

# UCSF

## UC San Francisco Previously Published Works

### Title

De novo protein identification in mammalian sperm using in situ cryoelectron tomography and AlphaFold2 docking.

### Permalink

<https://escholarship.org/uc/item/5qf6t8cf>

### Journal

Cell, 186(23)

### Authors

Shiozaki, Momoko

Haas, Kelsey

Skinner, Will

et al.

### Publication Date

2023-11-09

### DOI

10.1016/j.cell.2023.09.017

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Published in final edited form as:

Cell. 2023 November 09; 186(23): 5041–5053.e19. doi:10.1016/j.cell.2023.09.017.

## De novo protein identification in mammalian sperm using *in situ* cryo-electron tomography and AlphaFold2 docking

Zhen Chen<sup>1,2,\*</sup>, Momoko Shiozaki<sup>3</sup>, Kelsey M. Haas<sup>2,4,5</sup>, Will M. Skinner<sup>6</sup>, Shumei Zhao<sup>3</sup>, Caiying Guo<sup>3</sup>, Benjamin J. Polacco<sup>2,5</sup>, Zhiheng Yu<sup>3</sup>, Nevan J. Krogan<sup>2,4,5</sup>, Polina V. Lishko<sup>6,7</sup>, Robyn M. Kaake<sup>2,4,5</sup>, Ronald D. Vale<sup>2,3,\*</sup>, David A. Agard<sup>1,5,8,\*</sup>

<sup>1</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA.

<sup>2</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA.

<sup>3</sup>Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA.

<sup>4</sup>J. David Gladstone Institutes, San Francisco, CA, USA.

<sup>5</sup>Quantitative Biosciences Institute (QBI), University of California San Francisco, San Francisco, CA, USA.

<sup>6</sup>Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA.

<sup>7</sup>New address: Department of Cell Biology and Physiology, Washington University in St. Louis, School of Medicine, St. Louis, MO, USA.

<sup>8</sup>Lead Contact

### SUMMARY

To understand molecular mechanisms of cellular pathways, contemporary workflows typically require multiple techniques to identify proteins, track their localization, and determine their structures *in vitro*. Here, we combined cellular cryo-electron tomography (cryoET) and

This work is licensed under a Creative Commons Attribution 4.0 International License, which allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

\*Correspondence: david@agard.ucsf.edu; valer@janelia.hhmi.org; zhen.chen@ucsf.edu.

#### AUTHOR CONTRIBUTIONS

Conceptualization: Z.C., R.D.V., D.A.A.; Mouse sample preparation: S.Z., C.G., W.M.S.; CryoEM sample preparation: Z.C., M.S.; Data processing: Z.C.; Biochemical extraction: Z.C.; Mass spectroscopy: K.M.H., R.M.K., B.J.P., N.J.K.; Sperm analyses: W.M.S., Z.C., P.V.L.; Writing – original draft: Z.C., K.M.H., R.M.K.; Writing – review & editing: Z.C., R.D.V., D.A.A.

#### DECLARATION OF INTERESTS

We declare that one or more authors have a competing interest as defined by Nature Portfolio. The Krogan Laboratory has received research support from Vir Biotechnology, F. Hoffmann-La Roche, and Rezo Therapeutics. Nevan Krogan has previously held financially compensated consulting agreements with the Icahn School of Medicine at Mount Sinai, New York and Twist Bioscience Corp. He currently has financially compensated consulting agreements with Maze Therapeutics, Interline Therapeutics, Rezo Therapeutics, and GEN1E Lifesciences, Inc.. He is on the Board of Directors of Rezo Therapeutics and is a shareholder in Tenaya Therapeutics, Maze Therapeutics, Rezo Therapeutics, and Interline Therapeutics.

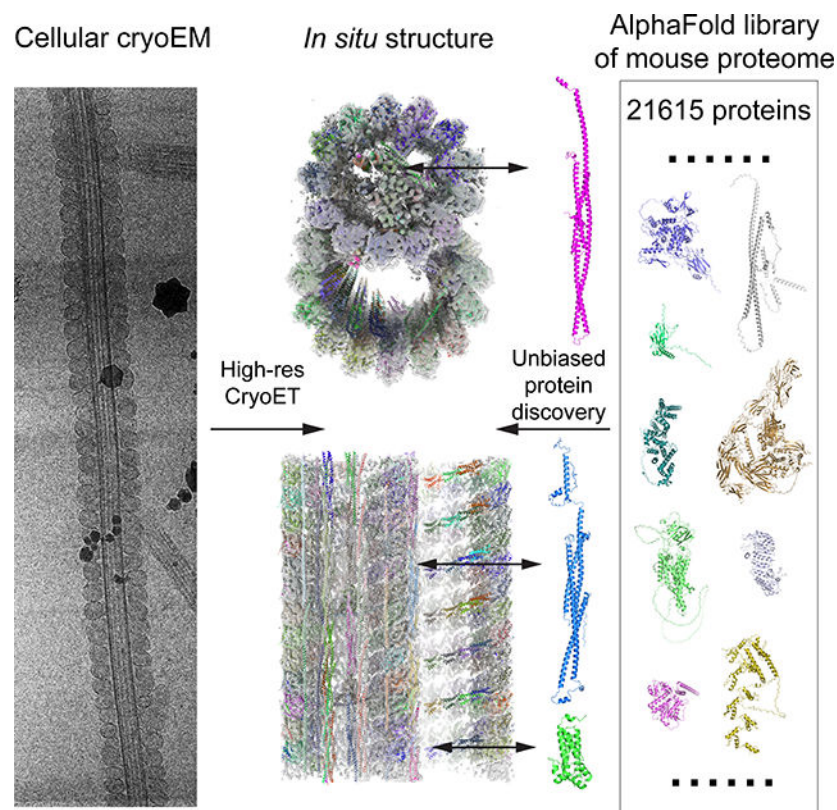
**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

AlphaFold2 modeling to address these questions and understand how mammalian sperm are built *in situ*. Our cellular cryoET and subtomogram averaging provided 6.0 Å reconstructions of axonemal microtubule structures. The well-resolved tertiary structures allowed us to unbiasedly match sperm-specific densities with 21,615 AlphaFold2-predicted protein models of the mouse proteome. We identified Tektin 5, CCDC105 and SPACA9 as novel microtubule-associated proteins. These proteins form an extensive interaction network crosslinking the lumen of axonemal doublet microtubules, indicating their roles in modulating the mechanical properties of the filaments. Indeed, *Tekt5*<sup>-/-</sup> sperm possess more deformed flagella with 180° bends. Together, our studies presented a cellular visual proteomics workflow and shed light on the *in vivo* functions of Tektin 5.

## eTOC blurb

Chen et al. reported reconstruction of microtubule doublets in mammalian sperm using *in situ* cryo-electron tomography. Unbiased matching of sperm-specific structures to the AlphaFold2 library allowed *de novo* protein identification of microtubule inner proteins. A transgenic mouse model revealed the functional importance and partial redundancy of Tektin 5 in sperm.

## Graphical Abstract



## INTRODUCTION

Natural fertilization requires the rhythmic beating motion of sperm flagella to propel the cell toward the egg<sup>1,2</sup>. This coordinated and repetitive bending of sperm flagella relies on macromolecular machinery to generate periodic force and endure mechanical stresses. Genetic analyses of infertility have so far offered only an incomplete list of protein candidates in sperm<sup>1</sup>. Additionally, we currently lack high-resolution information of sperm macromolecular complexes to understand their assemblies and functions at the molecular level.

Eukaryotic motile cilia and flagella share a conserved filamentous structure, the axoneme that has an overall architecture of nine doublet microtubules (doublets) surrounding two singlet microtubules<sup>3-6</sup>. These cytoskeletal filaments are extensively decorated externally and internally by proteins required for the various beating motions and the structural integrity of flagella<sup>3,4</sup>. Notably, sperm from different species can differ substantially in their morphologies, functions (*e.g.*, swimming behaviors), and genetics<sup>7-9</sup>. In particular, mammalian sperm flagella are much longer, wider, and must withstand larger bending torques compared to other motile cilia<sup>10,11</sup>. Despite their crucial roles of sperm axonemes in fertility and speciation, our understanding of their unique adaptations remains limited.

Isolation of axonemal complexes from non-sperm motile cilia combined with single-particle cryoEM analyses have provided high-resolution reconstructions (better than 4 Å)<sup>12-16</sup>. The rich structural information on the tertiary folds and side chains has allowed confident assignments of protein identities in the EM reconstructions. However, careful optimization of purification strategies is required to avoid partial loss of components<sup>14,16,17</sup>. Thus, the truly intact structures are not guaranteed by the end of purification in general. On the other hand, direct visualization of macromolecular complexes in mammalian sperm using cryogenic focused ion beam-scanning electron microscopy (cryoFIB-SEM) and *in situ* cryo-electron tomography (cryoET) indicated that there are indeed mammalian sperm-specific features<sup>18-20</sup>. However, the current cryoET subtomogram averaging of axonemal microtubule structures is limited to ~10–20 Å resolutions, mainly due to alignment inaccuracies. At such resolutions, tertiary structures of the proteins are rarely resolved and multi-protein complexes appear as blobs, making it challenging to determine the identities of individual sperm proteins.

Here, our *in situ* cryoET and subtomogram averaging has achieved up to 6.0 Å reconstructions of native microtubule structures in mouse and human sperm samples. The well-resolved tertiary structures in our cryoEM maps allowed us to survey the 21,615 AlphaFold2-predicted protein models of the mouse proteome and unbiasedly identify matching ones. Such a visual proteomics approach helped us to discover novel microtubule-associated proteins in mammalian sperm, localize them in cells, and determine their native structures and interaction network without cell disruption and biochemical purification. We also generated CRISPR-knockout mouse lines and showed that the newly identified Tektin 5 is important for the structural integrity of the flagella. Our studies established a cellular visual proteomics workflow and provided the structural and functional basis of sperm proteins in the microtubules.

## RESULTS

### *In situ* structures of sperm doublets at subnanometer resolutions

Freshly extracted mouse sperm were treated with the dynein inhibitor EHNA (erythro-9-(2-hydroxy-3-nonyl)adenine; 10 mM), which immediately stopped the beating motion of sperm flagella<sup>21</sup>. Subsequently, the inhibited sperm were vitrified on EM grids. To facilitate cryoET imaging which is limited by sample thickness, lamellae of ~300 nm-thickness were generated by cryo FIB-SEM milling. Tilt series were recorded using a 300 kV Krios cryo-electron microscope (cryoTEM) and a dose-symmetric scheme with a tilting increment of 4°. The 4° tilt increment, instead of the commonly used 1–3°<sup>18,19,22,23</sup>, was used to improve the signal-to-noise ratio of each tilt image while keeping the same angular range and total dose (Figure S1). Three-dimensional classification and refinement of subtomograms corresponding to the 96 nm-repeating units were performed as reported previously<sup>19</sup>. We then performed local refinement on the 48 nm-repeating structures of doublets focusing on the microtubules, aiming to reach the highest possible resolution (see the workflow shown in Figure S1). The newly developed RELION4 was used to refine the 3D reconstructions and achieved 7.7 Å overall resolution (FSC = 0.143) (Figures 1A and S2A–B, see Methods)<sup>24</sup>. The improvement of resolution compared to RELION3 comes from more accurate CTF estimation and alignment of tilt series as these parameters of each tilt image were iteratively refined relative to the 3D reconstructions (Figure S2A)<sup>24</sup>. Densities of microtubule inner proteins (MIPs) from mouse axonemes that repeat every 16 nm were observed despite of the overall periodicity of 48 nm (Figures 1C to E). In addition, we re-processed a previous human sperm dataset and achieved 10.3 Å for the 48 nm-repeating structures of the doublets (FSC = 0.143) (Figures 1B and S3A–B)<sup>19</sup>. Individual  $\alpha$ -helices for the tubulins and MIPs are resolved in both maps. These maps were then compared to the published cryoEM map of isolated doublets from bovine trachea reconstructed by single-particle cryoEM<sup>13</sup>, which were low-pass filtered to comparable resolutions of 7.5 Å and 10 Å, respectively (Figures S2C and S3C). Similar levels of detail in secondary and tertiary structures were resolved, validating the resolution estimates of our 3D reconstructions. To our knowledge, these resolutions are currently the highest achieved by cryoET for any *in situ* axonemal structures (12 Å maps were reported previously for equivalent structures from *Tetrahymena* cilia<sup>23,25</sup>).

Our 3D reconstructions of mouse and human sperm doublets reveal densities similar to the ones from bovine trachea doublets<sup>13</sup>, as well as sperm-specific densities (colored densities in Figure 1). Inside the A-tubule of mouse sperm axonemes, twelve helical bundles form a filamentous core parallel to the longitudinal axis, whereas only eight helical bundles, identified as Tektins 1–4, are present in bovine trachea cilia (Figures 1A and S2D)<sup>13</sup>. Among the four mouse sperm-specific helical bundles, only one is a continuous 3-helix bundle that runs along the entire length of the doublets (Figures 1A and 1C). This continuous bundle is also found in human sperm doublets (Figures 1A–C). The other three bundles, the two broken bundles and the curved bundles, all have breaks within the 48-nm periodic structure and appear different in mouse and human sperm (Figures S2D and S3D–G). In particular, the two broken straight 3-helix bundles have very low occupancy in the human sperm doublet (Figures 1B and S3D–E), while one of the curved helical bundles is connected to the microtubule lumen in human but not mouse sperm (Figures

S2D and S3F–G). In the mouse sperm doublet, we also observed unique “oblique” helical densities oriented  $\sim 45^\circ$  relative to the filament axis and a globular domain next to it every 16 nm (Figure 1C). These comparisons indicate there are sperm-specific MIPs compared to mammalian trachea cilia and also diversifications among mammalian sperm in the A-tubule of the doublets.

