

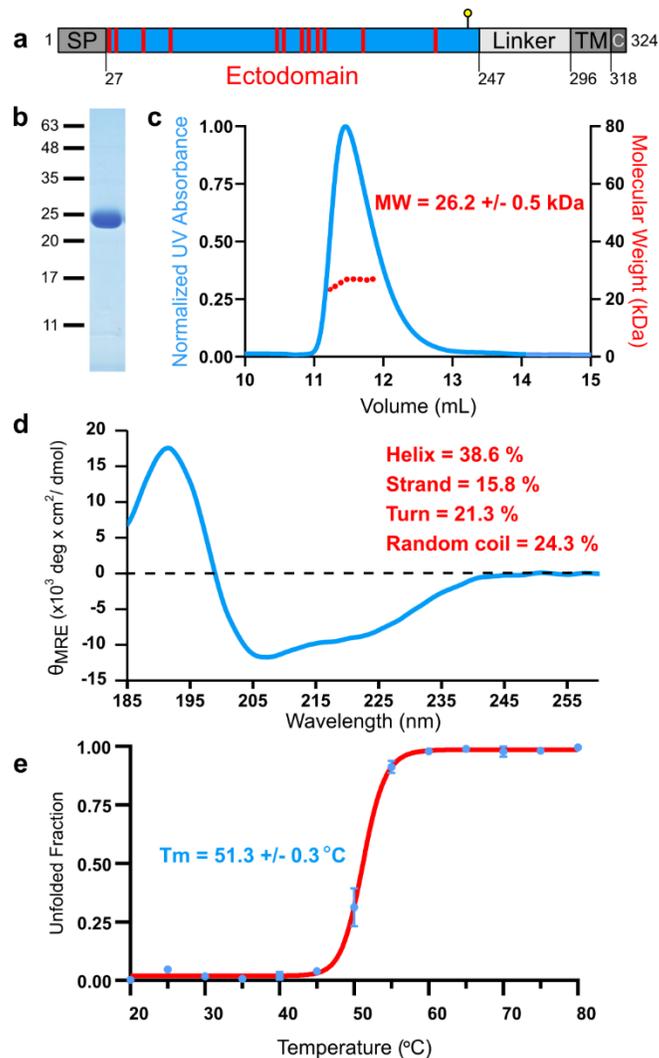
**Supplementary Information for**  
SPACA6 ectodomain structure reveals a conserved superfamily of  
gamete fusion-associated proteins

Tyler D.R. Vance, Patrick Yip, Elisabet Jiménez, Sheng Li, Diana Gawol, James Byrnes, Isabel Usón, Ahmed Ziyat, and Jeffrey E. Lee\*.

Correspondence: Jeffrey E. Lee  
Email: : [jeff.lee@utoronto.ca](mailto:jeff.lee@utoronto.ca)

**This PDF file includes:**

Figures S1 to S19  
Tables S1



**Figure S1. SPACA6 ectodomain expresses as a folded, monomeric protein.** **a** Domain architecture of full-length SPACA6. The signal peptide (SP), transmembrane helix (TM), and cytoplasmic tail (C) are indicated in the schematic. Red lines indicate cysteine residues, and yellow lollipop indicates predicted N-linked glycosylation site. **b** Coomassie-stained SDS-PAGE analysis of purified recombinant SPACA6 ectodomain. Molecular weights based on protein ladder bands are shown on the left. **c** SEC-MALS analysis of recombinant SPACA6 ectodomain. Size-exclusion chromatogram with normalized UV absorbance (280 nm) is shown in the blue curve with calculated molecular weight values derived from light scattering shown in red dots. The presented molecular weight is the average of calculated molecular weights from each individual light-scattering measurements ( $n=108$ ) taken within the peak,  $\pm$  standard deviation. **d** Far-UV CD spectra of recombinant SPACA6 ectodomain at 0.16 mg mL<sup>-1</sup>. Spectra is an average of ten accumulations that are buffer corrected and smoothed. **e** Thermal melt of recombinant SPACA6 ectodomain. Mean residue ellipticity was recorded at 207 nm. Values are averages of independent triplicates ( $n=3$ ) with error bars for standard deviation shown in blue.

>SPACA6\_Native

MALLALASAVPSALLALAVFRVPAWA**CLLCFTTYSERLRICQMFVGMRS**PKLEECEEAFTAA**FQGLSDTE**IN**YDERSHLHDTFTQ**M**THALQELAAAQGSFEVAF**PDAAEKMKK**VITQLKEAQACIPPCGLQEFARRFLCSGCYSRVCDLPLDCPVQDVTVTRGDQAMFSCIVNFQLPKEEITYSWKFAGGGLRTQDLSYFRDMPRAEGLARIRPAQLTHRGT**FSCVIK**QDQRPLARLYFFLNVTG**PPPPRAETELQASFR  
EVLRWAPRDAELIEPWRPSL**GELLARPEALT**PSNL**FLLAVL**GALASASAT**VLAWMFF**RWYCSGN

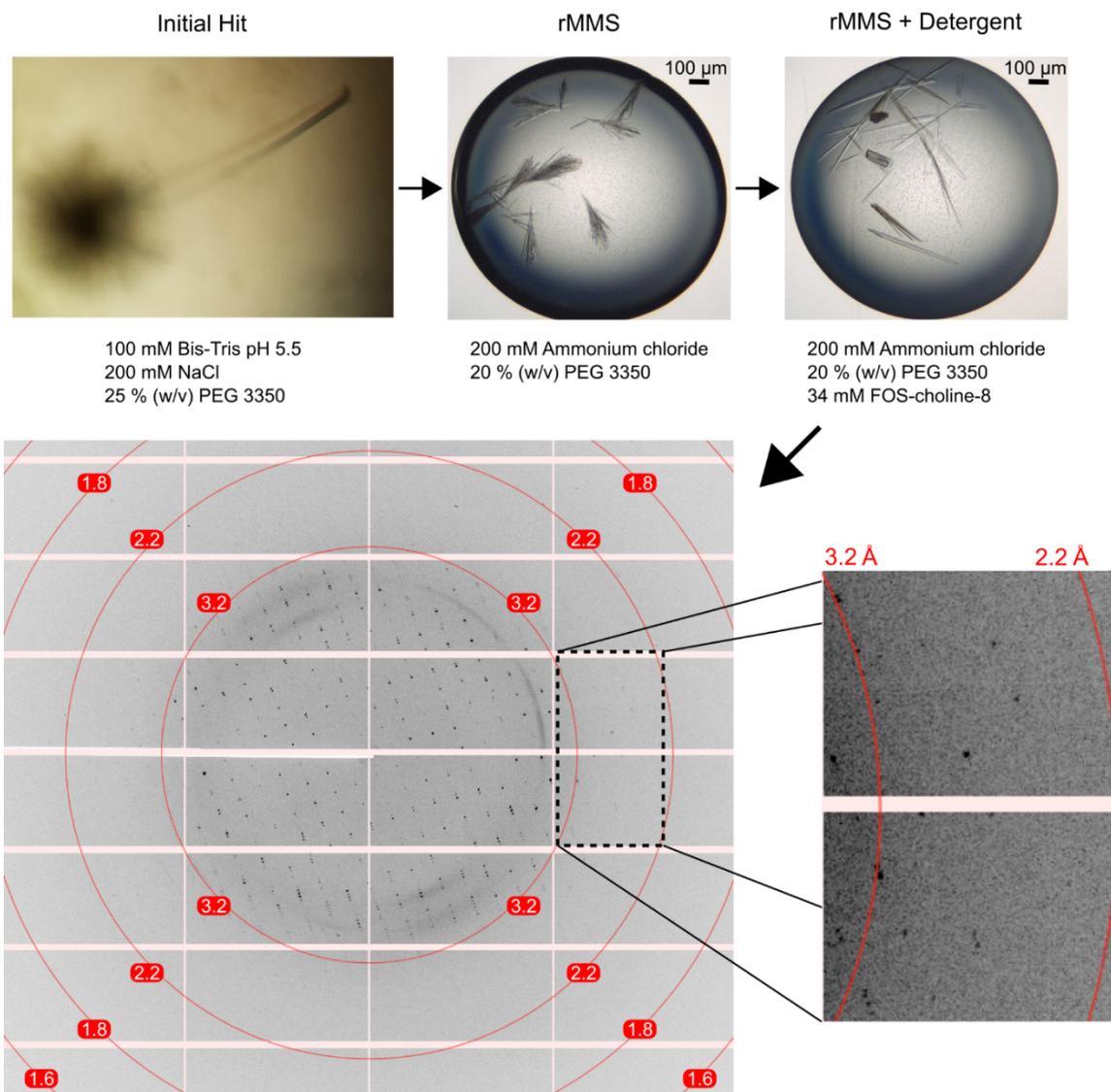
>SPACA6\_Recombinant

MKLCILLAVVAFVGLSLG**CLLCFTTYSERLRICQMFVGMRS**PKLEECEEAFTAA**FQGLSDTE**IN**YDERSHLHDTFTQ**M**THALQELAAAQGSFEVAF**PDAAEKMKK**VITQLKEAQACIPPCGLQEFARRFLCSGCYSRVCDLPLDCPVQDVTVTRGDQAMFSCIVNFQLPKEEITYSWKFAGGGLRTQDLSYFRDMPRAEGLARIRPAQLTHRGT**FSCVIK**QDQRPLARLYFFLNVTGGR****LVPRGSHHHHHHHHHH**

