Supplementary Material

Structural basis for the interaction between the first SURP domain of the SF3A1 subunit in U2 snRNP and the human splicing factor SF1

Nobukazu Nameki, Masayuki Takizawa, Takayuki Suzuki, Shoko Tani, Naohiro Kobayashi, Taiichi Sakamoto, Yutaka Muto, and Kanako Kuwasako

This file contains Tables S1-S4 and Figures S1-S7.

Table S1. SURP-containing proteins and SURP domains in eukaryotic species

Protein name	SURP domain	Type[1)]	Human	Chicken, Lizard, Frog	Fish	Fly	Nematode	Fission yeast
SF3A1	SURP1	Ia	✓	✓	✓	✓	✓	✓
	SURP2	II	✓	✓	✓	✓	✓	✓
SFSWAP	SURP1	Ib	✓	✓	✓	✓	✓	-
	SURP2	Ia	✓	✓	✓	✓	✓	-
SF140	SURP	I	✓	✓	✓	✓	✓	-
CHERP	SURP	Ia	✓	✓	✓	✓	-	-
SF4	SURP1	Ib	✓	✓	✓	-	-	-
	SURP2	Ib	✓	✓	-	-	-	-
SFRS140	SURP2	Ib	✓	✓	-	-	-	-
	SURP1[2)]	III	✓	-	-	-	-	-

1) SURP domains are classified into three types, as described in the text.

2) The SURP domain have a 3-residue loop instead of the G bend.

Table S2. CYANA structural statistics for 40 conformers of the chimera of SURP1 and S1BRp (res. 295–327) with the lowest CYANA target function values

CYANA target function (Å^2)	0.23 ± 0.02
Residual NOE Violations	
Number > 0.20 Å	1
Average (Å)	0.07
Maximum (Å)	0.23
Residual dihedral angle violations	
Number > 2.5º	0
Average (º)	0.17
Maximum (º)	0.70

Table S3. Summary of the putative salt bridges and hydrogen bonds between SURP1 and S1BR

SURP1	S1BR	(%)
K55 H^{ζ}	E319 O^{ε}	100
R62 H^{η}	E312 O^{ε}	100
N63 H^{δ}	E312 O^{ε}	95
R70 H^{η}	D310 O^{δ}	100
N74 H^{δ}	Y313 O^{η}	95
K80 H^{ζ}	E322 O^{ε}	100

Hydrogen bonds and salt bridges were calculated using MOLMOL, according to each of the definitions described in the text. Residues listed are those that form each interaction only in 10 or more of the 20 structures, or equivalently in more than 50% of the structures.

Table S4. Summary of NMR measurements for structural analysis of SURP1 in three forms

NMR measurements	A free form	A bound form	The chimera protein
2D [^{1}H,^{15}N]-HSQC	✓	✓	✓
3D HNCO	✓	*Not measured*	✓
HN(CA)CO	✓	*Not measured*	*Not measured*
HNCA	✓	✓	✓
HN(CO)CA	✓	*Not measured*	✓
HNCACB	✓	✓	✓
CBCA(CO)NH	✓	✓	✓
2D [^{1}H,^{13}C]-HSQC	✓	✓	✓
3D HBHA(CO)NH	✓	✓	✓
H(CCCO)NH	✓	✓	✓
(H)CC(CO)NH	✓	✓	✓
HCCH-COSY	✓	✓	✓
HCCH-TOCSY	✓	✓	✓
3D ^{15}N-edited [^{1}H,^{1}H]-NOESY	✓	✓	✓
3D ^{13}C-edited [^{1}H,^{1}H]-NOESY	✓	✓	✓

