# Main Manuscript for

SPACA6 structure reveals a conserved superfamily of gamete fusion-associated proteins.

Tyler D.R. Vance[1], Patrick Yip[1], Elisabet Jiménez[2], Sheng Li[3], Diana Gawol[1], James Byrnes[4], Isabel Usón[2,5], Ahmed Ziyyat[6,7], and Jeffrey E. Lee[1]*

[1]Department of Laboratory Medicine and Pathobiology, Temerty Faculty of Medicine, University of Toronto, Toronto ON, Canada.
[2]Institute of Molecular Biology of Barcelona (IBMB-CSIC), 08028 Barcelona, Spain.
[3]Department of Medicine, University of California San Diego, La Jolla, CA USA.
[4]National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY USA.
[5]ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain
[6]Université de Paris, Institut Cochin, INSERM, CNRS, Paris, France.
[7]Service d'Histologie, d'Embryologie, Biologie de la Reproduction, AP-HP, Hôpital Cochin, Paris, France.

*Correspondence: Jeffrey E. Lee
1 King's College Circle, Medical Sciences Building Room 6314, University of Toronto, Toronto ON Canada, M5S 1A8
Email: jeff.lee@utoronto.ca
1-416-978-1060

**Author Contributions:** J.E.L. and A.Z. conceived the project; T.D.R.V. and J.E.L. designed the project, discussed all results, and wrote the manuscript; T.D.R.V. collected the X-ray diffraction data, completed and refined the crystal structure, performed biochemical characterization, determined the SAXS reconstructions, and analyzed the structure; T.D.R.V. and P.Y. performed the expression, purification, and crystallization experiments; I.U. and E.J. obtained the initial ARCIMBOLDO SPACA6 model; S.L. performed the H/DXMS experiments; D.G. performed the pull-down experiments; J.B. collected the SEC-SAXS data; all authors edited and approve the manuscript.

**Competing Interest Statement:** All authors declare no competing financial or conflicts of interest.

**Classification:** Biological Sciences (Developmental Biology).

**Keywords:** Sperm-egg fusion, sperm-egg adhesion, fertilization, fusogen, sperm protein,

**This PDF file includes:**

Main Text
Figures 1 to 7

## Abstract

SPACA6 is a sperm-expressed surface protein that is critical for gamete fusion during mammalian sexual reproduction. Despite this fundamental role, little is known about how SPACA6 specifically functions. We elucidated the crystal structure of SPACA6 at 2.2-Å resolution, revealing a two-domain protein containing a four-helix bundle and Ig-like β-sandwich connected via a quasi-flexible linker. Based on the structural analysis, we propose SPACA6 is a founding member of a superfamily of gamete fusion-associated proteins, herein dubbed the IST superfamily. The IST superfamily is defined structurally by its distorted four-helix bundle and a pair of disulfide-bonded CXXC motifs. A structure-based search of the AlphaFold human proteome identified more protein members to this superfamily; remarkably, many of these proteins are linked to gamete fusion. The SPACA6 structure and its connection to other IST-superfamily members provide a missing link in our knowledge of mammalian gamete fusion.

## Significance Statement

SPACA6 is a human sperm protein vital for the fusion of gametes, though its exact function remains a mystery. We present the first solved structure of SPACA6: a two-domain fold comprised of an Ig-like domain and a distorted four-helix bundle. Dali searches of the PDB and AlphaFold reveal a family of structurally related proteins, several of which are also known to play a role in gamete fusion; as such, SPACA6 is a founding member of a conserved protein superfamily, dubbed the IST superfamily. Evolutionary analysis to ascertain functionally relevant structural elements in SPACA6 show a conservation of flexibility between the two domains and several conserved surfaces that could function as protein-protein interfaces.

## Main Text

### Introduction

Every human life begins with two separate haploid gametes: a sperm from the father and an oocyte from the mother. This sperm was the winner of an intensive selection process in which millions of sperm traversed the female reproductive tract, passed through various barriers (1), and underwent capacitation processes that augmented their motility and surface composition (2–4). Even once the sperm and oocyte have found each other, the process is not over. Oocytes are surrounded by a layer of cumulus cells as well as a glycoprotein barrier called the zona pellucida, both of which the sperm must pass through to gain access to the oocyte. Sperm use a combination of surface adhesion molecules and membrane-associated and secreted enzymes to break through these final barriers (5). These molecules and enzymes are predominantly stored within the inner membrane and acrosome matrix and are revealed through the dissolution of the outer membrane of the sperm during the acrosome reaction (6). The final step of this intensive journey is the sperm-egg fusion event, where the two cells merge their cell membranes and become a single diploid organism (7). Despite this process being a seminal one in human reproduction, little is known about the molecular interactions required.

Outside of gamete fertilization, the chemical process of fusing two lipid bilayers has been studied extensively. In general, membrane fusion is an energetically unfavorable process, requiring protein catalysts that undergo changes in structural conformation to draw two membranes close together, disrupt their continuity, and induce fusion (8, 9). Dubbed fusogens, these protein catalysts have been found in a myriad of fusion systems. They are necessary for viral entry into host cells (e.g., gp160 in HIV-1, spike in coronaviruses, hemagglutinin in influenza viruses) (10–12), the formation of the placenta (syncytins) (13–15), and in gamete fusion of lower eukaryotes (HAP2/GCS1 in plants, protists, and arthropods) (16–19). The fusogen for human gametes has yet to be discovered, although several proteins have been shown to be vital for gamete attachment and fusion (20). The

first discovered was the oocyte-expressed CD9, a transmembrane protein necessary for gamete fusion in both mice and humans (21–23). Though its exact function remains unclear, roles in adhesion and/or structuring of adhesive focal points on egg microvilli appear likely (24, 25). The two best characterized proteins critical for gamete fusion are the sperm protein IZUMO1 (26) and the oocyte protein JUNO (27), which bind each other as an essential step in gamete recognition and pre-fusion adhesion. *Izumo1* knockout male and *Juno* knockout female mice are completely infertile; in these models, sperms penetrate the perivitelline space, but gametes are unable to fuse (26, 27). Likewise, in human *in vitro* fertilization experiments, there is a reduction in fusion when gametes are treated with antibodies against IZUMO1 or JUNO (26, 28).

Recently, a collection of newly discovered, sperm-expressed proteins with similar phenotypes to IZUMO1 and JUNO have been discovered (20, 29–33). Sperm Acrosome Membrane-Associated protein 6 (SPACA6) was identified as essential in fertilization during a large-scale mutagenesis study in mice. Transgene insertion into the *Spaca6* gene produced sperm unable to fuse, although these sperm penetrated into the perivitelline space (34). Subsequent knockout studies in mice confirmed that *Spaca6* is essential for gamete fusion (29, 31). SPACA6 is expressed almost exclusively in the testis and has a localization pattern similar to that of IZUMO1, i.e. within the inner membrane of sperm prior to the acrosome reaction, followed by relocation to the equatorial region post-acrosome reaction (29, 31). Homologs to *Spaca6* are found in a variety of mammals and other eukaryotes (29), and its importance for human gamete fusion has been established through anti-SPACA6 inhibition of human *in vitro* fertilization (29). Unlike IZUMO1 and JUNO, the specifics regarding the structure, interactions, and function of SPACA6 remain unclear.

In the interest of better understanding the fundamental processes behind human sperm-egg fusion – thereby informing future advances in both family planning and infertility treatment – we undertook structural and biochemical studies of SPACA6. The crystal structure of the SPACA6 ectodomain revealed a four-helix bundle (4HB) and immunoglobulin-like (Ig-like) domain that are connected by a quasi-flexible region. Interestingly, the domain architecture of SPACA6 is similar to that of human IZUMO1, with both proteins sharing an uncommon motif: a 4HB with a triangular face of helices and pair of disulfide-bonded CXXC motifs. We propose that IZUMO1 and SPACA6 now define a larger, structurally related superfamily of gamete fusion-associated proteins. Using the hallmark features specific to the superfamily, we carried out an exhaustive search of the AlphaFold structural human proteome, revealing additional members of this superfamily that are all linked to gamete fusion. It now appears that there is a common structural fold and superfamily of proteins that are associated with gamete fusion, with our structure providing a molecular picture of this important aspect of the human gamete-fusion machinery.

## Results

### A soluble monomeric SPACA6 ectodomain

SPACA6 is a single-pass transmembrane protein with one N-linked glycan and six predicted disulfide linkages (*SI Appendix*, Fig. S1A and Fig. S2). We expressed the extracellular domain of human SPACA6 (residues 27-246) in *Drosophila* S2 cells and purified the protein using nickel-affinity, cation-exchange, and size-exclusion chromatographies (*SI Appendix,* Fig. S1B). The purified SPACA6 ectodomain was highly stable and homogeneous. Analysis with size-exclusion chromatography coupled multi-angle light scattering (SEC-MALS) revealed a single peak with a calculated molecular weight of 26.2 ± 0.5 kDa (*SI Appendix,* Fig. S1C). This is consistent with the size of a monomeric SPACA6 ectodomain, indicating that no oligomerization occurred during purification. Furthermore, circular dichroism (CD) spectroscopy revealed mixed α/β structure with

a melting temperature 51.3 °C (*SI Appendix,* Fig. S1D and S1E). Deconvolution of the CD spectra showed 38.6% α-helix and 15.8% β-strand elements (*SI Appendix,* Fig. S1D).

**SPACA6 has two domains: a four-helix bundle and an Ig-like β-sandwich**

The SPACA6 ectodomain was crystallized using a random matrix microseeding approach(35), yielding a 2.2-Å resolution dataset (*SI Appendix,* Fig. S3 and Table S1). The structure was determined using a combination of fragment-based molecular replacement and SAD phasing data from bromide soaks (*SI Appendix,* Fig. S4 and Table S1), with the final refined model consisting of residues 27-246. The SPACA6 ectodomain, with dimensions of 20 Å x 20 Å x 85 Å, is made up of seven helices and nine β-strands and adopts an elongated tertiary fold stabilized by six disulfide bonds (**Fig. 1A**). The structure consists of two domains: an N-terminal four-helix bundle (4HB) and a C-terminal Ig-like domain, with an intermediary hinge region between the two (**Fig. 1B**).

The 4HB domain of SPACA6 includes four main helices (Helices 1-4) arranged in a coiled-coil fashion (**Fig. 2A**) that alternate between antiparallel and parallel interactions (**Fig. 2B**). A small additional single-turn helix (Helix 1') packs perpendicularly with the bundle, forming a triangular shape with Helices 1 and 2. This triangle produces a slight distortion in the coiled-coil packing relative to the tight packing of Helices 3 and 4 (**Fig. 2A**).

The 4HB is centered around an internal hydrophobic core made up predominantly of aliphatic and aromatic residues (**Fig. 2C**). The core accommodates a disulfide bond between Cys41 and Cys55, which pinches Helices 1 and 2 together at the top, accentuating the triangular shape (**Fig. 2D**). Two additional disulfide bonds are formed between the CXXC motif of Helix 1' and another CXXC motif found at the tip of a β-hairpin in the hinge region (**Fig. 2D**). A conserved arginine residue (Arg37) of unknown function resides within the triangular hollow produced by Helices 1', 1 and 2. The $C_\beta$, $C_\gamma$, and $C_\delta$ aliphatic carbons of Arg37 interact with the hydrophobic core, and its guanidium group makes contacts with the loop between Helices 1' and 1 via Thr32 main and side chain interactions (*SI Appendix,* Fig. S5A and S5B). Tyr34 stretches over the hollow, leaving two small cavities through which Arg37 can interact with solvent.

