Deep learning insights into the architecture of the mammalian egg-sperm fusion synapse

Arne Elofsson[1,*], **Ling Han**[2], **Enrica Bianchi**[3], **Gavin J. Wright**[3] and **Luca Jovine**[2,*]

[1] Science for Life Laboratory and Department of Biochemistry and Biophysics, Stockholm University, Box 1031, 171 21 Solna, Sweden

[2] Department of Biosciences and Nutrition, Karolinska Institutet, 141 83 Huddinge, Sweden

[3] Department of Biology, Hull York Medical School, York Biomedical Research Institute, University of York, York, UK

* To whom correspondence should be addressed (arne@bioinfo.se and luca.jovine@ki.se)

Abstract

Fertilization is a crucial event in sexual reproduction culminating in the fusion of the haploid sperm and egg to form a new diploid organism. In mammals, egg receptor JUNO and sperm IZUMO1 directly interact to mediate adhesion between gamete membranes; however, their role in fusion remains enigmatic. Here, we used AlphaFold to predict the structure of other extracellular proteins that are essential for fertilization and determine if they could form a fusogenic complex. We first identified TMEM81, whose gene is expressed by mouse spermatids, as a protein that has structural homologies with both IZUMO1 and another essential sperm molecule, SPACA6. Using a set of proteins known to be important for fertilization, we systematically searched for predicted direct binary interactions using an unguided approach. We identified a tetrameric complex of two sperm (IZUMO1 and SPACA6) and two egg (JUNO and CD9) proteins, and further modeling using this core complex identified TMEM81 as an additional component. We show that the complex is structurally consistent with both the expected topology on opposing gamete membranes and the location of predicted N-linked glycans that were not considered during prediction with AlphaFold-Multimer. Our model suggests that these proteins could organize to form a synapse-like complex between the sperm and egg membrane at the point of fertilization. Finally, the structural modeling approach described here could be more generally useful to gain structural insights into highly transient protein complexes that are difficult to detect experimentally.

Introduction

By merging the plasma membranes of egg and sperm and combining genetic material to initiate the development of a new individual, gamete fusion is the culmination of fertilization and a fundamental event in the life cycle of sexually reproducing species. Significant advances during the last twenty years have started to unravel the molecular basis of this phenomenon by identifying proteins essential for this process in organisms ranging from unicellular algae to mammals (Clark, 2018; Deneke and Pauli, 2021). In particular, recognition between egg glycosylphosphatidylinositol-anchored protein JUNO and sperm type I transmembrane protein IZUMO1 was found to be essential for the fusion of mouse gametes by mediating the juxtaposition of their plasma membranes (Bianchi et al., 2014; Inoue et al., 2005). In agreement with such a docking function, structural studies showed that, although the architecture of the ectodomain of IZUMO1 is reminiscent of *Plasmodium* invasion proteins, neither molecule resembles known fusogens (Aydin et al., 2016; Han et al., 2016; Kato et al., 2016; Nishimura et al., 2016; Ohto et al., 2016); at the same time, mouse IZUMO1 was recently reported to have fusogenic activity *in vitro* (Brukman et al., 2023), but whether this reflects a comparable function *in vivo* remains to be determined (Bianchi and Wright, 2023).

Despite the evident importance of the JUNO/IZUMO1 interaction for gamete fusion, gene ablation experiments in the mouse have identified several other egg and sperm molecules essential for this process. On the female side, these include two phylogenetically close tetraspanin membrane proteins: CD9 (Kaji et al., 2000; Miller et al., 2000; Miyado et al., 2000), which concentrates to the gamete adhesion area concomitantly with IZUMO1 (Chalbi et al., 2014) and is thought to facilitate fusion by reshaping the oocyte's plasma membrane (Jégou et al., 2011; Umeda et al., 2020), and 44%-sequence identical CD81 (Rubinstein et al., 2006), which can partially rescue the infertility of CD9-deficient mouse eggs (Kaji et al., 2002; Ohnami et al., 2012). On the male side, additional surface-expressed molecules required for mouse gamete fusion are sperm acrosome membrane-associated protein 6 (SPACA6) (Barbaux et al., 2020; Lamas-Toranzo et al., 2020; Lorenzetti et al., 2014; Noda et al., 2020) and transmembrane protein 95 (TMEM95) (Lamas-Toranzo et al., 2020), both of which are type I-transmembrane proteins with an IZUMO1-like ectodomain structure (Lamas-Toranzo et al., 2020; Nishimura et al., 2016; Vance et al., 2022); sperm dendrocyte expressed seven transmembrane protein domain-containing proteins 1 and 2 (DCST1/2) (Inoue et al., 2021), which interact with each other and are also required for gamete fusion in fish (Noda et al., 2022), in addition to being orthologues of molecules essential for fusion in worm (SPE-49/42) (Kroft et al., 2005; Wilson et al., 2018) and fly (SNEAKY/DCST-2) (Wilson et al., 2006); fertilization influencing membrane protein (FIMP), the transmembrane domain-containing isoform of 4930451I11RIK (Fujihara et al., 2020); and sperm-oocyte fusion required 1 (SOF1) (Noda et al., 2020). In addition to this mouse gene knockout-derived information, there is biochemical evidence that Izumo1 is part

of sperm multiprotein complexes, including structurally related molecules IZUMO2-4 (Ellerman et al., 2009) and, more recently, egg Fc receptor-like 3 (FCRL3/MAIA) was also suggested to be involved in gamete adhesion and fusion by replacing JUNO as an IZUMO1-binding partner (Vondrakova et al., 2022).

The relatively large number of proteins that these studies collectively identified as required for mammalian egg-sperm fusion, together with the lack of conclusive evidence supporting a direct role of the JUNO/IZUMO1 complex in the fusion process itself, suggest that a larger macromolecular complex may orchestrate the latter. However, perhaps because such an assembly exists only transiently due to the need to prevent polyspermy, the identification of additional protein-protein interactions between the aforementioned factors has frustrated efforts by multiple laboratories.

Here, we show that, despite the clear centrality of the JUNO/IZUMO1 interaction, its mouse components have such a low affinity that, unlike their human homologs, they cannot be purified as a stable complex. Because the biochemical identification of other egg/sperm fusion factor complexes may be hindered by the fact that their binary affinities also vary significantly among different species, we attack the problem by taking advantage of the momentous advances in protein complex structure prediction using AlphaFold-Multimer (Burke et al., 2023; Evans et al., 2021). The rationale for using this approach lies in the fact that the availability of a significant number of sequences for the proteins of interest predicts not only possible complexes thereof highly accurately (Jumper et al., 2021; Lee et al., 2023; Mirdita et al., 2022), but also largely insensitive to the species-specific properties of a given protein-protein interaction. Consistent with these considerations, the analysis of AlphaFold-Multimer predictions supports the suggestion that JUNO and IZUMO1 are part of a complex that includes additional fusion factors.

Results

Mouse JUNO and IZUMO1 do not form a biochemically stable complex

Whereas mammalian cell-expressed human JUNO and IZUMO1 ectodomains form a stable complex ($JUNO_E/IZUMO1_E$) that can be detected by size-exclusion chromatography (SEC), their murine homologs do not (Figure 1). This is consistent with the low affinity of the interaction between the mouse proteins, whose 0.6-12 μM K_D is significantly higher than the ~50-90 nM K_D reported for the human $JUNO_E/IZUMO1_E$ complex expressed in insect cells (Aydin et al., 2016; Bianchi et al., 2014; Nishimura et al., 2016; Ohto et al., 2016). Notably, the K_D of wild-type $JUNO_E/IZUMO1_E$ is also higher than the 360 nM K_D of the complex between $IZUMO1_E$ and $JUNO_E$ W62A (Aydin et al., 2016; Ohto et al., 2016) — an interface mutation whose introduction into mouse JUNO abolishes its ability to rescue the sperm-fusion impairment of *Juno* null eggs, as well as halves its ability to support sperm binding to JUNO-expressing HEK293T cells (Kato et al., 2016). The low affinity of mouse $JUNO_E/IZUMO1_E$ could, in principle, be partially compensated by

the avidity resulting from a high local concentration of receptors at the egg/sperm contact point. At the same time, consistent with the considerations above, the binary interaction between JUNO and IZUMO1 may be stabilized within the context of a larger macromolecular complex.