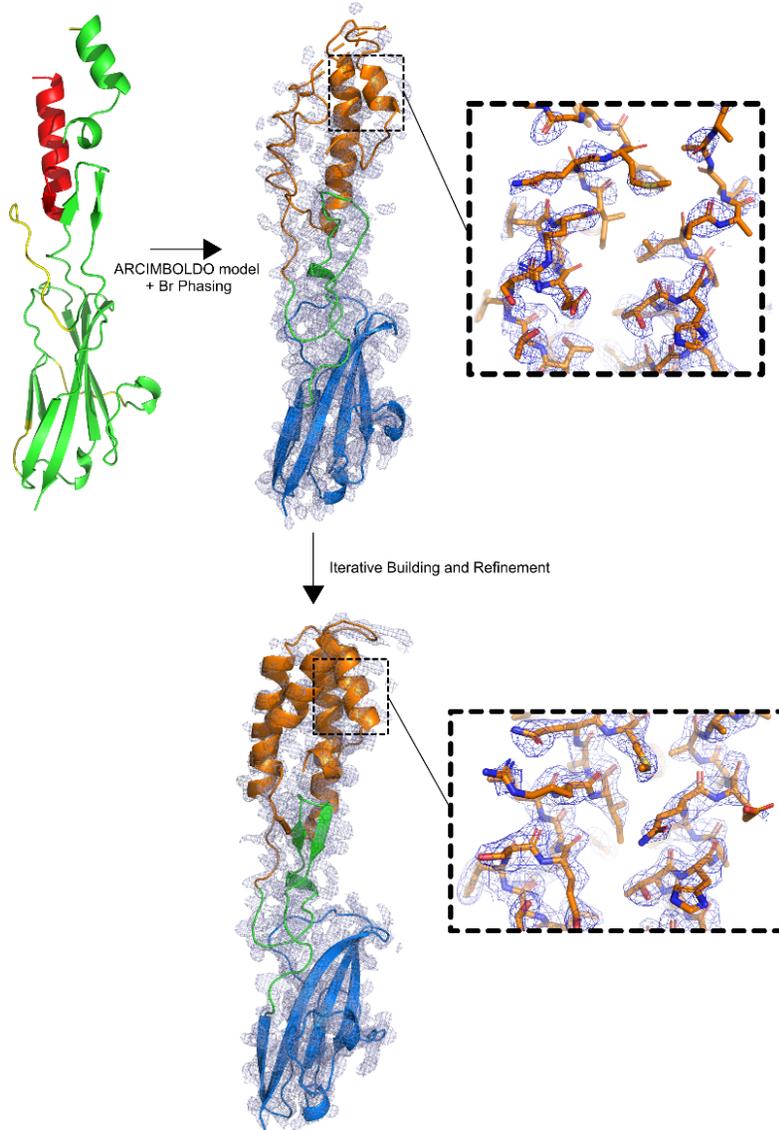
>SPACA6\_Recombinant\_Processed

**CLLCFTTYSERLRICQMFVGMRS**PKLEECEEAFTAA**FQGLSDTE**IN**YDERSHLHDTFTQ**M**THALQELAAAQGSFEVAF**PDAAEKMKK**VITQLKEAQACIPPCGLQEFARRFLCSGCYSRVCDLPLDCPVQDVTVTRGDQAMFSCIVNFQLPKEEITYSWKFAGGGLRTQDLSYFRDMPRAEGLARIRPAQLTHRGT**FSCVIK**QDQRPLARLYFFLNVTGGR****LVPR**

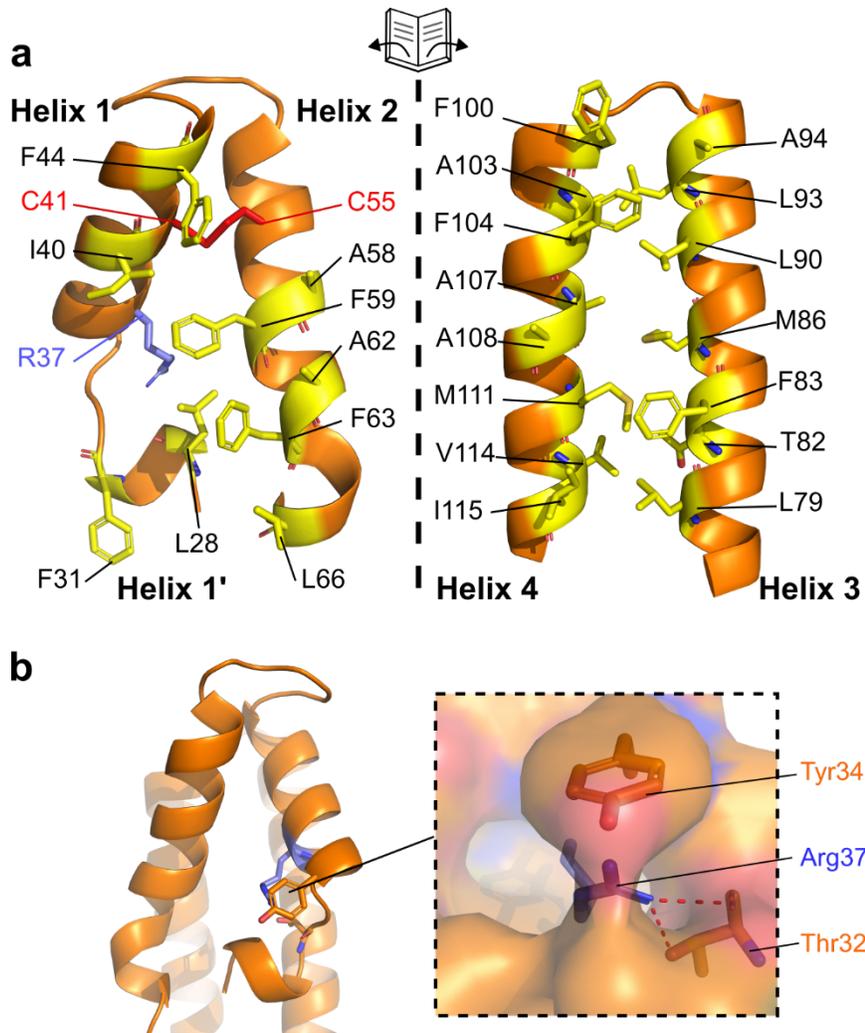
**Figure S2.** Sequences of native SPACA6 and recombinant SPACA6 ectodomain. One-letter amino acid sequences for SPACA6 Native, Recombinant, and the Recombinant construct after secretion and tag-removal with thrombin. Orange indicates amino acids of the four-helix bundle; green indicates hinge region amino acids; blue indicates Ig-like domain amino acids; grey indicates native signal peptide; red indicates native transmembrane helix; black indicates non-ectodomain native sequences; grey highlight indicates BiP signal peptide for *Drosophila* S2 cell secretion; green highlight indicates thrombin cleavage site; black highlight indicates 10x-His tag; and purple indicates putative N-linked glycosylation site.



