A. Acidic
838 residues and basic residues are highlighted in red and blue, respectively. The amino acid
839 residues that were biased in the library are highlighted as follows: Tyr and Ser are in
840 magenta; Phe, Trp, and Gly are in green. **(B)** Surface plasmon resonance analysis of the
841 GFPuv binding of scMonellin wild-type (WT) (*left*) and a scMonellin variant, GFP-40 (*right*).
842 Responses at equilibrium (R_{eq}) values calculated from the sensorgram (inset) were plotted
843 against the concentration of injected GFPuv to calculate the dissociation constants (K_D). **(C)**
844 Surface plasmon resonance analysis of the interaction between ySUMO and SUMO-31. The
845 R_{eq} values calculated from the sensorgram (inset) were plotted against the concentration of
846 injected ySUMO to estimate the K_D value. The errors indicated in panels B and C are the
847 95% confidence intervals for fitting.
848
849 **Figure 4. The target-binding scMonellin variants isolated by sorting loop library B. (A)**
850 Loop sequences of the target-binding scMonellin variants derived from loop library B. **(B, C)**
851 Surface plasmon resonance analysis of the interaction of the scMonellin variants with GFPuv
852 **(B)** and ySUMO **(C)**. The R_{eq} values calculated from the sensorgram (inset) were plotted
853 against the concentration of injected target protein to estimate the K_D values. The errors
854 indicated are the 95% confidence intervals for fitting.
855
856 **Figure 5. Size-exclusion chromatograms of the scMonellin variants.** The scMonellin
857 variants and wild-type (WT) purified from *Escherichia coli* (20 μM) were subjected to size

858 exclusion chromatography on an Enrich SEC 70 10×300 column, equilibrated with 20 mM
859 Tris-HCl, 150 mM NaCl, pH 7.5. The chromatographs are shown with vertical offsets. The
860 labels indicate the analyzed scMonellin variants. The void volume (V_0) and elution positions
861 for bovine serum albumin (molecular mass: 67 kDa), ovalbumin (44 kDa), carbonic
862 anhydrase (29 kDa), myoglobin (17 kDa), cytochrome *c* (12.4 kDa), and vitamin B_{12} (1.35
863 kDa) are indicated by vertical lines.
864
865 **Figure 6. Crystal structure of the SWEEPin GFP-40/GFPuv complex. (A)** Overall
866 structure of the GFP-40:GFPuv complex. GFP-40 and GFPuv polypeptide chains are
867 illustrated by a ribbon model (GFP-40 is colored in cyan and GFPuv in green) with a
868 translucent molecular surface. Two loops, loop 1 and loop 2, are colored in magenta. The N-
869 and C- termini of GFP-40 are shown in spheres and labeled. **(B)** The binding interface
870 between GFP-40 and GFPuv. Key interacting residues are shown as stick models, and labeled
871 with underlined and italicized for those of GFP-40 and GFPuv, respectively. Hydrogen bonds
872 are indicated by black dashed lines. **(C)** Water-mediated hydrogen bond network found at the
873 binding interface. Water molecules are shown as red spheres. The hydrogen bonds that
874 involve water molecules are indicated by yellow dashed lines.
875
876 **Figure 7. Comparison of the GFP-binding modes among the GFP binders. (A)** The
877 crystal structures of the GFP-binding proteins in complex with GFP or its variants are
878 superimposed on the GFPuv molecule of the SWEEPin GFP-40/GFPuv complex (*upper left*).
879 **(B)** Comparison of binding sites on GFP molecules of GFP-40 and α-Rep. The EGFP
880 (enhanced green fluorescent protein) and GFPuv in the crystal structure of the α-Rep/EGFP
881 (*upper*) and the SWEEPin GFP-40/GFPuv (*lower*) complexes are shown as surface models
882 with the same orientation as in panel A (*left*). The surfaces in contact with α-Rep and GFP-40,

within 4 Å, are shown in magenta and green, respectively, and those in contact with both proteins are in blue.

Supplementary Material

Supplementary Tables

Table S1. Sequence of oligonucleotides used for library construction.

Supplementary Figures

Figure S1. Size-exclusion chromatography analysis of the target binding for the scMonellin variants.

Figure S2. Thermal stability of the scMonellin variants.

Figure S3. Structural analyses of the complexes of SWEEPin GFP-40 with GFPuv determined in this study.

Acknowledgments

We thank Mr. Hiroki Maruhashi for help in preparing the purified scMonellin protein and in the preliminary experiments regarding phage display; Dr. Motoyuki Hattori at Fudan University for sharing the PDB data with us prior to data release; Dr. Kazuya Hasegawa and Dr. Nobuhiro Mizuno at SPring-8 BL41XU for X-ray diffraction data collection support. The authors would like to thank Enago (www.enago.jp) for the English language review. The synchrotron radiation experiments were performed at the BL41XU of SPring-8, with approvals of the Japan Synchrotron Radiation Research Institute (JASRI) (Proposal No. 2018A2534). This work was performed in part under the Cooperative Research Program of Institute for Protein Research, Osaka University, CR-18-05 and CR-19-05.

Funding

908 This work was supported by Japan Society for the Promotion of Science (JSPS)
909 KAKENHI Grants 15H05370 and 19H02841 to NY and 17H03644 to AY.

910

911 ## Conflict of interest

912 The authors declare no potential conflicts of interest with respect to the research,
913 authorship, and/or publication of this article.

914

Tables

Table 1. Thermal stability of scMonellin and its variants.

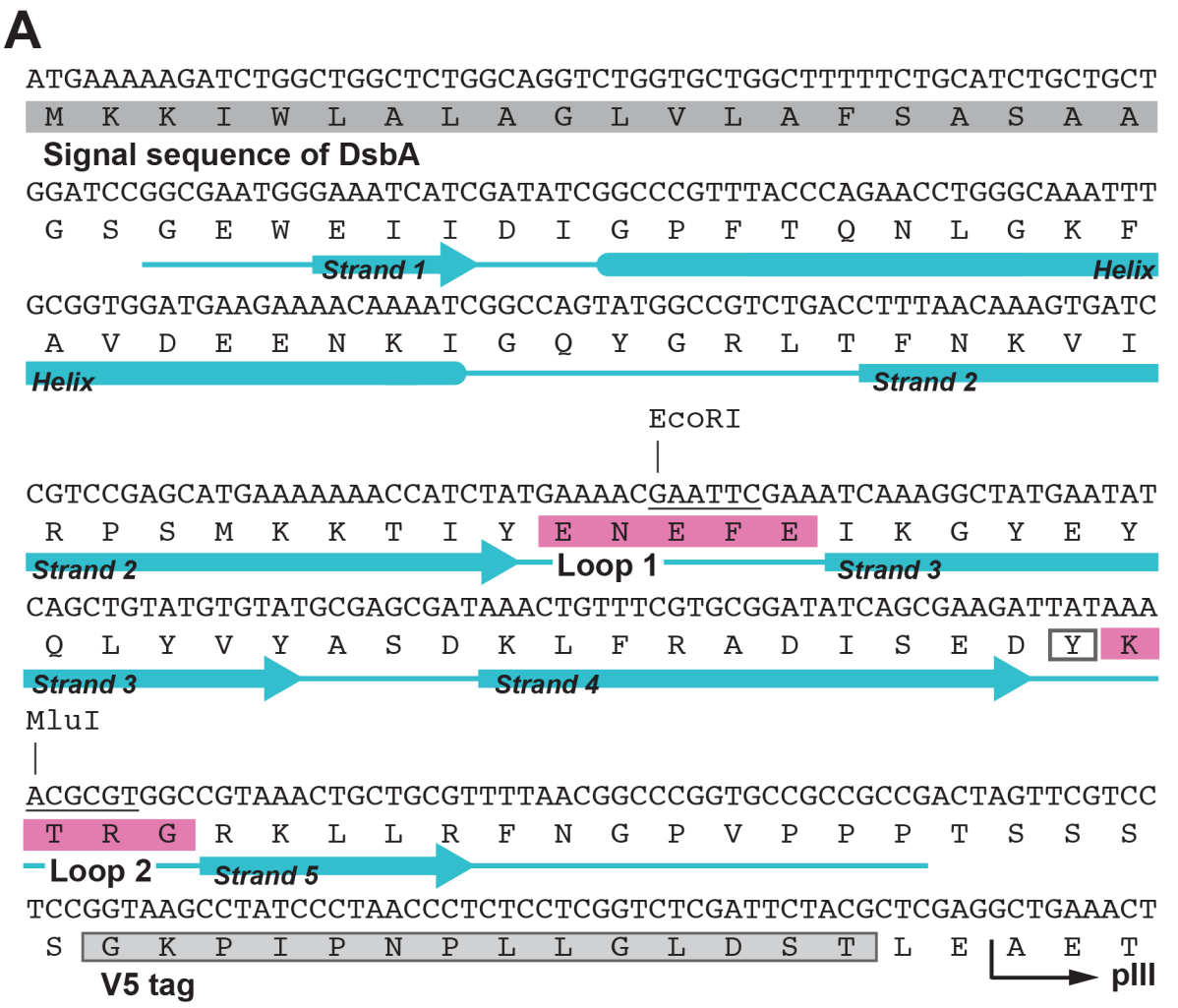
Protein	T_m (°C) *
scMonellin WT	74.2 ± 0.03
GFP-40	63.5 ± 0.11
SUMO-31	59.9 ± 0.12
GFP-kz02	70.5 ± 0.13
GFP-kz06	69.5 ± 0.09
GFP-kz09	63.8 ± 0.05
SUMO-kz03	64.6 ± 0.05
SUMO-kz11	61.1 ± 0.09