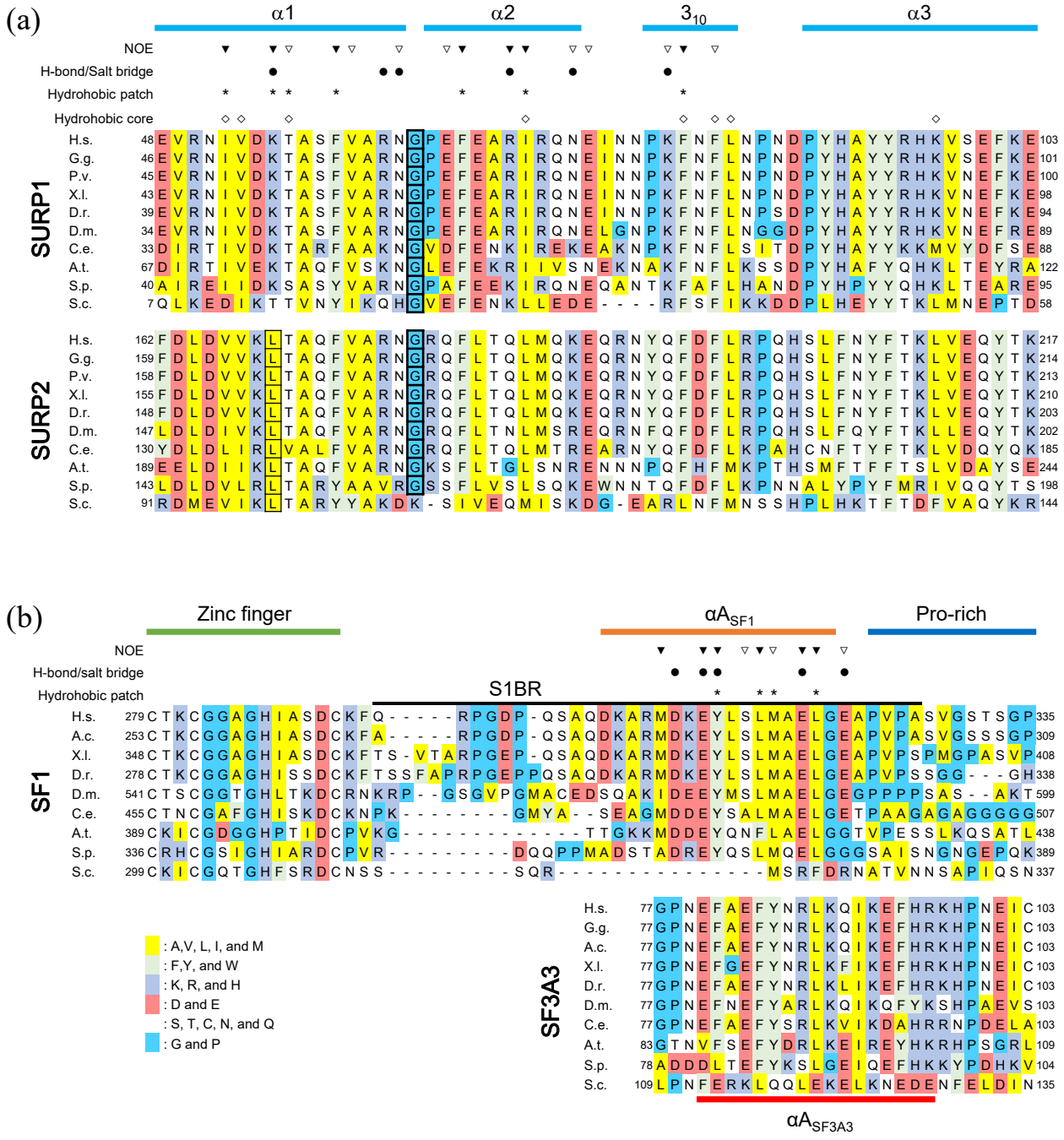


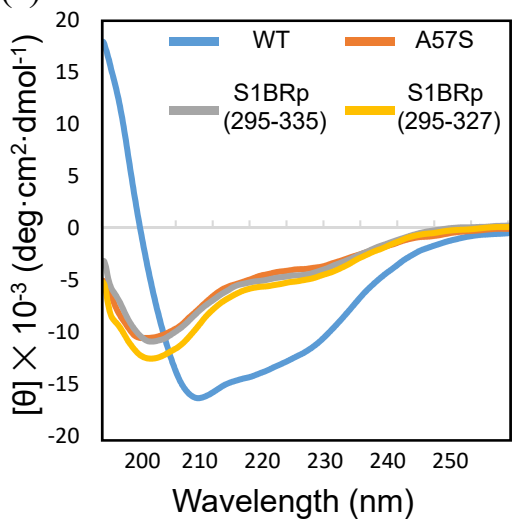
Figure S1 (legend on next page)

Figure S1. Structure-based multiple sequence alignment of SURP1 and SURP2 in SF3A1 orthologs, and that of S1BR in SF1 orthologs and S2BR in SF3A3 orthologs