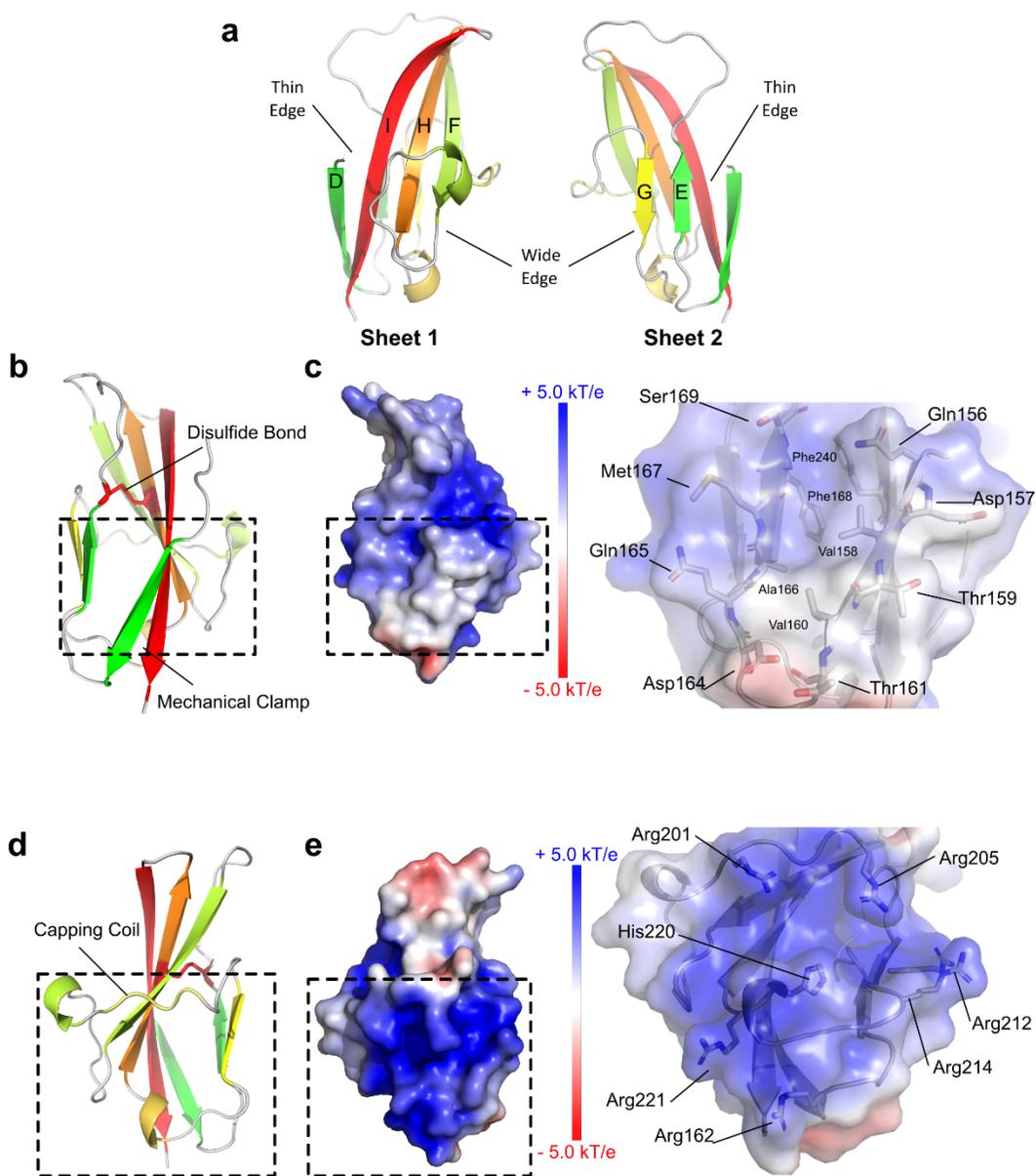
**Figure S3. Crystal growth and diffraction of SPACA6.** Upper images were taken during optimization; Optimization of SPACA6 crystals (upper images) with the compositions of the reservoir solutions stated the images. Upper left: Initial crystals obtained by vapor-diffusion; these crystals diffracted poorly. Upper middle: Crystals obtained when initial crystals were used as seeds for random microseed matrix screening (rMMS). Upper right: Larger crystals obtained through screening detergents as additives. Lower left: Diffraction image produced from the detergent-stabilized crystals, collected at the NE-CAT (24-ID-C) beamline at the Advanced Photon Source (Argonne National Laboratory, Lemont, IL). Red circles indicate (from innermost to outermost) 3.2 Å, 2.2 Å, 1.8 Å, and 1.6 Å resolution rings. Lower right insets: Higher magnification of boxed region showing high-resolution reflections to 2.2 Å resolution.



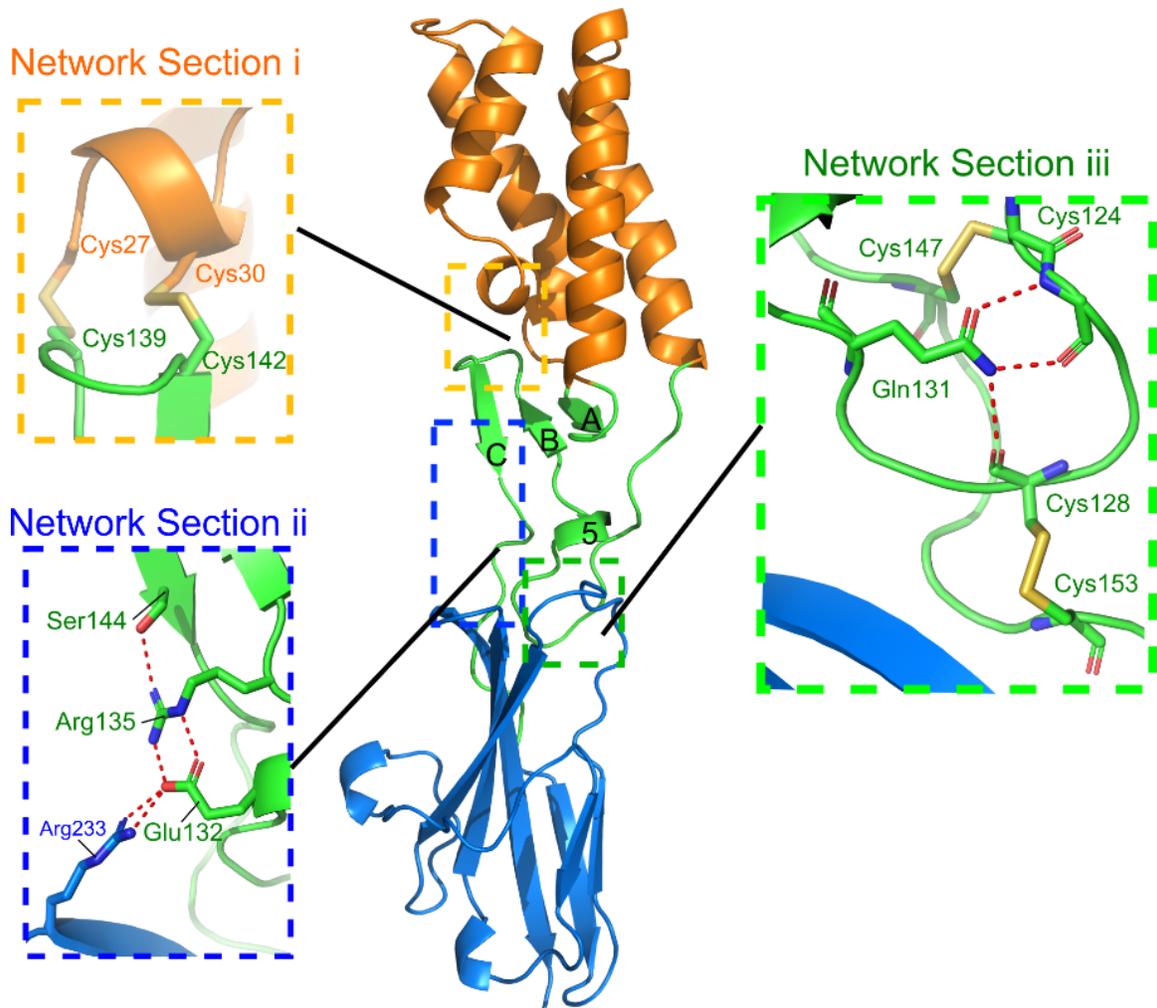
**Figure S4. Initial ARCIMBOLDO\_SHREDDER model and electron density map.** Upper left: Initial SPACA6 model produced via ARCIMBOLDO\_SHREDDER (65% complete). Portions of the model built into strong electron density are colored green; portions of the model built into strong electron density but modeled as poly-alanine are colored yellow; and portions of the model unconnected to other parts and made up of poly-alanine are colored red. Upper right: Following Br-SAD phasing and ARCIMBOLDO\_SHREDDER molecular replacement, an initial phased combined experimental electron density map ( $|F_o|$ ) map shown at  $1\sigma$  was obtained. The inset shows the poor electron density of the 4HB. Lower: Following iterative building/refinement, the final electron density map ( $2|F_o|-|F_c|$ ) at  $1\sigma$  superimposed with the refined SPACA6 structure is shown. The inset shows the same region of the 4HB as the upper inset. Structures are colored orange for 4HB, green for hinge, and blue for Ig-like domain.