AlphaFold-Multimer produces high-confidence predictions for both mouse and human JUNO/IZUMO1 ectodomain complexes

To assess whether the significant difference in affinity between the mouse and human complexes was reflected by the confidence of the corresponding AlphaFold-Multimer predictions, we compared the output of AlphaFold runs performed without using templates. This computational experiment showed that AlphaFold-Multimer not only generates a high-confidence model of human $JUNO_E$/$IZUMO1_E$ that accurately reproduces the corresponding crystal structure but also yields a model of mouse $JUNO_E$/$IZUMO1_E$ of comparable confidence (Figure 2). This is consistent with the expectation that, as long as a significant number of sequences can be aligned to those of a protein complex of interest, the quality of AlphaFold-Multimer predictions is independent of the specific affinity of the latter.

TMEM81 is a structural homolog of IZUMO1 and SPACA6

Since IZUMO1-4, SPACA6, and TMEM95 are part of a distinct superfamily of extracellular proteins implicated in gamete fusion (Lamas-Toranzo et al., 2020; Nishimura et al., 2016; Vance et al., 2022), we used Foldseek (van Kempen et al., 2023) to scan the AlphaFold/Swiss-Prot database for further proteins of similar structure. Despite insignificant sequence identities (15-27%), this search also identified transmembrane protein 81 (TMEM81) as a clear structural homolog of the conserved immunoglobulin (Ig)-like domain of IZUMO1 and SPACA6 (E-values 1.40e-8 - 1.39e-6) (Figure 3A, B). The TMEM81 hit confirmed the result of a search of the PDB database, that we had previously carried out by generating an AlphaFold model of the protein's ectodomain and using it as input for Dali (Holm, 2020), which matched it to the crystal structure of human IZUMO1 (PDB 5JK9 (Ohto et al., 2016)) with a Z-score of 11.3. Notably, TMEM81 is known to be expressed in mouse spermatids (Jung et al., 2019) and, like IZUMO1-3, SPACA6, and TMEM95, is predicted to be a type I transmembrane protein with a large extracellular domain. Moreover, TMEM81 was previously anonymously suggested to be a β-sheet-rich molecule that may be structurally related to IZUMO1 (Wikipedia contributors, 2020). Accordingly, the characteristic four-helix bundle (4HB) of IZUMO1 and SPACA6 is replaced by a three-stranded β-sheet in the AlphaFold model of TMEM81; however, the positioning of two invariant disulfide bonds that orient these highly different elements relative to the conserved Ig-like domain is remarkably similar in the three molecules (Figure 3C).

Prediction of interactions between JUNO/IZUMO1, SPACA6, and CD9

To infer whether a larger macromolecular complex may be involved in gamete fusion, we used AlphaFold-Multimer in template-free mode to examine pairwise interactions

of the egg and sperm proteins mentioned above and TMEM81. These prediction runs were performed using both full-length protein sequences (excluding N-terminal signal peptide regions) and sequence variants that only included ectodomain residues or lacked disordered protein regions; in parallel, we also generated models for different combinations of up to 12 sequences. Collectively, this resulted in an extensive set of predictions that were analyzed by considering pDockQ confidence scoring (Burke et al., 2023; Zhu et al., 2023), as well as consistency. These criteria led to a potential four-protein assembly consisting of egg JUNO and CD9 and sperm IZUMO1 and SPACA6. The 25 predictions for this complex shared the same subunit arrangement and had a mean ranking confidence (rc = 0.8*ipTM + 0.2*pTM) of 0.71, with the top-ranked prediction having rc = 0.76 (Figure 4). Consistent with their central role in interfacing the egg and sperm plasma membranes, JUNO and IZUMO1 constitute the core of this putative assembly, where they interact in the same way that was observed crystallographically (Aydin et al., 2016; Ohto et al., 2016) and reproduced computationally (Figure 2). On the opposite side of the JUNO/IZUMO1 interface, the hinge region and 4HB of SPACA6 wrap around the 4HB of IZUMO1, generating a concave surface that interacts with the long extracellular loop (LEL) of CD9.

TMEM81 as a fifth subunit of the predicted egg/sperm fusion complex

Using the 4-subunit model as a reference, we then set out to model a potentially larger unit by carrying out predictions where the sequences of JUNO, CD9, IZUMO1, and SPACA6 were combined with one additional protein at the time (the exception being DCST1 and DCST2, which were added together since they are known to form a complex (Noda et al., 2022)). The possibility that a further component binds to the reference complex was evaluated by calculating the pDockQ of the additional subunit considering interaction with all other subunits. As shown in Figure 5A, only DCST1/2 and TMEM81 had a high pDockQ score; however, the overall rc score for the complex with DCST1/2 was low (Figure 5B), reflecting the fact that, although the two proteins were included, they mainly interacted with each other to form a ~260 Å-long mace-shaped heterodimer (Figure 5C). On the other hand, the combination of high pDockQ and rc scores for the complex with TMEM81 suggests that the latter may positively interact with the reference assembly. In agreement with this possibility, in the corresponding 5-subunit complex predictions, TMEM81 adopts the same N-to-C orientation of IZUMO1 and SPACA6 and, by inserting its Ig-like module between the two proteins, bridges their C-terminal regions (Figure 5D and Figure E).

Discussion

Since the only specific information about the target proteins used as input to AlphaFold-Multimer is their primary sequence, the neural network model does not contain any knowledge of data associated with the system's biology. As a result,

biological information on egg-sperm fusion can be used as an independent criterion to validate the predictions.

Considering that the majority of proteins involved in this process are either C-terminally membrane-anchored or transmembrane proteins, a basic feature expected in a gamete fusion synapse is that the C-termini of its egg subunits or the egg subunits themselves should all be located on the opposite side of the corresponding elements from sperm, relative to the gamete interface. This is indeed the case both for the 4- and the 5-component complex predictions (Figure 4B and Figure 5E). On the egg plasma membrane side, JUNO and CD9 are positioned so that the GPI anchor attached to the C-terminus of the former (which is not modeled by AlphaFold, whose predictions are currently restricted to amino acids) would be located in correspondence with the transmembrane domains of CD9. Similarly, the general orientation and high flexibility of the juxtamembrane regions of IZUMO1, SPACA6, and TMEM81 are compatible with the fact that, in the context of the full-length proteins, these elements are connected to the single-spanning transmembrane helices that anchor the corresponding molecules to the sperm plasma membrane (Figure 4A and Figure 5D).

A second feature common to all the subunits included in the modeled complexes is the presence of N-glycosylation sites. Because AlphaFold has no explicit knowledge of sequons and does not model carbohydrates, the N-glycans that decorate the native molecules could, in principle, interfere with predicted interfaces. As shown for the 5-subunit assembly in Figure 6, the predicted complex architecture is compatible with the location of all the possible N-glycosylation sites of JUNO, IZUMO1, SPACA6, and TMEM81, for both human and mouse homologs (amounting to a total of ten sites). One possible exception is a sequon within the short extracellular loop (SEL) of CD9, which is conserved in both species (corresponding to human N52 and mouse N50, respectively) but whose glycosylation remains to be experimentally verified. Interestingly, this site is located in close proximity to where loop 3 of JUNO protrudes towards the region between the LEL and SEL of CD9 (Figure 7A). This suggests that if the conserved sequon of CD9 is glycosylated, this may interfere with the only minor contact that the protein makes with JUNO within the predicted complex. Importantly, a fusion synapse architecture where CD9 makes little or no contact with JUNO but interacts with the 4HB of IZUMO would immediately explain the experimental observation that, in mouse oocytes, CD9 is recruited to the gamete fusion site only upon binding of JUNO to IZUMO1 (Chalbi et al., 2014). Moreover, the predicted CD9/IZUMO1 interface agrees with the previous suggestions that the two proteins interact based on the observation that their sequences co-evolve (Claw et al., 2014; Vicens and Roldan, 2014).