(a) Amino acid sequence alignments of the two SURP domains in SF3A1 orthologs from various eukaryotes were performed using the ClustalW program and then manually modified based on the structures of the two SURP domains. Accession codes used in the sequence alignment were as follows: *Homo sapiens* (H.s., UniProt accession No. Q15459), *Gallus gallus* (G.g., F1NU16), *Pogona vitticeps* (P.v., A0A6J0U6T9), *Xenopus laevis* (X.l., Q6GPA9), *Danio rerio* (D.r., Q90X41), *Drosophila melanogaster* (D.m., Q9VEP9), *Caenorhabditis elegans* (C.e., G5ECL3), *Arabidopsis thaliana* (A.t., Q8RXF1), *Schizosaccharomyces pombe* (S.p., O13900), and *Saccharomyces cerevisiae* (S.c., P32524). Regions of secondary structure elements are shown by cyan bars at the top. Alignments are colored according to the amino acid type shown in the alignment legend. Each box with thick black lines indicates a Gly residue connecting α1 and α2 that forms the G bend. Each box with thin black lines indicates a Leu residue in α1 of SURP2, which is the determinant of the SURP2 complex formation. Inverted triangles indicate residues of SURP1 for which intermolecular NOEs from S1BR were observed (open inverted triangle: less than ten NOEs, filled inverted triangle: ten or more NOEs). Filled round marks indicate residues of SURP1 that participate in hydrogen bonds or salt bridges with S1BR. Asterisks indicate residues of SURP1 that form a hydrophobic patch interacting with S1BR. Diamonds indicate residues that form the hydrophobic core common to SURP domains.

(b) Amino acid sequence alignments of S1BR in SF1 orthologs and S2BR in SF3A3 orthologs from various eukaryotes were performed using the ClustalW program and then manually modified based on the α-helix structures of the two ligand peptides. Accession codes used in the sequence alignment are as follows: for S1BR; *H. sapiens* (H.s., Q15637), *Anolis carolinensis* (A.c., H9GN87), *X. laevis* (X.l., A0A1L8GD42), *D. rerio* (D.r., Q6TNQ8), *D. melanogaster* (D.m., Q6TNQ8), *C. elegans* (C.e., G5EF97), *A. thaliana* (A.t., Q9LU44), *S. pombe* (S.p., O74555), and *S. cerevisiae* (S.c., O74555) and for S2BR; *H. sapiens* (H.s., Q12874), *G. gallus* (G.g., Q5F387), *A. carolinensis* (A.c., H9GNY9), *X. laevis* (X.l., Q5U561), *D. rerio* (D.r., Q568L3), *D. melanogaster* (D.m., O46106), *C. elegans* (C.e., Q22469), *A. thaliana* (A.t., Q9FG01), *S. pombe* (S.p., O59706), and *S. cerevisiae* (S.c., P19736). The zinc finger domain and proline-rich region are shown by light green and blue bars, respectively, at the top of the figure. The α-helix of S1BR (αA_{SF1}) and that of S2BR (αA_{SF3A3}) are shown by orange (at the top) and red bars (below), respectively. Inverted triangles indicate residues of S1BR that were observed in the intermolecular NOEs from SURP1 (open inverted triangle: less than ten NOEs, filled inverted triangle: ten or more NOEs). Filled round marks indicate the residues of S1BR that participate in hydrogen bonds or salt bridges with SURP1. Asterisks indicate residues of S1BR that form a hydrophobic cluster interacting with SURP1.

(a)



Secondary Structure (%)	WT	A57S	S1BRp (295-335)	S1BRp (295-327)
Helix	96.8	26.1	24.6	26.5
Strand	3.2	1.2	0.0	0.0
Turns	0.0	9.1	6.5	0.0
Unordered	0.0	63.6	68.9	73.5

(b)

