Integrative modeling reveals the molecular architecture of the Intraflagellar Transport A (IFT-A) complex

Caitlyn L. McCafferty[1], Ophelia Papoulas[1], Mareike A. Jordan[2], Gabriel Hoogerbrugge[1], Candice Nichols[1], Gaia Pigino[3], David W. Taylor[1], John B. Wallingford[1], Edward M. Marcotte[1]

[1]Department of Molecular Biosciences, Center for Systems and Synthetic Biology,
University of Texas, Austin, TX 78712, USA
[2]Max Planck Institute of Molecular Cell Biology and Genetics,
D-01307 Dresden, Germany
[3]Human Technopole, Via Cristina Belgioioso 171, 20157 Milan, Italy

Correspondence to: C.L.M., caitie.mccafferty@gmail.com; E.M.M., marcotte@utexas.edu

ORCIDs: C.L.M., 0000-0002-0872-4527; O.P., 0000-0002-6370-0616; M.A.J., 0000-0001-6248-8863; G.P., 0000-0002-2295-9568; D.W.T., 0000-0002-6198-1194; J.B.W., 0000-0001-8701-4293; E.M.M., 0000-0001-8808-180X

Summary

The 3D structure of the six-subunit complex and its polymeric assembly gives insights into cargo transport in cilia and how specific mutations in these genes lead to ciliopathy birth defects.

Abstract

Intraflagellar transport (IFT) is a conserved process of cargo transport in cilia that is essential for development and homeostasis in organisms ranging from algae to vertebrates. In humans, variants in genes encoding subunits of the cargo-adapting IFT-A and IFT-B protein complexes are a common cause of genetic diseases known as ciliopathies. While recent progress has been made in determining the atomic structure of IFT-B, little is known of the structural biology of IFT-A. Here, we combined chemical cross-linking mass spectrometry and cryo-electron tomography with AlphaFold2-based prediction of both protein structures and interaction interfaces to model the overall architecture of the monomeric six-subunit IFT-A complex, as well as its polymeric assembly within cilia. We define monomer-monomer contacts and membrane-associated regions available for association with transported cargo, and we also use this model to provide insights into the pleiotropic nature of human ciliopathy-associated genetic variants in genes encoding IFT-A subunits. Our work demonstrates the power of integration of experimental and computational strategies both for multi-protein structure determination and for understanding the etiology of human genetic disease.

Introduction

Cilia are microtubule-based organelles that extend from many eukaryotic cells and play key roles in motility and signaling (1). Most motile cilia have a 9 + 2 microtubule axoneme arrangement (2) and are responsible for propelling the cells or adjacent fluid. In contrast, primary cilia lack the central pair of microtubules and serve a key role in cellular sensory functions (3,4). In humans, ciliary defects are linked to a number of developmental diseases, broadly known as ciliopathies (5,6).

Functional cilia require proper assembly of the ciliary axoneme and movement of cargos, both of which are implemented through a process known as intraflagellar transport (IFT). First defined in the flagellated, unicellular green algae *Chlamydomonas* (7), IFT comprises transport of proteins and protein complexes from the cell body to the tip of the cilia by kinesin motors (anterograde), followed by the movement of molecules back to the cell body by dynein motors (retrograde) (8,9). IFT has been shown to transport transmembrane proteins (10–13), tubulins (14,15), and chaperones (16), among other cargoes necessary for proper ciliary function, although some ciliary proteins also move by diffusion (17–19). Much of our understanding of IFT continues to emerge from work with simple organisms such as *Chlamydomonas,* trypanosomes, and the model single-celled ciliate *Tetrahymena*, but crucially, ciliary assembly and regulation by IFT are strongly conserved across evolution, from unicellular organisms to complex animals (20).

Accordingly, ciliopathies are a broad class of human disorders arising from dysfunction in cilia and affecting almost all organ systems (6). Due to the vital role of IFT in ciliogenesis and maintenance, variants in genes encoding IFT proteins are commonly linked to ciliopathies (21). For example, IFT is critical for primary cilia assembly in mice, where assembly defects are linked to renal disorders, such as polycystic kidney disease (22), and IFT-dependent defects in Hedgehog signal transduction produce neural tube and skeletal abnormalities (23,24). Likewise, in clinical studies, variants in human genes encoding IFT components are associated with several skeletal ciliopathies, including Jeune asphyxiating thoracic dystrophy, cranioectodermal dysplasia, and short-rib polydactyly (25–28). Curiously, variants within a single IFT-A gene can be associated with multiple distinct ciliopathies (29). The nature of such pleiotropy remains unclear.