Although there is a general agreement that the CD9 LEL plays an important role in gamete fusion, which of its residues are responsible for this is debated; in particular, an early suggestion that the SFQ motif of mouse CD9 LEL is required for fusion was recently challenged (Umeda et al., 2020; Zhu et al., 2002). Against this background, it is interesting to note that, in our predictions, the conserved CD9 Phe at the center of the SFQ tripeptide (TFT in human) stacks against α-helix 2 of the

IZUMO 4HB (Figure 7B). Not far from this interaction, the third α-helix of CD9 LEL makes hydrophobic contacts with IZUMO1 α2 and α4; notably, these involve two conserved IZUMO1 residues that have been previously implicated in fusion, W113 and L115 (Brukman et al., 2023; Inoue et al., 2013). By inserting between the LEL of CD9 and the double CXXC motif of SPACA6, IZUMO W113 bridges the two proteins; moreover, another IZUMO1 residue that has been suggested to be important for fusion, W88 (Brukman et al., 2023), also interacts hydrophobically with SPACA6 at the opposite side of IZUMO1's 4HB (Figure 7B).

Whereas all the data above is in good agreement with the structural predictions described in this manuscript, two aspects should be considered with caution. First, it remains unclear why, despite the fact that IZUMO1 complementation rescues the disappearance of SPACA6 from the mature sperm of IZUMO1 null mice (Inoue et al., 2021), attempts to biochemically identify a complex between IZUMO1 and SPACA6 have been met with limited success (Noda et al., 2020; Vance et al., 2022). Based on the predicted complex architecture, one obvious possibility would be that, in order to be stable, the interaction also requires the presence of TMEM81. One reason could be the low affinity of the interactions between these proteins, which is typical of extracellular receptor-ligand interactions and makes them difficult to detect experimentally (Wright and Bianchi, 2016). Also, to avoid any inappropriate membrane fusion events, the individual components of the complex may be purposefully spatially segregated until brought together at the moment of fusion, again making it difficult to experimentally detect this complex *in vivo*. We suggest that the type of structural modeling approach described here could play an important role in understanding the function of dynamic and transiently formed protein complexes in a range of biological processes that would otherwise be difficult to identify. Second, because it is difficult to assess the confidence of the interface between CD9 and IZUMO1+SPACA6 due to its relatively limited extent (combined interface area ~730 $Å^2$), it cannot be excluded that some protein other than CD9 may be the true counterpart of IZUMO1+SPACA6. In other words, it is, in principle, also possible that AlphaFold-Multimer simply recognizes that the concave surface generated by the combined 4HBs of IZUMO1 and SPACA6 is likely to engage in additional interactions and thus tries to fill it with another suitable input subunit.

Of direct relevance to these questions is an independent study, submitted back to back with the present manuscript, which provides experimental data for the existence of a trimeric IZUMO1/SPACA6/TMEM81 complex in zebrafish and suggests that this interacts with egg Bouncer (Deneke et al., 2023). Considering that the mammalian orthologue of Bouncer is expressed on sperm instead of the egg (Fujihara et al., 2021) and that role of CD9 in egg-sperm fusion is much more important in mammals than in fish (Greaves et al., 2022), the combination of our studies raises the intriguing possibility that, during the course of evolution, CD9 may have substituted Bouncer as a binding partner of the IZUMO1/SPACA6/TMEM81 complex.

Methods

DNA constructs

For expression of human JUNO, a synthetic gene encoding the protein's ectodomain (residues G20-S228) followed by a 2x GGGS linker sequence (ATUM) was cloned into the *Age*I and *Xho*I restriction sites of mammalian expression vector pHLsec3 (Raj et al., 2017), in frame with 5' and 3' sequences encoding a CRYPα signal peptide/ETG tripeptide and an 8His-tag, respectively. pHLsec3 was also used to express a C-terminally Myc-tagged version of the ectodomain of human IZUMO1, preceded by its signal peptide (residues M1-L283). The ectodomains of mouse JUNO and IZUMO1 were expressed using previously described constructs (Han et al., 2016; Nishimura et al., 2016).

Protein expression and purification

Polyethyleneimine-mediated transient transfection of HEK293 cells and protein purification by immobilized metal affinity chromatography (IMAC) and size-exclusion chromatography (SEC) was carried out following published protocols (Bokhove et al., 2016).

AlphaFold predictions

Predictions were generated with local copies of AlphaFold2 (Jumper et al., 2021), installed using versions 2.2-2.3.2 of the open-source code available at https://github.com/deepmind/alphafold, or by taking advantage of the Berzelius supercomputing resource (National Supercomputer Centre, Linköping University). All runs were performed using the full_dbs preset and excluding PDB templates.

Structure analysis and comparison

Model coordinates were visualized, inspected, and superimposed with PyMOL (Schrödinger, LLC), which was also used to generate all structural figures. Database searches were carried out using Dali (Holm, 2020) and Foldseek (van Kempen et al., 2023); structure-based alignments were generated with UCSF Chimera (Meng et al., 2006) and manually edited with Belvu (Barson and Griffiths, 2016).

Acknowledgments

This study was supported by the Knut and Alice Wallenberg Foundation project grant 2018.0042 (to L.J.), and Swedish Research Council project grants 2020-04936 (to L.J.) and 2021-03979 (to A.E.). Computations and data handling were enabled by the supercomputing resource Berzelius provided by National Supercomputer Centre at Linköping University, the Knut and Alice Wallenberg Foundation, and SNIC (grants Berzelius-2021-29 and SNIC 2021/5-297).

Author contributions

L.J. conceived the study. A.E. and L.J. performed structure predictions, analyzed the data, and wrote the manuscript with input from G.J.W and E.B. L.H. expressed proteins in mammalian cells and purified them.

Conflict of interest

The authors declare that there are no conflicts of interest.

ORCID

Arne Elofsson https://orcid.org/0000-0002-7115-9751
Ling Han https://orcid.org/0000-0001-9310-4789
Enrica Bianchi https://orcid.org/0000-0001-8124-7328
Gavin J. Wright https://orcid.org/0000-0003-0537-0863
Luca Jovine http://orcid.org/0000-0002-2679-6946

References

Aydin H, Sultana A, Li S, Thavalingam A, Lee JE. 2016. Molecular architecture of the human sperm IZUMO1 and egg JUNO fertilization complex. *Nature* **534**:562–565.

Barbaux S, Ialy-Radio C, Chalbi M, Dybal E, Homps-Legrand M, Do Cruzeiro M, Vaiman D, Wolf J-P, Ziyyat A. 2020. Sperm SPACA6 protein is required for mammalian Sperm-Egg Adhesion/Fusion. *Sci Rep* **10**:5335.

Barson G, Griffiths E. 2016. SeqTools: visual tools for manual analysis of sequence alignments. *BMC Res Notes* **9**:39.

Bianchi E, Doe B, Goulding D, Wright GJ. 2014. Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **508**:483–487.

Bianchi E, Wright GJ. 2023. Mammalian fertilization: Does sperm IZUMO1 mediate fusion as well as adhesion? *J Cell Biol* **222**:e202301035.

Bokhove M, Sadat Al Hosseini H, Saito T, Dioguardi E, Gegenschatz-Schmid K, Nishimura K, Raj I, de Sanctis D, Han L, Jovine L. 2016. Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins. *J Struct Biol* **194**:1–7.

Brukman NG, Nakajima KP, Valansi C, Flyak K, Li X, Higashiyama T, Podbilewicz B. 2023. A novel function for the sperm adhesion protein IZUMO1 in cell-cell fusion. *J Cell Biol* **222**:e202207147.

Burke DF, Bryant P, Barrio-Hernandez I, Memon D, Pozzati G, Shenoy A, Zhu W, Dunham AS, Albanese P, Keller A, Scheltema RA, Bruce JE, Leitner A, Kundrotas P, Beltrao P, Elofsson A. 2023. Towards a structurally resolved human protein interaction network. *Nat Struct Mol Biol* **30**:216–225.

Chalbi M, Barraud-Lange V, Ravaux B, Howan K, Rodriguez N, Soule P, Ndzoudi A, Boucheix C, Rubinstein E, Wolf JP, Ziyyat A, Perez E, Pincet F, Gourier C. 2014. Binding of sperm protein Izumo1 and its egg receptor Juno drives Cd9 accumulation in the intercellular contact area prior to fusion during mammalian fertilization. *Development* **141**:3732–3739.

Clark T. 2018. HAP2/GCS1: Mounting evidence of our true biological EVE? *PLoS Biol* **16**:e3000007.

Claw KG, George RD, Swanson WJ. 2014. Detecting coevolution in mammalian sperm-egg fusion proteins. *Mol Reprod Dev* **81**:531–538.