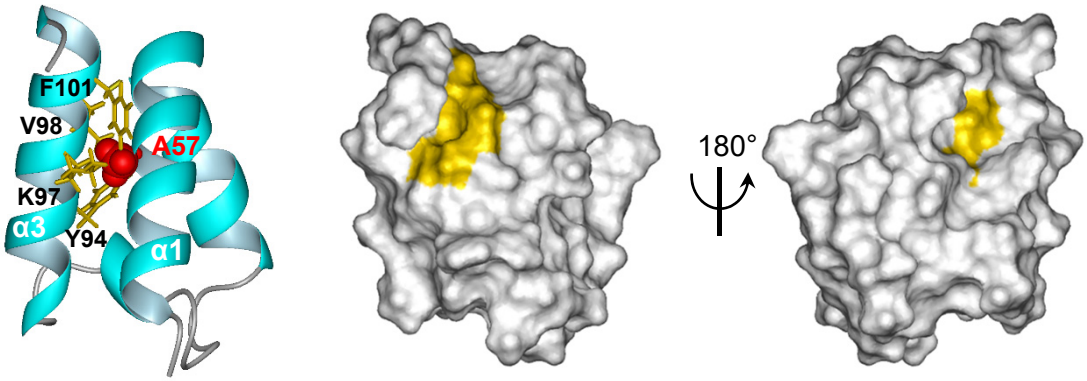


Figure S2. CD spectroscopy of wild-type SURP1, the A57S mutant, and the SF1 fragments

(a) CD spectroscopy of wild-type SURP1, the A57S mutant, and two ligand peptides, S1BRp (295–335) and S1BRp (295–327) (left panel), and secondary structure predictions based on spectra by a deconvolution analysis using Contin-LL software (right panel).

(b) Location of Ala57 in the SURP1 structure. Ribbon representation of SURP1 (left panel). The α- and 3_{10}- helices of SURP1 are colored cyan. The side chain of Ala57 is represented in red, while the side chains of four residues involved in hydrophobic interactions with Ala57 (i.e., Tyr94, Lys97, Val98, and Phe101) are represented in yellow. Molecular surface representations showing the locations of Ala57 and the four residues (middle and right panels). The middle view is in the same orientation as that in the left panel. The color scheme of these residues is the same as in the left panel. The absence of a red region on the surface shows that Ala57 is completely buried inside the structure.

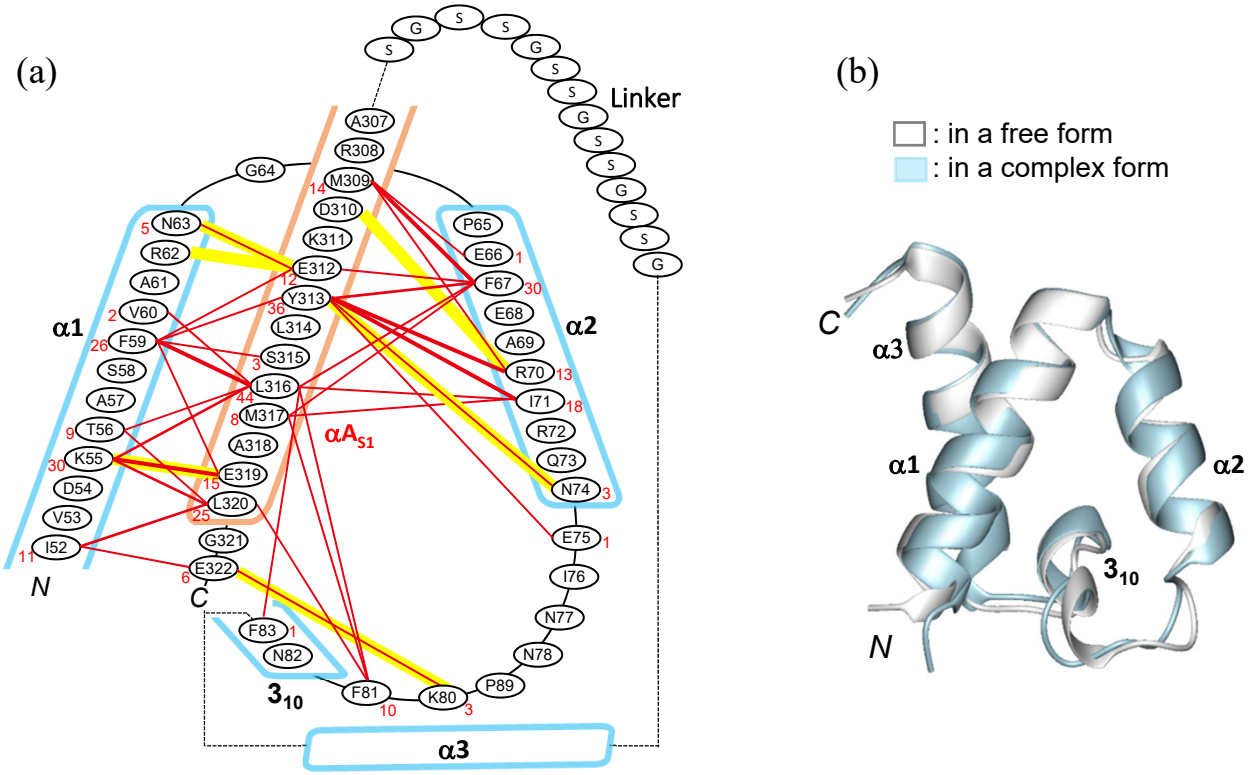


Figure S3. Intermolecular NOEs between SURP1 and S1BR, and comparison of SURP1 structures between free and complex forms