Ig-like β-sandwich domains are a large superfamily of proteins that share the common characteristic of two or more multi-stranded, amphipathic β-sheets interacting via a hydrophobic core(36). The C-terminal Ig-like domain of SPACA6 follows this same pattern; it consists of two sheets (*SI Appendix,* Fig. S6A). Sheet 1 is a four-stranded β-sheet (Strands D, F, H, and I) in which strands F, H, and I form an anti-parallel arrangement, and strands I and D adopt a parallel interaction. Sheet 2 is a small antiparallel two-stranded β-sheet (Strands E and G). An internal disulfide bond is observed between the C-terminal end of Strand E and the center of Strand H (Cys170-Cys226) (*SI Appendix,* Fig. S6B). This disulfide bond is akin to those in β-sandwich domains from immunoglobulin proteins(37, 38).

The four-stranded β-sheet twists significantly throughout its length, producing asymmetric edges that are distinct in shape and electrostatics. The thinner edge presents a flat hydrophobic surface to the environment, which stands out against the rest of the uneven and electrostatically diverse surface in SPACA6 (*SI Appendix,* Fig. S6B and S6C). A halo of exposed backbone carbonyl/amino groups and polar side chains surrounds the hydrophobic surface (*SI Appendix,* Fig. S6C). The wider edge is partially covered by a capping coiled segment that blocks the N-terminal portion of the hydrophobic core and forms three hydrogen bonds with the exposed backbone polar groups of Strand F (*SI Appendix,* Fig. S6D). The C-terminal portion of this edge produces a large pocket with a partially exposed hydrophobic core. The pocket is surrounded by positive charges due to three

sets of dual arginine residues (Arg162-Arg221, Arg201-Arg205, and Arg212-Arg214) and a central histidine (His220) (*SI Appendix,* Fig. S6E).

**SPACA6 hinge region connects and orients the two domains**

The hinge region is a short segment between the helical and Ig-like domains that is made up of a single antiparallel three-stranded β-sheet (Strands A, B, and C), a small $3_{10}$ helix, and several long random coil segments (*SI Appendix,* Fig. S7). A network of covalent and electrostatic contacts in the hinge region appear to stabilize the orientation between the 4HB and Ig-like domains. This network can be broken up into three sections. The first section involves the two CXXC motifs ($^{27}$CXXC$^{30}$ and $^{139}$CXXC$^{142}$) that form a pair of disulfides bonds between a β-hairpin in the hinge and Helix 1' in the 4HB. The second section involves an electrostatic interaction between the Ig-like domain and the hinge. Glu132 in the hinge forms salt bridges to Arg233 in the Ig-like domain and Arg135 in the hinge. The third section involves a covalent linkage between the Ig-like domain and the hinge region. Two disulfide bonds (C124-C147 and C128-C153) connect a loop in the hinge region, which is stabilized by electrostatic interactions between Gln131 and main chain functional groups, to a linker that leads into the first strand of the Ig-like domain.

**SPACA6 is structurally similar to IZUMO1**

The SPACA6 ectodomain structure and the separate 4HB and Ig-like domain structures were used to search the Protein Data Bank for structurally similar entries (39). We identified matches with high Dali Z-scores, small root mean square deviations, and large LALI scores (the latter indicates the number of structurally equivalent residues). While the top 10 hits from the full ectodomain search (SI Appendix, Table S2) have reasonable Z scores of >8 (39), searches of only the 4HB or Ig-like domains revealed that the majority of these hits align only to the β-sandwich, a ubiquitous fold in many proteins. Only one hit was present in all three Dali searches: IZUMO1.

The structure of SPACA6 bears a striking similarity to IZUMO1 (40–42). Both SPACA6 and IZUMO1 share a two-domain architecture (*SI Appendix,* Fig. S8A) with similar 4HB and Ig-like β-sandwich domains connected by a hinge region (*SI Appendix,* Fig. S8B). Although the specifics of each domain vary – as expected for proteins with 21% sequence identity (*SI Appendix,* Fig. S9A) – the disulfide bonding patterns in the ectodomain are conserved. SPACA6 does have a pair of cysteine residues that are absent in IZUMO1: the aforementioned Cys41 and Cys55 that covalently link Helices 1 and 2 in the SPACA6 4HB.

Previous studies have noted the potential for structural similarities between SPACA6 and IZUMO1 (7, 31, 41) – an early attempt at a homology model even predicted an N-terminal 4HB in mouse SPACA6 (41). Our structure reveals the true extent of this similarity, especially in the deviations from traditional helical bundles. The canonical 4HB, like those found in the SNARE protein complexes involved in endosomal fusion (43, 44), has uniformly distanced helices that maintain a consistent curvature around a central axis (45). In contrast, the coiled-coil domains in both IZUMO1 and SPACA6 are distorted with inconsistent curvatures and uneven packing (*SI Appendix,* Fig. S9B). The distortion is likely caused by the triangular shape formed by Helices 1', 1, and 2, conserved in IZUMO1 and SPACA6 and stabilized by the same CXXC motif on Helix 1'. However, the extra disulfide found in SPACA6 produces a much sharper vertex at the top of the triangle, making SPACA6 even more distorted than IZUMO1 with a more pronounced cavity at the center of the triangle. In addition, IZUMO1 lacks the Arg37 observed in the center of this cavity in SPACA6. IZUMO1 instead has a more typical hydrophobic core of aliphatic and aromatic residues.

IZUMO1 has an Ig-like domain composed of a two-stranded and a five-stranded β-sheet (40). The extra strand in IZUMO1 replaces the coil in SPACA6 that interacts with Strand F to cap the backbone hydrogen bonds in the strand. An interesting point of comparison is in the predicted surface charges for the Ig-like domains of these two proteins. The IZUMO1 surface is more negatively charged than that of SPACA6. The additional charges are located near the C-terminal end, which faces the sperm membrane. In SPACA6, the same areas are more neutral or positively

charged (*SI Appendix,* Fig. S9C). For example, both the hydrophobic surface (thinner edge) and positively charged pocked (wider edge) in SPACA6 are negatively charged in IZUMO1.

Whereas connectivities and secondary structure elements are well conserved between IZUMO1 and SPACA6, a structural alignment of the Ig-like domains revealed that the overall orientations of the two domains relative to each other are different (*SI Appendix,* Fig. S10). The helical bundle of IZUMO1 is bent relative to the β-sandwich, producing a previously described "boomerang" shape that deviates by about 50° from the central axis (40). In contrast, the helical bundle in SPACA6 has an approximately 10° lean in the opposite direction. These differences in orientation likely result from differences within the hinge region. At the primary sequence level, IZUMO1 and SPACA6 share almost no sequence similarity in the hinge save for the cysteine residues, a glycine, and an aspartate. As a result, the hydrogen-bonding and electrostatic networks are completely different. The secondary structure element of the β-sheet is shared between IZUMO1 and SPACA6, although the strands are much longer in IZUMO1, and the $3_{10}$ helix (Helix 5) is unique to SPACA6. These discrepancies result in different domain orientations of the two otherwise similar proteins.

**SPACA6 and IZUMO1 are founding members of a conserved protein superfamily**

Our Dali server search revealed that SPACA6 and IZUMO1 are the only two experimentally determined structures deposited in the Protein Data Bank that share this particular 4HB fold (*SI Appendix,* Table S2). Recently, DeepMind (Alphabet/Google) developed AlphaFold, a neural network-based system that accurately predicts protein 3D structure from a primary sequence (46). Shortly after we solved the SPACA6 structure, the AlphaFold Database was released, providing predicted structural models that cover 98.5% of all proteins in the human proteome (46, 47). Using our solved structure of SPACA6 as a search model, structural homology searching of the models in the AlphaFold human proteome identified candidates with structures potentially similar to those of SPACA6 and IZUMO1.

Previously, PSI-BLAST searches clustered IZUMO1 with three other sperm-associated proteins: IZUMO2, IZUMO3, and IZUMO4 (48). AlphaFold predicts that these IZUMO-family proteins fold into 4HB domains with the same disulfide patterns as IZUMO1 (**Fig. 3A** and *SI Appendix,* Fig. S11), though they lack the Ig-like domain. IZUMO2 and IZUMO3 are predicted to be single-pass membrane proteins like IZUMO1, whereas IZUMO4 appears to be secreted. Functions of IZUMO proteins 2, 3, and 4 in gamete fusion have not been established. IZUMO3 is known to play a role in the biogenesis of the acrosome during sperm development (49), and the IZUMO proteins have been observed to form complexes (48). The conservation of IZUMO proteins in mammals, reptiles, and amphibians signals a potential function aligned with those of other known gamete fusion-associated proteins like DCST1/2, SOF1, and FIMP.

Unlike IZUMO proteins, the other SPACA proteins (i.e., SPACA1, SPACA3, SPACA4, SPACA5, and SPACA9) are predicted to be structurally divergent from SPACA6 (*SI Appendix,* Fig. S12). Only SPACA9 has a 4HB, but it is not predicted to have the same parallel-antiparallel orientation as SPACA6 or the same disulfide linkages. Only SPACA1 has a similar Ig-like domain. SPACA3, SPACA4, and SPACA5 are predicted by AlphaFold to have completely different structures from SPACA6. Interestingly, SPACA4 is also known to play a role in fertilization but further upstream than SPACA6, instead aiding in the interactions between sperm and the oocyte zona pellucida (50).

Another match to the SPACA6 4HB, as predicted by AlphaFold, is TMEM95. TMEM95 contains the pair of CXXC motifs and the additional disulfide between Helices 1 and 2 (**Fig. 3A** and *SI Appendix,* Fig. 11). Whereas TMEM95 lacks an Ig-like domain, it has a region with the same disulfide bonding patterns as the hinge regions of both SPACA6 and IZUMO1 (**Fig. 3B**). Interestingly, TMEM95 is a sperm-specific, single-pass transmembrane protein that when ablated leaves male mice infertile (31, 32). Sperm lacking TMEM95 have normal morphology, motility, and ability to penetrate the zona pellucida and bind the oolemma but are not able to fuse with oocyte membranes. TMEM95,

much like SPACA6 and IZUMO1, is evolutionary conserved as far back as amphibians (**Fig. 4** and *SI Appendix,* Fig. S13).

Thus, the striking overall structural similarities between SPACA6 and IZUMO1 suggests that these are the founding members of a conserved structural superfamily of gamete fusion-associated proteins that includes TMEM95 and IZUMO proteins 2, 3, and 4. We propose the name *IST* superfamily after the initials of the three members known to be associated with gamete fusion so far: IZUMO1, SPACA6, and TMEM95. As only certain members possess an Ig-like domain, the hallmark feature of the IST superfamily is the 4HB domain, which has unique characteristics shared by all these proteins: 1) the distorted 4HB has helices packed in an alternating anti-parallel/parallel fashion (**Fig. 5A**), 2) the bundle has a triangular face made from two helices within the bundle and a third perpendicular helix (**Fig. 5B**), and 3) a double CXXC motif connects the perpendicular helix in the 4HB to a flexible hinge region via dual disulfide bonds (**Fig. 5C**). The CXXC motif, found in thioredoxin-like proteins, are known to act as redox sensors (51–53), and the motifs in IST family members might be linked to the role protein disulfide isomerases like ERp57 play in gamete fusion (54, 55).