Outside the A-tubule, we found novel densities that are conserved in both mouse and human sperm doublets. A continuous three-helix bundle with multiple protrusions is situated at the external interface between A11 and A12 protofilaments, previously named the “ribbon” of doublets (Figure 1A, 1B and 1D)<sup>26</sup>. Inside the B-tubule, there are groups of four-helix bundles lining the inner surface of tubulins from the B4–B9 protofilaments. These four-helix bundles are stacked next to one another along the helical pitch of the microtubule, consistent with the previously reported striation density at lower resolutions (Figure 1E)<sup>19</sup>. Together, these data reveal that the mammalian sperm doublets have the most extensive MIP network of any microtubule structure observed to date. While the B-tubule appears similar, mouse sperm have more MIPs than human sperm in the A-tubule.

### ***De novo* protein identification using AlphaFold2**

The clearly resolved secondary and tertiary structures allowed us to interpret maps and build pseudo-atomic models. First, we were able to identify densities corresponding to the 29 MIPs observed in bovine trachea cilia (Figure S4)<sup>13</sup>, suggesting their orthologs or homologs are likely present in sperm axonemes. We then sought to identify proteins contributing to the conserved sperm densities in mouse and human doublets (highlighted in Figure 1C–E). Since most of these features repeat every 16 nm along the axoneme axis, we performed focused refinement of the 16-nm repeating structures in the A- and B-tubules of mouse doublets separately and the larger number of subtomograms further improved the resolutions of averages to 6.0 Å and 6.7 Å, respectively (Figure S5).

We then aimed to develop a general strategy to assign or narrow down the protein identities of the densities in the 6–7 Å reconstructions. Unassigned densities from our maps were manually isolated and unbiasedly matched to the predicted tertiary structures from the AlphaFold2 mouse proteome library (21,615 proteins) using the COLORES program from the SITUS package (Figure 2A)<sup>27,28</sup>. The best poses for 21,615 mouse proteins were scored and ranked by the cross-correlation scores calculated by COLORES<sup>28</sup>. We envisioned that AlphaFold2 may not be able to predict the inter-domain orientations of multi-domain proteins with no or limited inter-domain contacts. Thus, we tested this workflow using densities corresponding to known single-domain MIPs (CFAP20 and PACRG) and a multi-domain MIP (NME7) as controls. The correct PDBs corresponding to the selected densities all came up as top hits using this unbiased proteome-wide search (Data S1), indicating this visual proteomic approach can reliably identify proteins with matching tertiary structures.

For the continuous 3-helix densities in the A-tubule, the best hit was Tektin 5 (Figure S6A), a Tektin found only in the mammalian testis and sperm in previous proteomic studies (Figures 2B and S6A–B, see methods for details)<sup>29–31</sup>. Although Tektin 5 has no reported structure, AlphaFold2 predicted that it possesses single-helix, 3-helix and 2-helix segments from the N- to C-termini (Figure 2B)<sup>27</sup>, a tertiary structure that is almost identical to the

ones reported for Tektin 1–4 in bovine trachea cilia<sup>13</sup>. The ColabFold, an AlphaFold2-based Google notebook, was then used to model how two copies of Tektin 5 molecules interact<sup>32</sup>. The resulting complexes suggest the single-helix N-terminal region of one Tektin 5 could interact with the 2-helix C-termini of the other molecule (Figure 2C), indicating its potential to self-polymerize and form a quasi-continuous 3-helix bundle. Indeed, multiple Tektin 5 could be fitted into the continuous 3-helix densities with 16-nm periodicity, with minor adjustments of the orientations of individual  $\alpha$ -helices of the original AlphaFold2 model (Figure 2D). Upon manual inspection of the hit list, Tektin 1–4 were also among the top 10 hits (Figures S6A and S6B); this finding corroborated the robustness of our search method in finding proteins with matching tertiary structures. We assigned these densities as Tektin 5 since it is uniquely present in mammalian sperm based on previous proteomic studies<sup>30,31</sup> and such densities are absent in bovine trachea cilia that only contain Tektin 1–4<sup>13</sup>.

For the continuous 3-helix densities with protrusions at the ribbon, CCDC105 (coiled-coil domain containing protein 105) was identified in the top 20 hits from the unbiased search in AlphaFold2 mouse proteome library, along with Tektin 1–5 (see Data S2 for the top 30 hits). Previous proteomic studies revealed that CCDC105 was found in mammalian sperm and the testis but not other tissues<sup>29–31</sup>. CCDC105 adopts a similar overall tertiary structure as Tektins 1–5 based on AlphaFold2 prediction (Figure 2E)<sup>27</sup>, suggesting it is a yet uncharacterized Tektin homolog. However, there are three proline-rich loops in CCDC105 that are uniquely conserved across CCDC105 orthologs (Figures 2E, S6C and S6D) and they likely form structured loops like the ones observed in other axonemal complexes<sup>33,34</sup>. We also modeled how two copies of CCDC105 would interact using AlphaFold2/ColabFold<sup>35</sup>. The predicted interface again involves coiled-coil interactions between the single-helix segment of one CCDC105 and the two-helix segment of the other (Figure 2F). Furthermore, CCDC105 fits well into the continuous 3-helix densities at the ribbon, with their characteristic proline-rich loops matching the protrusions in our density maps (Figure 2G). Notably, we could not swap the fitting of Tektin 5 and CCDC105 into these two 3-helix bundles after extensive trials, mostly due to the different orientations and lengths of the  $\alpha$ -helices (Figure S6F).

We also extracted the 4-helix bundles at the B-tubule striations and performed an unbiased search against the AlphaFold2 library. SPACA9 (Sperm Acrosome-Associated Protein 9) was found to be the best hit (Figure 2H and 2I, see Data S3 for the top 30 hits). SPACA9 was previously found in various ciliated organs in humans (testis, fallopian tubes and lung)<sup>29</sup> and its tertiary fold is so unique that no other homologous protein was found in the top 30 ranked structures from the unbiased search. Interestingly, no match was found when the search was done against the CATH library that curates non-redundant domains of published PDBs<sup>36</sup>. Thus, the capability of AlphaFold2 to predict protein structures accurately, especially for the ones without published homologous structures, is critical to carrying out the unbiased proteome-wide survey.

We next focused on the mouse sperm-specific densities. There are two 3-helix bundles that appear to be similar to the ones formed by Tektins 1–5, apart from the discontinuous sections (Figures 3A and S2D). We also applied the unbiased search method to the slanted and curved helical bundles (Figures 1C and S2D). Intriguingly, Tektin 1–5 and CCDC105

were found to be among the top 30 hits in both cases while no other PDB among the top 200 fits better, albeit only parts of the structures are observed for the densities (see Data S4). Other hits among the top 200 do not match the secondary structures of the target densities upon visual inspections, suggesting Tektin 1–5 and CCDC105 are the only proteins in mouse proteome that adopt such conformations. For the slanted helical densities, there is an additional  $\alpha$ -helix connecting to the position where the missing single helix was expected to originate and is folded back by  $\sim 180^\circ$  (Figures 3B and 3C). Interestingly, Tektin 5, but not Tektin 1–4, has multiple conserved Gly residues among its orthologs at this turning region, making it plausible that Tektin 5 could adopt the bent-helix conformation. At a lower threshold, this bent helix is connected to a nearby globular domain that also repeats every 16 nm (Figure S7A). The unbiased proteome-wide search suggests this globular domain matches the tertiary structure of multiple DUSP proteins (Dual Specificity Phosphatase 3, 13, 14, 18, 21 and 29) (Figures 3C and S7B).

For the curved helical bundles, there are three 16-nm groups of densities within every 48-nm repeat (Figures 3D and S2D). These densities can be explained by three modified Tektin 5 molecules, in which the two intermolecular interfaces near the two NME7s (a previously known MIP shared with bovine trachea cilia) are disrupted (Figure 3D). The first and second Tektin 5s lack densities for the single-helix segment beyond the conserved Gly137 (mouse), while the second and third Tektin 5s possess curved 2-helix segments (Figure 3D). Both modifications of Tektin 5s are necessary to avoid direct steric clashes with the two NME7s, which adopt similar conformations in bovine trachea and mouse sperm doublets. As curved bundles were not observed in the bovine trachea cilia that only contain Tektins 1–4<sup>13</sup>, we hypothesize that Tektin 5 has evolved to adopt multiple conformations and positions within sperm axonemes (Figure 4E).

In summary, the various helical conformers of Tektin 5, together with the more uniform 3-helix bundles of Tektin 1–4, are arranged with different polarities (Figures 4A and 4B) and orientations (Figure 4C), forming the most extensive MIP network inside microtubules discovered to date.

### Additional validation and redundancy of Tektin5

To further validate our *de novo* protein assignments, we used mass spectrometry-based proteomics. Mouse sperm were isolated and extracted using salt buffers with increasing denaturing capabilities (E1: 0.1% Triton, E2: NaCl, E3: KSCN, E4: Urea, and E5: 10% SDS) and the extractions were analyzed using SDS-PAGE and western blotting (Figure S8A).  $\alpha$ -Tubulins could be detected in KSCN and Urea extractions but not in others (Figure S8B), suggesting the microtubule doublets are disassembled and the MIP candidates are likely present in these two extractions. After analyzing the E1-E5 fractions by MS, proteins with significant changes in abundance between fractions were clustered into six distinct groups based on the correlation of intensity profile (Figures S8C and S8D and Tables S1 and S2)<sup>37,38</sup>. Gene ontology (GO) analyses suggest that cluster 4 is enriched for proteins involved in cilium, cilium assembly, cytoskeleton and the axoneme, and shows increased intensities in fraction E3, while cluster 6 is enriched for cilia assembly and shows increased intensities in fraction E4 (Figure S8e and table S3)<sup>39</sup>. The overall change in

protein abundance is consistent with the idea that E3 and E4 buffers extracted microtubule-associated proteins in the axonemes. Indeed, 28 of the 29 previously identified MIPs in bovine trachea doublets were reproducibly identified in all three biological replicates of mouse sperm extractions (table S4). The almost complete list of MIPs highlighted the coverage of our biochemical and MS analyses. Importantly, SPACA9 was reproducibly identified in fraction E3, while Tektin 1–5 and CCDC105 were reproducibly identified in fractions E3 and E4, with high protein intensity and a range of 3 to 34 unique peptide identifications per replicate (Figure S8F, table S4). Only one DUSP protein, DUSP3, was identified in two of three replicates in fraction E4 fraction (Figure S8F, table S4). However, additional analyses are required to identify these proteins. Moreover, AlphaFold2 models for the other candidates from fractions E1–5 were inspected but no additional candidates could explain the various densities of helical bundles described above.

### Tektin 5 strengthens the flagella and is partially redundant

In order to analyze the functions of Tektin 5, we generated knockout mice carrying null alleles of *Tekt5* using CRISPR technologies. To our surprise, the F2 homozygous knockout males are still fertile when they are mated with the WT females (litter size:  $7.3 \pm 1.4$ ,  $N=6$  mating trials), suggesting some levels of functional redundancy. However, the sperm extracted from the mutant males have a lower fraction of motile cells compared to WT controls ( $64 \pm 3\%$  vs.  $77 \pm 4\%$ ) and have a higher percentage of defective flagella with  $180^\circ$  bends ( $30 \pm 3\%$  vs.  $13 \pm 3\%$ ) (Figures 5A and 5B), suggesting the mechanical integrity of the cellular structures inside flagella is compromised in the mutant sperm.

We then analyzed the doublets structure of the mutant sperm and compared it to the WT counterpart. We first focused on the various densities that were assigned to be Tektin 5 in both our cellular cryoET (as shown in Figure 3). The continuous 3-helix bundle remains in the mutant (Figure 5C), suggesting a Tektin homolog could substitute on the same position in the absence of Tektin 5. In contrast, the occupancy of densities corresponding to broken and curved helical bundles are much lower compared to the surrounding proteins, such as the tubulins, in the mutant sperm (Figures 5D and 5E), while in wild-type sperm the occupancies are comparable. These comparisons suggest the compensation for the lack of Tektin 5 by other Tektin homologs is low at these sites. For the slanted helical densities that repeat every 16 nm in WT sperm, we observed that two of the slanted helical densities are partially occupied while the last one is almost absent (Figure 5C). Interestingly, the densities corresponding to the DUSP domains next to the slanted helical densities were not resolved, suggesting the lack of Tektin 5 would decrease the recruitment of the neighboring MIP. We did not observe additional differences in densities corresponding to other MIPs. Together, these results suggest that Tektin homologs could partially refill docking sites of Tektin 5 in the mutant sperm.

## DISCUSSION

Our *in situ* cryoET studies have provided high-resolution reconstructions of native structures within mouse and human sperm, enabling us to identify Tektin 5, SPACA9 and CCDC105 as novel components in sperm doublets.