(a) Schematic diagram of intermolecular NOEs between SURP1 and S1BR in the chimera. The numbers of intermolecular NOEs are indicated by red numbers and are represented schematically by line widths. Salt bridges and hydrogen bonds are depicted with yellow lines. In total, 163 NOEs were observed between residues 48–110 of SURP1 and residues 295–327 of S1BR. Note that there are no intermolecular NOEs between S1BR and the linker or between S1BR and α3.

(b) Superposition of structures of SURP1 in a free form (white) and SURP1 of the chimera (blue).

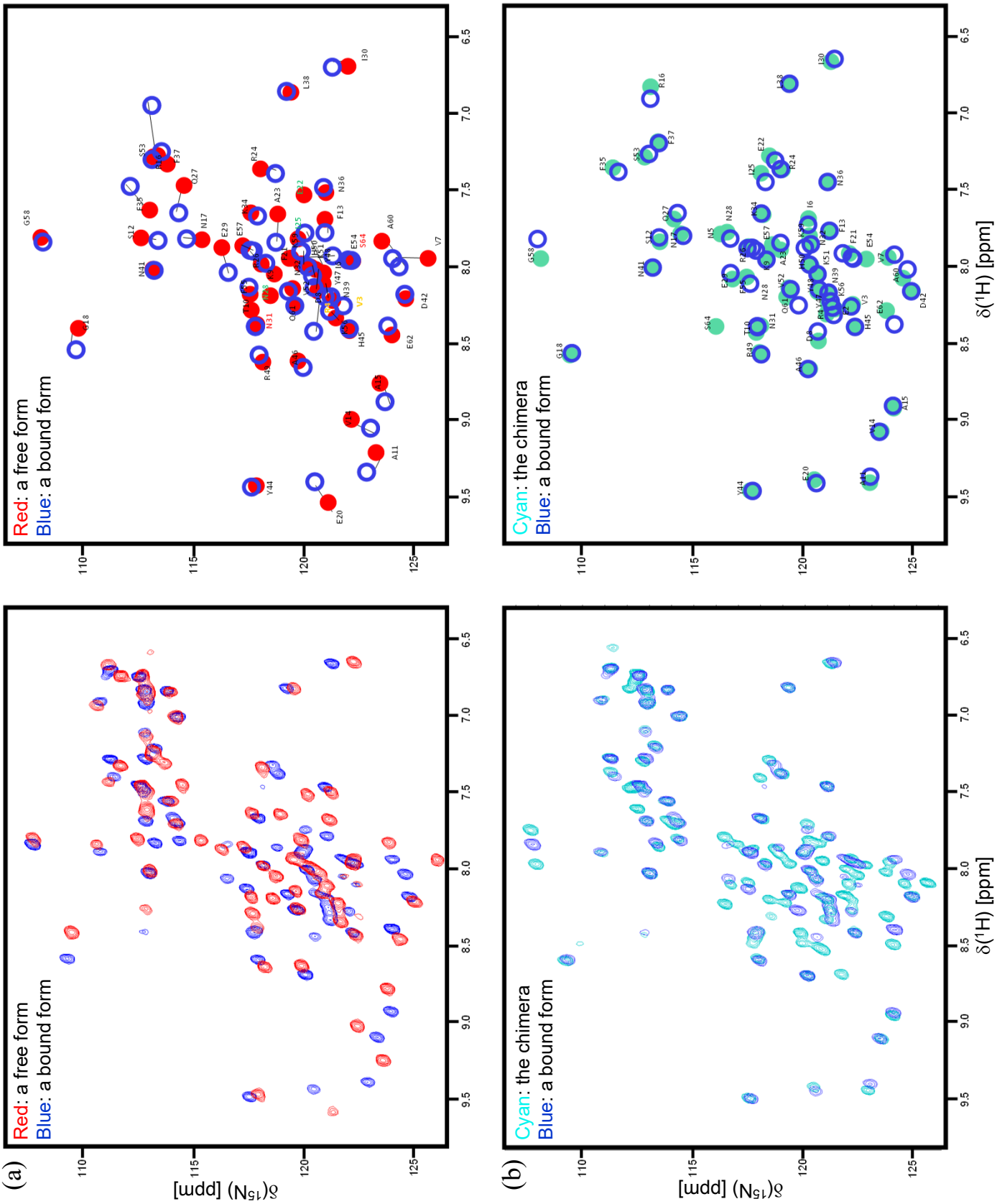


Figure S4. Comparisons of [^{1}H,^{15}N]-HSQC spectra between individual SURP1 in free and complex forms, and those between SURP1 of the chimera and individual SURP1 in a complex form

(a) Overlay of [^{1}H,^{15}N]-HSQC spectra of individual SURP1 in free and complex forms (left), and that of the corresponding schematic spectra in which cross peaks are indicated by filled or open circles labeled with the assigned residues (right).

(b) Overlay of [^{1}H,^{15}N]-HSQC spectra of SURP1 of the chimera and individual SURP1 in a complex form (left), and that of the corresponding schematic spectra in which cross peaks are indicated by filled or open circles labeled with the assigned residues (right). Note that for clarity, only the cross peaks from residues of SURP1 of the chimera are shown.