*Average ± s.e.m., $n = 4$.

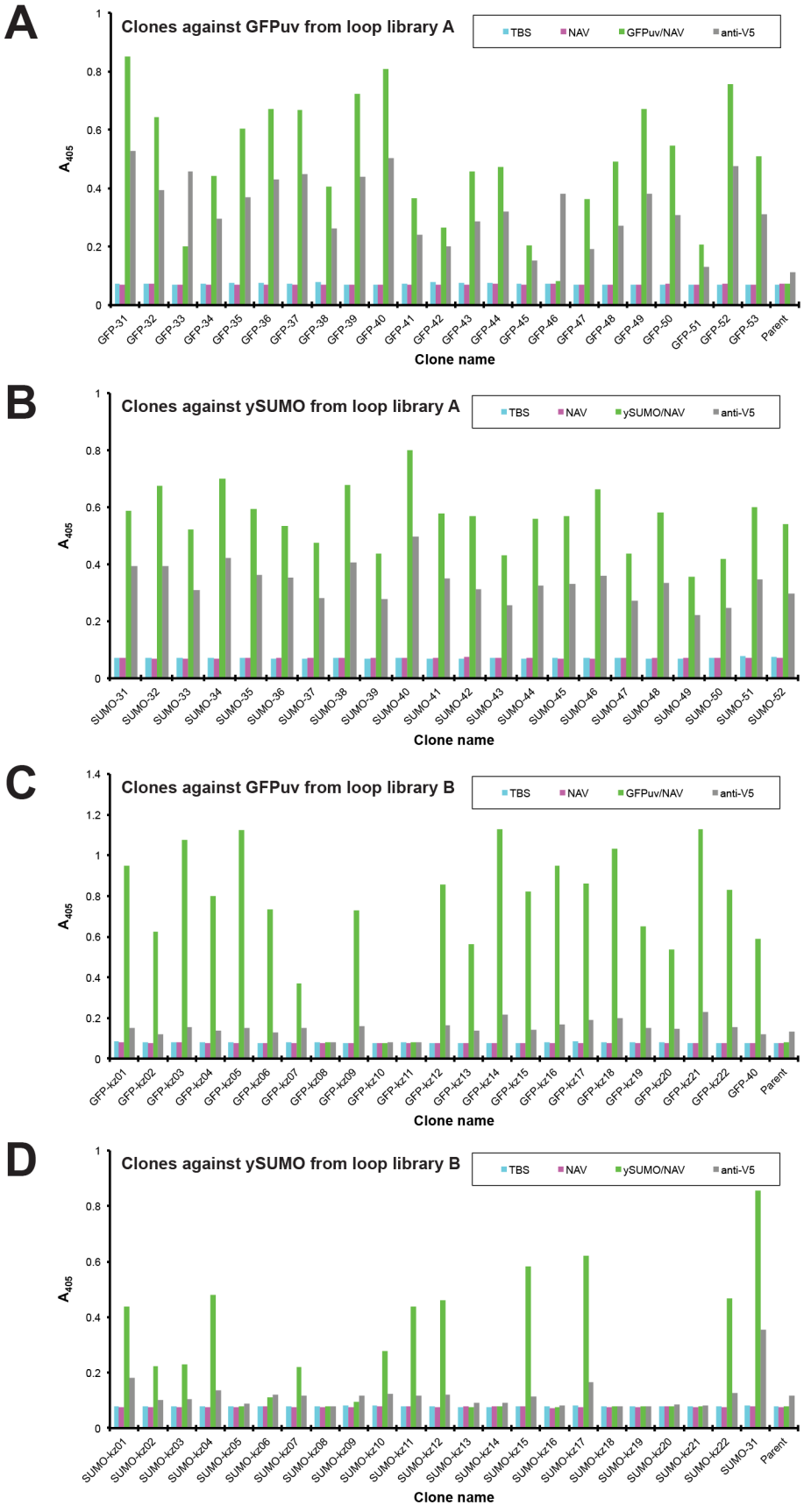
920 **Table 2. Data collection and refinement statistics.**

	Crystal form I (PDB 7CD7)	Crystal form II (PDB 7CD8)
Data collection*		
Space group	$P2_1$	$P2_12_12_1$
Unit cell		
a, *b*, *c* (Å)	54.7, 103.2, 67.9	51.4, 68.0, 115.9
β (˚)	106.3	90
No. of complexes/a.s.u	2	1
X-ray source	SPring-8 BL41XU	SPring-8 BL41XU
Wavelength (Å)	1.0	1.0
Resolution (Å)	50.00–1.70 (1.73–1.70)	50.00–2.0 (2.03–2.00)
Total reflections	249657	171132
Unique reflections	78215 (3848)	28608 (1403)
Completeness (%)	99.5 (99.1)	100 (100)
Redundancy	3.2 (2.9)	6.0 (6.2)
R_{sym}	0.066 (0.371)	0.074 (0.476)
$I/\sigma I$	24.5 (2.1)	22.3 (2.7)
$CC_{1/2}$	0.990 (0.838)	0.998 (0.911)
Refinement		
Resolution (Å)	47.98–1.70	46.98–2.0
Reflections used		
Working set/test set	78166/3863	28550/1389
R_{work}	0.182	0.178
R_{free}	0.226	0.215
Number of atoms	6274	2941
Protein	5370	2622
Ligands	42	21
Water	862	298
Average B-factor (Å^2)	28.0	31.7
macromolecules	26.6	31.0
ligands	19.8	20.3
solvent	37.5	38.7
Rmsd from ideality		
Bond length (Å)	0.011	0.007
Bond angles (˚)	1.30	1.06
Ramachandran plot		
Favored (%)	98.5	98.1
Outliers (%)	0	0