**SPACA6 ectodomain does not bind IZUMO1 or JUNO**

Given the similarities between SPACA6 and IZUMO1, the ability of the former to bind to either IZUMO1 or JUNO was tested. Biolayer interferometry (BLI) is a kinetics-based binding technique that was used previously to quantify the interaction between IZUMO1 and JUNO. Upon incubation of a biotin-labelled, sensor-bound IZUMO1 as bait with high concentrations of the JUNO analyte, a strong signal was detected (*SI Appendix,* Fig. S14A), indicating a binding-induced change in the thickness of the biomaterial attached to the sensor tip. A similar signal was detected with the inverse experiment (i.e., sensor-bound JUNO as bait against IZUMO1 analyte) (*SI Appendix,* Fig. S14B). No signal was detected when SPACA6 was used as the analyte against either sensor-bound IZUMO1 or sensor-bound JUNO (*SI Appendix,* Fig. S14A and S14B). This lack of signal provides evidence that the SPACA6 ectodomain does not interact with the ectodomains of IZUMO1 or JUNO.

Since the BLI as an assay relies on biotinylation of free lysine residues on the bait protein, this modification may prevent binding if lysine residues are involved in the interaction. In addition, the binding orientation relative to the sensor may create steric hindrances; thus, traditional pull-down assays were also performed with recombinant SPACA6, IZUMO1, and JUNO ectodomains. Regardless, SPACA6 was not precipitated with either His-tagged IZUMO1 or His-tagged JUNO (*SI Appendix,* Fig. S14C and S14D), indicating an agreement with the lack of interaction witnessed in BLI experiments. As a positive control, we confirmed interaction of JUNO with His-tagged IZUMO1 (*SI Appendix,* Fig. S15 and Fig. S16).

**SPACA6 surface has three patches of highly conserved residues**

Despite the known necessity of SPACA6 for gamete fusion and its similarity to IZUMO1, SPACA6 does not appear to perform the equivalent function of binding JUNO. Therefore, we sought to combine our structural data with evidence of importance provided by evolutionary biology. Sequence alignments of the SPACA6 homologs suggest a conservation of the general structure beyond mammals. For example, the cysteine residues are present even in distantly related amphibian animals (**Fig. 6A**). Using the ConSurf server, the conservation data from a multiple-sequence alignment of 66 sequences was mapped onto the surface of SPACA6. This type of analysis can reveal those residues that have been maintained throughout the protein's evolution and can suggest which surface areas play a role in function.

The SPACA6 structure has three highly conserved surface patches (**Fig. 6B**). Patch 1 spans the 4HB and the hinge region, and contains the two conserved CXXC disulfide bridges, the Arg233-Glu132-Arg135-Ser144 hinge network (*SI Appendix,* Fig. S7), as well as three outward facing conserved aromatic residues (Phe31, Tyr73, Phe137). Patch 2 encompasses the wider edge of the

Ig-like domain (*SI Appendix,* Fig. S6E), which presents several positively charged residues toward the sperm surface. Interestingly, this patch holds an antibody epitope previously shown to prevent SPACA6 from functioning (29). Patch 3 spans the hinge and one side of the Ig-like domain; this region has conserved prolines (Pro126, Pro127, Pro150, Pro154) and outward facing polar/charged residues. Strangely, the majority of the residues on the 4HB surface are quite variable (**Fig. 6B)**, despite the fold's conservation throughout the SPACA6 homologs (as indicated by the bundle's hydrophobic core being conserved) and beyond into the IST superfamily.

**Conformational dynamics of SPACA6**

Although it is the smallest region of SPACA6 with the fewest definable secondary structure elements, many hinge region residues (including Patch 3) are highly conserved amongst SPACA6 homologs, perhaps indicating that the orientation of the helical bundle and β-sandwich serves a conserved purpose. Yet, despite the extensive hydrogen bonding and electrostatic networks within the hinge regions of both SPACA6 and IZUMO1, evidence of inherent flexibility can be seen in an alignment of the multiple solved IZUMO1 structures (40–42). Alignments of the individual domains overlap well, but the orientation of the domains relative to each other varies between 50° and 70° from the central axis (*SI Appendix,* Fig. S17). To understand the conformational dynamics of SPACA6 in solution, SAXS experiments were performed (*SI Appendix,* Fig. S18A and S18B). *Ab initio* reconstructions of the SPACA6 ectodomain were consistent with the rod-like crystal structure (*SI Appendix,* Fig. S19), though the Kratky plot reveals a level of flexibility (*SI Appendix,* Fig. S18B). This conformation contrasts with IZUMO1 where the unbound protein adopts a boomerang shape in the crystal lattice and in solution (40).

To specifically identify regions of flexibility, hydrogen-deuterium exchange mass spectrometry (H-DXMS) was performed on SPACA6 and compared to previously acquired data on IZUMO1 (40) (**Fig. 7A and 7B**). SPACA6 is clearly more flexible than IZUMO1, as shown by the higher deuterium exchange over the entire structure after 100,000 seconds of exchange. In both structures, the C-terminal portions of the hinge region show high levels of exchange, likely allowing for limited pivoting of the 4HB and Ig-like domains relative to each other. Interestingly, the C-terminal portions of hinge section in SPACA6, comprised of residues $^{147}$CDLPLDCP$^{154}$, is the highly conserved Patch 3 (**Fig. 6B**). In agreement with the analysis of flexibility, CD thermal melt data showed that SPACA6 ($T_m$ = 51.2 °C) was less stable than IZUMO1 ($T_m$ = 62.9 °C) (*SI Appendix,* Fig. S1E and Fig. S20).

## Discussion

The use of CRISPR/Cas9 and genetic mouse knockout strategies have led to the identification of several factors important for sperm-egg engagement and fusion (21–23, 26, 27, 29–33). With the exceptions of the well-characterized IZUMO1-JUNO interaction and the structure of CD9, most gamete fusion-associated proteins remain a mystery in terms of both structure and function. The biophysical and structural characterization of SPACA6 is another piece in the molecular puzzle of adhesion/fusion during fertilization.

SPACA6 and its fellow members of the IST superfamily appear to be highly conserved in mammals, as well as select birds, reptiles, and amphibians; indeed, SPACA6 is even known to be essential for fertilization in zebrafish (56). This distribution is similar to those of other known gamete fusion-associated proteins such as DCST1 (33), DCST2 (33), FIMP (30), and SOF1 (31), suggesting that these factors are part of a conserved molecular mechanism for fertilization used by the higher eukaryotes that lack the HAP2 (also known as GCS1) protein, which is the fusion protein responsible for catalyzing fertilization in many protists, plants, and arthropods (57, 58). Despite the strong structural similarity between SPACA6 and IZUMO1, knockouts of genes encoding either protein alone results in infertility of male mice, indicating that they are not redundant in their functions in gamete fusion (29). More broadly, none of the known sperm proteins essential at the adhesion stage of fusion are redundant.

An open question is whether proteins in the IST superfamily participate in inter-gamete connections, form intra-gamete networks that recruit essential proteins to the point of fusion, or perhaps even function as the elusive fusogen. Co-immunoprecipitation studies in HEK293T cells suggested an interaction between IZUMO1 and SPACA6(31); however, our recombinant ectodomains failed to interact *in vitro*, indicating that interactions may be mediated through the transmembrane helices or linkers, or require a particular context that we do not reproduce *in vitro*, such as a physiologically specific conformation or a molecular complex containing other proteins (known or yet to be discovered). Although the IZUMO1 ectodomain is thought to mediate an attachment of the sperm to the egg in the perivitelline space, the purpose of the SPACA6 ectodomain is unknown. The structure of SPACA6 reveals several areas that could participate in protein-protein interactions. For example, the wide edge of the Ig-like domain forms a positively charged trough with highly conserved Arg and His residues, and an antibody against this region was previously used to block gamete fusion (29). The antibody recognizes the linear epitope $^{212}$RIRPAQLTHRGTFS$^{225}$, which possesses three of the six arginine residues and the highly conserved His220. Whether the function disruption resulted from occluding these specific residues or the region as a whole is unclear. The position of this cleft near the C-terminal end of the β-sandwich would suggest a *cis*-interaction with an adjacent sperm protein rather than an interaction with an oocyte protein.

SPACA6 possesses characteristics of cell-cell adhesion proteins including the Ig-like β-sandwich. Many adhesion proteins (e.g., cadherins, integrins, adhesin, and IZUMO1) have one or more β-sandwich domains that extend the protein away from the cell membrane and toward its environmental target (59–61). The SPACA6 Ig-like domain also contains a motif common amongst adhesion and cohesion β-sandwiches: a doublet of parallel strands at the termini of the β-sandwich known as a mechanical clamp (62). This motif is thought to impart increased resistance to shear forces, valuable for proteins involved in interactions between cells. Yet despite these similarities to adhesins, no evidence currently exists of SPACA6 interacting with ovum proteins. The SPACA6 ectodomain is unable to bind to JUNO, as shown here, and SPACA6-expressing HEK293T cells do not form a substantial interaction with zona pellucida-free oocytes (31). If SPACA6 does make inter-gamete connections, these interactions may require post-translational modifications or be stabilized by other sperm proteins. In support of the latter hypothesis, sperm lacking IZUMO1 bind to oocytes, demonstrating that molecules other than IZUMO1 are involved in gamete adhesion step (26).

Many viral, cellular, and developmental fusion proteins have characteristic features that are predictive of their function as fusogens. For example, the viral fusion glycoproteins (class I, II, and III) have hydrophobic fusion peptides or loops at the tip of the protein that insert into the host membrane. Hydropathy plots of IZUMO1 (40) and structures (both solved and predicted) for the IST superfamily reveal no obvious hydrophobic fusion peptides. Thus, if any of the proteins in the IST superfamily function as fusogens, they do so in a manner distinct from other known examples. In summary, the functions of the members of the IST superfamily of gamete fusion-associated proteins remains an enticing mystery. The structure of SPACA6 provides insight into the next steps that will connect these shared structures to gamete attachment and fusion.

## Materials and Methods

### Cloning and expression of recombinant SPACA6 ectodomain

The DNA sequence corresponding to the predicted ectodomain of human SPACA6 (NCBI Accession Number NP_001303901.1; residues 27-246) was codon optimized for expression in *Drosophila melanogaster* S2 cells and synthesized as a gene fragment (Eurofins Genomics) with an encoded Kozak sequence, BiP secretion signal, and appropriate 5' and 3' overhangs for ligation-independent cloning of the gene into the metallothionein promoter-based pMT expression vector

modified for puromycin selection (pMT-puro). The pMT-puro vector encodes for a thrombin cleavage site followed by a C-terminal 10x-His tag (*SI Appendix,* Fig. S2).