Alignment of gold beads has traditionally been the method of choice to align tomographic tilt series but the positions of gold beads undergo heterogeneous motions due to the sample deformation induced by electrons during cryoET imaging<sup>40</sup>. The significant improvement in resolutions made possible by RELION4 (Figure S1 and S2A) highlight the benefits of aligning 3D reconstructions of protein complexes with their individual 2D projection views on different tilt images to refine tilt series alignment<sup>24</sup>. Thus, new methods of aligning the molecular features directly from tilt series while considering their local motions have the potential to further improve the initial alignment of tilt series, and set an even better starting point for subtomogram averaging<sup>41</sup>.

The application of AlphaFold2 has facilitated structural modeling based on cryoEM reconstructions below 10 Å, where secondary structures are resolved<sup>42</sup>. However, most studies have focused on protein complexes with known components identified through other approaches such as mass spectrometry analyses of purified complexes. However, this information may not be readily available in other less-studied cell biology processes. Our studies provided the first demonstration that high-resolution cellular cryoET combined with unbiased proteome-wide searches could identify previously unknown components of cellular complexes in their native context. This integrative structural modeling approach offers a powerful alternative to the conventional genetic and cell biology approaches to identify participating protein components and localize them inside cells.

The comparison of MIPs in WT and mutant sperm doublets with bovine tracheal doublets suggest that the assembling of MIPs are modular, and novel MIPs discovered in this study (SPACA9, CCDC105, Tektin 5 and DUSP) are recruited after the commonly shared MIPs bind to the doublets (common MIPs). First, homologs or orthologs of MIPs identified from bovine tracheal doublets are all present in WT and mutant mouse sperm and they adopt similar conformations (Figure S4)<sup>13</sup>. Second, SPACA9 and CCDC105 crosslink the tubulin dimers only at the exposed lumen sites observed in tracheal doublets, without displacing any common MIPs (Figures 2G, 2I and S4). Interestingly, the remarkable conformational plasticity of Tektin 5 is partially molded by the doublets and common MIPs, as shown by the bent single helix that would otherwise clash with the microtubule wall (Figures 3B and 3C), as well as the missing single helix and curved 2-helix segment that would otherwise clash with NME7, a common MIP (Figure 3D). Lastly, knockout of Tektin 5 decreases the occupancies of the bent Tektin 5 and the nearby DUSP densities (Figure 5C). Such dependency or modularity could potentially be harnessed for modifications during evolution, as we observed that both the bent Tektin 5 and DUSP densities are absent in human sperm doublets.

Bending of the axonemes would stress the nine microtubule filaments in nine different directions and mammalian axonemes have to bend to various directions to generate the 3D beating waveforms<sup>43</sup>. The non-uniform arrangement of helical bundles in sperm doublets could be built to reinforce the doublets to withstand mechanical stress from different directions (Figures 4 and 6). From a structural perspective, the intramolecular and intermolecular coiled-coil interaction interfaces of Tektins are parallel to the microtubule axis so that bending would not expand the gap of these interfaces (Figure 6A). Instead, the bending force would distort the straight helical bundles and the ideal bond angles/lengths.

The transition of releasing such molecular strains could provide a restoring force that allows the curved filament to return to the straight conformation. In contrast, the interface between tubulin dimers along the protofilaments is at a plane perpendicular to the filament axis; bending would open the interface and lower the affinity or the potential restoring force. Therefore, the helical bundles are arranged to provide an effective means to bear the bending force, stabilizing the axoneme and flagella.

From a functional perspective, we discovered that mutant mice lacking Tektin 5 have more deformed sperm flagella, yet they remain fertile. This observation aligns with previous studies on knock-out mouse models lacking Tektin-3 or Tektin-4, which also showed deformed sperm but normal fertility<sup>44,45</sup>, underscoring the general functional redundancy of Tektins. Our structural analyses of *Tekt5*<sup>-/-</sup> sperm doublets uncovered the compensatory mechanisms of Tektin 5 at the molecular level. Moreover, we discovered that different Tektin 5 conformers were compensated to varying extents, likely due to the differential dependencies of the distinctive interaction interfaces.

The human sperm doublets revealed that the densities formed by Tektin 5 are significantly less resolved compared to the other MIPs (Figure 1), adopt different conformations (Figure S3F), or are completely absent compared to mouse sperm doublets (Figures 1C). This is also consistent with the idea that the bundles formed by Tektin 5 are partially redundant and plastic during evolution so that some extent of degeneration is tolerable. Still, the lower percentages of motile sperm would be selected against in the wild, particularly in species where sperm from multiple males in the female reproductive tracts competing for fertilization.

Our studies combine high-resolution *in situ* cryoET and AlphaFold2 modeling and define a visual proteomic approach of precisely placing proteins in their native cellular environment without the need for labeling, cellular disruption or purification. This workflow has allowed us to uncover the cellular locations and interaction networks of several MIPs, providing insights into how they contribute to the mechanics of flagellar bending. Moreover, this visual proteomic workflow could potentially be applied to other cell biology problems, such as membrane remodeling by viruses and identification of their *in situ* interactors.

### Limitations of the Study

The majorities of the densities in sperm doublets feature well-defined domains that could be isolated and identified using our visual proteomics approach. However, it is possible that there are unknown MIPs that are composed of coiled coils without substantial intramolecular interactions. These proteins may adopt conformations that depend on intermolecular interactions in the context of native complexes, which are not accounted for by AlphaFold2 predictions. Furthermore, map segregation for these types of proteins is challenging at ~6–10 Å resolutions. Improvement of resolution using cellular cryoET is also desirable to resolve the side chain densities in the EM reconstructions and distinguish the proteins with similar tertiary structures. Notably, while this manuscript was under revision, two single-particle cryoEM studies of splayed mammalian sperm doublets were reported<sup>46,47</sup>. The single-particle cryoEM reconstructions of microtubule doublets from mouse and bovine sperm are indistinguishable from our cryoET reconstruction at our resolutions, suggesting

that the sperm doublets are robust enough to withstand gentle biochemical treatments. Importantly, their assignments of protein identity based on side chain densities are consistent with our identification of CCDC105, SPACA9 and the different forms of Tektin 5, validating our cellular visual proteomics approach for *de novo* protein discovery. In the future, a more systematic characterization of MIPs using transgenic mice will be needed to elucidate the functions of individual MIPs. Additionally, genetic analyses of patients with infertility are likely to identify more essential and redundant components.

## STAR METHODS

### RESOURCE AVAILABILITY

**Lead contact**—Further information and requests for resources and reagents should be directed to the lead contact, David A. Agard (david@agard.ucsf.edu).

**Materials availability**—Experimental reagents generated in this study are available from the lead contact with a completed material transfer agreement.

**Data and code availability**—Cryo-EM maps of 48 nm-repeating structures of doublets from wildtype mouse, Tekt5 <sup>-/-</sup> mouse and human sperm have been deposited in the Electron Microscopy Data Bank (EMDB) with accession codes: EMD-41431, EMD-41320 and EMD-41317, respectively. The EMD-41431 is a composite map with its two submaps deposited with accession codes: EMD-41450 and EMD-41451. Maps of focused refinement of 16 nm-repeating structures of A- and B-tubules from wildtype mouse have been deposited also: EMD-41315 and EMD-41316. The atomic model of the 48-nm repeat of the mouse sperm doublets has been deposited in the Protein Data Bank (PDB) with accession codes 8TO0. MS data are shared and available through the ProteomeXchange Consortium via the PRIDE partner repository under the dataset identifier: PXD036885 (username: reviewer\_pxd036885@ebi.ac.uk; password: tMEZ90MC) <sup>57</sup>. R package source materials for MSstats (version 3) are publicly available through the Krogan Lab GitHub: <https://github.com/kroganlab>.

After downloading the AlphaFold2 library of the mouse proteome, this code is used to distribute PDB files into subdirectories.

```
i=0; for f in *;
do
## Splitting 50 PDBs in each subdirectory
d=dir_$(printf %03d $((i/50+1)));
mkdir -p $d;
mv "$f" $d;
let i++;
done
```

This code is used to unbiasedly match all PDBs with the target densities in each subdirectory:

```

for file in *
do
    echo $file
    ## CCDC105_flipped_b150.mrc is the target densities, the options could be
    found in the situs website
    colores ../CCDC105_flipped_b150.mrc ${file} -res 6.0 -cutoff 0.0048 -deg
    15.0
    mkdir ../output/${file}_out
    mv col_* ../output/${file}_out/.
    mv
done

```

The cross-correlation scores could then be extracted using the following script:

```

for f in *.out
do
    echo $f
    grep structure $f/*.pdb >> TheResultFile
    grep Unnormalized $f/*.pdb >> TheResultFile
done
grep "correlation" TheResultFile > JustCCResults

```

The final output could then be sorted based on the cross-correlation scores in Excel. Note each PDB would be matched to the target densities with multiple orientations, resulting in multiple entries with the same PDB but different cross-correlation scores. The duplicate items for each PDB could be deleted in Excel.

Any additional information required to reanalyze the data reported in this work paper is available from the Lead Contact upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

**Mouse models**—Wild-type and transgenic male C57BL/6J mice at ages of 10 to 16 weeks were used for imaging in this study. All mice were cared for in compliance with the guidelines outlined in the Guide for the Care and Use of Laboratory Animals. All experiments were approved by the Janelia Research Campus (JRC) IACUC. JRC is an AAALAC-accredited institution. Mice were maintained under SPF conditions.

**Human sample**—A man aged 25–39 years old was recruited and consented to participate in this study. We did not bias ancestry, race or ethnicity throughout the recruitment process. We only checked the samples under the microscope to make sure the sperm were normozoospermic with a cell count of at least 30 million sperm cells per milliliter. All experimental procedures using human-derived samples were approved by the Committee

on Human Research at the University of California, Berkeley, under IRB protocol number 2013-06-5395.

## METHOD DETAILS

**Sample preparation.**—Mouse sperm were collected from 10 to 16-week-old C57Bl/6J mice based on the published protocol<sup>58</sup>. Briefly, the sperm were extracted from vasa deferentia by applying pressure to cauda epididymides in 1x Krebs buffer (1.2 mM  $\text{KH}_2\text{PO}_4$ , 120 mM NaCl, 1.2 mM  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 14 mM dextrose, 1.2 mM  $\text{CaCl}_2$ , 5 mM KCl, 25 mM  $\text{NaHCO}_3$ ). The sperm were washed and resuspended in ~100  $\mu\text{L}$  Krebs buffer for the following experiments. For human sperm samples, freshly ejaculated semen samples were obtained by masturbation.

**Grid preparation.**—EM grids (Quantifoil R 2/2 Au 200 mesh) were glow discharged to be hydrophilic using an easiGlow system (Pelco). The grid was then loaded onto a Leica GP2 plunge freezer (pre-equilibrated to 95% relative humidity at 25 °C). The mouse sperm suspension was then mixed with 10 nm gold beads (Electron Microscopy Science, cat #25487) to achieve final concentrations at 2–6 million cells/mL. EHNA (erythro-9-(2-hydroxy-3-nonyl)adenine) (Santa Cruz Biotechnology, CAS 51350-19-7) was added to a final concentration of 10 mM. Next, 3.5  $\mu\text{L}$  of the sperm mixture was loaded onto each grid, followed by a 15-second incubation period. The grids were then blotted for 4 sec and plunge-frozen in liquid ethane.

**Cryogenic focused ion beam (cryoFIB) milling.**—CryoFIB was performed using an Aquilos II cryo-FIB/SEM microscope (Thermo Fisher Scientific). A panorama SEM map of the whole grid was first taken at 377x magnification using an acceleration voltage of 5 kV with a beam current of 13 pA, and a dwell time of 1  $\mu\text{s}$ . Targets with appropriate thickness for milling were selected on the grid. A platinum layer (~10 nm) was sputter coated and a gas injection system (GIS) was used to deposit the precursor compound trimethyl(methylcyclopentadienyl) platinum (IV). The stage was tilted to 15–20°, corresponding to a milling angle of 8–13° relative to the plane of grids. FIB milling was performed using stepwise decreasing current as the lamellae became thinner (1.0 nA to 30 pA, final thickness: ~300 nm). The grids were then stored in liquid nitrogen before data collection.

**Image acquisition and tomogram reconstruction.**—Tilt series of mouse sperm were collected on a 300-kV Titan Krios transmission electron microscope (Thermo Fisher Scientific) equipped with a high brightness field emission gun (xFEG), a spherical aberration corrector, a Bioquantum energy filter (Gatan), and a K3 Summit detector (Gatan). The images were recorded at a nominal magnification of 26,000x in super-resolution counting mode using SerialEM<sup>49</sup>. After binning over  $2 \times 2$  pixels, the calibrated pixel size was 2.612 Å on the specimen level. For each tilt series, images were acquired using a modified dose-symmetric scheme between –48° and 48° relative to the lamella with 4° increments and grouping of two images on either side (0°, 4°, 8°, –4°, –8°, 12°, 16°, –12°, –16°, 20°...) <sup>59</sup>. At each tilt angle, the image was recorded as movies divided into fourteen

subframes. The total electron dose applied to a tilt series was  $100 \text{ e}^-/\text{\AA}^2$ . The defocus target was set to be  $-2$  to  $-5 \text{ }\mu\text{m}$ .