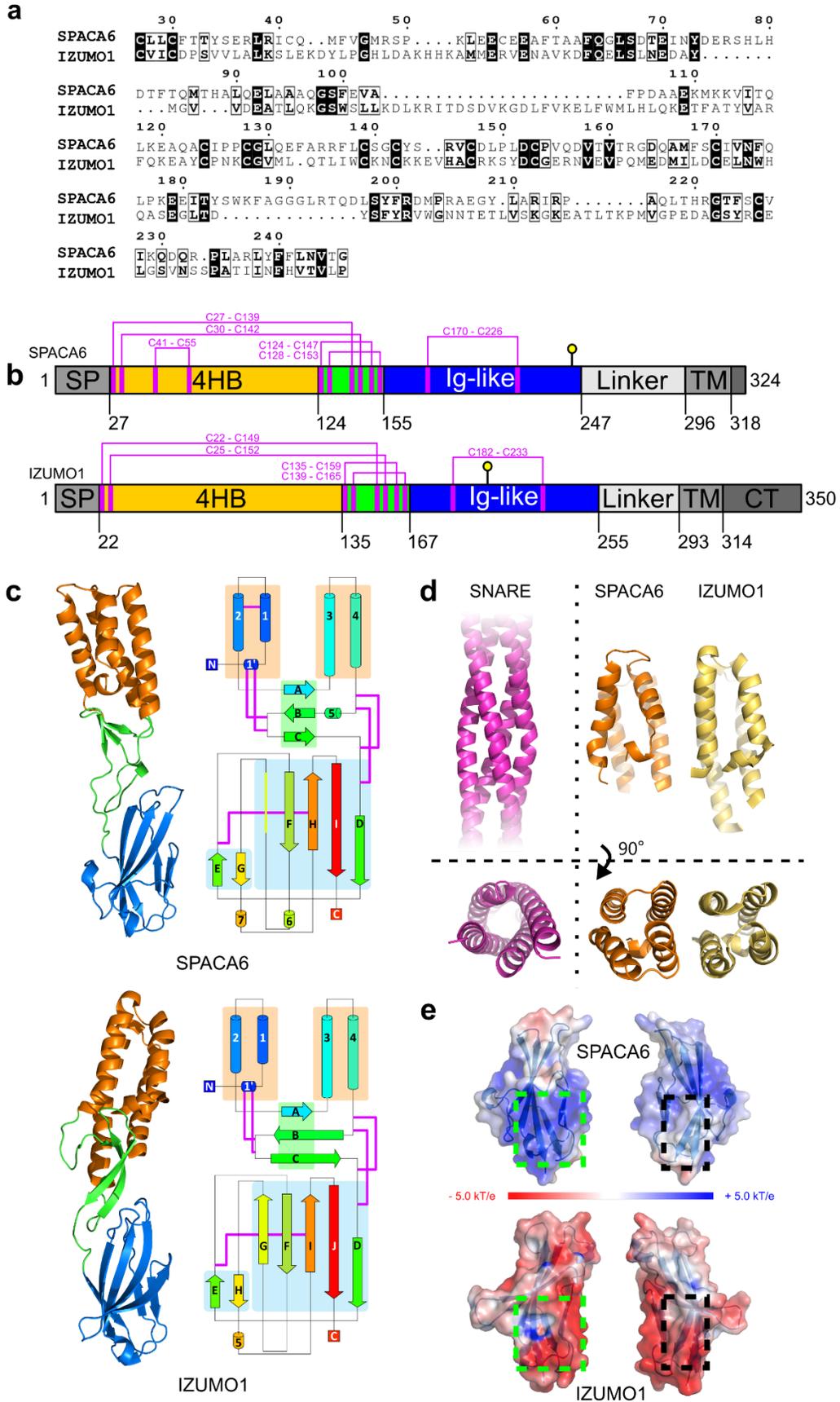
**Figure S5. Conserved arginine within 4HB hydrophobic core.** **a** Cross-section of SPACA6 four-helix bundle. The bundle was split in the middle and opened to reveal the hydrophobic core. Hydrophobic components of the 4HB core are colored yellow and displayed as sticks. Components of the 4HB core include Arg37 (purple) and a Cys41-Cys55 disulfide bond (red). **b** Triangular face of the 4HB, with Tyr34, Arg37 (purple), and Thr32 shown as sticks. Inset shows the Arg37 interaction with Thr32 and its occlusion by Tyr34.



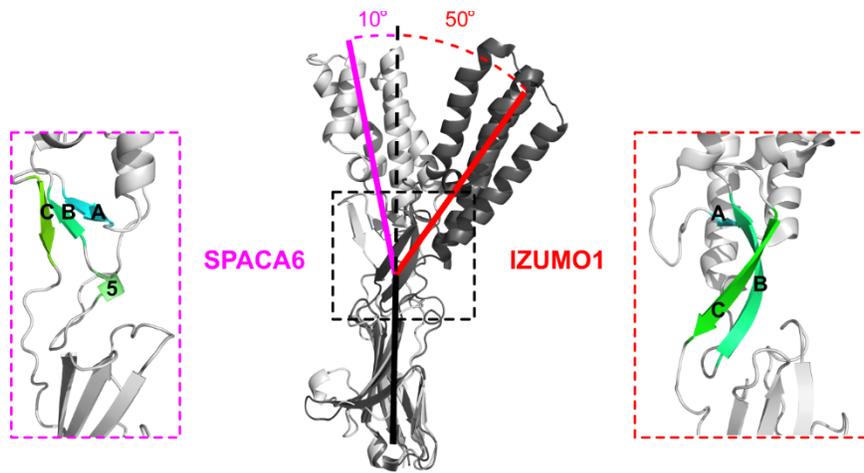
**Figure S6. C-terminal Ig-like domain of SPACA6.** **a** Ribbon diagram of the SPACA6 Ig-like domain with its six strands labelled. **b** Ribbon diagram of the thinner edge of the Ig-like domain. The disulfide bond is shown as sticks. **c** Surface electrostatic potential of the thinner edge of the Ig-like domain. Hydrophobic patch outlined by dashed box is shown as a zoomed image (right) with the residues shown as sticks. **d** Ribbon diagram of the wider edge of the Ig-like domain. Capping coil that covers the hydrogen bonds in Strand F is shown. **e** Surface electrostatic potential of the wider edge of the Ig-like domain. Positively charged cavity outlined by dashed box is shown as a zoomed image (right) with labeled residues shown as sticks.



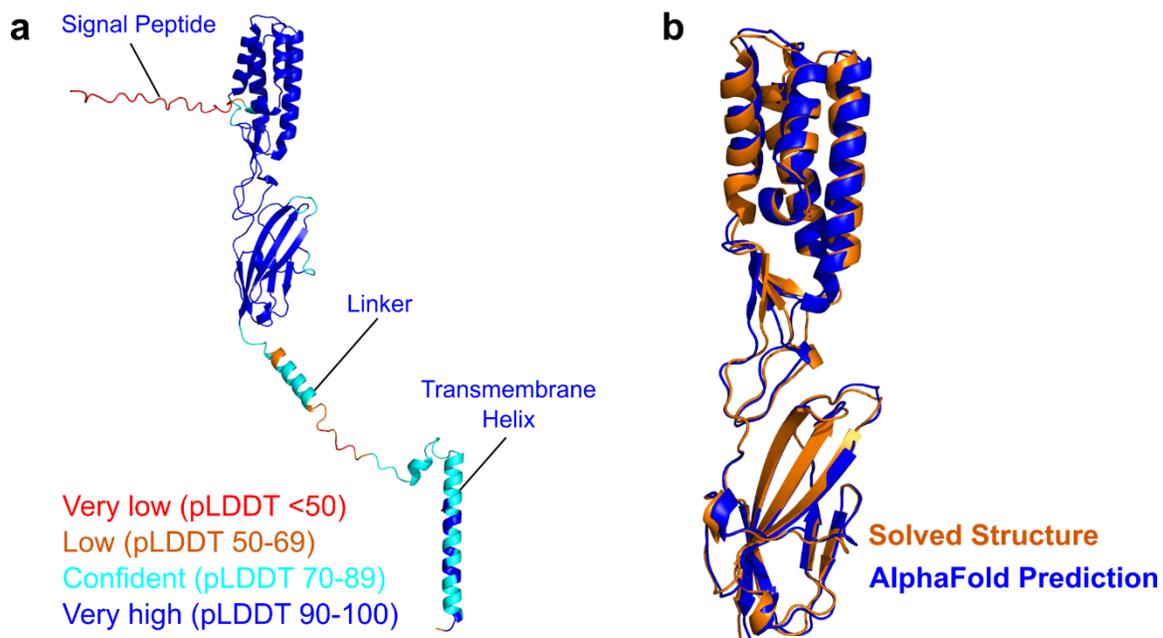
**Figure S7. SPACA6 hinge region connects the 4HB and Ig-like domains.** Ribbon diagram of the hinge region (green), which connects the 4HB (orange) and the Ig-like domain (blue). The three sections of the covalent/electrostatic network that rigidifies the region are shown in detail.