Stable transfection of the SPACA6 pMT-puro vector into *D. melanogaster* S2 cells (Gibco) was performed similar to protocols used for IZUMO1 and JUNO(40). Briefly, S2 cells were thawed and grown in Schneider's medium (Gibco) supplemented with a final concentration of 10% (v/v) heat-inactivated fetal bovine serum (Gibco) and 1X antibiotic-antimycotic (Gibco). Early passage cells were seeded (3.0 x$10^6$ cells) into a single well of a 6-well plate (Corning). After 24 h incubation at 27 °C, the cells were transfected with a mixture of 2 mg SPACA6 pMT-puro vector and Effectene transfection reagent (Qiagen), according to the manufacturer's protocol. Transfected cells were incubated for 72 h before the addition of 6 mg $mL^{-1}$ puromycin for selection. Subsequently, the cells were weaned off complete Schneider's medium and into serum-free Insect-XPRESS (Lonza) for large-scale protein production. Cultures of S2 cells in 1-L batches were grown to between 8-10 x $10^6$ cells $mL^{-1}$ in 2-L vented flat-bottom polypropylene Erlenmeyer flasks and then induced with a final concentration of sterile-filtered 500 µM $CuSO_4$ (Millipore Sigma). Induced cultures were incubated for four days at 27 °C with shaking at 120 rpm.

**Purification of recombinant SPACA6 ectodomain**

The SPACA6-containing conditioned media was separated from cells by centrifugation at 5660 x *g* at 4 °C prior to concentration and buffer exchange into Buffer A (10 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 20 mM imidazole) using a Centramate tangential flow filtration system (Pall Corp) with a 10-kDa MWCO membrane. The concentrated SPACA6-containing media was applied onto a 2-mL column of Ni-NTA agarose resin (Qiagen). The Ni-NTA resin was washed with 10 column volumes (CV) Buffer A and then 1 CV of Buffer A supplemented with a final concentration of 50 mM imidazole. SPACA6 was eluted with 10 mL of Buffer A supplemented with a final concentration of 500 mM imidazole. Restriction-grade thrombin (Millipore Sigma) was added at 1 unit per mg of SPACA6 directly into dialysis tubing (12-14 kDa MWCO) and dialyzed against 1-L of 10 mM Tris-HCl, pH 7.5, and 150 mM NaCl (Buffer B) for 48 h at 4 °C. Thrombin-cleaved SPACA6 was then diluted three-fold to reduce the salt concentration prior to loading onto a 1-mL MonoS 5/50 GL cation-exchange column (Cytiva/GE) equilibrated with 10 mM Tris-HCl, pH 7.5. The cation exchanger was washed with 3 CV 10 mM Tris-HCl, pH 7.5 prior to eluting SPACA6 using a linear gradient of 0 to 500 mM NaCl in 10 mM Tris-HCl, pH 7.5 over 25 CV. Following ion-exchange chromatography, SPACA6 was concentrated to 1-mL and isocratically eluted from an ENrich SEC650 10x300 column (BioRad) equilibrated with Buffer B. Based on the chromatogram, fractions containing SPACA6 were pooled and concentrated. Purity was monitored via Coomassie-stained 16% SDS-polyacrylamide gel electrophoresis. Protein concentration was quantified based on absorbance at 280 nm using the Beer-Lambert Law and theoretical molar extinction coefficient.

**Circular dichroism spectroscopy**

Purified SPACA6 was dialyzed overnight into 10 mM sodium phosphate, pH 7.4, and 150 mM NaF and diluted to 0.16 mg $mL^{-1}$ prior to analysis by CD spectroscopy. CD wavelength spectral scans were collected at 25 °C between 185 to 260 nm at a rate of 50 nm/min using a 1-mm pathlength quartz cuvette (Helma) in a Jasco J-1500 spectropolarimeter. CD spectra were baseline corrected, averaged over 10 accumulations, and converted to mean residue ellipticity ($\theta_{MRE}$) in units of degree $cm^2 \cdot dmol^{-1}$:

$$\theta_{MRE} = \left(\frac{MW}{N-1}\right).\left(\frac{\theta}{10.d.c}\right)$$

where *MW* is the molecular weight for each sample in Da; *N* is the number of amino acids; $\theta$ is the ellipticity in millidegree; *d* corresponds to the optical pathlength in cm; and *c* is the protein concentration in mg $mL^{-1}$.

Thermal denaturation assays were performed at a wavelength of 207 nm by increasing the temperature from 20 to 80 °C in 5 °C intervals with 2-min equilibration between temperature points. Four scans were taken per temperature point, averaged, and baseline corrected. The resultant change in ellipticity was normalized between 0 (folded) and 1 (unfolded) and fit to a non-linear biphasic sigmoidal curve using the program GraphPad (version 8.4.3) to determine the apparent melting temperature ($T_m$). Secondary structural composition (helices, strands, turns and coils) was estimated using the CONTIN method and the Set3 library (185-240 nm) using the Dichroweb server (63).

**SEC-MALS analysis**

The oligomeric state of tag-removed, fully glycosylated SPACA6 was assessed by SEC-MALS. 0.14 mg Bovine Serum Albumin (BSA) and 0.14 mg SPACA6 were prepared in 1X PBS at a concentration of 1.2 mg $mL^{-1}$. A Superdex 75 10/300 GL size-exclusion column (Cytiva/GE) was equilibrated overnight with 5 CV PBS. Monomeric BSA (MW = 66,432 Da) was used as a reference calibration standard. Prior to SEC-MALS analysis, each sample was centrifuged at 15000 x *g* for 15 min at 4 °C and then the supernatant was loaded onto the size-exclusion column on an AKTA Pure FPLC (Cytiva) at 0.2 mL $min^{-1}$. Triple detection was performed by measuring absorbance at 280 nm using the integrated UV monitor on the AKTA Pure, three-angle light scattering using the miniDAWN TREOS MALS detector (Wyatt) and refractive index (RI) using Optilab T-rEX RI detector (Wyatt). The data were processed, and weight-averaged molecular mass was calculated using the ASTRA software package (version 7.0.2.11).

.

**Sparse matrix crystallization and data collection**

Sitting-drop sparse matrix crystallization screening was performed using a Douglas Instrument Oryx8 liquid handling system. SPACA6 concentrated to 7 mg $mL^{-1}$ in Buffer B was mixed in a 1:1 volumetric ratio (0.3:0.3 mL) with JCSG+ (Qiagen), Cryos (Qiagen), and PACT (Qiagen) sparse matrix crystallization screens in 96-well 2-drop Art Robbins Intelliplates. Protein crystal clusters were obtained when SPACA6 was mixed with 100 mM Bis-Tris, pH 5.5, 200 mM NaCl, 25% (w/v) PEG 3350. To improve crystal morphology, the crystal clusters were crushed into microcrystals using the Seed Bead kit (Hampton Research) and used for random microseed matrix screening(35). SPACA6 (7 mg $mL^{-1}$), microseeds, and sparse matrix crystallization conditions from the JCSG+, and Cryos suites were mixed in a 3:1:2 ratio. From this procedure, larger clusters were consistently obtained in 200 mM ammonium chloride and 20% (w/v) PEG 3350. As initial X-ray diffraction was poor, a detergent screen (Hampton Research) was used to identify conditions that yielded strongly diffracting crystals. The crystals used to solve the structure were obtained by the microseeding protocol using a precipitant solution composed of 200 mM ammonium chloride, 20% (w/v) PEG 3350, and 34 mM FOS-choline-8. Crystals were looped and coated with perfluoropolyether cryo oil (Hampton Research) before being flash cooled in liquid nitrogen (100 K). Multiple native datasets were collected at the AMX (17ID-1) beamline at the National Synchrotron Light Source-II (NSLS-II) at Brookhaven National Laboratory (Upton, NY). Data were indexed and integrated using DIALS (64) and averaged and scaled using Aimless in the CCP4 suite (65).

**Structure determination**

Attempts to solve the structure of SPACA6 via SAD phasing were made, with multiple datasets collected at the NE-CAT (24-ID-C) beamline at the Advanced Photon Source (Argonne National Laboratory, Lemont, IL); phasing attempts with tantalum clusters and/or intrinsic sulfurs were fruitless. Eventually, initial phases were determined using the human IZUMO1 model (PDB: 5JK9) as a starting template for ARCIMBOLDO_SHREDDER (66, 67). ARCIMBOLDO_SHREDDER carried out expected log-likelihood gain-guided molecular replacement of template-generated fragments using the program Phaser (68). During molecular replacement, additional refinement, such as gyre refinement against the rotation function, and gimble refinement after translation, was employed to refine the internal geometry of the fragments, allowing additional degrees of freedom by subdivision in rigid groups (69). Consistently placed and refined fragments were combined in

reciprocal space with ALIXE (70). Best-scored phase sets were subject to density modification and autotracing in SHELXE (71). The solution from ARCIMBOLDO_SHREDDER provided clear density for ~65% of the structure, successfully built with a *CC* of 41.6%. X-ray data collection statistics are presented in *SI Appendix,* Table S1.

To complete the model, Br-SAD phasing was performed and combined with the phases from the ARCIMBOLDO_SHREDDER model. SPACA6 crystals were soaked in 200 mM ammonium chloride, 20% (w/v) PEG 3350, 34 mM FOS-choline-8, 25% (v/v) glycerol, and 1 M sodium bromide for 10 s before flash cooling directly into liquid nitrogen. Data were collected on the NE-CAT (24-ID-C) beamline at the Advanced Photon Source (Argonne National Laboratory, Lemont, IL). Data were indexed and integrated using DIALS (64), followed by scaling and merging using Aimless from the CCP4 suite (65). X-ray data collection statistics are presented in *SI Appendix,* Table S1. The Br-SAD dataset is not isomorphous to the native dataset, likely due to cryo solution-induced dehydration of the crystal. PHENIX_Autosol was used to carry out a combination of molecular replacement and SAD phasing (72, 73). The atomic positions of ten bromide ions were found using a hybrid substructure search in PHENIX_Autosol. Model phases calculated from a molecular replacement solution using the ARCIMBOLDO model was combined with the anomalous signal from the Br-SAD data. PHENIX_AutoBuild was used to build the majority (~80%) of the SPACA6 model. The final PHENIX_AutoBuild model was subjected to an initial round of simulated annealing torsion angle refinement starting at 5000 K using PHENIX_refine (74). The complete SPACA6 model was built with iterative rounds of manual model building followed by coordinate refinement using Coot (75) and PHENIX_refine (74), respectively. Riding hydrogens and Translation/Liberation/Screw groupings were included to improve the structure. The program PDB-REDO (76) was used to provide a final geometry clean up prior to model validation using Ramachandran, geometry, and rotamer analysis, difference map peak searches in Coot (75), and MolProbity (77) via PHENIX. All structural images were generated using The PyMOL Molecular Graphics System (Version 2.0 Schrödinger, LLC.). Electrostatic calculations were done using the APBS web server (78).