All movie frames were corrected with a gain reference collected in the same EM session. Movement between frames was corrected using MotionCor2 without dose weighting<sup>60</sup>. All tilt series were aligned using AreTomo and the tomograms were inspected as a screening step to identify good tilt series<sup>41</sup>. Tilt series with crystalline ice and big ice blocks, or possessing less than five doublets of the axonemes were discarded. Alignment of the good tilt series was then performed in Etomo using the gold beads as fiducial markers<sup>50</sup>. The AreTomo is less labor-intensive for screening purposes, while the Etomo workflow allowed us to achieve high-resolution reconstructions. The aligned tilt series were then CTF-corrected using TOMOCTF<sup>51</sup> and the tomograms were generated using TOMO3D<sup>52</sup> (bin4, pixel size:  $10.448 \text{ }\text{\AA}$ ). In total, we started with eight milling grids of mouse sperm and obtained 159 lamellae. Ultimately, the final reconstructions of the consensus averages were based on 77 usable tomograms.

**Subvolume averaging.**—Subvolume extraction, classification and refinement were first performed using RELION3 as reported previously<sup>61</sup>. Briefly, subvolumes from the doublets were manually picked every  $24 \text{ nm}$  and extracted at binning of 6 (pixel size:  $15.672 \text{ }\text{\AA}$ , box size:  $80$  pixels, dimension:  $125.376 \text{ nm}$ ). These subvolumes were aligned to a map of non-treated mouse sperm doublet structure (EMDB-27444) lowpass filtered to  $80 \text{ }\text{\AA}$  and the resulting map was used as the reference for further processing. Supervised 3D classification on radial spokes gave rise to four class averages of the  $96\text{-nm}$  repeating units at four different registers. All four class averages were recentered at the base of Radial spoke 2 and re-extracted at the same point at the binning of 4 (pixel size:  $10.448 \text{ }\text{\AA}$ , box size:  $120$  pixels, dimension:  $125.376 \text{ nm}$ ). All subvolumes were combined and aligned to one reference and duplicate subvolumes were removed based on minimum distance ( $< 40 \text{ nm}$ ). The remaining subvolumes were aligned to yield the consensus average for all nine doublets. Subvolumes of the  $96\text{-nm}$  repeating units were recentered on MIP features that repeat every  $48 \text{ nm}$  or  $16 \text{ nm}$  to obtain the coordinates of these subvolumes. The tomograms and coordinates were imported in RELION4 without binning<sup>24</sup>. Pseudo-subtomograms were extracted (pixel size:  $2.612 \text{ }\text{\AA}$ , box size:  $220$  pixels, dimension:  $57.464 \text{ nm}$ ) and the first round of Refine3D jobs yield the initial reference for the following refinement of geometric and optical parameters of the tilt series. TomoFrameAlign and CtfRefineTomo jobs were executed alternatively for two rounds and new pseudo-subtomograms were extracted (see FSC curves in Figure S2A). The same “Refine3D-TomoFrameAlign-CtfRefineTomo-TomoFrameAlign-CtfRefineTomo” process was repeated again and new pseudo-subtomograms were extracted. The final Refine3D job yields the reported maps. Further refinement did not improve the resolutions and quality of maps. In order to generate a map covering the entire  $48 \text{ nm}$ -repeating structure of the doublets, the run\_data.star file from the final Refine3D job was shifted along the longitudinal axis and another round of refinement yielded averages with the shifted register. These reconstructions were aligned to the reported  $48 \text{ nm}$  repeating structure of doublets from bovine trachea cilia and a composite map was generated to match the register of the periodic structure.

The resolutions of the maps were estimated based on the FSC of two independently refined half datasets (FSC = 0.143). Local resolution maps for doublets of both human and mouse sperm were calculated by RELION4 and displayed in UCSF Chimera<sup>53</sup>. These local-resolution maps represent relative differences in resolution across the maps but the absolute values may not be precise. IMOD was used to visualize the tomographic slices<sup>50</sup>. UCSF Chimera was used to manually segment the maps for various structural features and these maps were colored individually to prepare the figures using UCSF ChimeraX<sup>53,54,62</sup>.

### **Model building and unbiased matching of density maps to protein candidates.**

—Model building was performed in Coot v0.9.8.1<sup>55</sup> and rigid body fitting was achieved using UCSF Chimera. The interpretation of the mouse sperm doublet map started with the atomic model of the bovine trachea doublet (PDB 7RRO)<sup>13</sup>. Densities matching tubulins and 29 bovine MIPs in the bovine trachea doublet were found in the mouse sperm doublet map so all of these densities were considered to be formed by *M. musculus* orthologs (Figure S4). We cannot exclude the possibilities that they are formed by sperm-specific homologs with similar tertiary structures. These orthologs were identified using UniProt<sup>63</sup> or the NCBI protein database<sup>64</sup> based on the sequences of bovine proteins. The atomic models of bovine trachea MIPs were mutated to match the sequence of the mouse proteins using the Chainsaw plugin in Coot. The resulting models of individual proteins were then fit into the mouse sperm doublet map as rigid bodies in Chimera.

MIP densities that are unique in sperm doublets were segmented from the corresponding maps using UCSF Chimera manually. At 6–8 Å resolutions,  $\alpha$ -helix is well-resolved and  $\beta$ -strands appear as curved sheets and we focused on the unassigned densities with well-defined tertiary structures or domains (as the various colored densities shown in Figure 1). Meanwhile, the PDB library of 21,615 mouse proteins based on AlphaFold2 prediction was downloaded<sup>27</sup>. The unbiased matching was carried out using the COLORES program (Situs package, see the code in the Star Methods)<sup>65</sup>. The matching was scored and ranked by the cross-correlation scores and the top 200 hits were inspected individually with the target densities in UCSF Chimera. We noticed that matching of densities to much larger PDBs could lead to unrealistic cross correlations ( $>1$ ). Setting the box sizes of the maps to be two or three times larger compared to the isolated densities does not solve the issue since COLORES would cut off the zero valued edges by default to reduce computational loads. We thus used “voledit” command from SITUS to edit the voxels at the corner of the cubic map. Specifically, we edited the first and the last values in the sit file to be slightly larger than the threshold values to avoid the cropping. To test the unbiased proteome-wide search approach, we used densities corresponding to known MIPs identified in bovine tracheal doublets, including PACRG, CFAP20 and NME7 as controls for the method. Indeed, PDB models corresponding to the respective mouse orthologs were identified as the top hits (Data S1), suggesting this visual proteomics approach can identify protein with matching tertiary structures. We then applied this method to identify candidates for the mouse sperm-specific densities (Data S2–4). For the 4-helix bundle densities, the CATH library, which curated non-redundant PDBs of published structural domains<sup>36</sup>, was also used and no homologous proteins of SPACA9 were found.

The AlphaFold2 predicted PDBs were then used as starting models and initial fitting pose discovered by COLORES were inspected in Chimera. The various Tektin 5 and CCDC105 models were built in Coot to match the corresponding densities. Unresolved loops were deleted. We observed densities corresponding to 24 copies of SPACA9 for the 48 nm-repeating units of the doublet microtubules, albeit with varied occupancies. We rigid-body fitted all 24 SPACA9 in these densities in Coot to reflect the stacking oligomerization in the deposited model. We also performed rigid-body fitting of DUSP3 into the globular densities next to the slanted Tektin 5s. These models were combined in ChimeraX and all side chains were stripped using the *phenix.pdbtools* command.

**Sequence alignment and search for homologous proteins.**—Sequence alignment was performed using Clustal Omega<sup>32</sup> server and displayed in Jalview<sup>66</sup>. *M. musculus* Tektin 1 sequence was used as input to search for Tektin homologs using the HHpred server<sup>67</sup>.

**Biochemical extractions of mouse sperm.**—For each of the three biological replicates, sperm from two mice were washed with PBS and pelleted at 2000x g for 5 min. Then, the E1-E5 buffers were used to extract proteins from the pellets [0.1 % Triton in PBS (E1), 0.6 M NaCl in PBS (E2), 0.6 M KCSN in PBS (E3), 8 M urea (E4) and 10% SDS (E5)]. For E1 to E4, 100  $\mu$ L of the buffer was added to the pellets and the resuspension was mixed by pipetting up and down using a p200 pipette. Then the solution was incubated at room temperature for 10 min and the pellet was spun down at 21,000x g for 10 min. For E5, after 10% SDS was added and mixed, the resuspension was heated at 95° for 5 min. After the pellets were spun down, the supernatant was taken as the extraction. 20  $\mu$ L and 2  $\mu$ L of the extractions were used for SDS-PAGE analyses, either stained with AcquaStain (Fisher Scientific, NCO170988) and blotted with an antibody against  $\alpha$ -tubulins (ThermoFisher Scientific, DM1A, #62204). The remaining extractions were used for mass spectrometry analysis.

**Mass spectrometry (MS)-based global protein abundance of mouse sperm.**—Proteins in biochemical fractions E1, E2, E3, E4 and E5 from three biological replicates were reduced and alkylated in 4 mM final concentration tris (2-carboxyethyl) phosphine (TCEP) and 10 mM final concentration iodoacetamide by 20-minute incubation in the dark, after which excess iodoacetamide was quenched with 10 mM final concentration dithiothreitol (DTT). Proteins were then subjected to methanol chloroform precipitation. Briefly, 1 part sample was combined and vortexed sequentially with 4 parts methanol, 1 part chloroform, and 3 parts water for phase separation, after which samples were spun for 2 minutes at top speed (14,000 g) in a bench-top centrifuge (Centrifuge 5424R, Eppendorf). The upper phase was removed and discarded, and 4 parts methanol were combined and vortexed with the interphase and lower phase and subsequently centrifuged for 3 minutes at 14,000 g. The supernatant was removed and discarded, and the pellet was washed three times in 80% ice-cold acetone followed by centrifugation for 3 minutes at 14,000 g. Extracted proteins were air dried, resuspended in 8 M urea buffer (8 M urea, 150 mM NaCl, 50 mM  $\text{NH}_4\text{HCO}_3$ , cOmplete Mini EDTA-free protease inhibitor (Roche, 11836170001)), and quantified using Bradford reagent (Sigma, B6916) following Coomassie (Bradford)

Protein Assay Kit's protocol (Thermo Fisher, 23200). Following quantification, protein samples were diluted 4-fold to 2 M urea concentration with 0.1 M  $\text{NH}_4\text{HCO}_3$  pH 8, digested with trypsin (Promega, V5111) at a protease:protein ratio of 1:100 (weight/weight), and incubated overnight at 37°C in a thermomixer at 750 rpm.

After tryptic digest, samples were acidified to pH <3 with 1% final concentration formic acid, and desalted for MS analysis using HPLC-grade reagents and 100  $\mu\text{L}$  OMIX C18 tips (Agilent Technologies, A57003100) according to the manufacturer's protocol with the following adjustments. Briefly, OMIX tips were conditioned by sequential washes of 100% acetonitrile and 50% acetonitrile, 0.1% formic acid, and equilibrated with two washes of 0.1% formic acid. Peptides were bound to the C18 polymer by repeated pipetting, subsequently washed three times with 0.1% formic acid, and sequentially eluted in 50% acetonitrile, 0.1% formic acid followed by 90% acetonitrile, 0.1% formic acid. Peptides were dried by vacuum centrifugation (CentriVap Cold Trap, Labconco) and stored at  $-80^\circ\text{C}$  until MS analysis.

Digested, desalted peptides were resuspended to 0.125–2  $\mu\text{g}/\mu\text{L}$  final concentration in 2% acetonitrile, 0.1% formic acid. 1–2  $\mu\text{L}$  were injected in technical singlet onto an Easy-nLC 1200 (Thermo Fisher Scientific) interfaced via a nanoelectrospray source (Nanospray Flex) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). Peptides were separated on a PepSep reverse-phase C18 column (1.9  $\mu\text{m}$  particles, 1.5  $\mu\text{m} \times 15\text{ cm}$ , 150  $\mu\text{m}$  ID) (Bruker) with a gradient of 5–88% buffer B (0.1% formic acid in acetonitrile) over buffer A (0.1% formic acid in water) over a 100-minute data acquisition. Spectra were acquired continuously in a data-dependent manner. One full scan in the Orbitrap (scan range 350–1350  $m/z$  at 120,000 resolution in profile mode with a custom AGC target and maximum injection time of 50 milliseconds) was followed by as many MS/MS scans as could be acquired on the most abundant ions in 2 seconds in the dual linear ion trap (rapid scan type with fixed HCD collision energy of 32%, custom AGC target, maximum injection time of 50 milliseconds, and isolation window of 0.7  $m/z$ ). Singly and unassigned charge states were rejected. Dynamic exclusion was enabled with a repeat count of 1, an exclusion duration of 25 seconds, and an exclusion mass width of  $\pm 10$  ppm. Liquid chromatography<sup>68</sup> and MS acquisition parameters are reported in (table S1).

Raw MS files were searched using MaxQuant (version 1.6.3.3) against a database of the mouse proteome (SwissProt Mus musculus reviewed protein sequences, downloaded 07 May 2022) with a manual addition to include mouse piercer of microtubule wall 2 protein (protein sequence from NCBI Reference Sequence NP\_001185718.1, manually assigned the UniProt identifier “ZCC15orf65” in our database after its bovine homolog)<sup>37</sup>. MaxQuant settings were left at default, with the following exceptions: LFQ was enabled with skip normalization enabled; and match between runs was enabled with a 1.5-minute matching time window and 20-minute alignment window. Trypsin (KR|P) was selected and allowed up to two missed cleavages, and variable and fixed modifications were assigned for protein acetylation (N-terminal), methionine oxidation and carbamidomethylation.