921 Statistics for the highest-resolution shell are shown in parentheses.


A
ATGAAAAAGATCTGGCTGGCTCTGGCAGGTCTGGTGCTGGCTTTTTCTGCATCTGCTGCT
M K K I W L A L A G L V L A F S A S A A
Signal sequence of DsbA
GGATCCGGCGAATGGGAAATCATCGATATCGGCCCGTTTACCCAGAACCTGGGCAAATTT
G S G E W E I I D I G P F T Q N L G K F
Strand 1
Helix
GCGGTGGATGAAGAAAACAAAATCGGCCAGTATGGCCGTCTGACCTTTAACAAAGTGATC
A V D E E N K I G Q Y G R L T F N K V I
Helix
Strand 2
EcoRI
CGTCCGAGCATGAAAAAAACCATCTATGAAAACGAATTCGAAATCAAAGGCTATGAATAT
R P S M K K T I Y E N E F E I K G Y E Y
Strand 2
Loop 1
Strand 3
CAGCTGTATGTGTATGCGAGCGATAAACTGTTTCGTGCGGATATCAGCGAAGATTATAAA
Q L Y V Y A S D K L F R A D I S E D Y K
Strand 3
Strand 4
MluI
ACGCGTGGCCGTAAACTGCTGCGTTTTAACGGCCCGGTGCCGCCGCCGACTAGTTCGTCC
T R G R K L L R F N G P V P P P T S S S
Loop 2
Strand 5
TCCGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCTCGAGGCTGAAACT
S G K P I P N P L L G L D S T L E A E T
V5 tag
pIII
B
Loop 1
Loop 2
N
β5 β4 β3 β2 β1
α
C
90°
Loop 1
Loop 2
N
Concave side
Convex side
C
C
Library name | Loop 1 | Loop 2
Loop library A | X7 | X5
Loop library B | X5-10 | X5-6
X: Tyrosine 30%
Serine 15%
Glycine 10%
Tryptophan 5%
Phenylalanine 5%
Others except cysteine 2.5% each

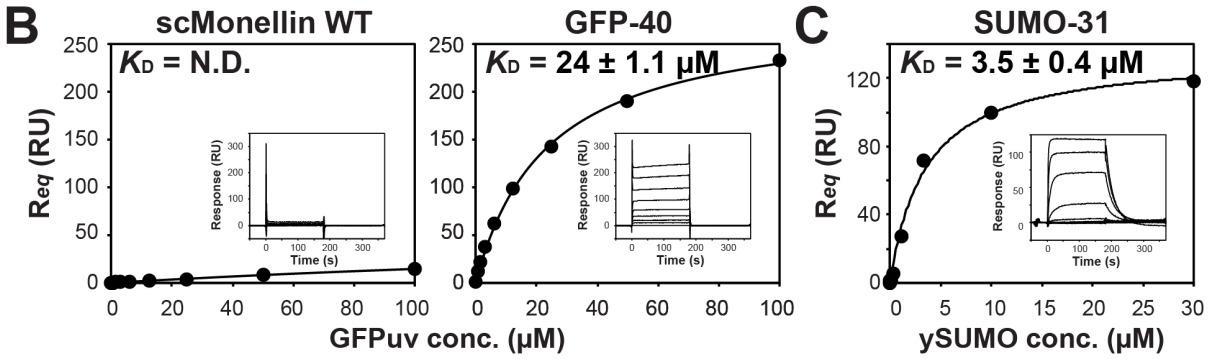
922 **Figure 1.**


A
Clones against GFPuv from loop library A
TBS
NAV
GFPuv/NAV
anti-V5
A405
Clone name
B
Clones against ySUMO from loop library A
TBS
NAV
ySUMO/NAV
anti-V5
A405
Clone name
C
Clones against GFPuv from loop library B
TBS
NAV
GFPuv/NAV
anti-V5
A405
Clone name
D
Clones against ySUMO from loop library B
TBS
NAV
ySUMO/NAV
anti-V5
A405
Clone name

923 **Figure 2.**

A

Target	Clone	K_D (μM)	Amino acid sequence													Frequency	Other redundant clones
			Loop 1 (X_7)							Loop 2 (X_5)							
	Library		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_1	X_2	X_3	X_4	X_5		
	Parent		E	N	E	F	E	–	–	K	T	R	G	–		
GFPuv	GFP-32		S	D	S	F	I	S	Y	S	G	G	S	–	9/20	GFP-35,37,39,43,44,48,49,52
GFPuv	GFP-33		D	D	K	F	Y	G	Y	Y	N	L	G	Q	1/20	
GFPuv	GFP-34		F	K	G	G	Q	T	G	W	P	H	Q	A	6/20	GFP-36,38,42,45,53
GFPuv	GFP-40	24 ± 1.1	Y	A	N	S	R	G	M	Q	H	Y	P	H	2/20	GFP-50
GFPuv	GFP-47		A	T	D	I	R	F	T	K	T	R	G	–	1/20	
GFPuv	GFP-51		Y	D	S	A	N	S	R	Q	I	Y	P	Y	1/20	
ySUMO	SUMO-31	3.5 ± 0.4	D	D	Y	Y	V	D	R	W	S	N	G	Y	22/22	SUMO-32~52

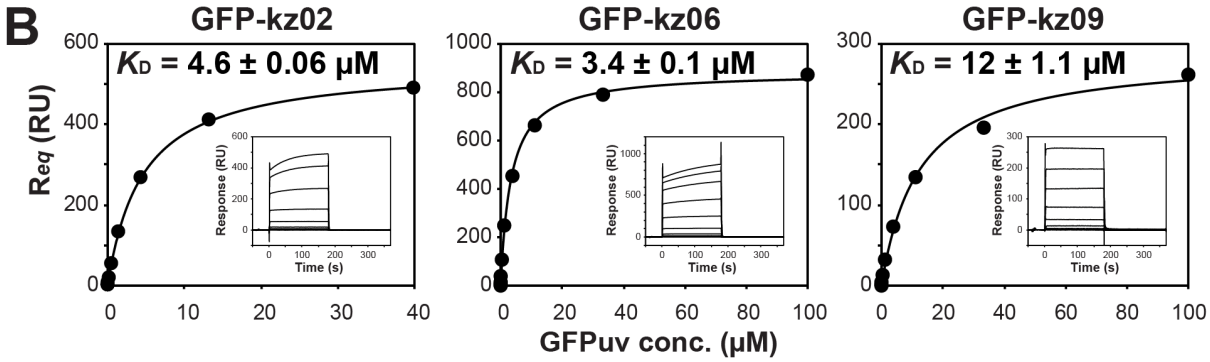


924 **Figure 3.**

A

Target	Clone	K_D (μM)	Amino acid sequence																Frequency	Other redundant clones
			Loop 1 (X_{5-10})										Loop 2 (X_{5-6})							
	Library		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_1	X_2	X_3	X_4	X_5	X_6		
	Parent		E	N	E	F	E	–	–	–	–	–	K	T	R	G	–	–		
GFPuv	GFP-kz02	4.6 ± 0.06	Y	K	N	G	T	–	–	–	–	–	Q	T	Y	P	G	–	6/15	GFP-kz04,05,16,19,22
GFPuv	GFP-kz03		Y	K	N	G	S	–	–	–	–	–	Q	N	Y	P	Y	–	3/15	GFP-kz12,21
GFPuv	GFP-kz06	3.4 ± 0.1	Y	S	N	G	S	–	–	–	–	–	Q	H	W	P	S	–	2/15	GFP-kz15
GFPuv	GFP-kz09	12 ± 1.1	K	S	D	R	H	D	–	–	–	–	Y	Y	G	S	T	Y	1/15	
GFPuv	GFP-kz14		S	Q	F	G	S	W	–	–	–	–	Y	S	G	Y	–	–	3/15	GFP-kz17,20
ySUMO	SUMO-kz01		G	V	G	D	G	S	Y	–	–	–	Y	E	N	Y	S	V	1/12	
ySUMO	SUMO-kz02		Y	Y	D	L	E	N	E	V	E	Y	S	Y	V	P	N	–	1/12	
ySUMO	SUMO-kz03	0.9 ± 0.02	E	Y	V	D	E	D	V	S	I	–	G	T	I	V	S	–	1/12	
ySUMO	SUMO-kz04		T	Y	D	V	D	S	G	V	Y	E	S	Y	N	K	A	Y	1/12	
ySUMO	SUMO-kz06		S	H	Y	Y	G	N	G	E	G	E	N	Y	Q	S	–	–	1/12	
ySUMO	SUMO-kz09		G	Y	D	V	D	E	H	D	H	Y	Y	H	G	T	N	–	1/12	
ySUMO	SUMO-kz10		E	G	Y	D	K	Y	G	E	–	–	S	G	N	Y	V	D	1/12	
ySUMO	SUMO-kz11	1.0 ± 0.02	V	Y	D	V	D	D	G	S	Y	E	K	Y	S	G	S	–	1/12	
ySUMO	SUMO-kz12		R	F	H	V	D	E	N	S	M	E	E	Y	G	Y	T	–	2/12	SUMO-kz22
ySUMO	SUMO-kz15		Y	V	D	D	D	E	E	D	–	–	T	Y	G	G	G	–	2/12	SUMO-kz17