**Biolayer Interferometry**

The binding affinities of SPACA6 to IZUMO1 and JUNO were measured by BLI using a single-channel BLItz instrument (FortéBio/Sartorius). Purified JUNO and IZUMO1 in PBS were biotinylated using the EZ link sulfo-NHS-LC-biotinylation kit (Thermo Pierce), according to the manufacturer's instructions, though altered to use a 1:1 protein to biotin ratio. Excess biotin reagent was removed using a 5-mL HiTrap Desalting column (Cytiva) equilibrated in PBS. All streptavidin-coated biosensors were hydrated in BLI rehydration buffer (PBS, 0.5 mg $mL^{-1}$ BSA, and 0.01% (v/v) Tween-20) for 10 min. Biotinylated bait proteins were diluted in BLI kinetics buffer (PBS, 0.1 mg $mL^{-1}$ BSA, and 0.01% (v/v) Tween-20) to a final concentration of 20 µg $mL^{-1}$ and immobilized onto a streptavidin-coated biosensor for 90 s. Analyte proteins were prepared in BLI kinetics buffer at high concentrations (IZUMO1, 12.5 µg $mL^{-1}$; JUNO, 12.5 µg $mL^{-1}$; SPACA6, 150 µg $mL^{-1}$); analyte association to bait was measured over 120 s at 20 °C. Subsequently, the streptavidin-coated biosensor was immersed into BLI kinetics buffer for 120 s to dissociate the analyte. BSA and BLI kinetics buffer only against streptavidin-coated biosensors loaded with biotinylated bait were used to quantify non-specific binding. The data were analysed and sensorgrams were step corrected, reference corrected, and fit to a 1:1 binding model. Reported results are representative of independent duplicate trials.

**Pull down assay**

Bait protein (IZUMO1 or JUNO) tagged with 10x-His tags were mixed with target protein (JUNO or SPACA6) in a 1:1 molar ratio in Buffer B at a total volume of 100 µL. The samples were incubated on a rotator for 1 h at room temperature and then incubated with 15 µL of a 25% slurry of nickel-charged MagBeads (Genscript) for an additional 1 h on the rotator at room temperature. Magnetic beads were captured using the PureProteome magnetic stand (Millipore Sigma), and the supernatant was removed. The MagBeads were then washed five times with 100 µL of Buffer B

plus 25 mM imidazole. Bound protein was then eluted with 100 µL of Buffer B plus 400 mM imidazole and separated on 20% polyacrylamide gels for SDS-PAGE analysis. Protein bands were detected by Coomassie Brilliant Blue staining and anti-His western blot. For the western blot, protein was transferred from the 20% polyacrylamide gels to PVDF Immobilon-P membranes (Millipore Sigma) and blocked with 5% (w/v) skim milk in PBS with 0.1% (v/v) Tween 20 (PBST). Transfer membranes were then incubated with a 1:10,000 dilution of primary mouse anti-6xHis antibody (Roche, cat. #11922416001) for 1 h, followed by 1:10,000 secondary HRP-conjugated goat anti-mouse IgG (H+L) antibody (Invitrogen, cat. #62-6520) for 1 h with three 10-min PBST washes in between. The membranes were developed using UltraScence Pico Western Substrate (BIO-HELIX), and its chemiluminescence signal was imaged using a G:Box gel documentation system (Syngene). Reported results are representative from independent duplicate trials.

**Small angle X-ray scattering**

SEC-SAXS was performed at NSLS-II using their mail-in service on the Life Sciences X-ray Scattering 16-ID beamline (79–81). The SPACA6 ectodomain was dialyzed into commercially prepared 1X PBS (Millipore Sigma) and concentrated to ~6.5 mg $mL^{-1}$ prior to shipment on ice. The protein was centrifuged at 20,000 x *g* for 10 min before loading 45 µL onto a Superdex 200 Increase 5/150 GL column (Cytiva) at 0.5 mL $min^{-1}$ on a Shimadzu bio-inert HPLC system for in-line SAXS measurements. Flow from the column was split 2:1 using a passive splitter between the X-ray scattering measurements and the UV/Vis and refractive index detectors. Subsequent buffer subtraction, peak selection, and profile analysis was undertaken using Lixtools (79) and the ATSAS (3.0.3) software package (82). *Ab initio* reconstructions of SPACA6 were performed using DAMMIN (83, 84). Twenty independent DAMMIN models were superimposed and averaged using DAMAVER (85) to obtain a consensus molecular envelope. The final SPACA6 structure was superimposed onto the SAXS molecular envelope using SUPALM function as part of the SASpy PyMOL plugin (86).

**Hydrogen/deuterium-exchange mass spectrometry**

Deuterium exchange was initiated by mixing 25 µL of untagged SPACA6 ectodomain (4 mg $mL^{-1}$) with 75 µL $D_2O$ buffer (8.3 mM Tris-HCl, pH 7.2, 150 mM NaCl in $D_2O$, $pD_{read}$ 7.2). Spectra were obtained after incubation at 0 °C for 10, 100, 1000, 10,000, and 100,000 s. At these times, 16 µL of exchange samples were added to 24 µL quench solution to stop the $D_2O$ exchange reaction. After 5 min incubation on ice, quenched samples were diluted by addition of 129 µL of ice-cold dilution buffer, and then immediately frozen on dry ice and stored at −80 °C. The non-deuterated control samples and equilibrium-deuterated control samples were also prepared by mixing protein with 8.3 mM Tris-HCl, pH 7.2, 150 mM NaCl in $H_2O$ or with 0.8% (v/v) formic acid in 99.9% $D_2O$. The frozen samples were then thawed at 5 °C and passed over an immobilized pepsin column (16 µL bed volume) at a flow rate of 25 µL $min^{-1}$. The resulting peptides were collected on a $C_{18}$ trap for desalting and separated by a nanoEase M/Z Peptide BEH C18 reverse phase column (Waters, 0.3x50 mm, 130 A, 17 µm) using a linear gradient of acetonitrile from 5% to 45% over 30 min. MS analysis was performed using the OrbiTrap Elite Mass Spectrometer (Thermo Scientific) with a capillary temperature of 200 °C. Data were acquired in both data-dependent MS/MS mode and MS1 profile mode, and the data were analysed by Proteome Discoverer software and HDExaminer (Sierra Analytics Inc.).

**Statistics and Reproducibility**

For CD thermal melts, the results are presented as mean ± standard deviation (SD). Independent triplicates, *n*=3, were performed for all CD spectroscopy studies. Independent duplicates, *n*=*2*, were performed for BLI and pull downs studies.

**Data and materials availability**

Further information and requests for resources and reagents are available from the corresponding author on reasonable requests. Atomic coordinates and structure factors for SPACA6 have been deposited in the Protein Data Bank (PDB) with the accession code: 7TA2. SAXS data have been deposited in the Small Angle Scattering Biological Data Bank (SASBDB) with the accession code: SASDNM3. Source data are provided with this paper.

## Acknowledgments

We thank Dr. Vitor Hugo B. Serrão and Dr. Karen K. Siu for their assistance with the biophysical studies and the crystallography, respectively. We thank the staff on Beamline 16ID (LiX) and 17ID-1 (AMX) at the National Synchrotron Light Source II (NSLS-II), and Beamline 24ID-C (NE-CAT) at the Advanced Photon Source (APS) for synchrotron access and support. This work was supported by funding from the Canadian Institutes of Health Research (CIHR; PJT-153281) and Canada Research Chair to J.E.L. and by a joint grant to J.E.L. and A.Z. from the New Frontiers in Research Fund (NFRFE-2019-00230). Biophysics and structural biology infrastructure were supported by funding from the Canada Foundation for Innovation-John R. Evans Leaders Fund. T.D.R.V. was supported by a CIHR Postdoctoral Fellowship. The phasing of the structure is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Support for work performed at the Center for Biomolecular Structure beamline LIX (16ID) | AMX (17ID-1) | FMX (17ID-2) at NSLS-II is provided by NIGMS-1P30GM133893 and BER-BO 070. NSLS-II is supported by DOE, BES-FWP-PS001.

## References

1. M. Ikawa, N. Inoue, A. M. Benham, M. Okabe, Fertilization: a sperm's journey to and interaction with the oocyte. *J Clin Invest* **120**, 984–994 (2010).

2. M. C. Chang, Fertilizing capacity of spermatozoa deposited into the fallopian tubes. *Nature* **168**, 697–698 (1951).

3. C. R. Austin, The capacitation of the mammalian sperm. *Nature* **170**, 326 (1952).

4. L. C. Puga Molina, *et al.*, Molecular Basis of Human Sperm Capacitation. *Front. Cell Dev. Biol.* **6** (2018).

5. E. Kim, *et al.*, Sperm penetration through cumulus mass and zona pellucida. *Int J Dev Biol* **52**, 677–682 (2008).

6. N. Hirohashi, R. Yanagimachi, Sperm acrosome reaction: its site and role in fertilization. *Biol Reprod* **99**, 127–133 (2018).

7. V. E. Deneke, A. Pauli, The Fertilization Enigma: How Sperm and Egg Fuse. *Annu Rev Cell Dev Biol* **37**, 391–414 (2021).

8. L. V. Chernomordik, M. M. Kozlov, Mechanics of membrane fusion. *Nat Struct Mol Biol* **15**, 675–683 (2008).

9. B. Podbilewicz, Virus and cell fusion mechanisms. *Annu. Rev. Cell Dev. Biol.* **30**, 111–139 (2014).

10. R. Wyatt, J. Sodroski, The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**, 1884–1888 (1998).

11. F. Li, Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* **3**, 237–261 (2016).

12. D. C. Wiley, J. J. Skehel, The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu Rev Biochem* **56**, 365–394 (1987).

13. S. Mi, *et al.*, Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789 (2000).

14. S. Blaise, N. de Parseval, L. Bénit, T. Heidmann, Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *PNAS* **100**, 13013–13018 (2003).

15. A. Dupressoir, *et al.*, Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *PNAS* **106**, 12127–12132 (2009).

16. M. A. Johnson, *et al.*, Arabidopsis hapless Mutations Define Essential Gametophytic Functions. *Genetics* **168**, 971–982 (2004).

17. T. Mori, H. Kuroiwa, T. Higashiyama, T. Kuroiwa, GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nature Cell Biology* **8**, 64–71 (2006).

18. K. von Besser, A. C. Frank, M. A. Johnson, D. Preuss, Arabidopsis HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development* **133**, 4761–4769 (2006).

19. Y. Liu, *et al.*, The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in Chlamydomonas and Plasmodium gametes. *Genes Dev.* **22**, 1051–1068 (2008).

20. K. K. Siu, V. H. B. Serrão, A. Ziyyat, J. E. Lee, The cell biology of fertilization: Gamete attachment and fusion. *J Cell Biol* **220**, e202102146 (2021).

21. F. Le Naour, E. Rubinstein, C. Jasmin, M. Prenant, C. Boucheix, Severely reduced female fertility in CD9-deficient mice. *Science* **287**, 319–321 (2000).

22. K. Miyado, *et al.*, Requirement of CD9 on the egg plasma membrane for fertilization. *Science* **287**, 321–324 (2000).

23. K. Kaji, *et al.*, The gamete fusion process is defective in eggs of Cd9-deficient mice. *Nat. Genet.* **24**, 279–282 (2000).

24. K. E. Runge, *et al.*, Oocyte CD9 is enriched on the microvillar membrane and required for normal microvillar shape and distribution. *Dev Biol* **304**, 317–325 (2007).

25. A. Jégou, *et al.*, CD9 tetraspanin generates fusion competent sites on the egg membrane for mammalian fertilization. *Proc Natl Acad Sci U S A* **108**, 10946–10951 (2011).

26. N. Inoue, M. Ikawa, A. Isotani, M. Okabe, The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* **434**, 234–238 (2005).

27. E. Bianchi, B. Doe, D. Goulding, G. J. Wright, Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **508**, 483–487 (2014).