Statistical analysis of protein quantitation was completed with R Bioconductor package artMS (version 1.14.0)<sup>56</sup> and its function artmsQuantification, which is a wrapper around

the R Bioconductor package Mass Spectrometry Statistics and Quantification (MSstats) (version 4.4.0) as follows<sup>38</sup> (table S2). Peptide intensities from the MaxQuant evidence file were summarized to protein intensities using the MSstats function `dataProcess` with default settings. The differences in log2-transformed intensity between biochemical fractions were scored using the MSstats function `groupComparison`, which fits a single linear model for each protein with a single categorical variable for condition, or fraction in our case. From these models, MSstats reports pairwise differences in means between conditions as log2 fold change (log2FC) with a p-value based on a t-test assuming equal variance across all conditions, and reports adjusted p-values using the false discovery rate (FDR) estimated by the Benjamini-Hochberg procedure. Proteins with significant changes in abundance between fractions were defined as: (1) absolute(log2FC) > 1; and (2) adjusted p-value < 0.05. Proteins with significant changes in abundance were tested for enrichment of Gene Ontology terms (table S3). The over-representation analysis was performed using the `enricher` function from R package `clusterProfiler` (version 4.4.1)<sup>39</sup>. Gene Ontology (GO Biological Process, Molecular Function and Cellular Component) terms and annotations were obtained from the R annotation package `org.Mm.eg.db` (version 3.15.0). From among all significantly enriched terms, we selected a set of non-redundant terms following a clustering procedure. We first constructed a term tree based on distances (1-Jaccard Similarity Coefficients of shared genes in KEGG or GO) between the significant terms. The term tree was cut at a specific level ( $h = 0.99$ ) to identify clusters of non-redundant gene sets (table S4). For results with multiple significant terms belonging to the same cluster, we selected the most significant (lowest adjusted p-value) term.

#### **Generation of Tektin-5 knockout mice and functional/structural analyses of mutant sperm.**

—To create an easily detected frameshift mutation in the *Tekt5* gene, we used two gRNAs located in exon 1 of the gene. There is a pseudogene located on chromosome 19 with 85.8% homology to the Tekin5 coding region. Three gRNAs were carefully selected to avoid cutting the pseudogene. They are gRNA1 (CGCTGGGTCTCCACGCGTTCAGG), gRNA2 (AGTTTCTGTGGCCCCAAGAAAGG), and gRNA3 (CCGAGGAATGCTCAGGCATCCGG). The gRNAs were in vitro transcribed using the MEGA shortscript T7 kit (Life Tech Corp AM1354).

Two combinations of the gRNAs were used: gRNA1 + gRNA2, and gRNA2 + gRNA3. The gRNAs and Cas9 protein (Invitrogen Truecut Cas9 protein V2, cat#A36498) with a concentration of 125 ng each were co-electroporated into 1-cell C57Bl/6J embryos using a BEX Genome Editor. A total of 75 pups were weaned and genotyped by PCR. Thirty-nine of them with an obviously shorter PCR band were sequenced, and nine of them with a frameshift deletion were selected for germline testing. Germline transmission was found in 7 out of 9, and 2 lines were selected for further breeding: line 6 with a 166 bp deletion and line 7 with a 181 bp deletion. Both lines were then mated with wild-type females to generate F1 animals. Mating of heterozygous F1 male and female led to F2 homozygotes, which were further confirmed by sequencing.

Sperm motility was recorded at 37°C on a Hamilton Thorne IVOS II CASA machine using a Zeiss 10x NH objective, at a frame rate of 60 Hz, in the presence of 1% polyvinyl alcohol

to prevent cell adhesion to glass. The motility and sperm morphologies were inspected and counted manually.

The structural analyses of the mutants were performed using the same workflow described for wild-type sperm, except omitting the EHNA treatment. Sperm from two different Tektin5-knockout lines were processed, imaged and analyzed independently. The two independent reconstructions show consistently low occupancies of specific Tektin 5 densities and only the higher-resolution reconstruction is shown.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses of *Tekt5*<sup>-/-</sup> mouse were prepared using Prism9 (GraphPad). We performed six mating trials by using six *Tekt5*<sup>-/-</sup> males (3 from each line of two lines) and six wildtype females. The average and standard deviation are presented ( $7.3 \pm 1.4$ , N=6). For the sperm analyses, we counted enough videos so the number of sperm from each mouse is > 200. Sperm from three *Tekt5*<sup>-/-</sup> knockout mutants and two wild-type mice were analyzed functionally.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

We are grateful to members of the Agard and Vale laboratories for the discussions and critical reading of the manuscript. We thank Xiaowei Zhao, Shixin Yang and Rui Yan from the CryoEM facility at Janelia Research Campus for their assistance with data collection. We thank Zanlin Yu for his suggestions on sample processing and model building. We thank Garrett Greenan, Shawn Zheng and Sam Li at UCSF for discussions on cryoET data processing. We thank Hao Wu at UCSF for suggestions on model building. We thank Willy Wrigger from Old Dominion University for his input and suggestions on SITUS package. EM data processing utilized computing resources at both the workstations at the CryoEM facility at Janelia Research Campus and UCSF HPC Wynton cluster. We also thank David Bulkley, Glenn Gilbert, and Matt Harrington from the UCSF CryoEM facility for their discussion on data collection and processing. We also thank Colin Morrow, Gillian Harris, Crystall Lopez and Catherine Lindsey from Janelia Vivarium for mouse experiments. Z.C. was supported by the Helen Hay Whitney Foundation Postdoctoral Fellowship. W.M.S. was supported by the National Science Foundation Graduate Research Fellowship Program under grant numbers DGE 1752814 and DGE 2146752. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. P.V.L. received funding from a Pew Biomedical Scholars Award and a GCRLE grant from the Global Consortium for Reproductive Longevity and Equality made possible by the Bia-Echo Foundation. D.A.A. received funding from NIH R35GM118099. R.D.V. received funding from NIH R35GM118106 and the Howard Hughes Medical Institute. The UCSF cryoEM facility was supported by NIH instrumentation grants 1S10OD026881, 1S10OD020054, and 1S10OD021741.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

## REFERENCE

1. Sironen A, Shoemark A, Patel M, Loebinger MR, and Mitchison HM (2020). Sperm defects in primary ciliary dyskinesia and related causes of male infertility. *Cell Mol Life Sci* 77, 2029–2048. 10.1007/s00018-019-03389-7. [PubMed: 31781811]
2. Fernandez-Lopez P, Garriga J, Casas I, Yeste M, and Bartumeus F (2022). Predicting fertility from sperm motility landscapes. *Commun Biol* 5, 1027. 10.1038/s42003-022-03954-0. [PubMed: 36171267]

3. Ishikawa T (2017). Axoneme Structure from Motile Cilia. Cold Spring Harbor perspectives in biology 9. 10.1101/cshperspect.a028076.
4. Linck RW, Chemes H, and Albertini DF (2016). The axoneme: the propulsive engine of spermatozoa and cilia and associated ciliopathies leading to infertility. J Assist Reprod Genet 33, 141–156. 10.1007/s10815-016-0652-1. [PubMed: 26825807]
5. Fawcett DW (1975). The mammalian spermatozoon. Dev Biol 44, 394–436. 10.1016/0012-1606(75)90411-x. [PubMed: 805734]
6. Beeby M, Ferreira JL, Tripp P, Albers SV, and Mitchell DR (2020). Propulsive nanomachines: the convergent evolution of archaella, flagella and cilia. FEMS Microbiol Rev 44, 253–304. 10.1093/femsre/uaa006. [PubMed: 32149348]
7. Pitnick S, Hosken DJ, and Birkhead TR (2009). Sperm morphological diversity in Sperm Biology - An Evolutionary Perspective (Academic Press, Cambridge, ed. 2, 2009), pp. 69–149.
8. Lindemann CB, and Lesich KA (2021). The many modes of flagellar and ciliary beating: Insights from a physical analysis. Cytoskeleton (Hoboken) 78, 36–51. 10.1002/cm.21656. [PubMed: 33675288]
9. Civetta A (2003). Positive selection within sperm-egg adhesion domains of fertilin: an ADAM gene with a potential role in fertilization. Molecular biology and evolution 20, 21–29. 10.1093/molbev/msg002. [PubMed: 12519902]
10. Lindemann CB (1996). Functional significance of the outer dense fibers of mammalian sperm examined by computer simulations with the geometric clutch model. Cell motility and the cytoskeleton 34, 258–270. 10.1002/(SICI)1097-0169(1996)34:4<258::AID-CM1>3.0.CO;2-4. [PubMed: 8871813]
11. Lindemann CB, and Lesich KA (2016). Functional anatomy of the mammalian sperm flagellum. Cytoskeleton (Hoboken) 73, 652–669. 10.1002/cm.21338. [PubMed: 27712041]
12. Ma M, Stoyanova M, Rademacher G, Dutcher SK, Brown A, and Zhang R (2019). Structure of the Decorated Ciliary Doublet Microtubule. Cell 179, 909–922 e912. 10.1016/j.cell.2019.09.030. [PubMed: 31668805]
13. Gui M, Farley H, Anujan P, Anderson JR, Maxwell DW, Whitchurch JB, Botsch JJ, Qiu T, Meleppattu S, Singh SK, et al. (2021). De novo identification of mammalian ciliary motility proteins using cryo-EM. Cell 184, 5791–5806 e5719. 10.1016/j.cell.2021.10.007. [PubMed: 34715025]
14. Khalifa AAZ, Ichikawa M, Dai D, Kubo S, Black CS, Peri K, McAlear TS, Veyron S, Yang SK, Vargas J, et al. (2020). The inner junction complex of the cilia is an interaction hub that involves tubulin post-translational modifications. eLife 9. 10.7554/eLife.52760.
15. Grossman-Haham I (2023). Towards an atomic model of a beating ciliary axoneme. Curr Opin Struct Biol 78, 102516. 10.1016/j.sbi.2022.102516. [PubMed: 36586349]
16. Kubo S, Black CS, Joachimiak E, Yang SK, Legal T, Peri K, Khalifa AAZ, Ghanaeian A, McCafferty CL, Valente-Paterno M, et al. (2023). Native doublet microtubules from Tetrahymena thermophila reveal the importance of outer junction proteins. Nature communications 14, 2168. 10.1038/s41467-023-37868-0.
17. Ichikawa M, Khalifa AAZ, Kubo S, Dai D, Basu K, Maghrebi MAF, Vargas J, and Bui KH (2019). Tubulin lattice in cilia is in a stressed form regulated by microtubule inner proteins. Proceedings of the National Academy of Sciences of the United States of America 116, 19930–19938. 10.1073/pnas.1911119116. [PubMed: 31527277]
18. Leung MR, Roelofs MC, Ravi RT, Maitan P, Henning H, Zhang M, Bromfield EG, Howes SC, Gadella BM, Bloomfield-Gadella H, and Zeev-Ben-Mordehai T (2021). The multi-scale architecture of mammalian sperm flagella and implications for ciliary motility. The EMBO journal 40, e107410. 10.15252/embj.2020107410. [PubMed: 33694216]
19. Chen Z, Greenan GA, Shiozaki M, Liu Y, Skinner W, Zhao X, Zhao S, Yan R, Guo C, Yu Z, et al. (2022). In situ cryo-electron tomography reveals the asymmetric architecture of mammalian sperm axonemes. Nature structural & molecular biology, in press.
20. Gadadhar S, Alvarez Viar G, Hansen JN, Gong A, Kostarev A, Ialy-Radio C, Leboucher S, Whitfield M, Ziyat A, Toure A, et al. (2021). Tubulin glycylation controls axonemal dynein activity, flagellar beat, and male fertility. Science 371. 10.1126/science.abd4914.