Deneke VE, Blaha A, Lu Y, Draper JM, Phan CS, Panser K, Schleiffer A, Jacob L, Humer T, Stejskal K, Krssakova G, Handler D, Kamoshita M, Vance TDR, Roitinger E, Lee JE, Ikawa M, Pauli A. 2023. A conserved fertilization complex of Izumo1, Spaca6, and Tmem81 mediates sperm-egg interaction in vertebrates. *bioRxiv*. doi:10.1101/2023.07.27.550750.

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

Ellerman DA, Pei J, Gupta S, Snell WJ, Myles D, Primakoff P. 2009. Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. *Mol Reprod Dev* **76**:1188–1199.

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. 2021. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. doi:10.1101/2021.10.04.463034.

Fujihara Y, Herberg S, Blaha A, Panser K, Kobayashi K, Larasati T, Novatchkova M, Theussl H-C, Olszanska O, Ikawa M, Pauli A. 2021. The conserved fertility factor SPACA4/Bouncer has divergent modes of action in vertebrate fertilization. *Proc Natl Acad Sci U S A* **118**:e2108777118.

Fujihara Y, Lu Y, Noda T, Oji A, Larasati T, Kojima-Kita K, Yu Z, Matzuk RM, Matzuk MM, Ikawa M. 2020. Spermatozoa lacking Fertilization Influencing Membrane Protein (FIMP) fail to fuse with oocytes in mice. *Proc Natl Acad Sci U S A* **117**:9393–9400.

Greaves S, Marsay KS, Monk PN, Roehl H, Partridge LJ. 2022. Tetraspanin Cd9b plays a role in fertility in zebrafish. *PLoS One* **17**:e0277274.

Han L, Nishimura K, Sadat Al Hosseini H, Bianchi E, Wright GJ, Jovine L. 2016. Divergent evolution of vitamin B_9 binding underlies Juno-mediated adhesion of mammalian gametes. *Curr Biol* **26**:R100–R101.

Holm L. 2020. Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2112**:29–42.

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

Inoue N, Hamada D, Kamikubo H, Hirata K, Kataoka M, Yamamoto M, Ikawa M, Okabe M, Hagihara Y. 2013. Molecular dissection of IZUMO1, a sperm protein essential for sperm-egg fusion. *Development* **140**:3221–3229.

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

Jégou A, Ziyyat A, Barraud-Lange V, Perez E, Wolf JP, Pincet F, Gourier C. 2011. CD9 tetraspanin generates fusion competent sites on the egg membrane for mammalian fertilization. *Proc Natl Acad Sci U S A* **108**:10946–10951.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589.

Jung M, Wells D, Rusch J, Ahmad S, Marchini J, Myers SR, Conrad DF. 2019. Unified single-cell analysis of testis gene regulation and pathology in five mouse strains. *Elife* **8**:e43966.

Kaji K, Oda S, Miyazaki S, Kudo A. 2002. Infertility of CD9-deficient mouse eggs is reversed by mouse CD9, human CD9, or mouse CD81; polyadenylated mRNA injection developed for molecular analysis of sperm-egg fusion. *Dev Biol* **247**:327–334.

Kaji K, Oda S, Shikano T, Ohnuki T, Uematsu Y, Sakagami J, Tada N, Miyazaki S, Kudo A. 2000. The gamete fusion process is defective in eggs of Cd9-deficient mice. *Nat Genet* **24**:279–282.

Kato K, Satouh Y, Nishimasu H, Kurabayashi A, Morita J, Fujihara Y, Oji A, Ishitani R, Ikawa M, Nureki O. 2016. Structural and functional insights into IZUMO1 recognition by JUNO in mammalian fertilization. *Nat Commun* **7**:12198.

Kroft TL, Gleason EJ, L'Hernault SW. 2005. The *spe-42* gene is required for sperm-egg interactions during *C. elegans* fertilization and encodes a sperm-specific transmembrane protein. *Dev Biol* **286**:169–181.

Lamas-Toranzo I, Hamze JG, Bianchi E, Fernández-Fuertes B, Pérez-Cerezales S, Laguna-Barraza R, Fernández-González R, Lonergan P, Gutiérrez-Adán A, Wright GJ, Jiménez-Movilla M, Bermejo-Álvarez P. 2020. TMEM95 is a sperm membrane protein essential for mammalian fertilization. *Elife* **9**:e53913.

Lee S, Kim G, Karin EL, Mirdita M, Park S, Chikhi R, Babaian A, Kryshtafovych A, Steinegger M. 2023. Petascale Homology Search for Structure Prediction. *bioRxiv*. doi:10.1101/2023.07.10.548308.

Lorenzetti D, Poirier C, Zhao M, Overbeek PA, Harrison W, Bishop CE. 2014. A transgenic insertion on mouse chromosome 17 inactivates a novel immunoglobulin superfamily gene potentially involved in sperm-egg fusion. *Mamm Genome* **25**:141–148.

Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. 2006. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* **7**:339.

Miller BJ, Georges-Labouesse E, Primakoff P, Myles DG. 2000. Normal fertilization occurs with eggs lacking the integrin α6β1 and is CD9-dependent. *J Cell Biol* **149**:1289–1296.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods* **19**:679–682.

Miyado K, Yamada G, Yamada S, Hasuwa H, Nakamura Y, Ryu F, Suzuki K, Kosai K, Inoue K, Ogura A, Okabe M, Mekada E. 2000. Requirement of CD9 on the egg plasma membrane for fertilization. *Science* **287**:321–324.

Nishimura K, Han L, Bianchi E, Wright GJ, de Sanctis D, Jovine L. 2016. The structure of sperm Izumo1 reveals unexpected similarities with *Plasmodium* invasion proteins. *Curr Biol* **26**:R661–R662.

Noda T, Blaha A, Fujihara Y, Gert KR, Emori C, Deneke VE, Oura S, Panser K, Lu Y, Berent S, Kodani M, Cabrera-Quio LE, Pauli A, Ikawa M. 2022. Sperm membrane proteins DCST1 and DCST2 are required for sperm-egg interaction in mice and fish. *Commun Biol* **5**:332.

Noda T, Lu Y, Fujihara Y, Oura S, Koyano T, Kobayashi S, Matzuk MM, Ikawa M. 2020. Sperm proteins SOF1, TMEM95, and SPACA6 are required for sperm-oocyte fusion in mice. *Proc Natl Acad Sci U S A* **117**:11493–11502.

Ohnami N, Nakamura A, Miyado M, Sato M, Kawano N, Yoshida K, Harada Y, Takezawa Y, Kanai S, Ono C, Takahashi Y, Kimura K, Shida T, Miyado K, Umezawa A. 2012. CD81 and CD9 work independently as extracellular components upon fusion of sperm and oocyte. *Biol Open* **1**:640–647.

Ohto U, Ishida H, Krayukhina E, Uchiyama S, Inoue N, Shimizu T. 2016. Structure of IZUMO1–JUNO reveals sperm–oocyte recognition during mammalian fertilization. *Nature* **534**:566–569.

Raj I, Sadat Al Hosseini H, Dioguardi E, Nishimura K, Han L, Villa A, de Sanctis D, Jovine L. 2017. Structural Basis of Egg Coat-Sperm Recognition at Fertilization. *Cell* **169**:1315–1326.e17.

Rubinstein E, Ziyyat A, Prenant M, Wrobel E, Wolf J-P, Levy S, Le Naour F, Boucheix C. 2006. Reduced fertility of female mice lacking CD81. *Dev Biol* **290**:351–358.

Umeda R, Satouh Y, Takemoto M, Nakada-Nakura Y, Liu K, Yokoyama T, Shirouzu M, Iwata S, Nomura N, Sato K, Ikawa M, Nishizawa T, Nureki O. 2020. Structural insights into tetraspanin CD9 function. *Nat Commun* **11**:1606.

Vance TDR, Yip P, Jiménez E, Li S, Gawol D, Byrnes J, Usón I, Ziyyat A, Lee JE. 2022. SPACA6 ectodomain structure reveals a conserved superfamily of gamete fusion-associated proteins. *Commun Biol* **5**:984.

van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. 2023. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* doi:10.1038/s41587-023-01773-0.

Vicens A, Roldan ERS. 2014. Coevolution of positively selected IZUMO1 and CD9 in rodents: evidence of interaction between gamete fusion proteins? *Biol Reprod* **90**:113.