**Figure S8. SPACA6 and IZUMO1 share a similar structure.** **a** Human SPACA6 (RefSeq: NP\_001303901) and IZUMO1 (RefSeq: NP\_872381) ectodomain sequences aligned using CLUSTAL OMEGA. Strictly conserved residues are displayed as white text on a black background. Residues with conservative substitutions are in bold. **b** Domain architecture schematic of SPACA6 and IZUMO1. Magenta lines indicate disulfide bonds and yellow lollipop indicates a predicted N-linked glycan site. Abbreviations: SP, signal peptide; TM, transmembrane helix; CT, cytoplasmic tail. **c** Ribbon and topology diagrams comparing ectodomain structures for SPACA6 and IZUMO1 (PDB: 5F4E). Domains are colored according to panel b. **d** Four-helix bundle comparison in two orientations. Ribbon diagrams for autophagic SNARE complex (PDB: 4WY4, magenta), SPACA6 (orange), and IZUMO1 (PDB: 5F4E, yellow). **e** Surface electrostatic potential of SPACA6 and IZUMO1 Ig-like domains. The positively charged pocket from SPACA6 is outlined with a dashed green line, and the hydrophobic surface from SPACA6 is outlined with a dashed black line.



**Figure S9. Differences in domain orientation between IZUMO1 and SPACA6.** SPACA6 (white) and IZUMO1 (PDB: 5F4E, black) structures were aligned by the Ig-like domain to compare the orientation of the 4HB. Insets show the hinge region for both SPACA6 (purple box) and IZUMO1 (red box).



**Figure S10. AlphaFold prediction of SPACA6.** **a** Full-length SPACA6 prediction from AlphaFold (AF-W5XKT8-F1) colored by confidence level (pLDDT). **b** Ectodomain comparison between the solved structure (orange) and the AlphaFold prediction (blue) of SPACA6. Structures align with an RMSD of  $\sim 1.1$  Å (aligning 1494 atoms).

```

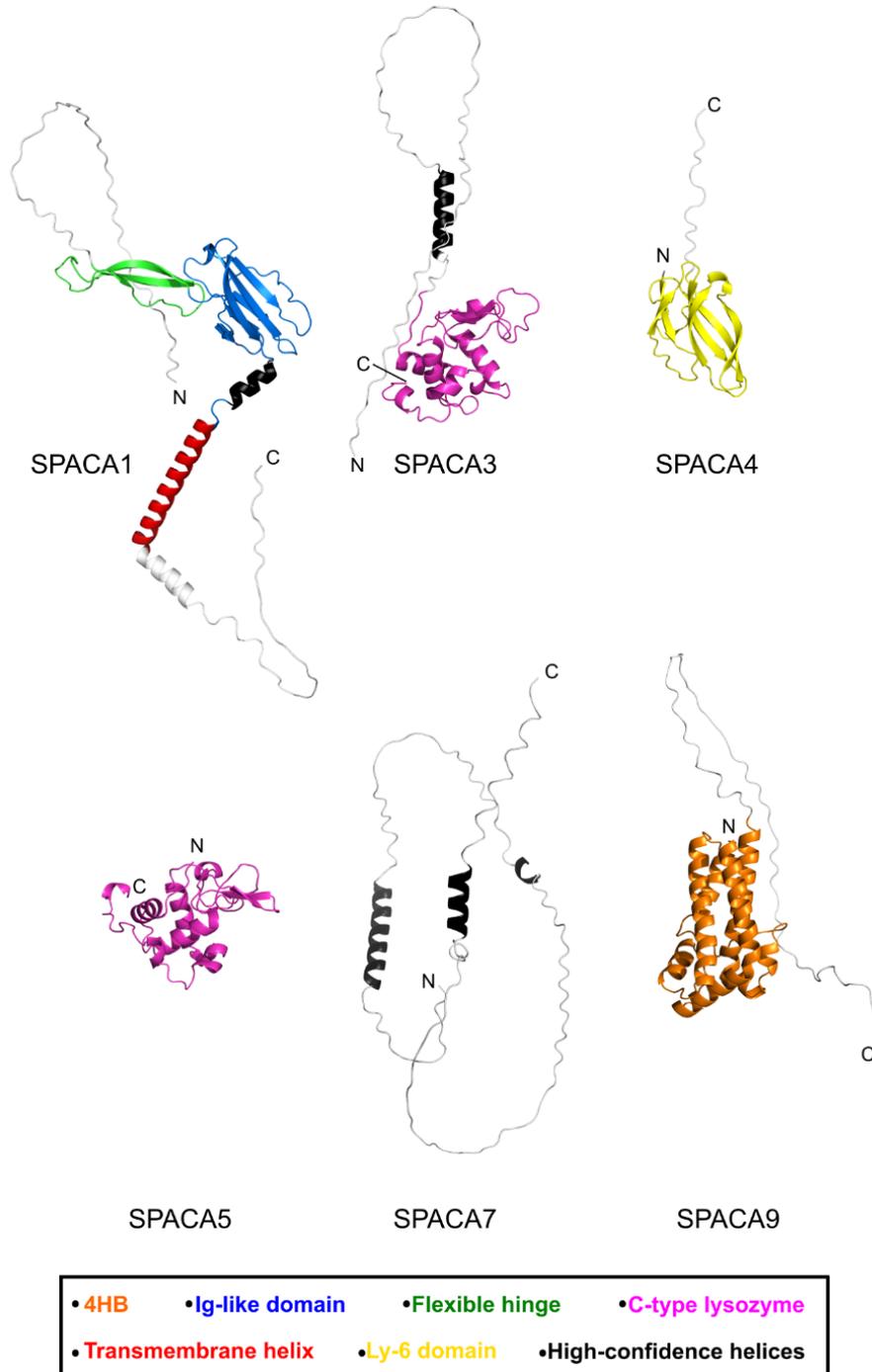
      1      10      20      30      40      50
IZUMO1 CVIC D P S V V L A L K S L E K D Y L P G . H L D A K H H K A M M E R V E N A V K D F Q E L S . . . L N E D . A Y M G
IZUMO2 CLQC D P L V L E A L G H L R S A L I P S . R F Q L E Q L Q A R A G A V . . . L M G M E G P F F R D Y A L N . V F V G
IZUMO3 CLEC D P K F I E D V G S L L G N L I P S . E V P G R T Q L L E R Q . . . . I K E M I H L S F K V S H S D . K R L R
IZUMO4 CLHC H S N F S K K F S F Y R H H V N F K S W W V G D I . P . V S G . . . A L L T D W S D D T M K E L H L A . . I P A
SPACA6 CLLC F T T Y S E R L R I . . . . . C Q . M F V G M R S P K L E E C E E A F T A A F Q G L S D T E I N Y D E . . . .
TMEM95 CVFC R L P A H D L S G R L A R . . L C S . Q M E A R Q K E . . . . C G . . . . . A S P D F S A F A L D E V S M N

      60      70      80      90      100      110
IZUMO1 V V D E A T L Q K . G S W S L L K D L K R I T D S D V K G D L F . V K E L F W M L H L Q K E T F A T Y V A R . . F Q K E
IZUMO2 K V E T N Q L D L V A S F V . K N Q T Q H L M G N S L K D E P L . L E E L V T L R A N V I K E F K K V L I S . . Y E L K
IZUMO3 V L A V Q Q V V K L R T W L . K N E F Y K L G N E T W K G V F I Y Q G K L L D V C Q N L E S K L K E L L K N . . F S E I
IZUMO4 K I T R E K L D Q V . . . . . A T A V Y Q M M D Q L Y Q G K M Y F P G Y F P N E L R N I F R E Q V H L I Q N A I I E S R
SPACA6 . . . . . R S H . . L H D T F T Q M T H A L Q E L A A A Q G S F E V A F P D A A E K M K K V I T Q . . L K E A
TMEM95 K V T E K T H R V L R V M E I K E . . . . . A V S S L P S . . . . Y W S W L . . . . . R K T K L P E . . Y T R E

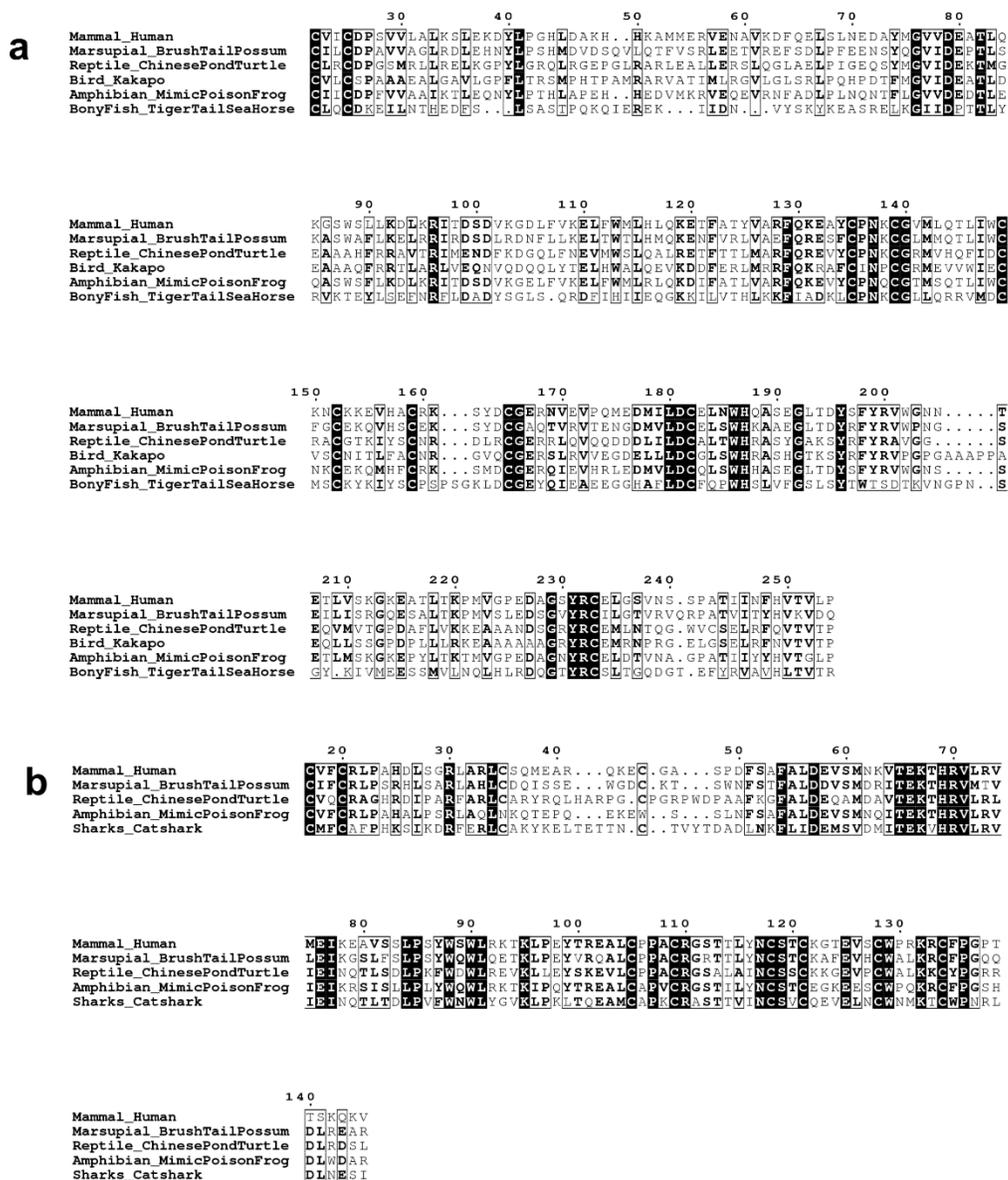
      120      130      140
IZUMO1 AYC . P N K CGVM L . Q T L I W CKNC K K E V H A CR K S Y D C
IZUMO2 A . C N P K L CR L L K . E E V L D CLHC Q R I T P K CI H K K Y C
IZUMO3 A . C S . E D C I V V E . G P I L D C W T CL R M T N R CF K G E Y C
IZUMO4 ID C . Q H R C G I F Q . Y E T I S C N N C T D S H V A CF . G Y N C
SPACA6 Q A C . I P P CG L Q E F A R R F L CS G C Y S R . . V C D L P L D C
TMEM95 AL C . P P A C R G S . . T T L Y N CS T C K G T E V S C W P R K R C