21. Bouchard P, Penningroth SM, Cheung A, Gagnon C, and Bardin CW (1981). erythro-9-[3-(2-Hydroxynonyl)]adenine is an inhibitor of sperm motility that blocks dynein ATPase and protein carboxylmethylase activities. *Proceedings of the National Academy of Sciences of the United States of America* 78, 1033–1036. 10.1073/pnas.78.2.1033. [PubMed: 6453342]
22. Nicastro D, Schwartz C, Pierson J, Gaudette R, Porter ME, and McIntosh JR (2006). The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* 313, 944–948. 10.1126/science.1128618. [PubMed: 16917055]
23. Li S, Fernandez JJ, Fabritius AS, Agard DA, and Winey M (2022). Electron cryo-tomography structure of axonemal doublet microtubule from *Tetrahymena thermophila*. *Life Sci Alliance* 5. 10.26508/lsa.202101225.
24. Zivanov J, Oton J, Ke Z, von Kugelgen A, Pyle E, Qu K, Morado D, Castano-Diez D, Zanetti G, Bharat TAM, et al. (2022). A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *eLife* 11. 10.7554/eLife.83724.
25. Song K, Shang Z, Fu X, Lou X, Grigorieff N, and Nicastro D (2020). In situ structure determination at nanometer resolution using TYGRESS. *Nature methods* 17, 201–208. 10.1038/s41592-019-0651-0. [PubMed: 31768058]
26. Linck R, Fu X, Lin J, Ouch C, Scheffter A, Steffen W, Warren P, and Nicastro D (2014). Insights into the structure and function of ciliary and flagellar doublet microtubules: tektins, Ca<sup>2+</sup>-binding proteins, and stable protofilaments. *The Journal of biological chemistry* 289, 17427–17444. 10.1074/jbc.M114.568949. [PubMed: 24794867]
27. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. 10.1038/s41586-021-03819-2. [PubMed: 34265844]
28. Wriggers W, Milligan RA, and McCammon JA (1999). Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *Journal of structural biology* 125, 185–195. 10.1006/jsbi.1998.4080. [PubMed: 10222274]
29. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419. 10.1126/science.1260419. [PubMed: 25613900]
30. Baker MA, Hetherington L, Reeves GM, and Aitken RJ (2008). The mouse sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. *Proteomics* 8, 1720–1730. 10.1002/pmic.200701020. [PubMed: 18340633]
31. Firat-Karalar EN, Sante J, Elliott S, and Stearns T (2014). Proteomic analysis of mammalian sperm cells identifies new components of the centrosome. *Journal of cell science* 127, 4128–4133. 10.1242/jcs.157008. [PubMed: 25074808]
32. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, and Lopez R (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic acids research*. 10.1093/nar/gkac240.
33. Han L, Rao Q, Yang R, Wang Y, Chai P, Xiong Y, and Zhang K (2022). Cryo-EM structure of an active central apparatus. *Nature structural & molecular biology* 29, 472–482. 10.1038/s41594-022-00769-9.
34. Rupp G, O'Toole E, and Porter ME (2001). The *Chlamydomonas* PF6 locus encodes a large alanine/proline-rich polypeptide that is required for assembly of a central pair projection and regulates flagellar motility. *Molecular biology of the cell* 12, 739–751. 10.1091/mbc.12.3.739. [PubMed: 11251084]
35. Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, and Steinegger M (2022). ColabFold: making protein folding accessible to all. *Nature methods* 19, 679–682. 10.1038/s41592-022-01488-1. [PubMed: 35637307]
36. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, Pang CSM, Woodridge L, Rauer C, Sen N, et al. (2021). CATH: increased structural coverage of functional space. *Nucleic acids research* 49, D266–D273. 10.1093/nar/gkaa1079. [PubMed: 33237325]
37. Cox J, and Mann M (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26, 1367–1372. 10.1038/nbt.1511. [PubMed: 19029910]

38. Choi M, Chang CY, Clough T, Broudy D, Killeen T, MacLean B, and Vitek O (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 30, 2524–2526. 10.1093/bioinformatics/btu305. [PubMed: 24794931]
39. Yu G, Wang LG, Han Y, and He QY (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. 10.1089/omi.2011.0118. [PubMed: 22455463]
40. Fernandez JJ, Li S, Bharat TAM, and Agard DA (2018). Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. *Journal of structural biology* 202, 200–209. 10.1016/j.jsb.2018.02.001. [PubMed: 29410148]
41. Zheng S, Wolff G, Greenan G, Chen Z, Faas FGA, Barcena M, Koster AJ, Cheng Y, and Agard DA (2022). AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. *J Struct Biol X* 6, 100068. 10.1016/j.yjsbx.2022.100068. [PubMed: 35601683]
42. Fontana P, Dong Y, Pi X, Tong AB, Hecksel CW, Wang L, Fu TM, Bustamante C, and Wu H (2022). Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. *Science* 376, eabm9326. 10.1126/science.abm9326. [PubMed: 35679401]
43. Gadelha H, Hernandez-Herrera P, Montoya F, Darszon A, and Corkidi G (2020). Human sperm uses asymmetric and anisotropic flagellar controls to regulate swimming symmetry and cell steering. *Sci Adv* 6, eaba5168. 10.1126/sciadv.aba5168. [PubMed: 32789171]
44. Roy A, Lin YN, Agno JE, DeMayo FJ, and Matzuk MM (2007). Absence of tektin 4 causes asthenozoospermia and subfertility in male mice. *FASEB J* 21, 1013–1025. 10.1096/fj.06-7035com. [PubMed: 17244819]
45. Roy A, Lin YN, Agno JE, DeMayo FJ, and Matzuk MM (2009). Tektin 3 is required for progressive sperm motility in mice. *Mol Reprod Dev* 76, 453–459. 10.1002/mrd.20957. [PubMed: 18951373]
46. Zhou L, Liu H, Liu S, Yang X, Dong Y, Pan Y, Xiao Z, Zheng B, Sun Y, Huang P, et al. (2023). Structures of sperm flagellar doublet microtubules expand the genetic spectrum of male infertility. *Cell* 186, 2897–2910 e2819. 10.1016/j.cell.2023.05.009. [PubMed: 37295417]
47. Leung MR, Zeng J, Wang X, Roelofs MC, Huang W, Zenezini Chiozzi R, Hevler JF, Heck AJR, Dutcher SK, Brown A, et al. (2023). Structural specializations of the sperm tail. *Cell* 186, 2880–2896 e2817. 10.1016/j.cell.2023.05.026. [PubMed: 37327785]
48. Gomez-Roman N, Felton-Edkins ZA, Kenneth NS, Goodfellow SJ, Athineos D, Zhang J, Ramsbottom BA, Innes F, Kantidakis T, Kerr ER, et al. (2006). Activation by c-Myc of transcription by RNA polymerases I, II and III. *Biochemical Society symposium*, 141–154. [PubMed: 16626295]
49. Mastronarde DN (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology* 152, 36–51. 10.1016/j.jsb.2005.07.007. [PubMed: 16182563]
50. Kremer JR, Mastronarde DN, and McIntosh JR (1996). Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology* 116, 71–76. 10.1006/jsbi.1996.0013. [PubMed: 8742726]
51. Fernandez JJ, Li S, and Crowther RA (2006). CTF determination and correction in electron cryotomography. *Ultramicroscopy* 106, 587–596. 10.1016/j.ultramic.2006.02.004. [PubMed: 16616422]
52. Agulleiro JI, and Fernandez JJ (2015). Tomo3D 2.0--exploitation of advanced vector extensions (AVX) for 3D reconstruction. *Journal of structural biology* 189, 147–152. 10.1016/j.jsb.2014.11.009. [PubMed: 25528570]
53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605–1612. 10.1002/jcc.20084. [PubMed: 15264254]
54. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, and Ferrin TE (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 27, 14–25. 10.1002/pro.3235. [PubMed: 28710774]

55. Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66, 486–501. 10.1107/S0907444910007493. [PubMed: 20383002]
56. Jimenez-Morales D, Rosa Campos A, Von Dollen J, Krogan N, and Swaney D (2023). artMS: Analytical R tools for Mass Spectrometry. R package version 1.18.0.
57. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic acids research* 47, D442–D450. 10.1093/nar/gky1106. [PubMed: 30395289]
58. van der Spoel AC, Jeyakumar M, Butters TD, Charlton HM, Moore HD, Dwek RA, and Platt FM (2002). Reversible infertility in male mice after oral administration of alkylated imino sugars: a nonhormonal approach to male contraception. *Proceedings of the National Academy of Sciences of the United States of America* 99, 17173–17178. 10.1073/pnas.262586099. [PubMed: 12477936]
59. Hagen WJH, Wan W, and Briggs JAG (2017). Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of structural biology* 197, 191–198. 10.1016/j.jsb.2016.06.007. [PubMed: 27313000]
60. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, and Agard DA (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods* 14, 331–332. 10.1038/nmeth.4193. [PubMed: 28250466]
61. Bharat TA, and Scheres SH (2016). Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature protocols* 11, 2054–2065. 10.1038/nprot.2016.124. [PubMed: 27685097]
62. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, and Ferrin TE (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* 30, 70–82. 10.1002/pro.3943. [PubMed: 32881101]
63. UniProt C (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic acids research* 49, D480–D489. 10.1093/nar/gkaa1100. [PubMed: 33237286]
64. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. (2022). Database resources of the national center for biotechnology information. *Nucleic acids research* 50, D20–D26. 10.1093/nar/gkab1112. [PubMed: 34850941]
65. Wriggers W (2012). Conventions and workflows for using Situs. *Acta crystallographica. Section D, Biological crystallography* 68, 344–351. 10.1107/S0907444911049791. [PubMed: 22505255]
66. Waterhouse AM, Procter JB, Martin DM, Clamp M, and Barton GJ (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. 10.1093/bioinformatics/btp033. [PubMed: 19151095]
67. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, and Alva V (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of molecular biology* 430, 2237–2243. 10.1016/j.jmb.2017.12.007. [PubMed: 29258817]
68. Cruz-Gonzalez I, Gonzalez-Ferreiro R, Freixa X, Gafoor S, Shakir S, Omran H, Berti S, Santoro G, Kefer J, Landmesser U, et al. (2020). Left atrial appendage occlusion for stroke despite oral anticoagulation (resistant stroke). Results from the Amplatzer Cardiac Plug registry. *Rev Esp Cardiol (Engl Ed)* 73, 28–34. 10.1016/j.rec.2019.02.013. [PubMed: 31036510]

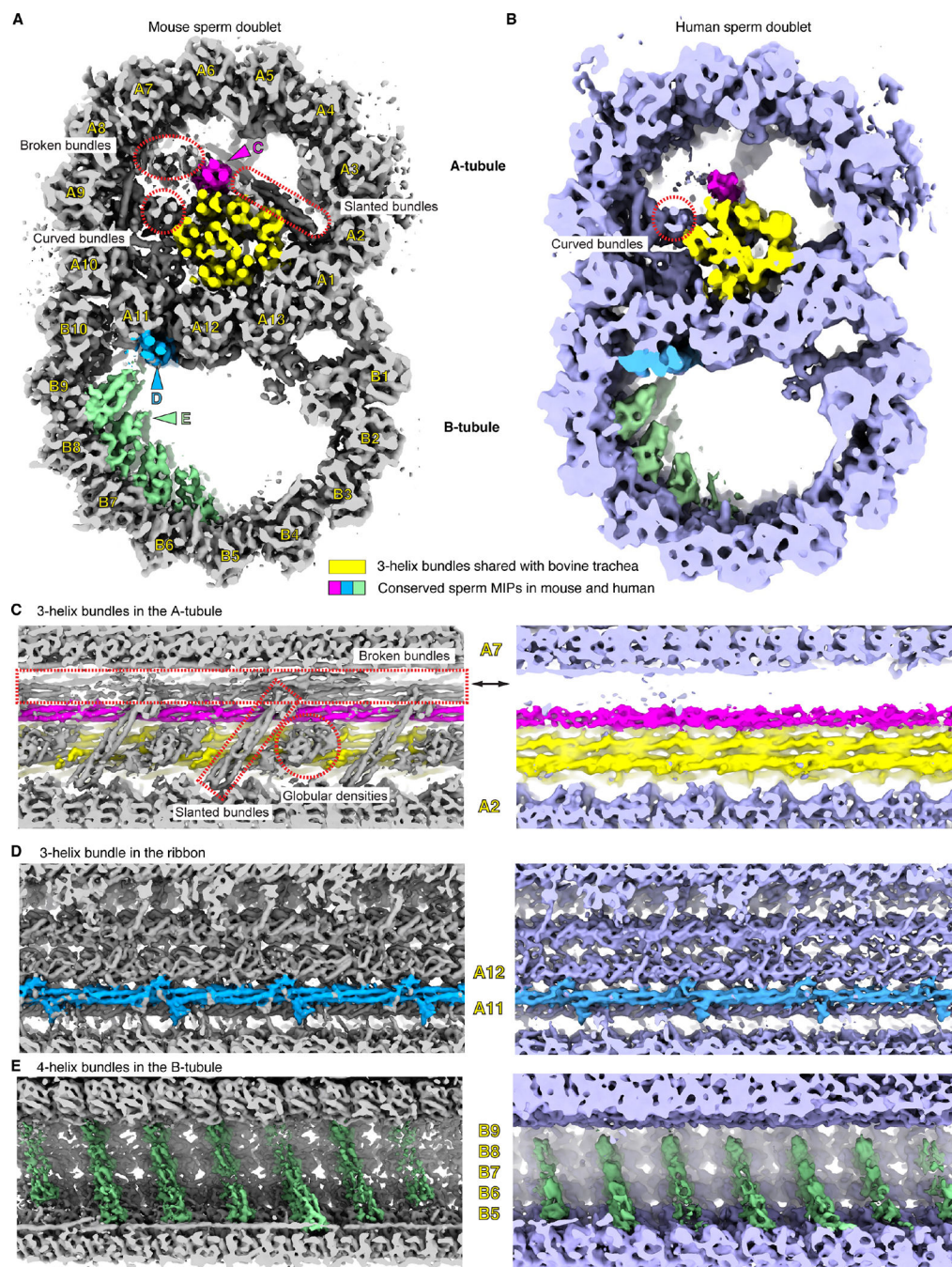
**HIGHLIGHTS**

*In situ* cryoET revealed native structures of doublet microtubules in mammalian sperm

Subtomogram averaging led to 6-Å reconstruction of microtubule doublets

Protein discovery by matching structures to the AlphaFold2-predicted mouse proteome

Sperm doublets feature a plastic and partially redundant Tektin 5 network



**Figure 1. The 3D reconstructions of mouse and human sperm doublets revealed novel MIPs.** (A), (B) Transverse cross-section views of the doublets of mouse (A) and human (B) sperm. Conserved sperm MIP densities are highlighted (pink, blue and green) and the corresponding viewing angles of (C)-(E) are indicated (colored arrowheads). The 3-helix densities in A-tubule shared with Bovine trachea doublets (EMD-24664) are colored (yellow)<sup>13</sup>. Divergent sperm densities are also indicated (red dashed shapes). Individual protofilaments of the doublets are labeled as A1–13 and B1–10. (C)-(E). Zoom-in views of the conserved sperm MIP densities along the longitudinal axis. In (C), mouse sperm-specific