B



C

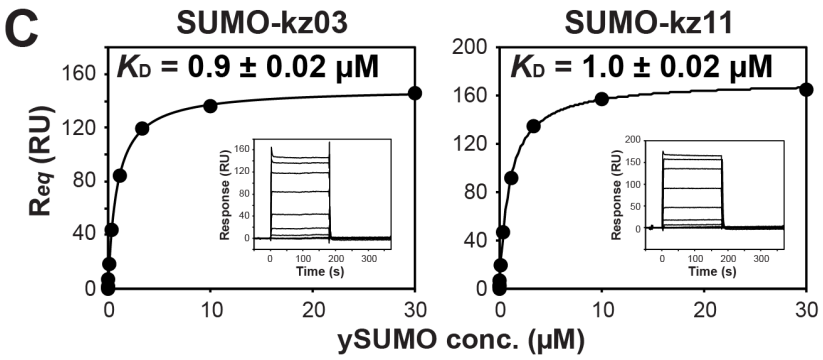
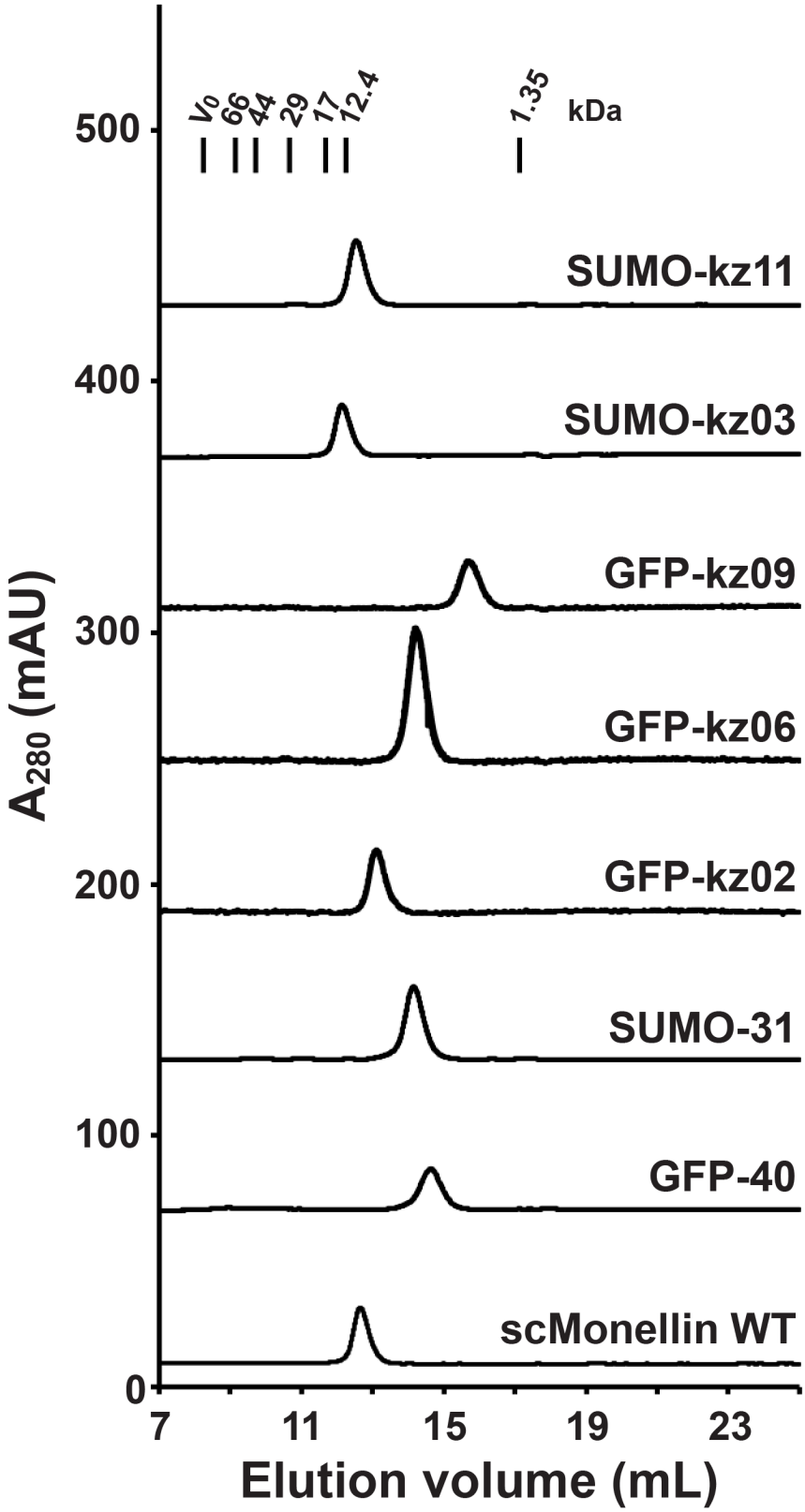
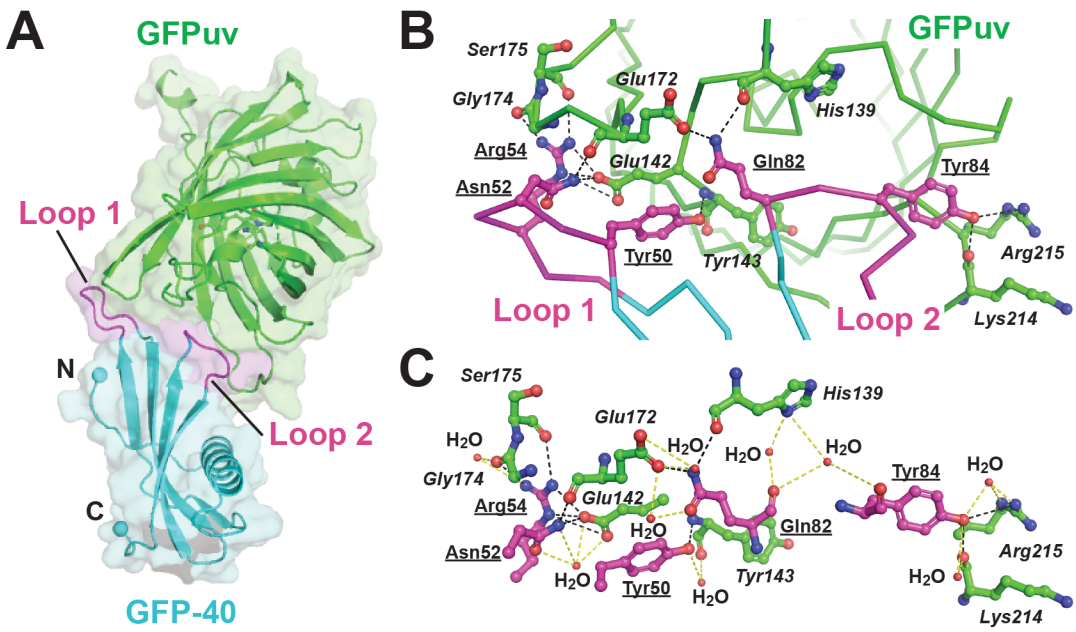


Figure 4.