28. C. Jean, *et al.*, JUNO, the receptor of sperm IZUMO1, is expressed by the human oocyte and is essential for human fertilisation. *Hum Reprod* **34**, 118–126 (2019).

29. S. Barbaux, *et al.*, Sperm SPACA6 protein is required for mammalian Sperm-Egg Adhesion/Fusion. *Scientific Reports* **10**, 5335 (2020).

30. Y. Fujihara, *et al.*, Spermatozoa lacking Fertilization Influencing Membrane Protein (FIMP) fail to fuse with oocytes in mice. *Proc Natl Acad Sci U S A* **117**, 9393–9400 (2020).

31. T. Noda, *et al.*, Sperm proteins SOF1, TMEM95, and SPACA6 are required for sperm−oocyte fusion in mice. *PNAS* **117**, 11493–11502 (2020).

32. I. Lamas-Toranzo, *et al.*, TMEM95 is a sperm membrane protein essential for mammalian fertilization. *eLife* **9**, e53913 (2020).

33. N. Inoue, Y. Hagihara, I. Wada, Evolutionarily conserved sperm factors, DCST1 and DCST2, are required for gamete fusion. *Elife* **10**, e66313 (2021).

34. D. Lorenzetti, *et al.*, A transgenic insertion on mouse chromosome 17 inactivates a novel immunoglobulin superfamily gene potentially involved in sperm-egg fusion. *Mamm. Genome* **25**, 141–148 (2014).

35. G. C. Ireton, B. L. Stoddard, Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr D Biol Crystallogr* **60**, 601–605 (2004).

36. C. Chothia, T. Hubbard, S. Brenner, H. Barns, A. Murzin, Protein folds in the all-beta and all-alpha classes. *Annu Rev Biophys Biomol Struct* **26**, 597–627 (1997).

37. A. F. Williams, A. N. Barclay, The Immunoglobulin Superfamily—Domains for Cell Surface Recognition. *Annual Review of Immunology* **6**, 381–405 (1988).

38. P. Bork, L. Holm, C. Sander, The Immunoglobulin Fold: Structural Classification, Sequence Patterns and Common Core. *Journal of Molecular Biology* **242**, 309–320 (1994).

39. L. Holm, Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2112**, 29–42 (2020).

40. H. Aydin, A. Sultana, S. Li, A. Thavalingam, J. E. Lee, Molecular architecture of the human sperm izumo1 and egg juno fertilization complex. *Nature* **534**, 562–565 (2016).

41. K. Nishimura, *et al.*, The structure of sperm Izumo1 reveals unexpected similarities with Plasmodium invasion proteins. *Current Biology* **26**, R661–R662 (2016).

42. U. Ohto, *et al.*, Structure of IZUMO1-JUNO reveals sperm-oocyte recognition during mammalian fertilization. *Nature* **534**, 566–569 (2016).

43. R. Jahn, R. H. Scheller, SNAREs — engines for membrane fusion. *Nat Rev Mol Cell Biol* **7**, 631–643 (2006).

44. J. Diao, *et al.*, ATG14 promotes membrane tethering and fusion of autophagosomes to endolysosomes. *Nature* **520**, 563–566 (2015).

45. L. Truebestein, T. A. Leonard, Coiled-coils: The long and short of it. *Bioessays* **38**, 903–916 (2016).

46. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

47. M. Varadi, *et al.*, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**, D439–D444 (2022).

48. D. A. Ellerman, *et al.*, Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. *Mol Reprod Dev* **76**, 1188–1199 (2009).

49. N. Inoue, Y. Satouh, I. Wada, IZUMO family member 3, IZUMO3, is involved in male fertility through the acrosome formation. *Mol Reprod Dev* **88**, 479–481 (2021).

50. Y. Fujihara, *et al.*, The conserved fertility factor SPACA4/Bouncer has divergent modes of action in vertebrate fertilization. *Proc Natl Acad Sci U S A* **118**, e2108777118 (2021).

51. P. T. Chivers, K. E. Prehoda, R. T. Raines, The CXXC Motif: A Rheostat in the Active Site. *Biochemistry* **36**, 4061–4066 (1997).

52. S. Quan, I. Schneider, J. Pan, A. V. Hacht, J. C. A. Bardwell, The CXXC Motif Is More than a Redox Rheostat *. *Journal of Biological Chemistry* **282**, 28823–28833 (2007).

53. A. P. Carvalho, P. A. Fernandes, M. J. Ramos, Similarities and differences in the thioredoxin superfamily. *Prog Biophys Mol Biol* **91**, 229–248 (2006).

54. D. A. Ellerman, D. G. Myles, P. Primakoff, A role for sperm surface protein disulfide isomerase activity in gamete fusion: evidence for the participation of ERp57. *Dev Cell* **10**, 831–837 (2006).

55. N. Inoue, Y. Hagihara, D. Wright, T. Suzuki, I. Wada, Oocyte-triggered dimerization of sperm IZUMO1 promotes sperm-egg fusion in mice. *Nat Commun* **6**, 8858 (2015).

56. M. I. Binner, *et al.*, The Sperm Protein Spaca6 is Essential for Fertilization in Zebrafish. *Front Cell Dev Biol* **9**, 806982 (2021).

57. T. Clark, HAP2/GCS1: Mounting evidence of our true biological EVE? *PLOS Biology* **16**, e3000007 (2018).

58. T. D. R. Vance, J. E. Lee, Virus and eukaryote fusogen superfamilies. *Curr Biol* **30**, R750–R754 (2020).

59. M. Takeichi, Cadherins: a molecular family important in selective cell-cell adhesion. *Annu. Rev. Biochem.* **59**, 237–252 (1990).

60. I. D. Campbell, M. J. Humphries, Integrin Structure, Activation, and Interactions. *Cold Spring Harb Perspect Biol* **3** (2011).

61. P. Klemm, M. A. Schembri, Bacterial adhesins: function and structure. *Int. J. Med. Microbiol.* **290**, 27–35 (2000).

62. A. Valbuena, *et al.*, On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13791–13796 (2009).

63. L. Whitmore, B. A. Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, W668-673 (2004).

64. G. Winter, *et al.*, DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85–97 (2018).

65. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214 (2013).

66. C. Millán, *et al.*, Exploiting distant homologues for phasing through the generation of compact fragments, local fold refinement and partial solution combination. *Acta Cryst D* **74**, 290–304 (2018).

67. D. D. Rodríguez, *et al.*, Crystallographic ab initio protein structure solution below atomic resolution. *Nat Methods* **6**, 651–653 (2009).

68. A. J. McCoy, *et al.*, Phaser crystallographic software. *J Appl Cryst, J Appl Crystallogr* **40**, 658–674 (2007).

69. A. Medina, *et al.*, ALEPH: a network-oriented approach for the generation of fragment-based libraries and for structure interpretation. *Acta Cryst D* **76**, 193–208 (2020).

70. C. Millán, E. Jiménez, A. Schuster, K. Diederichs, I. Usón, ALIXE: a phase-combination tool for fragment-based molecular replacement. *Acta Cryst D* **76**, 209–220 (2020).

71. I. Usón, G. M. Sheldrick, An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features. *Acta Cryst D* **74**, 106–116 (2018).

72. D. Liebschner, *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).

73. T. C. Terwilliger, *et al.*, Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 582–601 (2009).

74. P. V. Afonine, *et al.*, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* **68**, 352–367 (2012).

75. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

76. R. P. Joosten, F. Long, G. N. Murshudov, A. Perrakis, The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).

77. C. J. Williams, *et al.*, MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293–315 (2018).

78. E. Jurrus, *et al.*, Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112–128 (2018).

79. L. Yang, *et al.*, Tools for supporting solution scattering during the COVID-19 pandemic. *J Synchrotron Radiat* **28**, 1237–1244 (2021).

80. L. Yang, *et al.*, Solution scattering at the Life Science X-ray Scattering (LiX) beamline. *J Synchrotron Radiat* **27**, 804–812 (2020).

81. E. O. Lazo, *et al.*, Robotic sample changers for macromolecular X-ray crystallography and biological small-angle X-ray scattering at the National Synchrotron Light Source II. *J Synchrotron Radiat* **28**, 1649–1661 (2021).

82. K. Manalastas-Cantos, *et al.*, ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr* **54**, 343–355 (2021).

83. D. I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).

84. D. Franke, D. I. Svergun, DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342–346 (2009).

85. V. Volkov, D. Svergun, Uniqueness of ab initio shape determination in small-angle scattering (2003) https:/doi.org/10.1107/S0021889803000268.

86. A. Panjkovich, D. I. Svergun, SASpy: a PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data. *Bioinformatics* **32**, 2062–2064 (2016).

## Figures and Tables

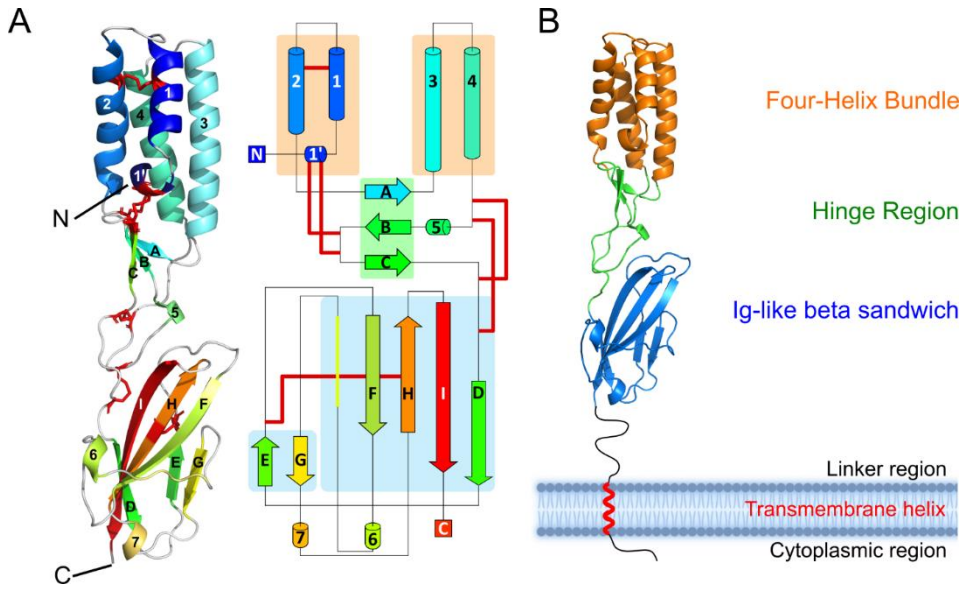



**Figure 1. SPACA6 contains two distinct domains. A)** Structure of the SPACA6 ectodomain. Ribbon diagram (left) of the SPACA6 ectodomain, with the chain from the N to C termini colored from dark blue to dark red. Cysteines involved in disulfide bonds are colored red. This same color scheme is kept in the topology diagram (right). **B)** Domains of the SPACA6 ectodomain. Ribbon diagram with the 4HB, hinge, and Ig-like domains colored orange, green, and blue, respectively. The membrane layer is not drawn to scale.