```

**Figure S11. Sequence alignment of the 4HB domains from the proposed IST-superfamily.** The 4HB domains from human IZUMO1 (NCBI: NP\_872381), IZUMO2 (NCBI: NP\_689571), IZUMO3 (NCBI: NP\_001351937), IZUMO4 (NCBI: NP\_001034935), SPACA6 (NCBI: NP\_001303901), and TMEM95 (NCBI: XP\_016880054) were aligned using CLUSTAL OMEGA. Strictly conserved residues are displayed as white text on a black background. Residues with conservative substitutions are in bold. Alignment numbering corresponds to the human IZUMO1 sequence.



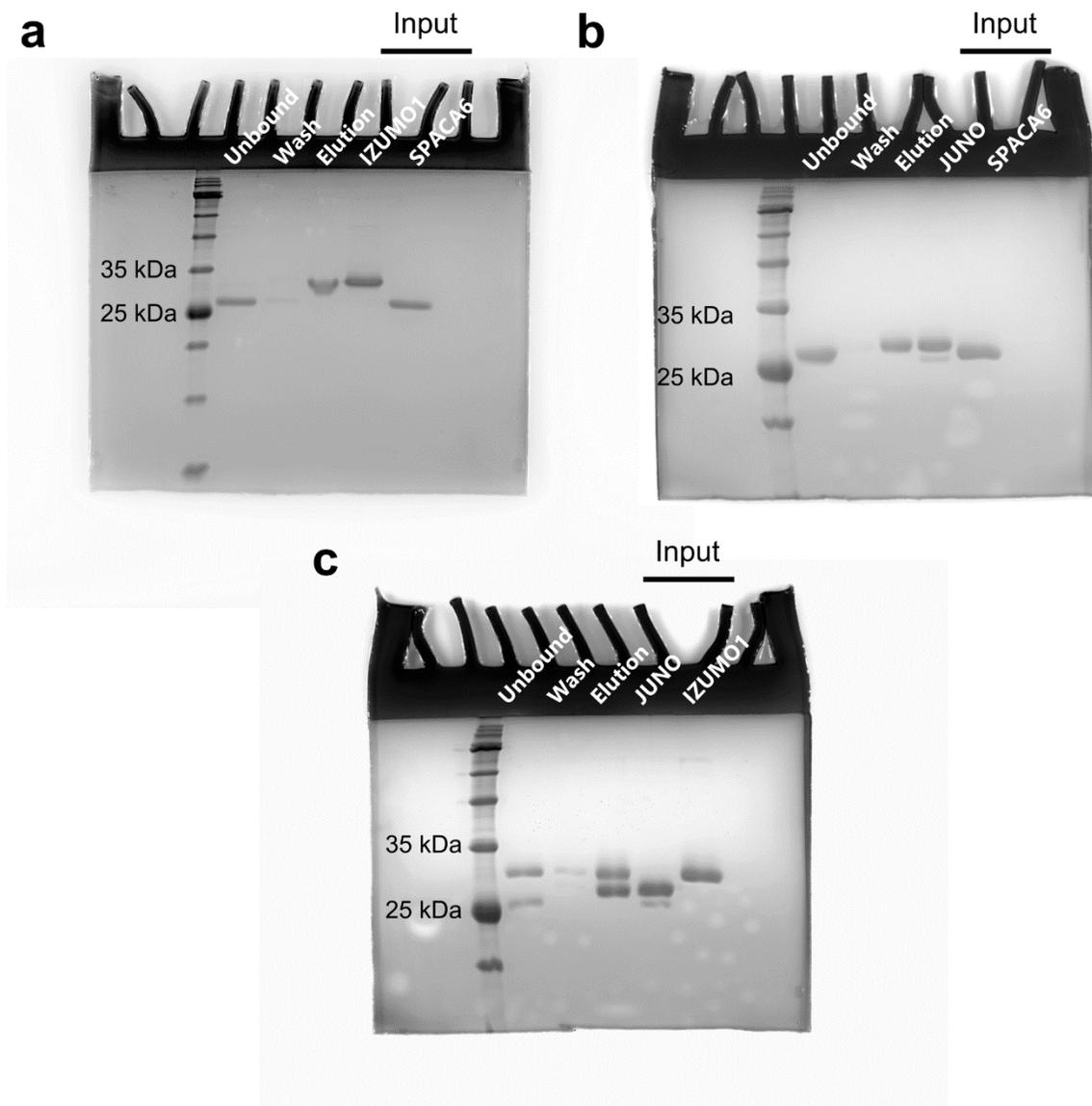
**Figure S12. AlphaFold predictions of structures of SPACA proteins.** Cartoon representations of AlphaFold structure predictions of human SPACA1 (AF-Q9HBV2-F1), SPACA3 (AF-Q8IXA5-F1), SPACA4 (AF-Q8TDM5-F1), SPACA5 (AF-Q96QH8-F1), SPACA7 (AF-Q96KW9-F1), and SPACA9 (AF-Q96E40-F1). Black helices are non-interacting with high predictive confidence. Areas of low predictive confidence are colored white. Signal peptides predicted for SPACA1, SPACA4, SPACA5, and SPACA7 were removed.