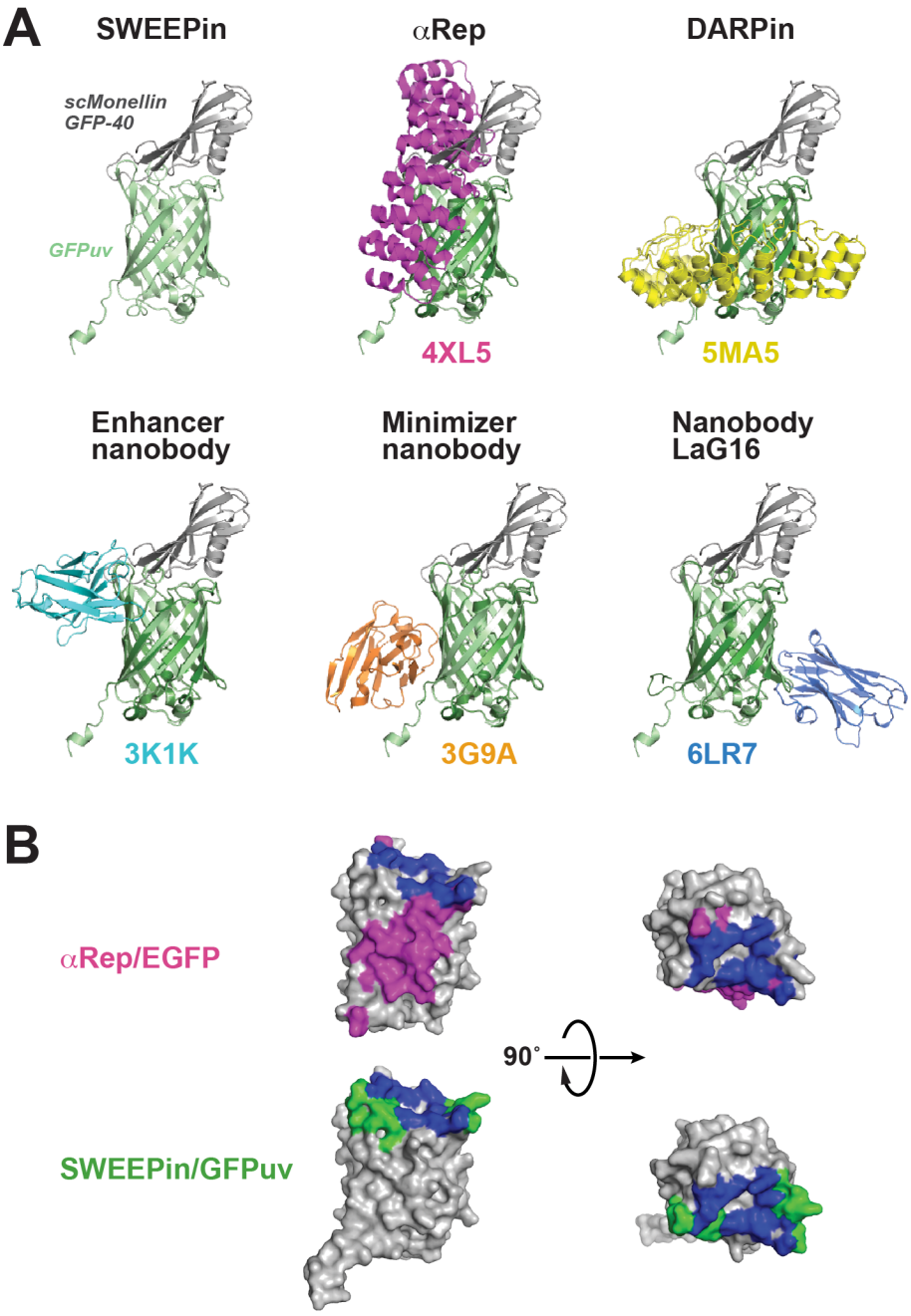
927 **Figure 5.**

928



929 **Figure 6.**

930


A
SWEEPin
αRep
DARPin
scMonellin
GFP-40
GFPuv
4XL5
5MA5
Enhancer
nanobody
Minimizer
nanobody
Nanobody
LaG16
3K1K
3G9A
6LR7
B
αRep/EGFP
90°
SWEEPin/GFPuv

931 **Figure 7.**

Supplementary Information for:

A sweet protein monellin as a non-antibody scaffold for synthetic binding proteins

Norihisa Yasui, Kazuaki Nakamura, Atsuko Yamashita

Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 1-1-1, Tsushima-naka, Kita-ku, Okayama, 700-8530, Japan

Corresponding Author

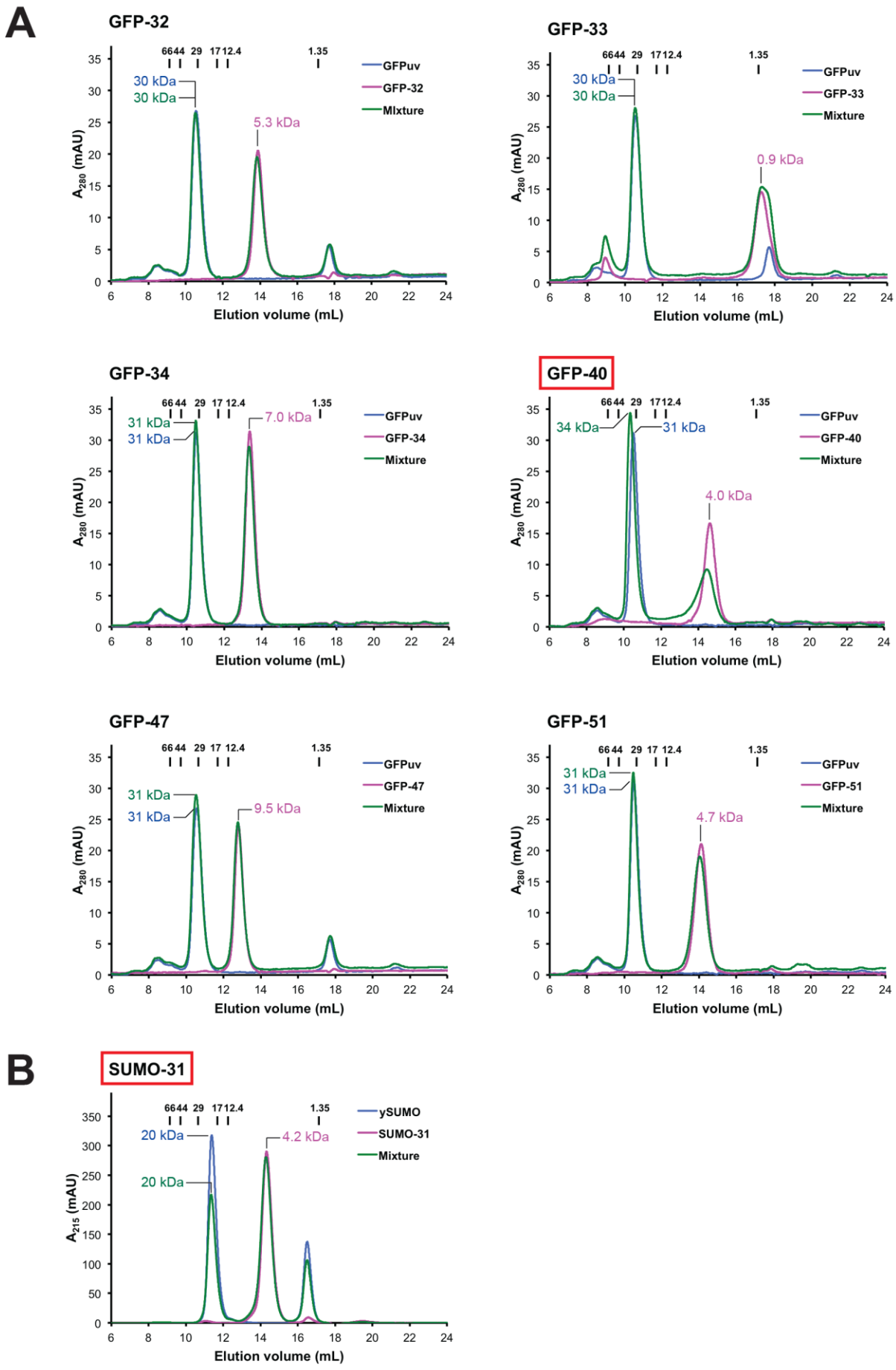
Norihisa Yasui
E-mail: nyasui@okayama-u.ac.jp