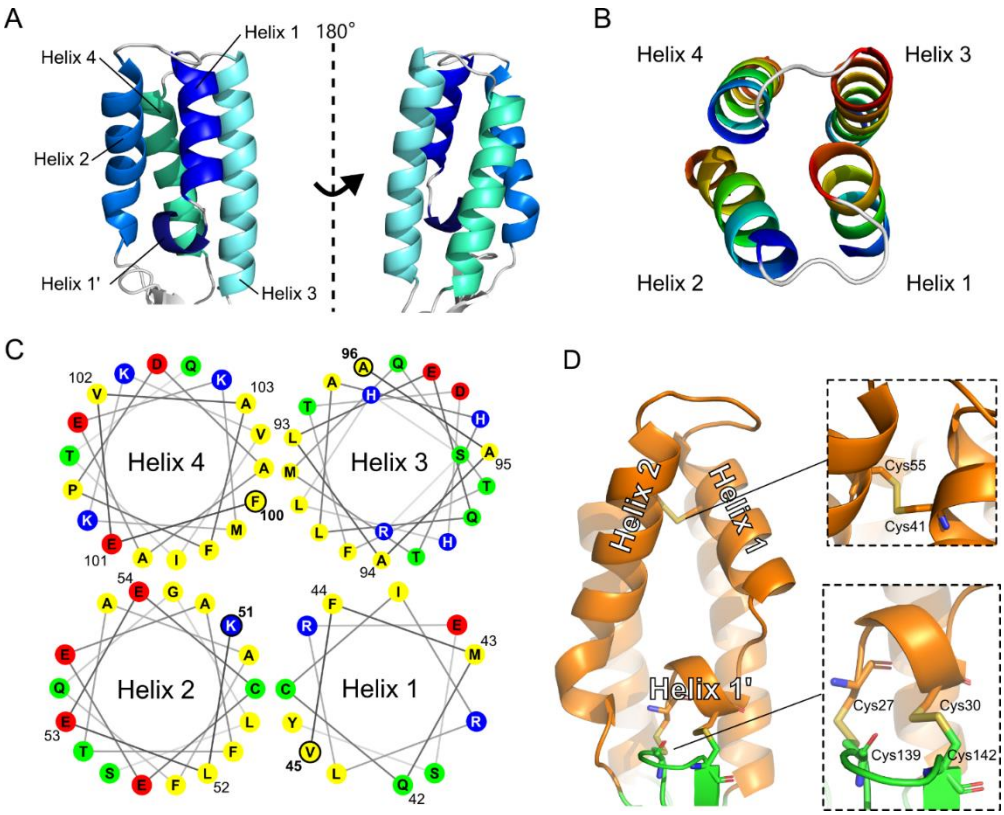
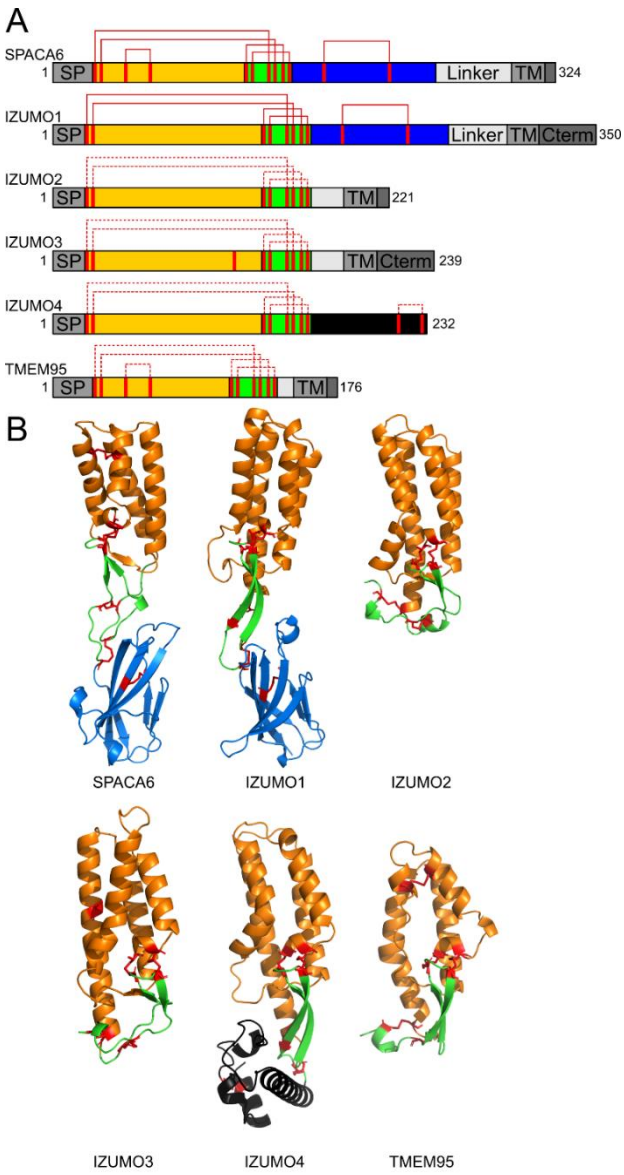



**Figure 2. N-terminal 4HB domain of SPACA6. A)** Ribbon diagram of the N-terminal 4HB. **B)** Top-down view of four-helix bundle, with each helix colored dark blue at N terminus and dark red at C terminus. **C)** Top-down helical wheel diagram of the 4HB with each residue presented as a circle marked with the one-letter amino-acid code; only the four amino acids that lie on the top of the wheel are numbered. Non-polar residues are colored yellow, polar non-charged residues are green, positively charged residues are blue, and negatively charged residues are red. **D)** Triangular face of the 4HB domain with the 4HB colored orange and hinge colored green. The two insets display the disulfide bonds as sticks.

872




873

874 **Figure 3. SPACA6 and IZUMO1 are founding members of a gamete fusion-associated**
875 **superfamily. A)** Domain architecture schematic of the IST superfamily with the 4HB, hinge, and
876 Ig-like domains colored orange, green, and blue, respectively. IZUMO4 has a unique C-terminal
877 region, which is colored black. Confirmed and putative disulfide bonds are shown in solid and
878 dashed red lines, respectively. **B)** Ribbon diagrams of the IST superfamily of gamete fusion-
879 associate proteins IZUMO1 (PDB: 5F4E), SPACA6, IZUMO2 (AlphaFold DB: AF-Q6UXV1-F1),
880 IZUMO3 (AlphaFold DB: AF-Q5VZ72-F1), IZUMO4 (AlphaFold DB: AF-Q1ZYL8-F1), and TMEM95
881 (AlphaFold DB: AF-Q3KNT9-F1) shown in the same color scheme as in panel A. Disulfide bonds
882 are shown in red. Transmembrane helices of TMEM95, IZUMO2, and IZUMO3 are not shown.
883
884

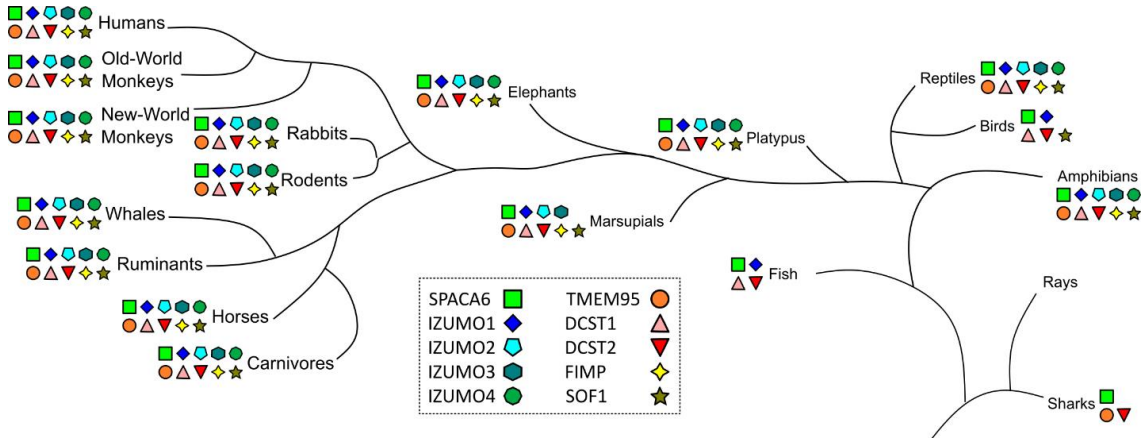



**Figure 4. Distribution of gamete fusion-associated proteins throughout the tree of life.** PSI-BLAST searches using SPACA6, IZUMO1-4, TMEM95, DCST1, DCST2, FIMP, and SOF1 of the NCBI database were used to determine where in the tree of life these sequences are found. Distances between branch points are not drawn to scale.

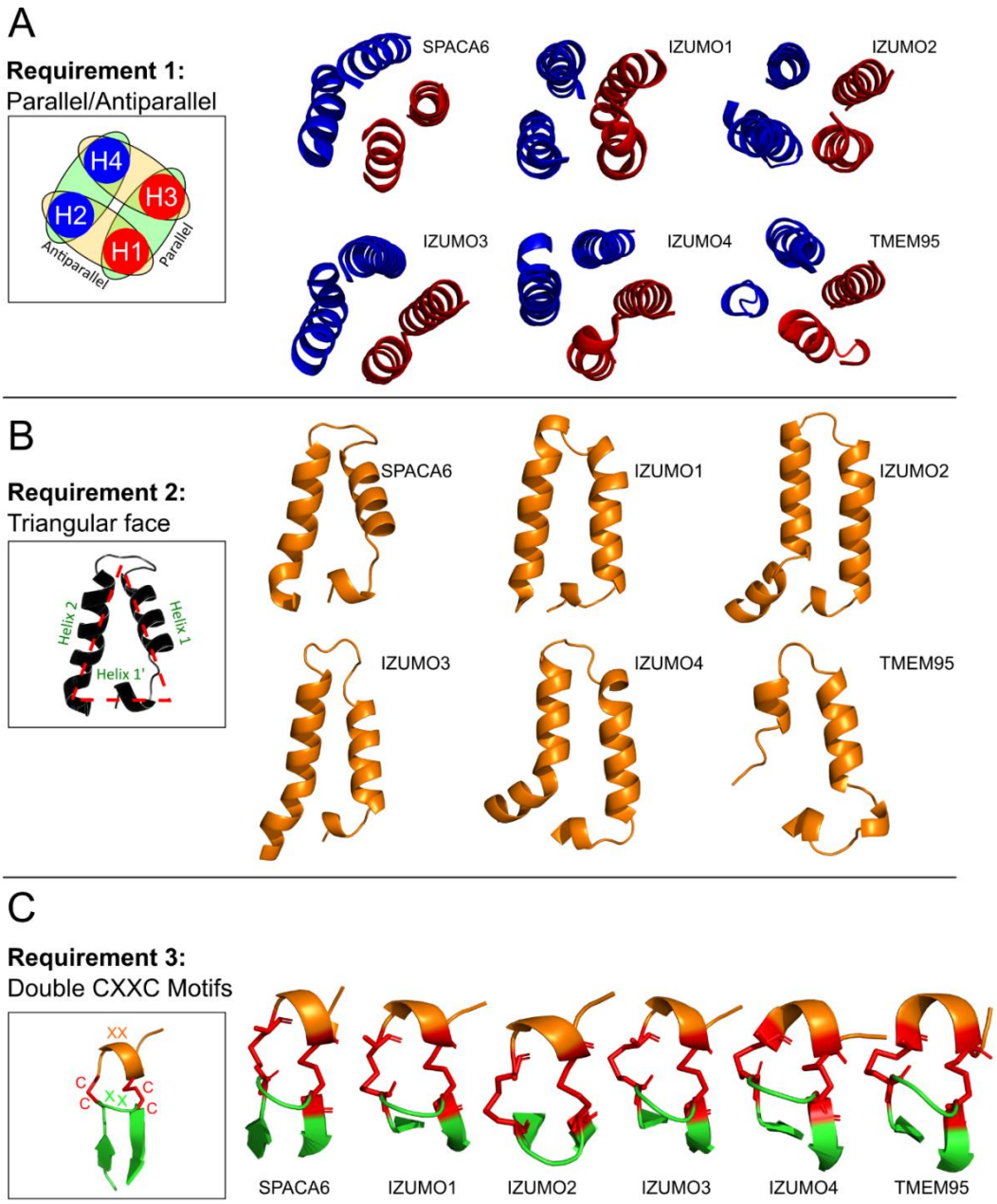



**Figure 5. Hallmark features of the IST gamete fusion-associated protein superfamily.** Members of the IST superfamily are defined by three hallmark features of the 4HB domain: **A)** four helices that alternate between parallel and antiparallel orientations, **B)** a triangular face of the helical bundle, and **C)** a double CXXC motif that forms two disulfides (red) between a small N-terminal helix (orange) and a β-hairpin (green) in the hinge region.