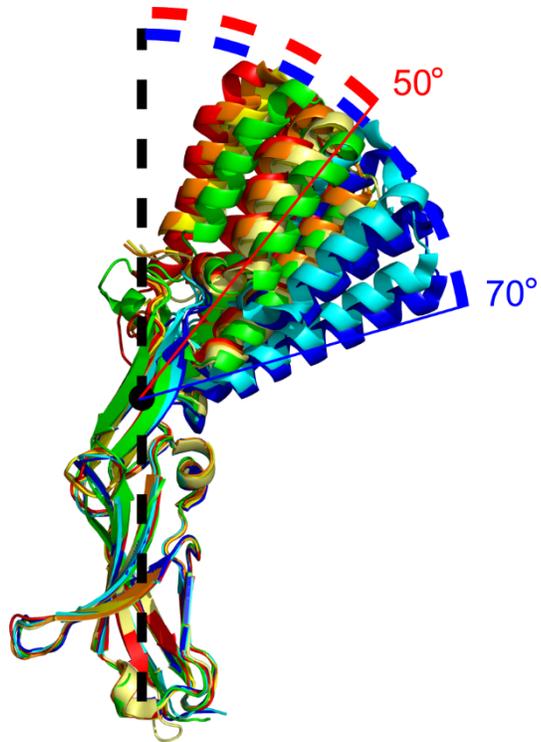
**Figure S13. Sequence alignment of IZUMO1 and TMEM95 from multiple species.** **a** Ectodomain sequences for IZUMO1 from human (*Homo sapiens*, NCBI: NP\_872381), the marsupial brush tail possum (*Trichosurus vulpecula*, NCBI: XP\_036602972), the Chinese pond turtle (*Mauremys reevesii*, NCBI: XP\_039366991), the flightless parrot Kākāpō (*Strigops habroptila*, NCBI: XP\_030366609), the mimic poison frog (*Ranitomeya imitator*, NCBI: CAF4957531), and the tiger tail seahorse (*Hippocampus comes*, NCBI XP\_019744032) were aligned using CLUSTAL OMEGA. **b** Ectodomain sequences for TMEM95 from human (*Homo sapiens*, NCBI: XP\_016880054), the marsupial brush tail possum (*Trichosurus vulpecula*, NCBI: XP\_036622079), the Chinese Pond Turtle (*Mauremys reevesii*, NCBI: XP\_039355851), the mimic poison frog (*Ranitomeya imitator*, NCBI: CAF5092145), and the catshark (*Scyliorhinus canicular*, NCBI: XP\_038642792) were aligned using CLUSTAL OMEGA. Strictly conserved residues are displayed as white text on a black background. Residues with conservative substitutions are in bold. Both alignments are numbered according to the human sequence.



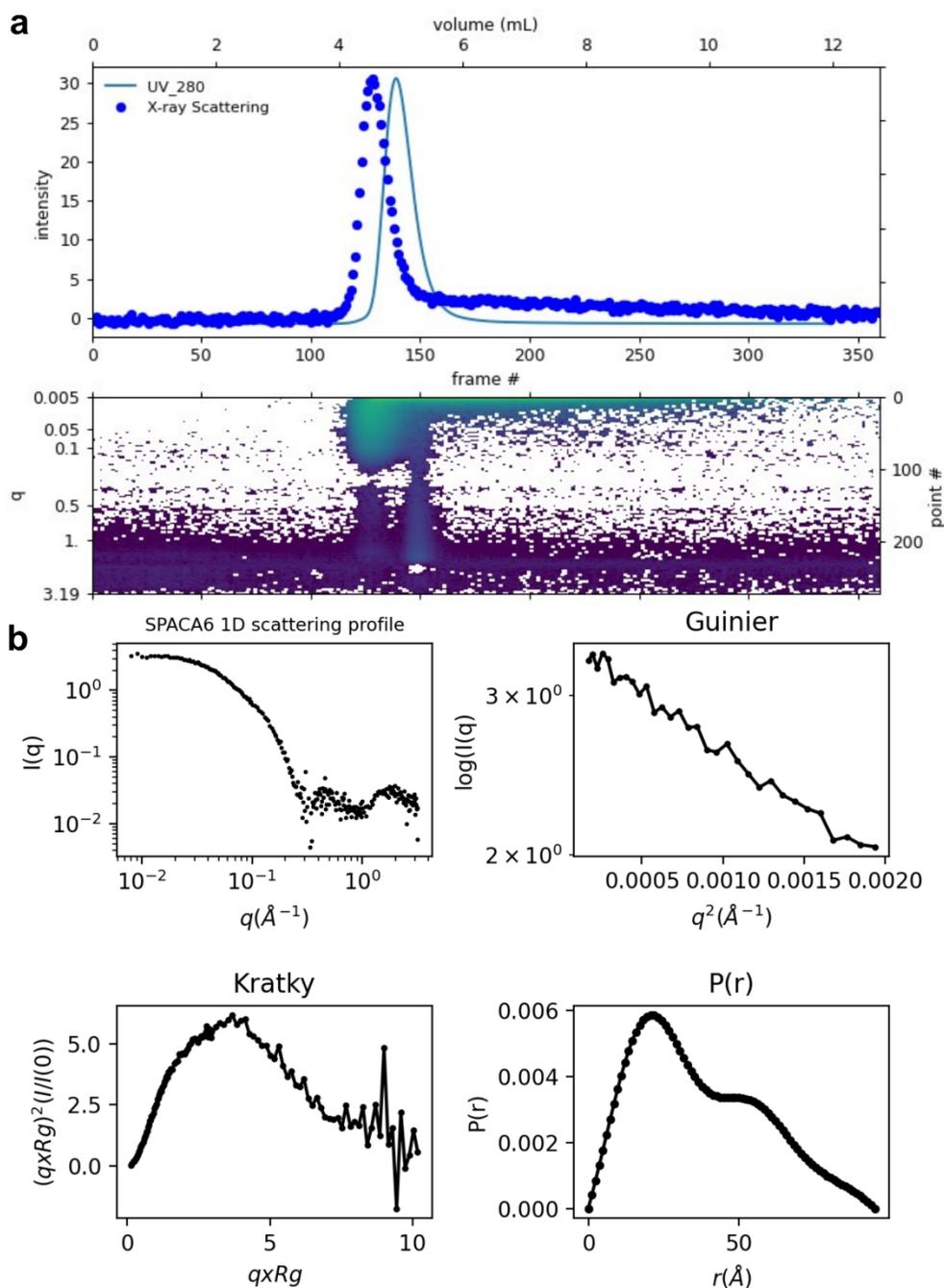
**Figure S14. SPACA6 does not bind to IZUMO1 and JUNO.** BLI sensorgrams of biotin-labelled **a** IZUMO1 or **b** JUNO as bait with high concentrations of JUNO, IZUMO1 or SPACA6 proteins as the analyte. Pull-down analyses of His-tagged **c** IZUMO1 or **d** JUNO as bait with equal concentrations of untagged SPACA6 as analyte protein. SDS-PAGE analyses stained with Coomassie brilliant blue are used to detect the protein. Both BLI and pull-downs results are representative of independent duplicate ( $n=2$ ) studies. **e** Positive control for pull-down assays. Coomassie stained SDS-PAGE (top) and Western blot of His-tagged JUNO incubated with untagged IZUMO1 (bottom). Mouse anti-His primary antibody was used with an HRP-conjugated anti-mouse secondary for detection of proteins. **f** Sequence alignments between the portions of IZUMO1 that interact with JUNO and their equivalent positions in SPACA6. IZUMO1 residues known to interact with JUNO are colored depending on the region in which they are found (green = hinge, orange = helix, blue = Ig-like). Cysteine residues are colored magenta. **g** Surface representation of the portion of IZUMO1 that interacts with JUNO and the equivalent position in SPACA6. Interacting surfaces are colored depending on the region in which they are found, as above. Residue names are colored depending on the biochemical characteristics of the residue (red = negatively charged, blue = positively charged, white = polar, black = non-polar).