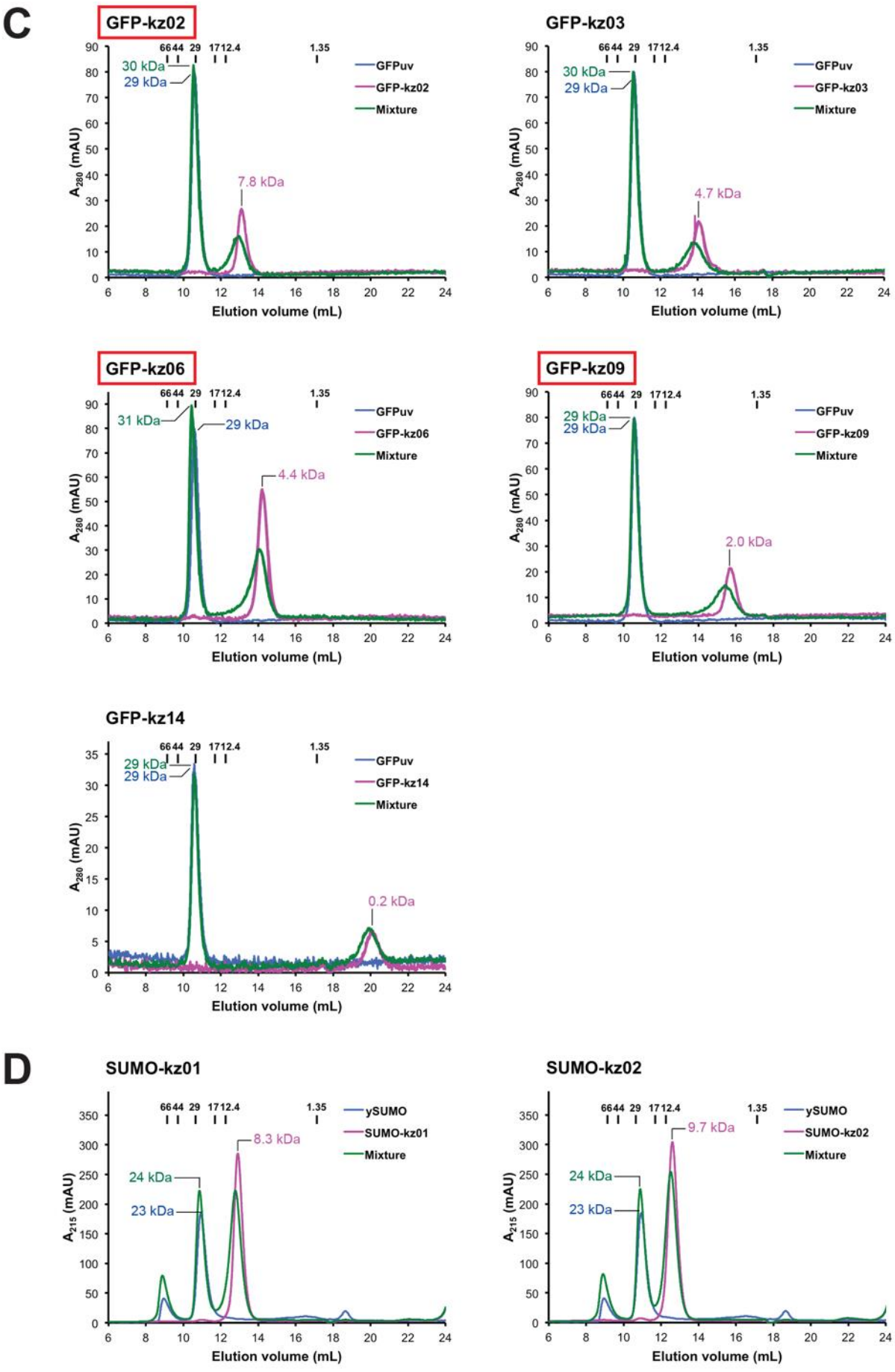
CONTENTS

Table S1. Sequence of oligonucleotides used for library construction

Name	Sequence (5' to 3')*
scMonellin_L1_Tri5	GAAAAAAACCATCTATNNNNNNNNNNNNNNNATCAAAGGCTATGAA
scMonellin_L1_Tri6	GAAAAAAACCATCTATNNNNNNNNNNNNNNNNNNATCAAAGGCTATGAA
scMonellin_L1_Tri7	GAAAAAAACCATCTATNNNNNNNNNNNNNNNNNNNNNATCAAAGGCTATGAA
scMonellin_L1_Tri8	GAAAAAAACCATCTATNNNNNNNNNNNNNNNNNNNNNNNNATCAAAGGCTATGAA
scMonellin_L1_Tri9	GAAAAAAACCATCTATNNNNNNNNNNNNNNNNNNNNNNNNNNNATCAAAGGCTATGAA
scMonellin_L1_Tri10	GAAAAAAACCATCTATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATCAAAGGCTATGAA
scMonellin_L2_Tri5	ATCAGCGAAGATTATNNNNNNNNNNNNNNNCGTAAACTGCTGCGT
scMonellin_L2_Tri6	ATCAGCGAAGATTATNNNNNNNNNNNNNNNNNNCGTAAACTGCTGCGT

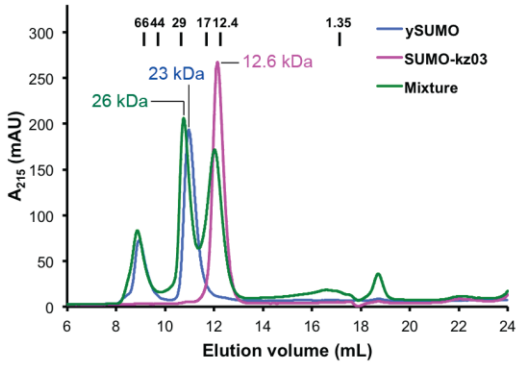
*NNN indicates the mixture of codons encoding a biased amino acid residue composition that includes Tyr (30%) Ser (15%), Gly (10%), Trp (5%), Phe (5%) and 2.5% of each of the other amino acids except for Cys. Codons for Cys residue and termination are excluded.