Vondrakova J, Frolikova M, Ded L, Cerny J, Postlerova P, Palenikova V, Simonik O, Nahacka Z, Basus K, Valaskova E, Machan R, Pacey A, Holubcova Z, Koubek P, Ezrova Z, Park S, Liu R, Partha R, Clark N, Neuzil J, Ikawa M, Erickson K, Lam KS, Moore H, Komrskova K. 2022. MAIA, Fc receptor-like 3, supersedes JUNO as IZUMO1 receptor during human fertilization. *Sci Adv* **8**:eabn0047.

Wikipedia contributors. 2020. TMEM81. https://en.wikipedia.org/w/index.php?title=TMEM81&oldid=954520237

Wilson KL, Fitch KR, Bafus BT, Wakimoto BT. 2006. Sperm plasma membrane breakdown during *Drosophila* fertilization requires sneaky, an acrosomal membrane protein. *Development* **133**:4871–4879.

Wilson LD, Obakpolor OA, Jones AM, Richie AL, Mieczkowski BD, Fall GT, Hall RW, Rumbley JN, Kroft TL. 2018. The *Caenorhabditis elegans spe-49* gene is required for fertilization and encodes a sperm-specific transmembrane protein homologous to SPE-42. *Mol Reprod Dev* **85**:563–578.

Wright GJ, Bianchi E. 2016. The challenges involved in elucidating the molecular basis of sperm-egg recognition in mammals and approaches to overcome them. *Cell Tissue Res* **363**:227–235.

Zhu G-Z, Miller BJ, Boucheix C, Rubinstein E, Liu CC, Hynes RO, Myles DG, Primakoff P. 2002. Residues SFQ (173-175) in the large extracellular loop of CD9 are required for gamete fusion. *Development* **129**:1995–2002.

Zhu W, Shenoy A, Kundrotas P, Elofsson A. 2023. Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. *Bioinformatics* **39**:btad424.

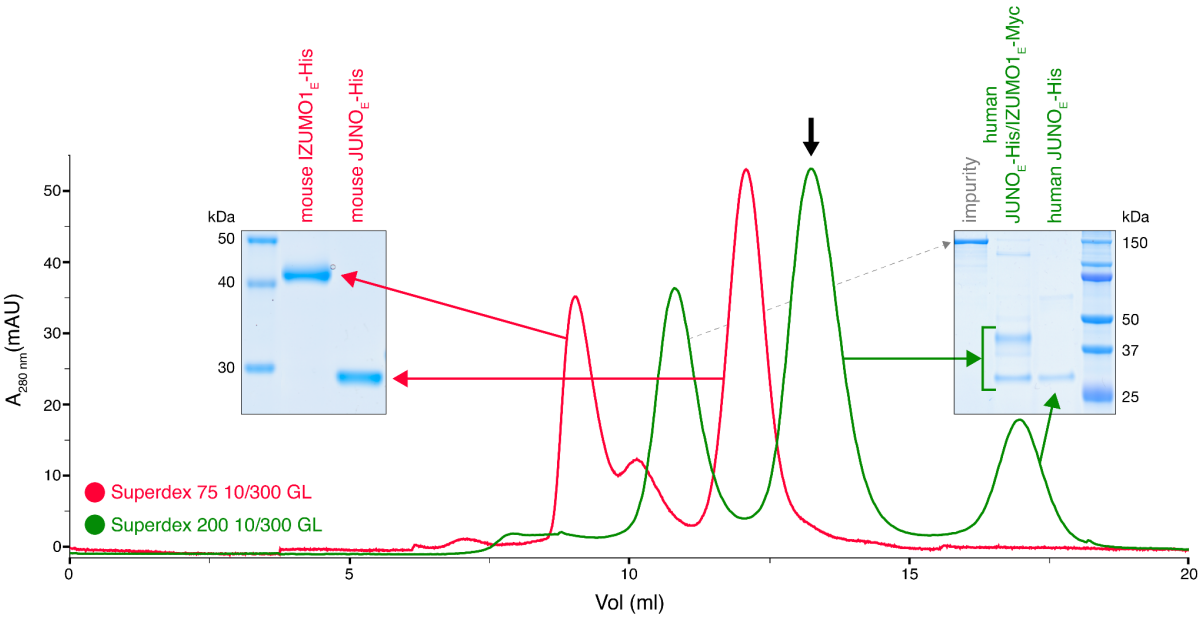

Figure 1. Human but not mouse JUNO and IZUMO1 ectodomains form a stable complex in solution.

The SEC elution profile of immobilized metal affinity chromatography (IMAC)-purified human $JUNO_E$-His/$IZUMO1_E$-Myc (green trace) shows a major peak that contains both proteins (black arrow), as well as a peak corresponding to unbound $JUNO_E$-His (SDS-PAGE analysis on the right). On the contrary, mouse $JUNO_E$-His and mouse $IZUMO1_E$-His elute separately (red trace and SDS-PAGE analysis on the left).

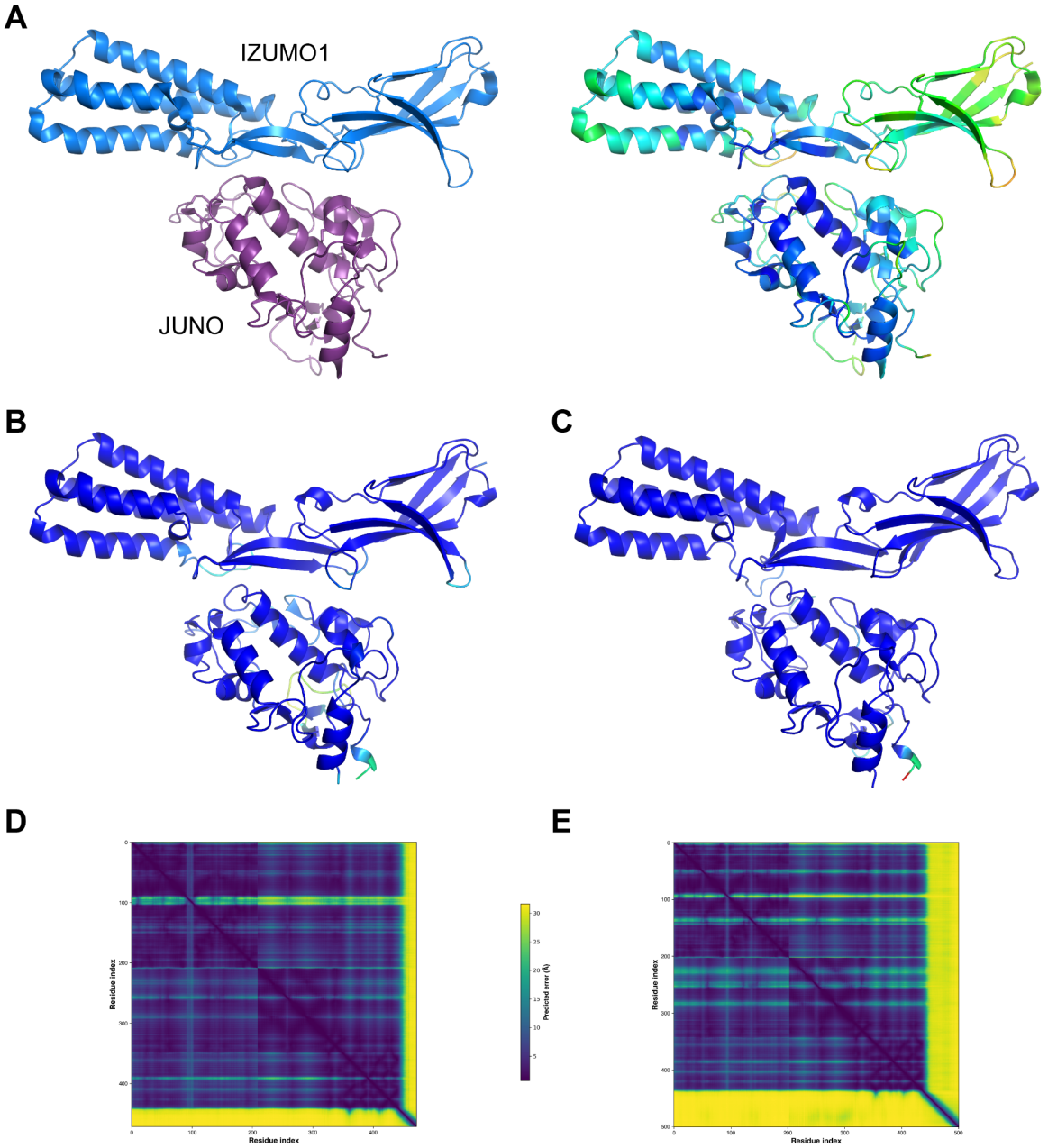

Figure 2. Mouse JUNO-IZUMO1 complex structure prediction.