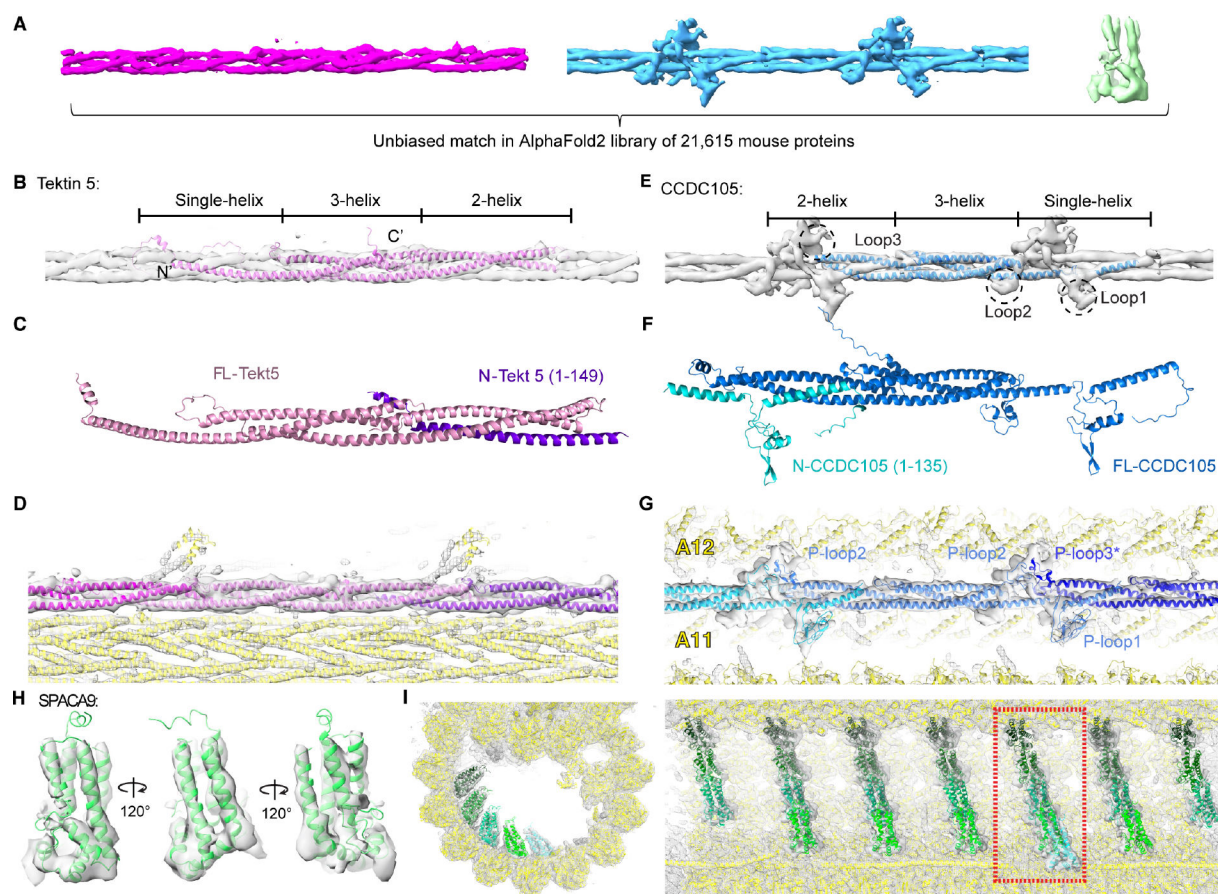
densities are indicated and labeled (red dashed shapes, see more in Figures S2 and S3). In (E), although the striations are 8 nm apart from one another, the overall periodicity is 48 nm.

Author Manuscript

Author Manuscript

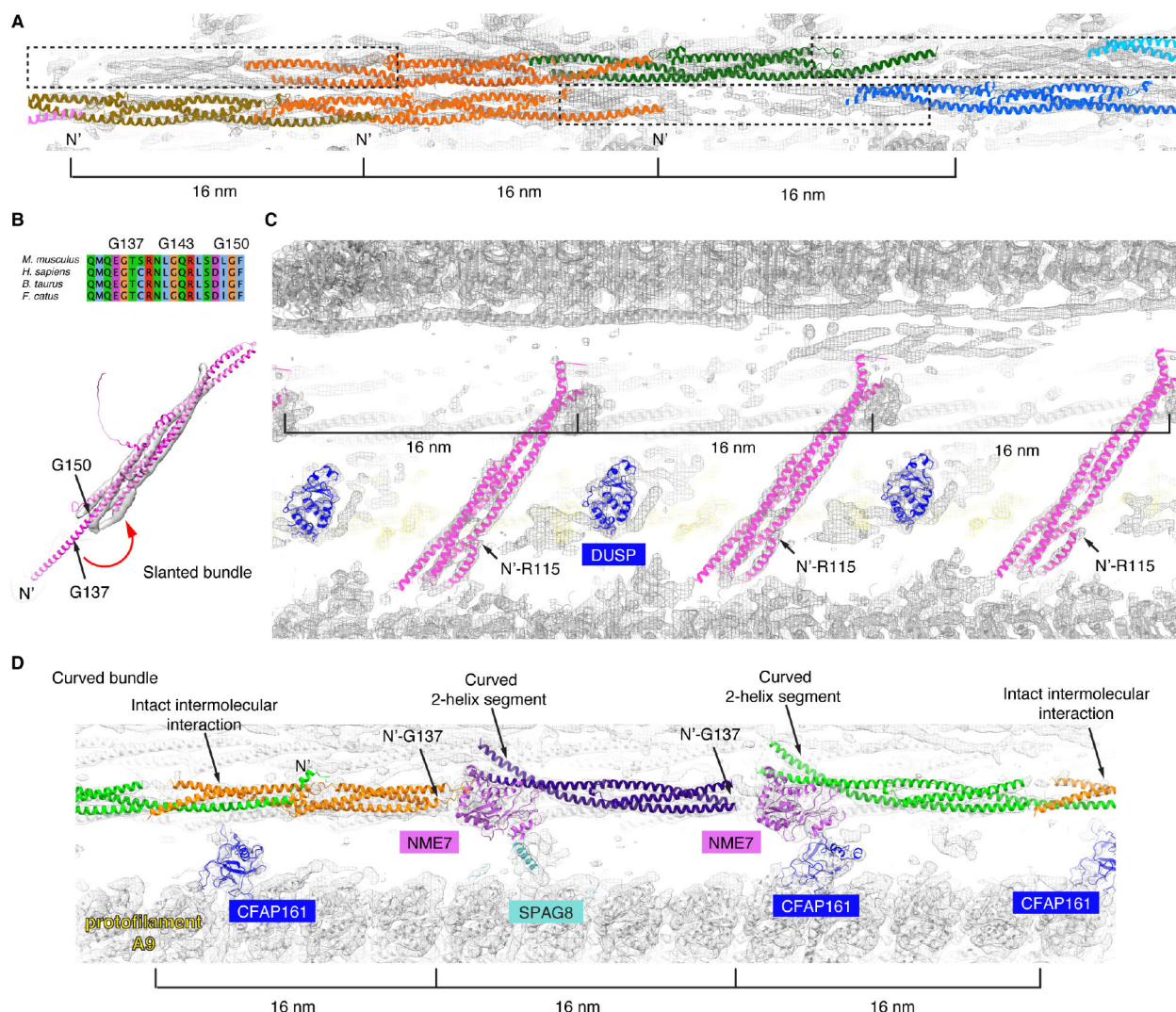
Author Manuscript

Author Manuscript



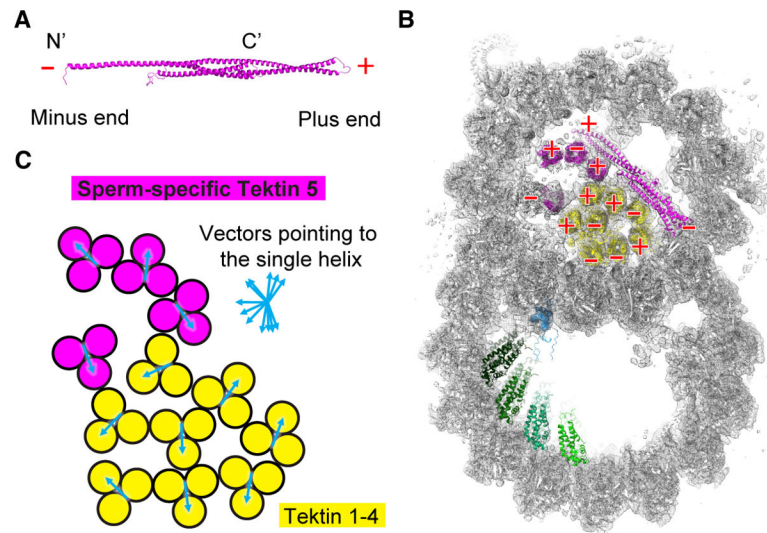
**Figure 2. *De novo* protein identification of sperm MIPs assisted by AlphaFold2.**

(A) Conserved densities in mouse and human sperm were segmented from the averages of 16-nm repeats of mouse sperm doublets and searched in the AlphaFold2 library of the mouse proteome (21,615 proteins). (B) The predicted structure of Tektin 5 based on AlphaFold2 was fitted into the continuous 3-helix bundle. (C) Modeling of a complex formed by a full-length Tektin 5 and a truncated one (N-Tekt 5: a.a. 1–149) using Colabfold<sup>35</sup>. (D) Fitting and modeling of Tektin 5s into the 3-helix bundle densities in the A-tubule. The nearby densities accounted for by other proteins are also shown (yellow ribbon). (E) An unbiased search in the AlphaFold2 library identified CCDC105 as the candidate for the continuous 3-helix density at the ribbon. The three conserved proline-rich loops among CCDC105 orthologs could account for the protrusion densities but were not modeled (See Figures S6C–D). (F) Modeling of a complex formed by a full-length CCDC105 and a truncated one (N-CCDC105: a.a. 1–135) using Colabfold<sup>35</sup>. (G) Fitting and modeling of CCDC105 into the 3-helix bundle density at the ribbon. The nearby densities are accounted for by other proteins (yellow ribbon). (H) The AlphaFold2 model for SPACA9 was directly fitted into the density and viewed from different angles. (I) Two orthogonal views of the striations of SPACA9 in the B-tubule. Different SPACA9 molecules are colored with different shades of green. The left panel showed a particular striation indicated in the right panel (the dashed rectangle).

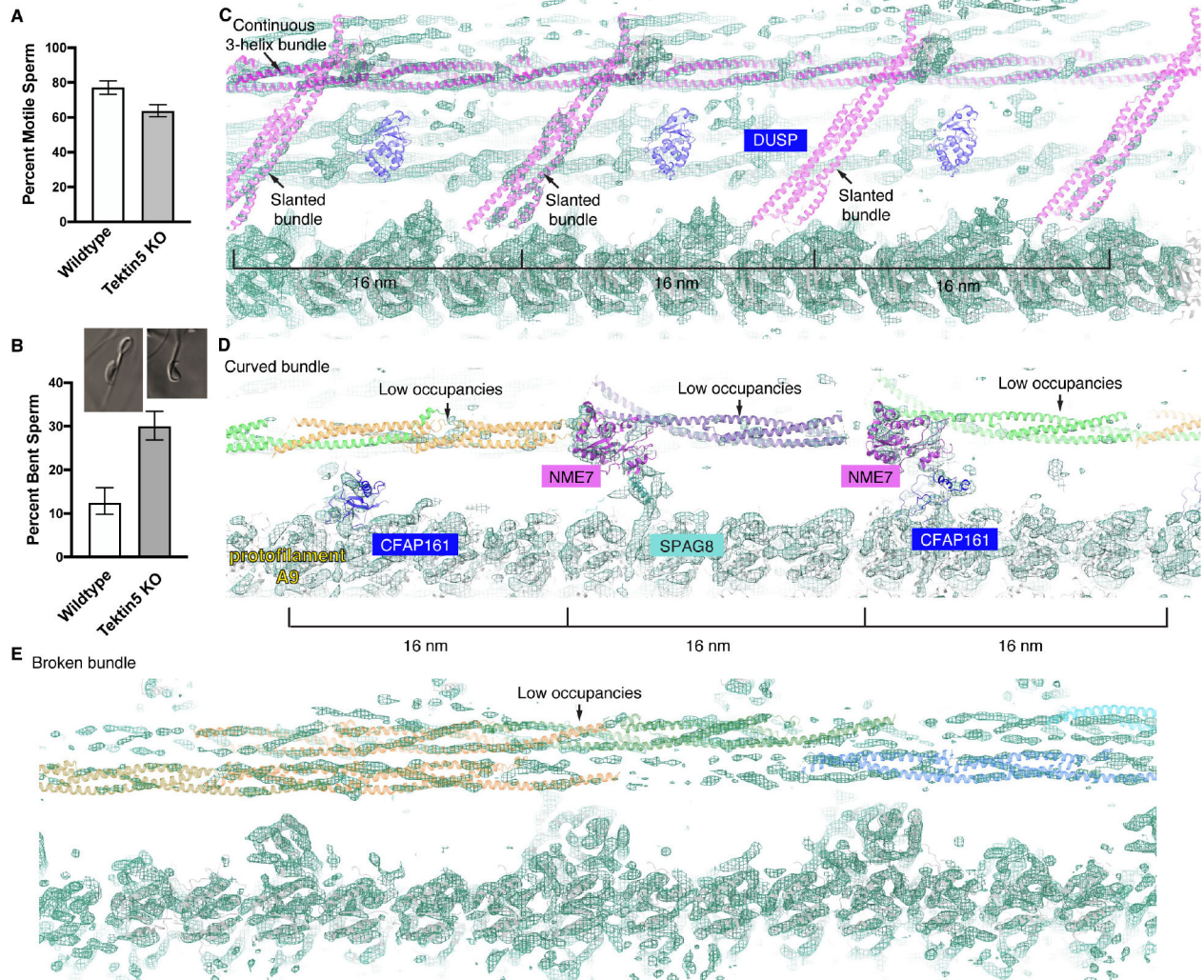


**Figure 3. Conformational plasticity of Tektin 5.**

(A) The two broken 3-helix bundles could be explained by two complete and a third partial copies of Tektin 5 (dashed rectangles) per 48-nm repeat, instead of three Tektins in the continuous 3-helix bundle. (B) The AlphaFold2 model of mouse Tektin 5 was fitted into the slanted helical densities. Sequence alignment of Tektin 5 from *M. musculus*, *H. sapiens*, *B. taurus* and *F. catus* is shown from Q133-F151 (the numbering of amino acids is based on *M. musculus* Tektin 5). The conserved Gly137, Gly143 and Gly150 are near the turning point of the bent  $\alpha$ -helix. (C) The fitting of Tektin 5 and DUSP3 protein (its homologs are also possible candidates) into the 16-nm repeating features, see the same view of the map in Figure 1C. (D) Three modified Tektin 5 were fitted into the densities of curved bundles in the mouse sperm doublet (as indicated in Figure 1A). The intact intermolecular interaction interface, N-termini of the Tektin 5s and curved 2-helix segments are indicated (arrows). Nearby MIPs shared between mouse sperm flagella and bovine trachea cilia are also colored and labeled (NME7, CFAP161, SPAG8). (E) The cross-section schematic is shown. The highlighted models of panels A-D are indicated using arrows.