A
GFP-32
GFP-33
GFP-34
GFP-40
GFP-47
GFP-51
B
SUMO-31
66 44 29 17 12.4 1.35
GFPuv
GFP-32
Mixture
GFP-33
GFP-34
GFP-40
GFP-47
GFP-51
ySUMO
SUMO-31
30 kDa
31 kDa
20 kDa
5.3 kDa
0.9 kDa
7.0 kDa
4.0 kDa
9.5 kDa
4.7 kDa
4.2 kDa
34 kDa
A280 (mAU)
A215 (mAU)
Elution volume (mL)

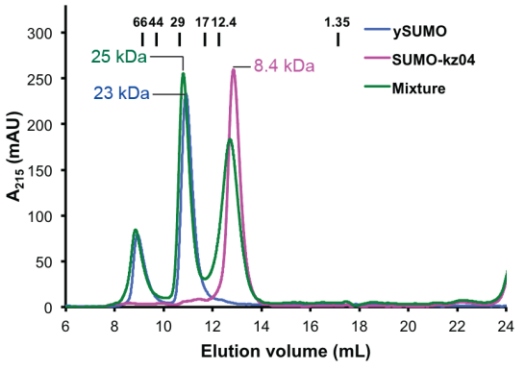

C
GFP-kz02
66 44 29 17 12.4 1.35
30 kDa
29 kDa
7.8 kDa
GFPuv
GFP-kz02
Mixture
A280 (mAU)
Elution volume (mL)
GFP-kz03
30 kDa
29 kDa
4.7 kDa
GFPuv
GFP-kz03
Mixture
GFP-kz06
31 kDa
29 kDa
4.4 kDa
GFPuv
GFP-kz06
Mixture
GFP-kz09
29 kDa
29 kDa
2.0 kDa
GFPuv
GFP-kz09
Mixture
GFP-kz14
29 kDa
29 kDa
0.2 kDa
GFPuv
GFP-kz14
Mixture
D
SUMO-kz01
8.3 kDa
24 kDa
23 kDa
ySUMO
SUMO-kz01
Mixture
A215 (mAU)
SUMO-kz02
9.7 kDa
24 kDa
23 kDa
ySUMO
SUMO-kz02
Mixture

D (continued)

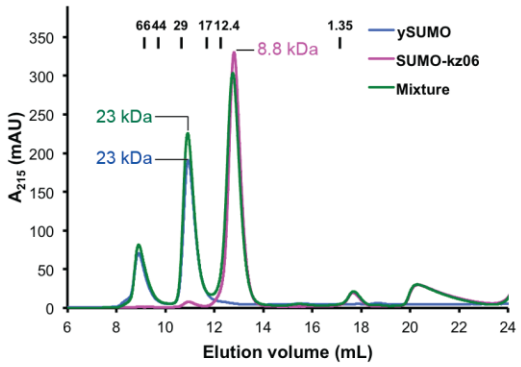
SUMO-kz03



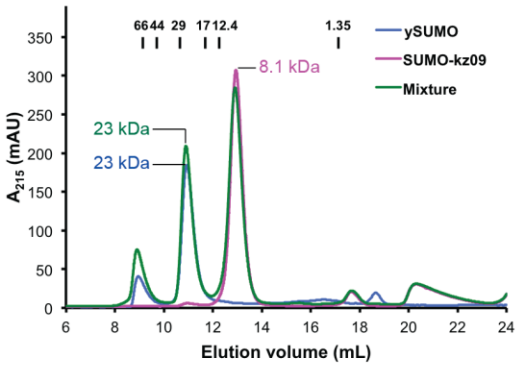
SUMO-kz04



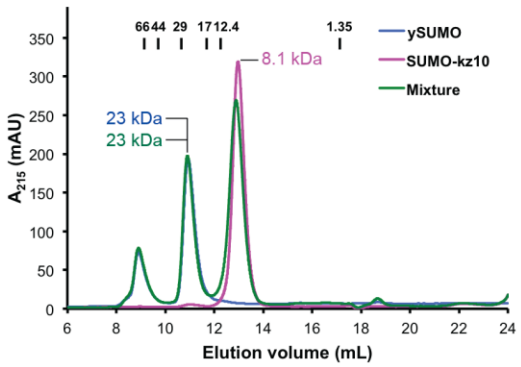
SUMO-kz06



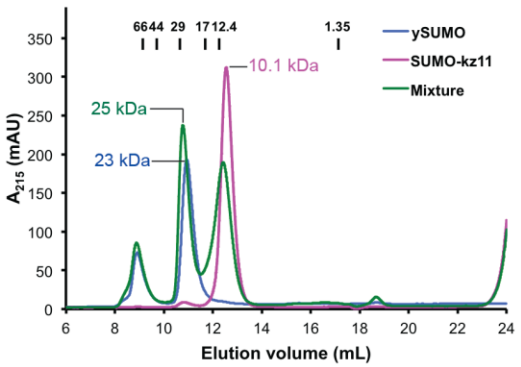
SUMO-kz09



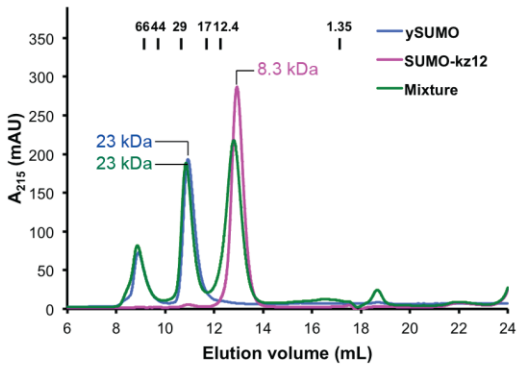
SUMO-kz10



SUMO-kz11



SUMO-kz12



SUMO-kz15

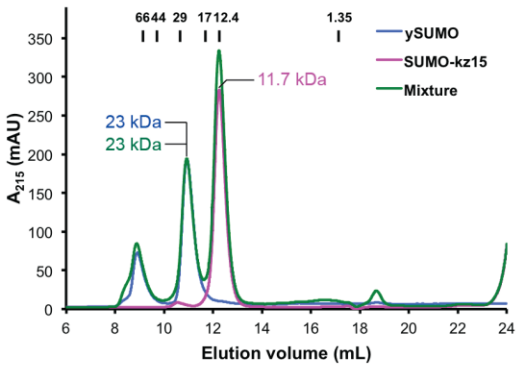


Figure S1. Size-exclusion chromatography analysis of the target binding for the scMonellin variants. Purified target protein (GFPuv or ySUMO) alone, scMonellin variant alone, or the mixture of them were analyzed by size-exclusion chromatography on an ENrich SEC 70 10 × 300 column, equilibrated with 20 mM Tris-HCl, 150 mM NaCl, pH 7.5. The overlaid chromatograms of the target protein alone (blue line), scMonellin variant alone (magenta line), and the mixture of them (green line) are shown for **(A)** the scMonellin variants targeted to GFPuv from loop library A, **(B)** the ySUMO-targeted variant, SUMO-31, from loop library A, **(C)** the scMonellin variants targeted to GFPuv from loop library B, and **(D)** the scMonellin variants targeted to ySUMO from loop library B. For the analyses of ySUMO-targeted scMonellin variants, the absorbance at 215 nm (A_{215}) was recorded because the ySUMO protein used in this study contained no tryptophan residue. The scMonellin variants further analyzed by surface plasmon resonance measurements are labeled with red boxes. The elution positions for bovine serum albumin (molecular mass: 67 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), myoglobin (17 kDa), cytochrome *c* (12.4 kDa), and vitamin B_{12} (1.35 kDa) are indicated by vertical lines on each panel. Relative molecular mass for the peaks corresponding to target protein alone, scMonellin variant alone, and the complex of them are estimated from the elution positions for standard proteins.