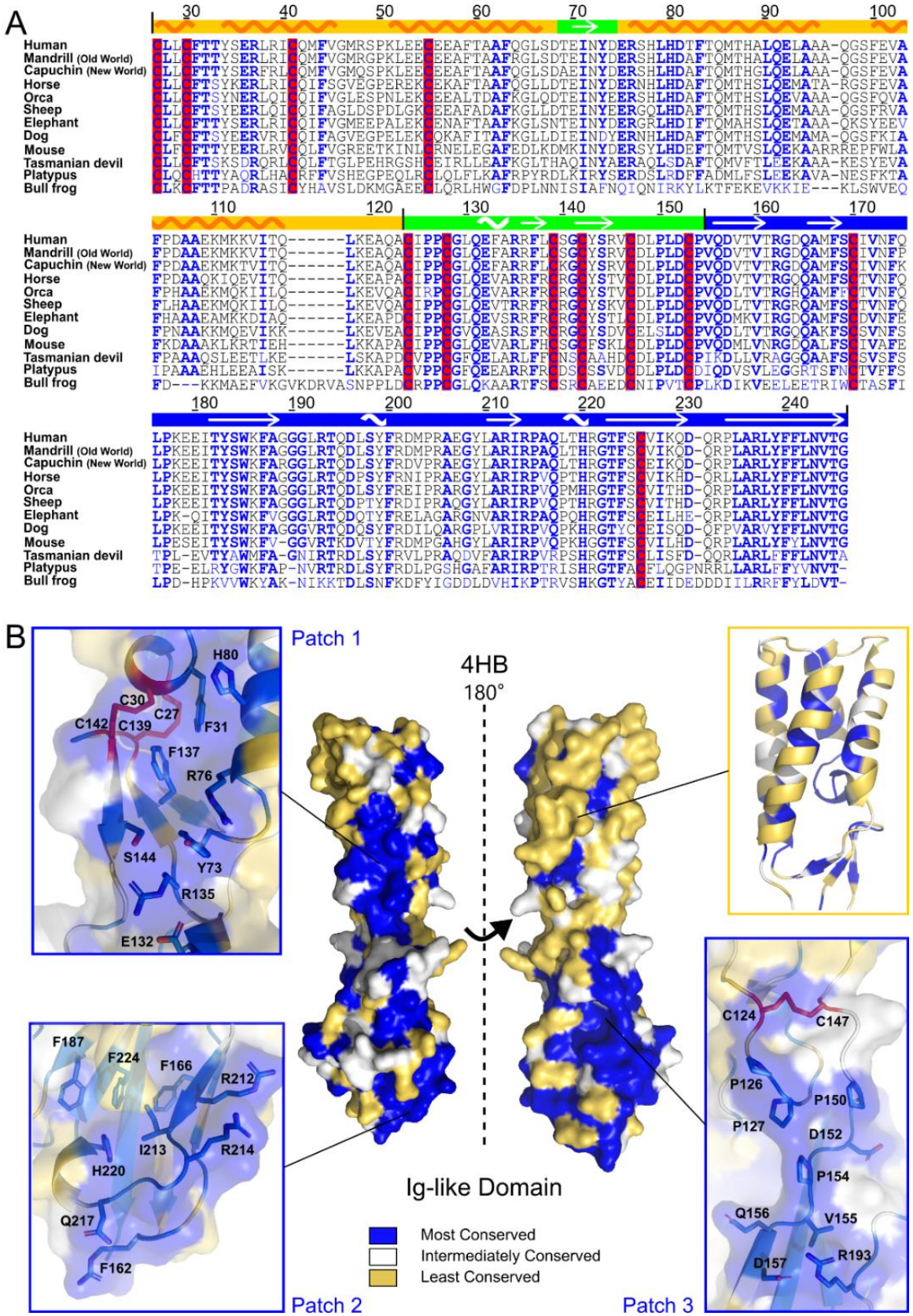


**Figure 6. Conservation of structural elements throughout SPACA6 homologs. A)** Sequence alignment of SPACA6 ectodomains from twelve different species prepared using CLUSTAL OMEGA. Most conserved positions according to ConSurf analysis are colored blue. Cysteine residues are highlighted in red. Domain boundaries and secondary structure elements are shown on top of the alignment with arrows indicating beta strands and a wave indicating helices. NCBI accession IDs for included sequences are as follows: human (*Homo sapiens*, NP_001303901), mandrill (*Mandrillus leucophaeus*, XP_011821277), capuchin (*Cebus imitator*, XP_017359366), horse (*Equus caballus*, XP_023506102), orca (*Orcinus orca*, XP_012394831), sheep (*Ovis aries*, XP_014955560), elephant (*Loxodonta africana*, XP_010585293), dog (*Canis lupus familiaris*, XP_025277208), mouse (*Mus musculus*, NP_001156381), Tasmanian devil (*Sarcophilus harrisii*, XP_031819146), platypus (*Ornithorhynchus anatinus*, XP_039768188), and bull frog (*Bufo bufo*, XP_040282113). Numbering is based on the human sequence. **B)** Surface representation of the SPACA6 structure oriented with the 4HB at the top and the Ig-like domain at the bottom and colored based on conservation scores from the ConSurf server. Most conserved portions are colored blue, portions with intermediate levels of conservation are white, and least conserved portions are yellow. Cysteines are colored red. Three surface patches showing high levels of conservation are shown in the insets labeled Patches 1, 2 and 3. A cartoon representation of the 4HB is shown in top-right inset (same color scheme).

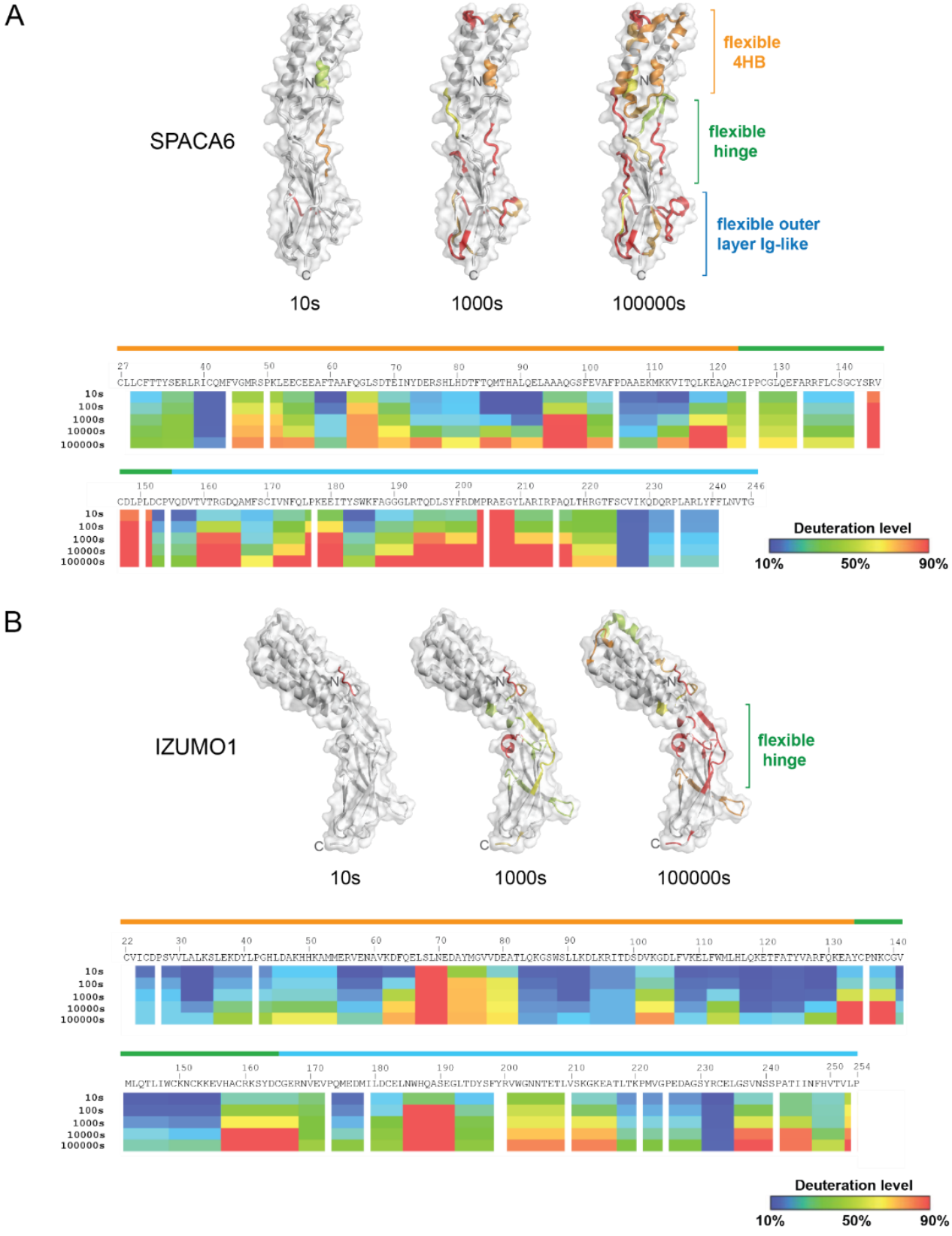


**Figure 7. Hydrogen-deuterium exchange mass spectrometry analysis of SPACA6.** H-DXMS profiles of **A)** SPACA6 and **B)** IZUMO1. Percentage deuteration exchange was determined at indicated time points. Hydrogen-deuterium exchange levels are color coded in a gradient scale from blue (10%) to red (90%). The 4HB, hinge and Ig-like domain boundaries observed in the crystal structures are shown above the primary sequence. Deuterium exchange levels at 10 s, 1000 s, and 100,000 s are mapped onto a ribbon diagram overlayed with the transparent molecular surface of SPACA6 and IZUMO1. Portions of the structure with deuterium exchange levels below 50% are colored white. Areas above 50% H-DXMS exchange are colored on the gradient scale.