A, The crystal structure of the human $JUNO_E$/$IZUMO1_E$ complex (PDB 5F4E) (Aydin et al., 2016), shown in cartoon representation and colored by chain (left) or by B-factor (right).

B, AlphaFold-Multimer top-ranked template-free prediction of the structure of the human $JUNO_E$/$IZUMO1_E$ complex (0.86; average root mean square deviation (RMSD) vs. 5F4E: 2.34 Å over 437 Cα (0.88 Å over 380 Cα after outlier rejection)). Only the residues that match those resolved in the crystal structure are shown; the model is colored by prediction confidence from blue to red, according to a 100-(per-residue confidence (predicted local distance difference test, pLDDT) (Jumper et al., 2021)) scale that ranges from 0 (blue; maximum confidence) to 100 (red; minimum confidence)), respectively.

C, AlphaFold-Multimer top-ranked template-free prediction of the structure of the mouse $JUNO_E$/$IZUMO1_E$ complex (RMSD vs. 5F4E: 2.53 Å over 435 Cα (1.73 Å over 389 Cα after outlier rejection)), depicted and colored as in panel B.
D, Predicted Aligned Error (PAE) plot for the human complex model shown in panel B. Residue indexes refer to the sequence of JUNO (amino acids G20-S228) followed by that of IZUMO1 (amino acids C22-Q284). The high PAE regions correspond to loop 2 of JUNO (residues V110-G123) and the C-terminal tail of $IZUMO1_E$ (residues K255-Q284), both of which have low pLDDT scores and are far away from the interface between the two proteins.
E, PAE plot of the mouse complex shown in panel C, with residue indexes referring to JUNO (amino acids G20-G222) followed by IZUMO1 (amino acids C22-R319).

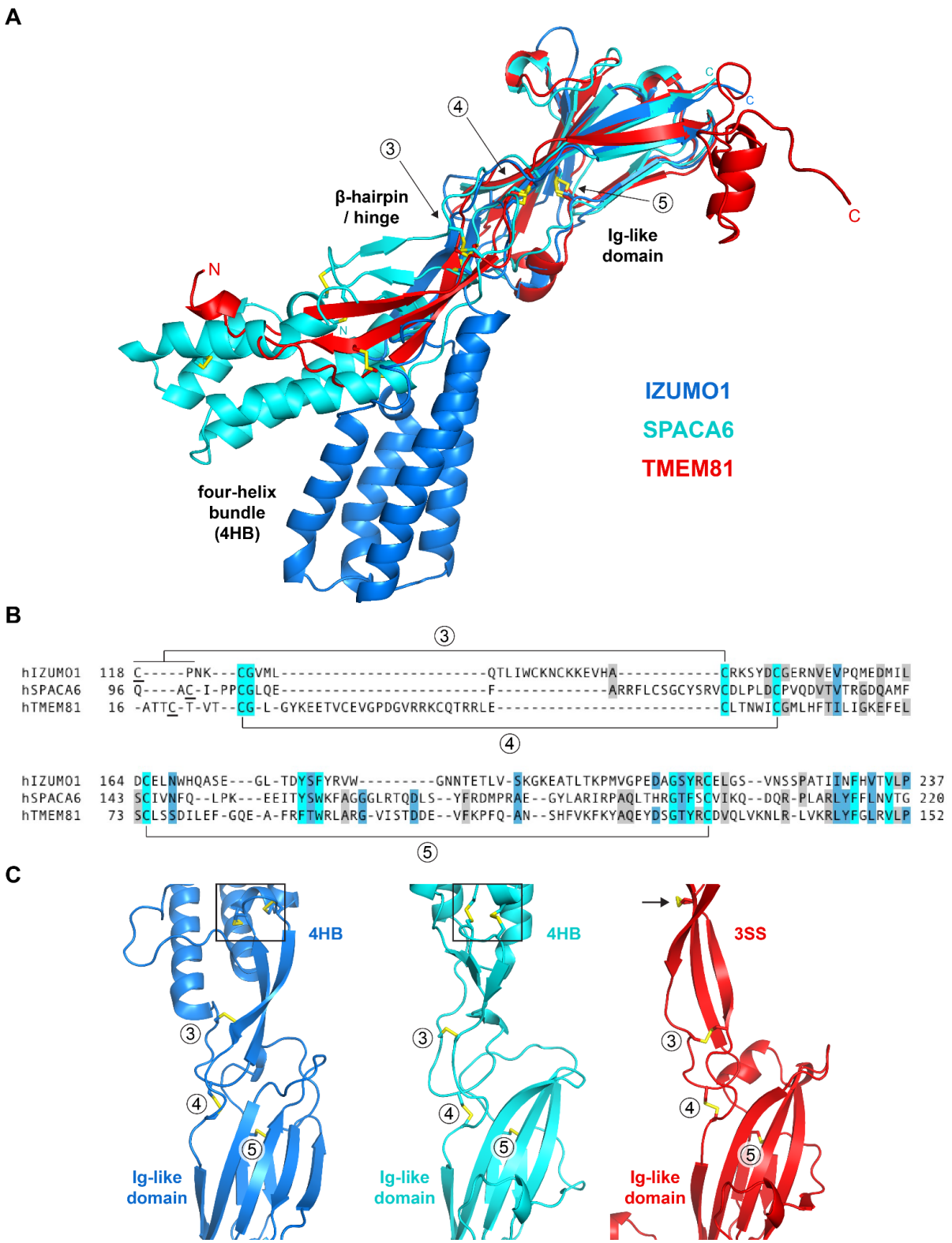

Figure 3. Structural homology between IZUMO1, SPACA6 and TMEM81.
A, Structural superposition of the ectodomains of human IZUMO1 (residues C22-K255 of PDB 5JK9 chain A (Aydin et al., 2016)), human SPACA6 (residues C27-G246 of PDB 7TA2 (Vance et al., 2022)) and an AlphaFold model of the ectodomain of human TMEM81 (corresponding to residues I31-P218 of UniProt entry Q6P7N7). The three different regions of IZUMO1 and SPACA6 are indicated in

black. Disulfide bonds are shown as yellow sticks, with arrows indicating disulfides 3-5 of IZUMO1 that are conserved in both SPACA6 and TMEM81. N- and C-termini are marked.

B, Structure-based alignment of the sequence regions includes conserved disulfides 3 and 4, followed by the Ig-like domain harboring conserved disulfide 5.

C, Partial grid view of the superposition shown in panel A, centered around the junction between three molecules' variable (top) and conserved (bottom) domains. Note the strikingly similar relative arrangement of invariant disulfides 3, 4, and 5, and how an additional disulfide within the three-stranded sheet (3SS) of TMEM81 (black arrow) roughly matches the position of the double CXXC motifs of IZUMO1 and SPACA6 (black boxes).