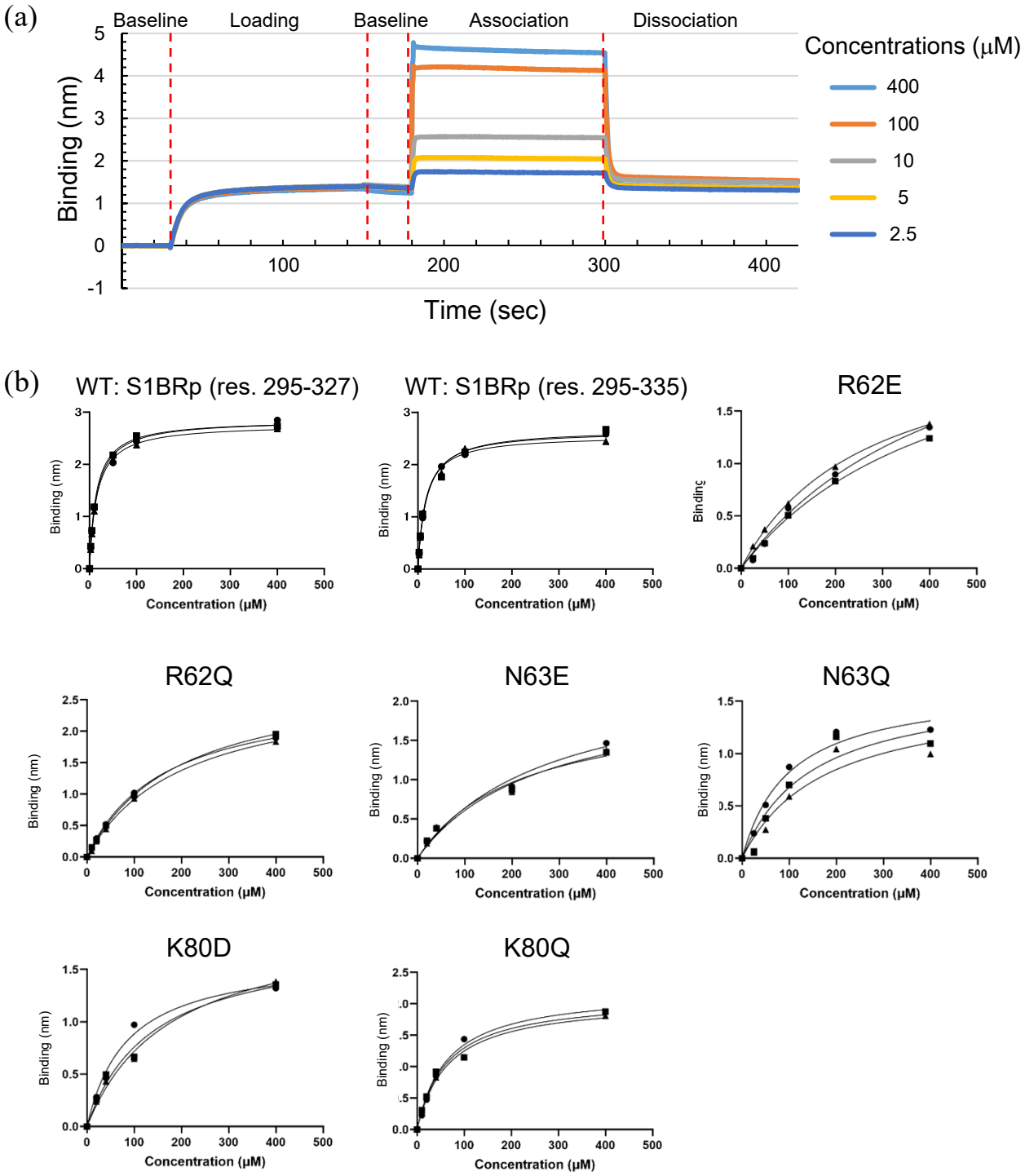


Figure S5. Equilibrium analysis of binding responses for interactions of SURP1 or mutants to S1BRp using BLI

(a) A sensorgram of the binding of wild-type SURP1 (analyte) to S1BRp (res. 295–327) (immobilized). (b) Binding curves showing the binding of wild-type SURP1 to pre-incubated S1BRp (res. 295–327) or S1BRp (res. 295–335), and those showing the binding of mutants to pre-incubated S1BRp (res. 295–327). The x-axis shows the concentrations of free wild-type SURP1 or each free mutant, and the y-axis shows plateau binding values or specific binding (nm). A saturation binding curve was generated by a non-linear curve fit using GraphPad Prism 9 for single-site binding. K_D values are reported in Table 1.