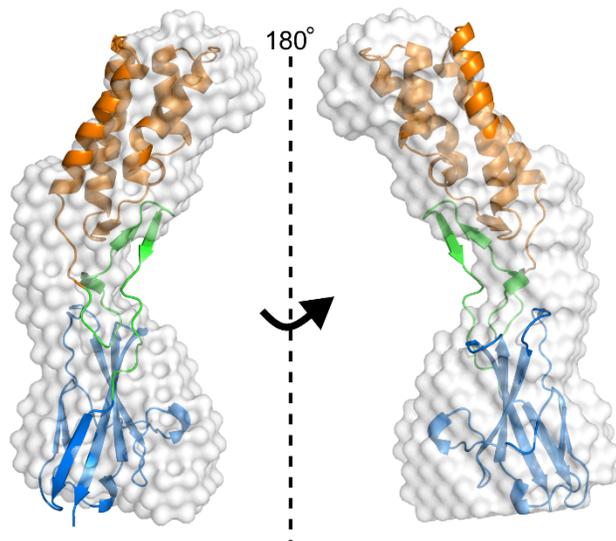
**Figure S15. Source data for pull-down assays.** Uncropped Coomassie stained SDS-PAGE source data for: **a** Supplementary Figure 14c, **b** Supplementary Figure 14d, and **c** Supplementary Figure 14e.



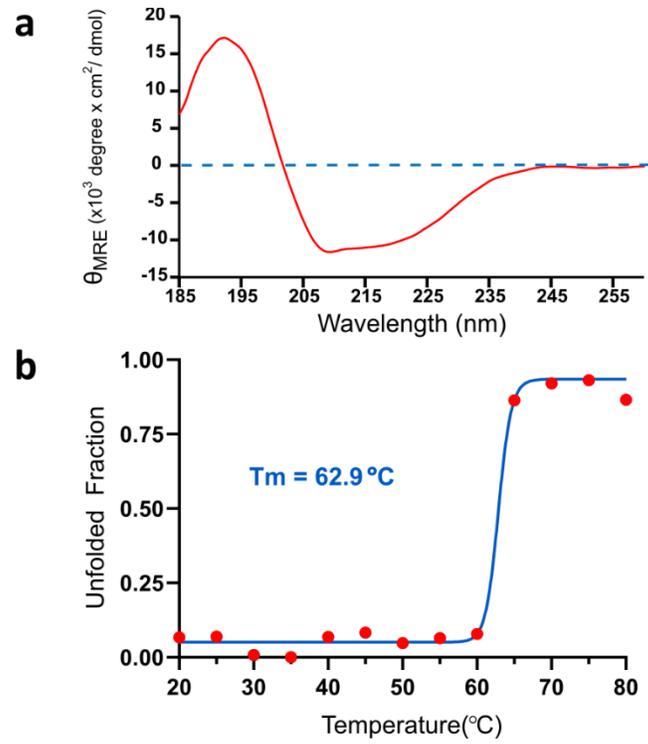
**Figure S16. Crystallographic evidence of IZUMO1 flexibility.** Alignments of the multiple IZUMO1 solved structures by the Ig-like domain to compare orientation of the 4HB; the smallest and largest angles between the bottom of the Ig-like domain, the center of the hinge region, and the tip of helix 4 in the 4HB are displayed. IZUMO1 structures include PDB: 5F4E (red), 5JKC (orange), 5JKD (yellow), 5JK9 (light yellow), 5B5K (green), 5F4T (cyan), and 5F4V (blue).



**Figure S17. SEC-SAXS data of SPACA6.** **a** Top panel shows X-ray scattering intensity (blue dots) and UV absorbance 280 nm trace (solid blue line) of SEC-SAXS elution profile (360 frames). Bottom panel represents a 2D heat map of the buffer subtracted X-ray intensity. Each vertical slice represents the X-ray intensity of the entire SAXS/WAXS  $q$  range from 0.005-3.1  $\text{\AA}^{-1}$  for each frame. **b** Average intensity of frames 90-100 was used to perform buffer subtraction on frames 122-125 of the SPACA6 elution peak. A  $R_g$  of 29.5  $\text{\AA}$  was calculated from the Guinier plot and  $D_{\max}$  of 95  $\text{\AA}$  was determined from the Distance Distribution function ( $P(r)$ ). Dimensionless Kratky plot indicates the protein is well folded but displays flexibility.



**Figure S18. SAXS reconstruction of SPACA6.** *Ab initio* reconstruction of SPACA6 ectodomain using SAXS data (white envelope) overlaid with the SPACA6 crystal structure, colored according to its three regions (4HB, orange; Ig-like domain, blue; hinge, green). DAMMIF was used to produce 20 envelope models, which were subsequently averaged with DAMAVER and filtered with DAMFILT.



**Figure S19. CD spectra and thermal melt of IZUMO1 ectodomain. a** Far-UV CD spectra of the IZUMO1 ectodomain at  $0.2 \text{ mg mL}^{-1}$ . **b** Thermal melt of IZUMO1 ectodomain in 10 mM sodium phosphate, pH 7.4, and 150 mM NaF.

**Table S1. Dali server search of PDB.**

Full Ectodomain*					
Protein Name	PDB	Z-score	RMSD	LALI	%ID
IZUMO SPERM-EGG FUSION PROTEIN 1	5jk9-A	12.1	3.9	130	15
DOWN SYNDROME CELL ADHESION MOLECULE	3dmk-A	9.2	6.7	92	13
MUCOSA-ASSOCIATED LYMPHOID TISSUE TRANSLOCATION PROTEIN	3bfo-A	8.8	1.7	75	15
COXSACKIE AND ADENOVIRUS RECEPTOR	3j6n-K	8.6	2.3	81	21
CELL SURFACE GLYCOPROTEIN CD200 RECEPTOR 1	4bf1-A	8.6	2.8	82	17
LEUCINE-RICH REPEAT AND IMMUNOGLOBULIN-LIKE DOMAIN	4oqt-A	8.6	13.4	118	14
INTERLEUKIN-33	5vi4-F	8.5	3.7	97	9
BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1	2cr3-A	8.5	3	80	18
CD2	1hng-A	8.3	2.2	78	13
ADVANCED GLYCOSYLATION END PRODUCT-SPECIFIC RECEPTOR	4lp5-A	8.2	2.7	80	15
Four-Helix Bundle†					
IZUMO SPERM-EGG FUSION PROTEIN 1	5jk9-A	7.2	3.5	116	8
DESIGNED HELICAL REPEAT PROTEIN	5cwc-A	6.5	2.9	86	14
PHOSPHOPROTEIN	3l32-B	6.2	1.2	43	5
55-KDA IMMEDIATE-EARLY PROTEIN 1	6tgz-F	6	5	95	7
DOLICHYL-DIPHOSPHOOLIGOSACCHARIDE PROTEIN	6s7o-E	5.7	3.7	66	9
SOLUBLE CYTOCHROME B562	4or2-B	5.5	4.5	68	7
GLUTAMYL-TRNA REDUCTASE 1	5che-A	5.5	6.7	62	3
DNA-DEPENDENT RNA POLYMERASE SUBUNIT RPO132	6rf1-C	5.5	7.5	56	4
SPOROZOITE MICRONEME PROTEIN	4u5a-B	5.5	3.2	90	9
RHUL123	4wid-A	5.4	7	94	7
Ig-like domain‡					
IZUMO SPERM-EGG FUSION PROTEIN 1	5jk9-A	11.9	3.5	113	18
DOWN SYNDROME CELL ADHESION MOLECULE	3dmk-A	9.2	4.4	89	12
INTERFERON ALPHA-5	3oq3-B	8.7	2.5	88	15
COXSACKIE AND ADENOVIRUS RECEPTOR	3j6n-K	8.6	2.4	81	20
MUCOSA-ASSOCIATED LYMPHOID TISSUE TRANSLOCATION PROTEIN	3bfo-A	8.6	1.7	74	15
BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1	2cr3-A	8.5	4.1	83	17
LEUCINE-RICH REPEAT AND IMMUNOGLOBULIN-LIKE DOMAIN	4oqt-A	8.5	8.5	88	17
CELL SURFACE GLYCOPROTEIN CD200 RECEPTOR 1	4bf1-A	8.5	2.9	81	20
TYROSINE-PROTEIN PHOSPHATASE	4yh7-B	8.5	2.5	89	18
IP13724P	6qp8-A	8.3	2.7	101	11

\* SPACA6 ectodomain structure, residues 27-246, was used to search the PDB25 subset of the Protein Data Bank.

† SPACA6 4HB + Hinge structure, residues 27-150, was used to search the PDB25 subset of the Protein Data Bank.

‡ SPACA6 Hinge + Ig-like domain structure, residues 121-246, was used to search the PDB25 subset of the Protein Data Bank.