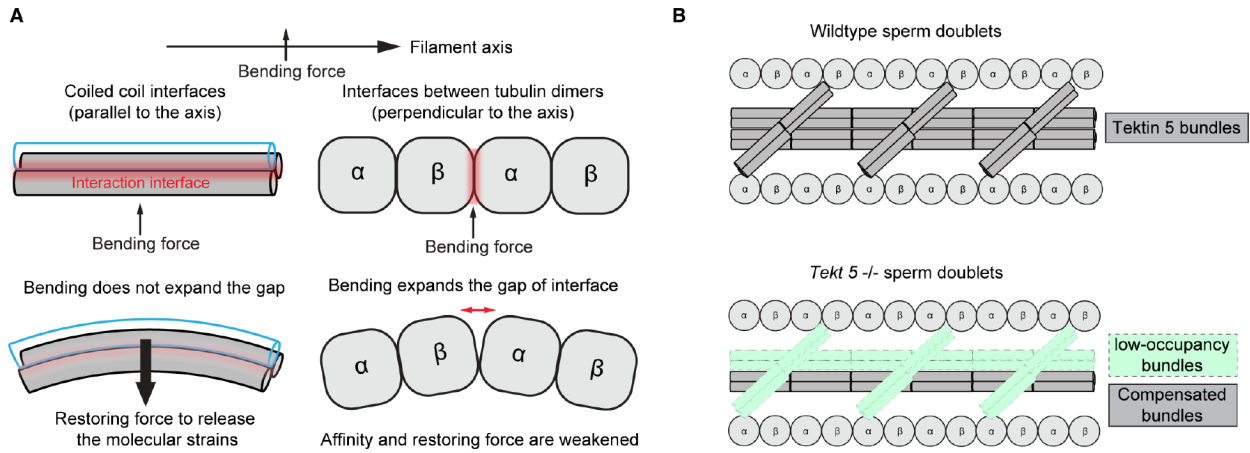
**Figure 4. Sperm doublets are composed of microtubules and extensive coiled-coil bundles.** (A) The plus and minus ends of Tektin 5 were named based on the N- and C-termini of the protein. (B) The cross-section view of the mouse sperm doublets shows the polarities of 3-helix bundles pointing toward the readers. (C) The orientations for each 3-helix bundle were represented by a vector starting from the middle point of the 2-helix segment and pointing toward the single-helix segment of the other Tektin molecule.



**Figure 5. Characterization of mutant *Tekt 5*  $-/-$  sperm.**

(A) The percentages of motile sperm from wild-type and *Tekt 5* knockout mice (> 200 cells were counted for each mouse and three knockout  $-/-$  mice and two wild-type mice were analyzed, the pool percentage and 95% Confidence Intervals (Wilson/Brown method) were shown). (B) The percentages of bent sperm from wild-type and *Tekt 5* knockout mice (> 200 cells were counted for each mouse and three knockout  $-/-$  mice and two wild-type mice were analyzed, the pool percentage and 95% Confidence Intervals by Wilson/Brown method were shown). Two examples of bent sperm are shown. (C) An overlay of wild-type models with the densities of *Tekt 5*  $-/-$  sperm around the slanted bundles. The continuous 3-helix bundle assigned as Tektin 5 (high occupancies) and slanted helical bundles (low occupancies) are shown. The densities corresponding to the DUSP proteins are barely resolved. Note there are substantially less densities for these models compared to Figure 3C. (D) An overlay of wildtype models with the densities of *Tekt 5*  $-/-$  sperm around the curved bundles. The occupancies of the curved bundles are lower than the other MIPs and tubulins. Note there are substantially less densities for these models compared to Figure 3D. (E) The

two broken 3-helix bundles have lower occupancies compared to the surrounding MIPs and tubulins. Note there are substantially less densities for these models compared to Figure 3A.



**Figure 6. Coiled-coil interfaces are suitable to withstand mechanical stress from orthogonal directions.**

(A) A model of how 3-helix bundles would be able to bear mechanical stress differently compared to the microtubules. The bending curvatures and gaps are exaggerated for illustration purposes. (B) A schematic of wild-type and mutant sperm doublets structures highlighting the Tektin 5 bundles and the partial redundancy.

## KEY RESOURCES TABLE

| REAGENT or RESOURCE                                                         | SOURCE                                                         | IDENTIFIER                                                                        |
|-----------------------------------------------------------------------------|----------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Biological samples                                                          |                                                                |                                                                                   |
| Mouse ( <i>Mus musculus</i> ) sperm                                         | Gene Targeting and Transgenics Center, Janelia Research Campus | N/A                                                                               |
| Human ( <i>Homo sapiens</i> ) sperm                                         | Lishko Laboratory, UC Berkeley                                 | N/A                                                                               |
| Chemicals, peptides, and recombinant proteins                               |                                                                |                                                                                   |
| NaCl                                                                        | Sigma-Aldrich                                                  | Cat #71376                                                                        |
| KH <sub>2</sub> PO <sub>4</sub>                                             | Sigma-Aldrich                                                  | Cat #P0662                                                                        |
| MgSO <sub>4</sub> ·7H <sub>2</sub> O                                        | Sigma-Aldrich                                                  | Cat #M3409                                                                        |
| Dextrose                                                                    | Sigma-Aldrich                                                  | Cat # D9434                                                                       |
| CaCl <sub>2</sub>                                                           | Sigma-Aldrich                                                  | Cat #21097                                                                        |
| KCl                                                                         | Sigma-Aldrich                                                  | Cat #60128                                                                        |
| NaHCO <sub>3</sub>                                                          | Sigma-Aldrich                                                  | Cat #S5761                                                                        |
| DTT                                                                         | Sigma-Aldrich                                                  | Cat #DTT-RO                                                                       |
| TCEP                                                                        | Sigma-Aldrich                                                  | Cat #C4706                                                                        |
| erythro-9-(2-hydroxy-3-nonyl)adenine                                        | Santa Cruz Biotechnology                                       | Cat #sc-201184                                                                    |
| KCSN                                                                        | Sigma-Aldrich                                                  | Cat #60178                                                                        |
| Urea                                                                        | Sigma-Aldrich                                                  | Cat #U5128                                                                        |
| SDS                                                                         | Sigma-Aldrich                                                  | Cat #62862                                                                        |
| Triton                                                                      | Sigma-Aldrich                                                  | Cat #X100PC                                                                       |
| Deposited data                                                              |                                                                |                                                                                   |
| Cryo-EM map of the 16 nm-repeating A-tubule (mouse)                         | This paper                                                     | EMD-41315                                                                         |
| Cryo-EM map of the 16 nm-repeating B-tubule (mouse)                         | This paper                                                     | EMD-41316                                                                         |
| Cryo-EM map of the 48 nm-repeating doublets (mouse), the composite map      | This paper                                                     | EMD-41431                                                                         |
| Cryo-EM map of the 48 nm-repeating doublets (mouse), submap 1 of EMD-41431  | This paper                                                     | EMD-41450                                                                         |
| Cryo-EM map of the 48 nm-repeating doublets (mouse), submap 2 of EMD-41431  | This paper                                                     | EMD-41451                                                                         |
| Cryo-EM map of the 48 nm-repeating doublets (human)                         | This paper                                                     | EMD-41317                                                                         |
| Cryo-EM map of the 48 nm-repeating doublets ( <i>Tekt5</i> <sup>-/-</sup> ) | This paper                                                     | EMD-41320                                                                         |
| Model of the mouse sperm doublets                                           | This paper                                                     | PDB: 8TO0                                                                         |
| PRIDE partner repository for MS data                                        | This paper                                                     | PXD036885                                                                         |
| R package source materials for MSstats from Krogan Lab                      | This paper                                                     | <a href="https://github.com/kroganlab">https://github.com/kroganlab</a>           |
| Experimental models: Organisms/strains                                      |                                                                |                                                                                   |
| Mouse: C57BL/6J                                                             | The Jackson Laboratory                                         | <a href="https://www.jax.org/strain/000664">https://www.jax.org/strain/000664</a> |
| <i>Tekt5</i> <sup>-/-</sup> mouse                                           | This paper                                                     | N/A                                                                               |

| REAGENT or RESOURCE                                 | SOURCE                                   | IDENTIFIER                                                                                                                      |
|-----------------------------------------------------|------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Software and algorithms                             |                                          |                                                                                                                                 |
| Prism v8                                            | GraphPad                                 | <a href="https://www.graphpad.com/">https://www.graphpad.com/</a>                                                               |
| SerialEM 3.8                                        | Mastronarde <sup>49</sup>                | <a href="https://bio3d.colorado.edu/SerialEM/">https://bio3d.colorado.edu/SerialEM/</a>                                         |
| Etomo                                               | Kremer et al. <sup>50</sup>              | <a href="https://bio3d.colorado.edu/imod/doc/UsingEtomo.html">https://bio3d.colorado.edu/imod/doc/UsingEtomo.html</a>           |
| TOMOCTF                                             | Fernandez et al. <sup>51</sup>           | <a href="https://sites.google.com/site/3demimageprocessing/tomocf">https://sites.google.com/site/3demimageprocessing/tomocf</a> |
| TOMO3D 2.0                                          | Agulleiro and Fernandez <sup>52</sup>    | <a href="https://sites.google.com/site/3demimageprocessing/tomo3d">https://sites.google.com/site/3demimageprocessing/tomo3d</a> |
| Chimera                                             | Pettersen et al. <sup>53</sup>           | <a href="https://www.cgl.ucsf.edu/chimera/">https://www.cgl.ucsf.edu/chimera/</a>                                               |
| ChimeraX                                            | Goddard et al. <sup>54</sup>             | <a href="https://www.rbvi.ucsf.edu/chimerax/">https://www.rbvi.ucsf.edu/chimerax/</a>                                           |
| RELION-4.0                                          | Zivanov et al. <sup>24</sup>             | <a href="https://relion.readthedocs.io/en/release-4.0/">https://relion.readthedocs.io/en/release-4.0/</a>                       |
| AlphaFold2                                          | Jumper et al. <sup>27</sup>              | <a href="https://alphafold.ebi.ac.uk/">https://alphafold.ebi.ac.uk/</a>                                                         |
| Situs                                               | Wriggers et al. <sup>28</sup>            | <a href="https://situs.biomachina.org/fguide.html">https://situs.biomachina.org/fguide.html</a>                                 |
| Coot 0.9.8.1                                        | Emsley et al. <sup>55</sup>              | <a href="http://www2.mrc-lmb.cam.ac.uk/personal/%20pemsley/coot">http://www2.mrc-lmb.cam.ac.uk/personal/%20pemsley/coot</a>     |
| MaxQuant 1.6.3.3                                    | Cox and Mann <sup>37</sup>               | <a href="https://www.maxquant.org/">https://www.maxquant.org/</a>                                                               |
| R Bioconductor package artMS 1.14.0                 | Jimenez-Morales, D. et al. <sup>56</sup> | <a href="https://doi.org/10.18129/B9.bioc.artMS">doi:10.18129/B9.bioc.artMS</a>                                                 |
| Other                                               |                                          |                                                                                                                                 |
| Quantifoil holey carbon grids (R2/2, 200-mesh gold) | Quantifoil MicroTools GmbH               | <a href="https://www.quantifoil.com/products">https://www.quantifoil.com/products</a>                                           |
| EM GP2 Automatic Plunge Freezer                     | Leica Microsystems                       | <a href="https://www.leica-microsystems.com/products">https://www.leica-microsystems.com/products</a>                           |