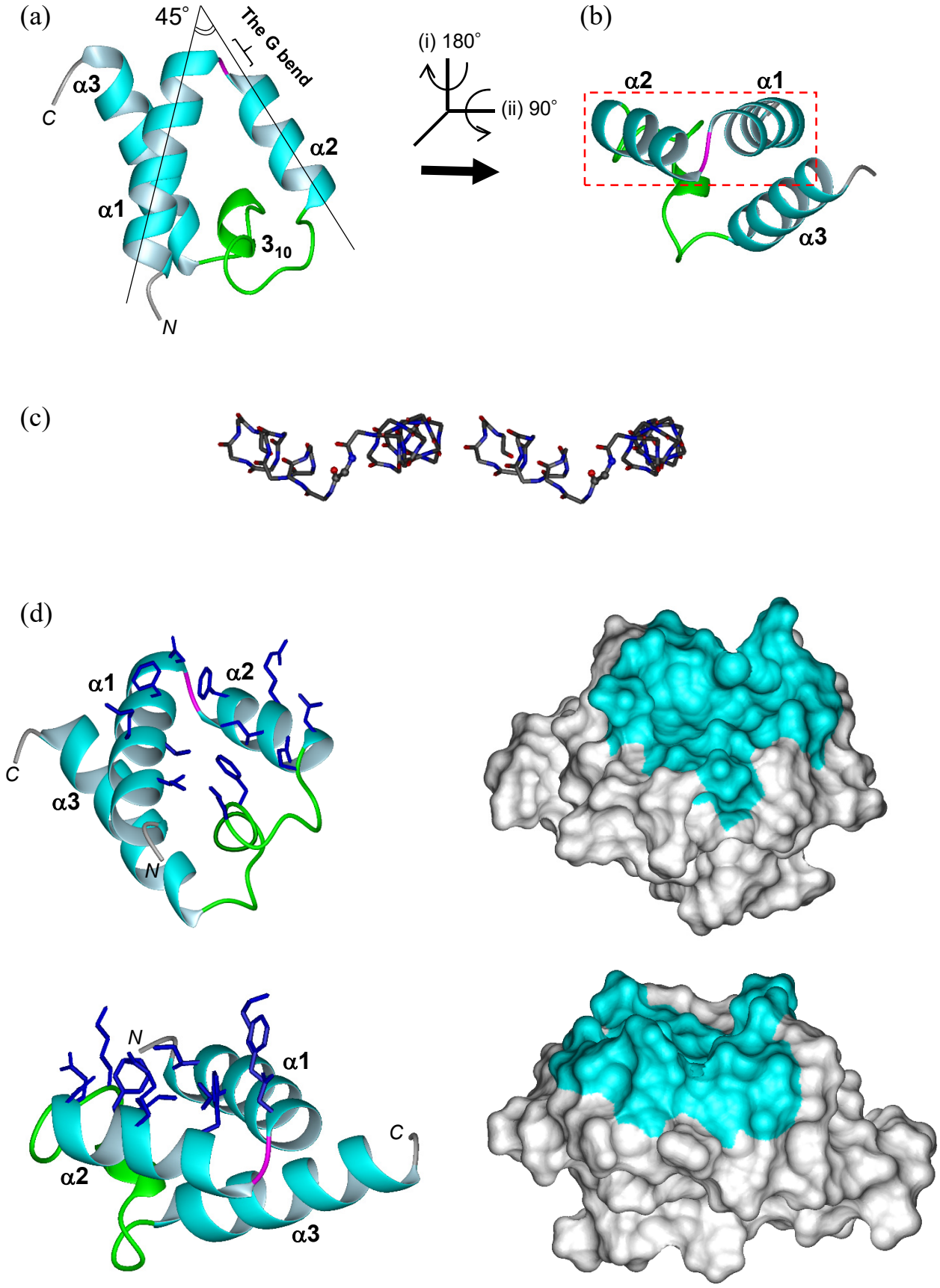


Figure S6 (legend on next page)

Figure S6. A unique topology of α1 and α2 characterized by the G bend.

(a) A ribbon representation of SURP1 of the chimera. The view is in the same orientation as that of Figure 5. The α-helices and the 3_{10} loop are colored cyan and green, respectively. The G bend is depicted in magenta.

(b) The view is rotated around the axes, as directed, to make the G bend more visible.

(c) Stereoview of the backbone of α1-[Gly]-α2 represented by a stick model. The corresponding region is surrounded by red dotted lines in (b).

(d) Different views of a binding surface formed by residues on α1, α2, and the 3_{10} loop. Only the side chains of the residues are represented in blue (left panels). Molecular surface representations showing the residues (both main and side chains) that are colored cyan (right panels).