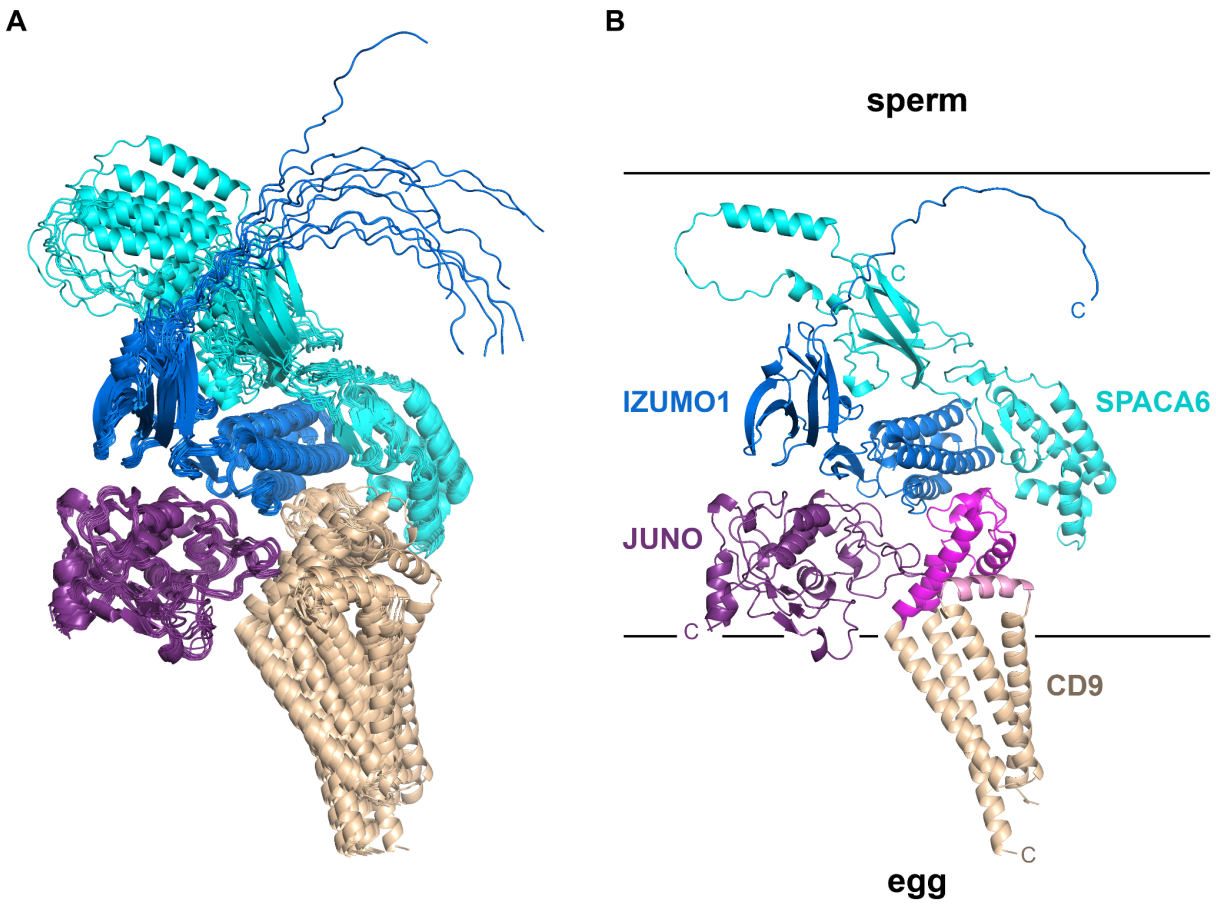

Figure 4. A predicted four-subunit complex at the egg/sperm plasma membrane interface.

A, Superposition of the top ten-ranked AlphaFold predictions for a four-subunit complex consisting of egg CD9 and the ectodomains of JUNO, IZUMO1 and SPACA6. Proteins are shown in cartoon representation and colored by chain.

B, Top-ranked model from the ensemble in panel A. Subunits are colored as in the previous panel, except for CD9 whose short extracellular loop (SEL) and long extracellular loop (LEL) are highlighted in pink and magenta, respectively. Protein C-termini are marked, with horizontal lines representing the approximate surfaces of the gamete plasma membranes.

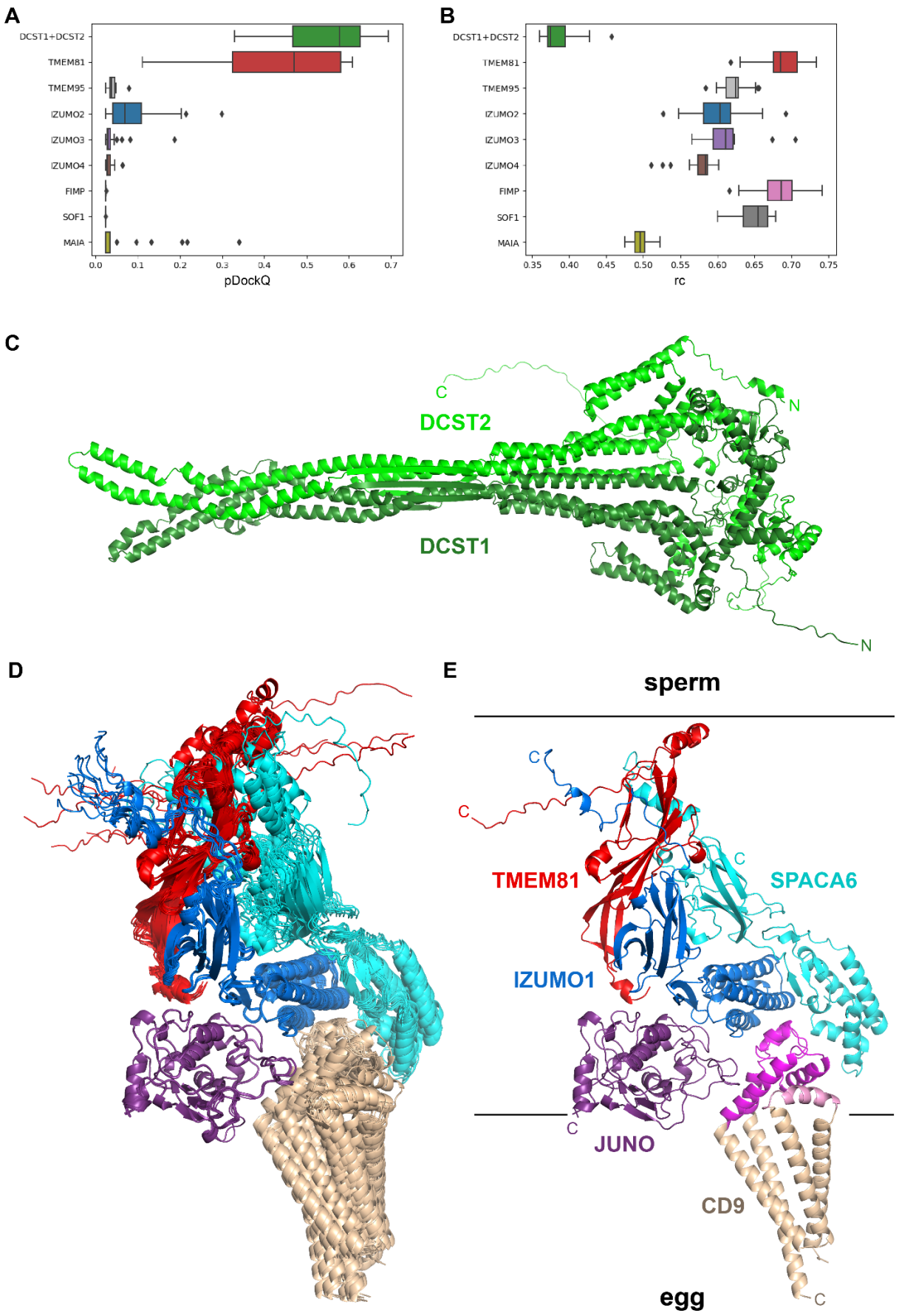
A
DCST1+DCST2
TMEM81
TMEM95
IZUMO2
IZUMO3
IZUMO4
FIMP
SOF1
MAIA
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
pDockQ
B
DCST1+DCST2
TMEM81
TMEM95
IZUMO2
IZUMO3
IZUMO4
FIMP
SOF1
MAIA
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
rc
C
C
N
DCST2
C
DCST1
N
D
E
sperm
C
C
C
TMEM81
SPACA6
IZUMO1
C
JUNO
CD9
egg
C

Figure 5. TMEM81, an additional component of the predicted egg/sperm plasma membrane protein complex.

A, Distribution of pDockQ scores for the chain(s) added as a fifth component to the JUNO/IZUMO1/SPACA6/CD9 complex. 25 models were generated for each chain, and the scores are depicted as a boxplot. Only TMEM81 and DCST1/DCST2 show pDockQ scores indicating binding.

B, Overall quality, as measured by ranking confidence (rc) for all models in panel A, shows that the overall quality of the complex with DCST1/DCST2 is low. In contrast, other complexes (including the one with TMEM81) appear to have an acceptable overall quality.

C, AlphaFold-Multimer model of the human DCST1/DCST2 complex (rc = 0.64).

D, Superposition of the top ten-ranked AlphaFold predictions for the five-subunit protein complex (mean rc = 0.71), shown as in Figure 4A.

E, Top-ranked model from the ensemble in panel D (rc = 0.74), depicted as in Figure 4B.