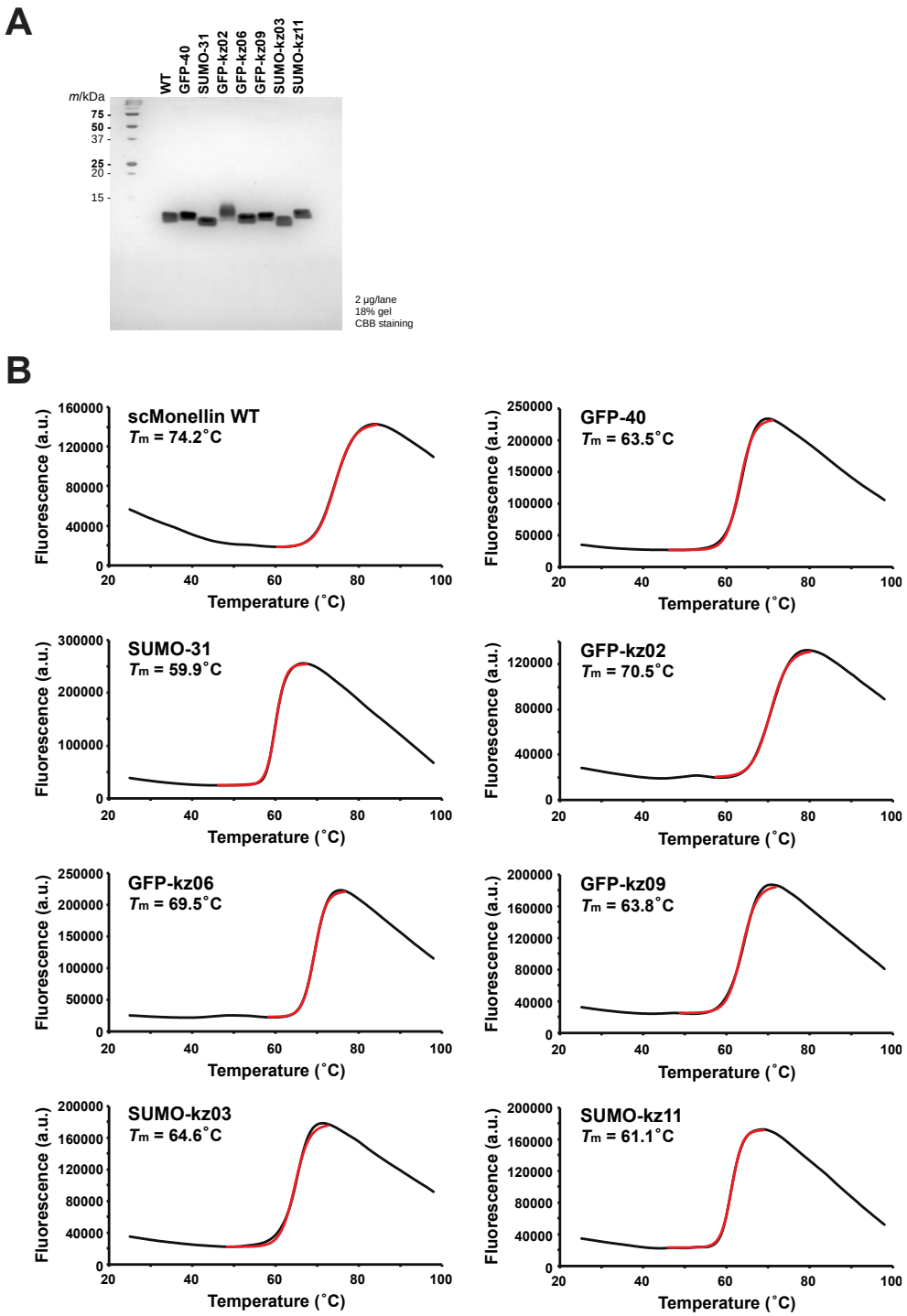


Figure S2. Thermal stability of the scMonellin variants. (A) SDS-PAGE analysis of the purified protein samples used in differential scanning fluorimetry. **(B)** Thermal stability of the scMonellin variants was analyzed by differential scanning fluorimetry, where thermal melt curves are shown in black, and the fitting of the two-state Boltzmann model in red.

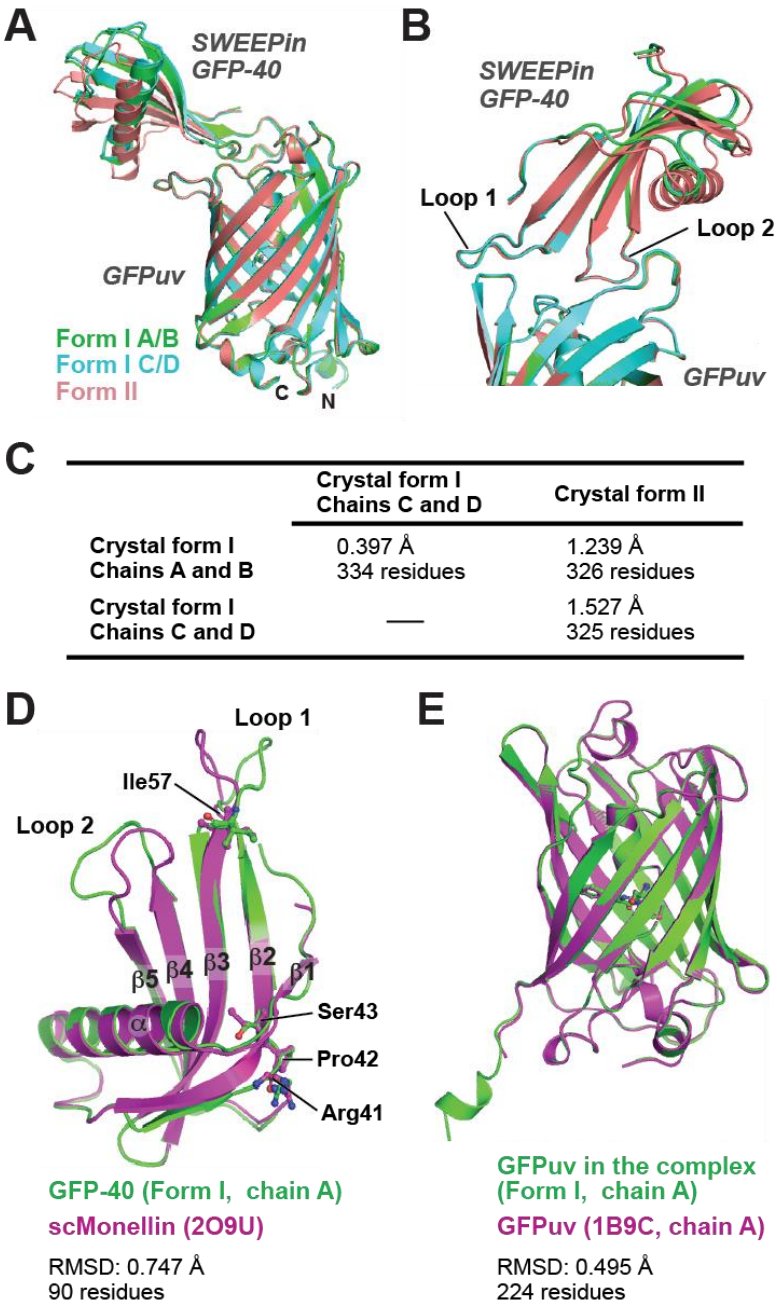


Figure S3. Structural analyses of the complexes of SWEEPin GFP-40 with GFPuv determined in this study. (A) Superposition of the three complexes of GFP-40 with GFPuv determined in this study. **(B)** Close-up view of the regions surrounding the binding interfaces in three complexes. **(C)** Pairwise structural comparison among three complexes of GFP-40 with GFPuv. Root mean square deviations of corresponding Cα atoms are shown in Å. The numbers of aligned residues when superposed are shown. **(D)** Superposition of SWEEPin GFP-40 with scMonellin (PDB ID: 2O9U). The amino acid residues in SWEEPin GFP-40 showing the differences in the backbone structures when compared with scMonellin are labeled. **(E)** Superposition of GFPuv in the complex with GFPuv alone (PDB ID: 1B9C).