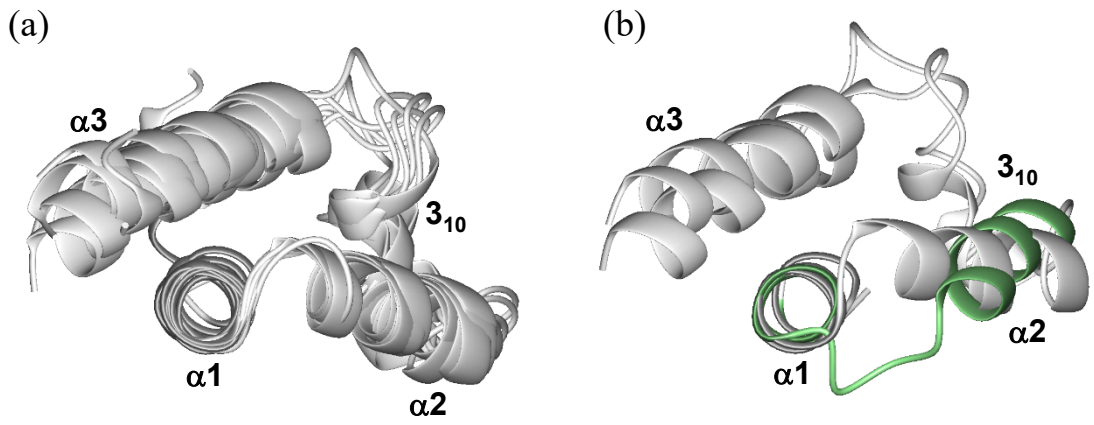


Figure S7. Comparison of structures among human SURP domains in terms of the G bend

(a) Superposition of ribbon representations of SF3A1 SURP1 (PDB ID 2DT6), SF3A1 SURP2 (2DT7), SFSWAP SURP1 (2E60), SFSWAP SURP2 (2E5Z), SF4 SURP1 (1UG0), and SF4 SURP2 (1X4O), all of which display a G bend, on α1 (res. 52-63 in SF3A1SURP1). All structural data were previously deposited in the PDB by our group.

(b) Superposition of ribbon representations of SF3A1 SURP1 (2DT6) and SFRS14 SURP1 (1X4P). In humans, only SFRS14 SURP1 has a 3-residue loop connecting α1 and α2 instead of the G bend. Only α1, α2, and the connecting loop in SFRS14 SURP1 are colored light green.