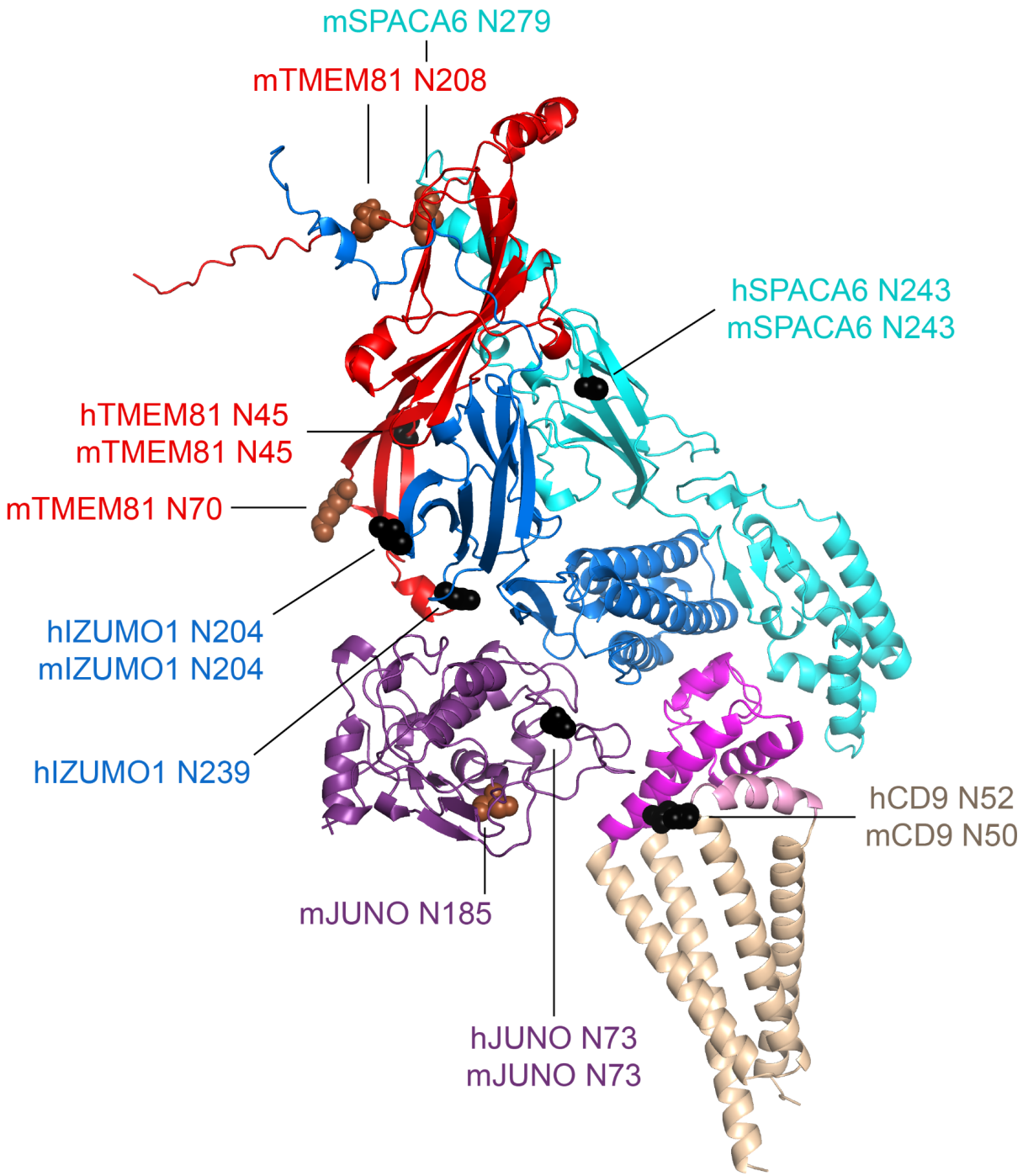

Figure 6. Mapping of sequon positions onto the 5-subunit complex prediction.

The positions of possible glycosylation sites are mapped onto the model of the 5-subunit assembly (depicted as in Figure 5E) by showing the corresponding Asn residues in sphere representation. Sequons found in human or in both human and mouse proteins are colored black, whereas sequons only found in mouse proteins are brown.

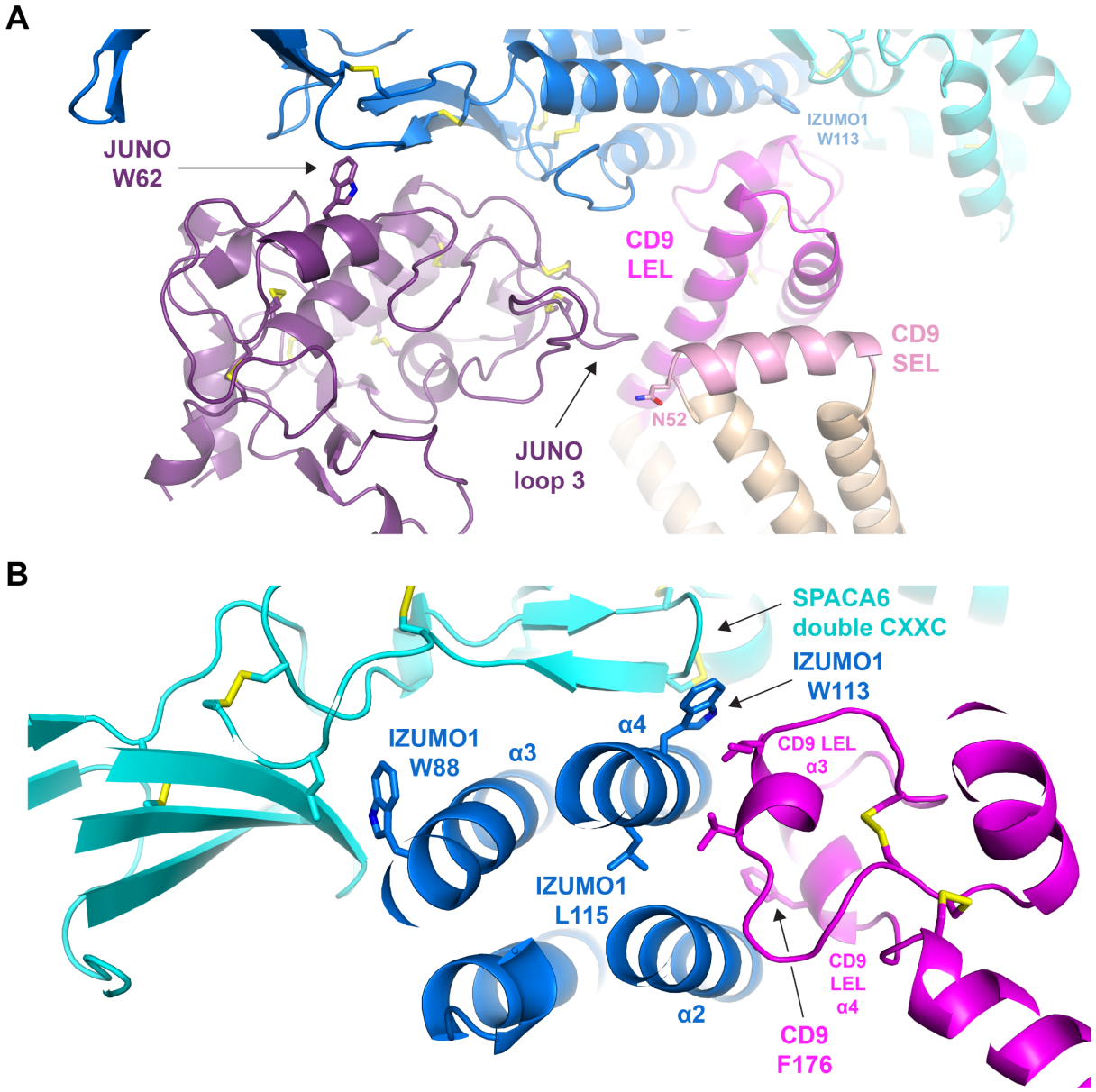

Figure 7. Subunit interfaces of the predicted complex involve protein elements previously implicated in fusion.

A, Detail of the prediction shown in Figure 4B, highlighting functionally important regions of JUNO and CD9, as well putatively N-glycosylated CD9 N52.

B, Different view of the same complex prediction, centered around the IZUMO1 4HB.