The IFT complex consists of two distinct subcomplexes, the six-subunit IFT-A complex and the larger, 16-subunit IFT-B (9). While crystal structures have been determined for several IFT-B subunits (30–35), no such structures exist for any IFT-A proteins. Recently, cryo-electron tomography (cryo-ET) has provided a valuable *in situ* view of the assembly of IFT complexes into polymeric structures, known as IFT trains, at the ciliary base, highlighting their stepwise association (36), and capturing snapshots of IFT trains moving along the axonemes (37). These cryo-ET studies have not only revealed the low-resolution structures of the IFT-A and IFT-B complexes, but also revealed their cellular context, the mode of IFT-A/IFT-B associations, and the stoichiometries between the complexes. However, this technique has not provided the high-resolution structures necessary to understand the molecular assembly of the individual IFT-A proteins into the overall IFT super-complex.

Here, we present a detailed molecular model of an intact IFT-A complex obtained from a combination of chemical cross-linking mass spectrometry, AlphaFold2 structure and interaction interface prediction, and integrative computational modeling, and by using cryo-ET, we model its assembly into polymeric IFT trains. We validate the model by comparison to experimental evidence across multiple organisms, consistent with the deep structural conservation of the IFT-A complex (38). Comparisons to other proteins with related domain architectures show new modes of protein assembly that are unique to IFT-A, as well as preferred interaction modes conserved across protein complexes. Finally, this IFT-A model provides insights into human IFT-A ciliopathy-causing mutations based on their potential to disrupt the IFT-A complex itself or its interaction with cargo. Together, this work highlights the power of integrative modeling in structural biology and provides a mechanistic framework in which to better understand both basic ciliary biology and the complex genotype/phenotype relationships in IFT-associated disease.

Results

Determination of individual IFT-A protein structures by AlphaFold2 and chemical cross-linking mass spectrometry

We first independently modeled each of the six individual proteins that constitute the IFT-A complex: IFT43, IFT121, IFT122, IFT139, IFT140, and IFT144 (for simplicity, we use the human gene nomenclature for all genes) (39). Four of these proteins–IFT121, IFT122, IFT140, and IFT144–are predicted to share the same general architecture of two Tryptophan/Aspartic Acid Repeat (WD40) head domains and a tetratricopeptide repeat (TPR) tail domain, while IFT139 is composed of 19 TPR repeats, and the IFT43 domain structure is largely uncharacterized (39). These proteins are broadly conserved across eukaryotes (38). While the individual domain structures have been predicted from sequence, with recent advances in structural predictions (40,41), we could model the 3D structures of the full-length proteins with high confidence.

Using the AlphaFold2 Google Colab notebook (42) and protein sequences from the ciliate *Tetrahymena thermophila*, we predicted the structure of each full-length IFT-A protein (**Figure 1—figure supplement 1**), as well as those of *Chlamydomonas reinhardtii* and *Homo sapiens*. The computed models scored well by the predicted local distance difference test (pLDDT) (43), with most of the residues falling within the confident prediction threshold (pLDDT > 70).

To experimentally validate these structure predictions, we characterized endogenous IFT-A complexes from *Tetrahymena* with cross-linking mass spectrometry (XL/MS), in which cross-linked amino acid residues in a protein sample are connected by covalent crosslinks (XLs) of a defined length and can be identified in MS analyses (44–47). Such data define protein interaction interfaces at amino acid resolution and provide distance restraints for structural modeling. Briefly, we purified cilia from *Tetrahymena* (48) (**Figure 1A**), solubilized the membrane and matrix fraction (M+M) which contains IFT-A and IFT-B (49), and then further enriched for monomeric IFT-A complexes using size-exclusion chromatography (SEC) (**Figure 1—figure supplement 2**). We cross-linked the proteins in the IFT-A-containing fractions using a mass spectrometer-cleavable crosslinker, disuccinimidyl sulfoxide (DSSO), and mapped the crosslinked residues using MS^2/MS^3 tandem mass spectrometry. DSSO covalently cross-links pairs of accessible lysine residues falling within a distance determined by the length of the DSSO linker itself and the linked lysine side chains (50). This is generally less than 30 Å, though we include an additional 5 Å margin of error to account for protein dynamics as in (51–54).

We identified a total of 69 intramolecular cross-links between amino acids spanning the length of each individual IFT-A protein, apart from IFT43 (**Figure 1B, purple arcs**). We then tested the concordance between these XL/MS data and our AlphaFold2-predicted structures by calculating the distance between each linked residue pair in each predicted IFT-A protein structure. Our XL/MS data strongly validated the predicted structures, as 97% of all cross-linked residue pairs fell within the 35 Å length restraint in the models (**Figure 1C, pink in violin plots**). Impressively, even intramolecular cross-linked residues separated by more than 500 residues in their primary sequence fell within 35 Å in the 3D structure (**Figures 1B, 1C**).

The structures of the IFT-A complex and its subunits are expected to be highly conserved across eukaryotes (38), so we explored this relationship by superimposing models of IFT-A proteins from *C. reinhardtii* on those for *T. thermophila.* As expected, our *Tetrahymena*-derived chemical cross-links were still strongly in agreement with the *C. reinhardtii* structures, though as expected the concordance was slightly weaker (**Figure 1C, green**). The overall structure was also largely conserved between structure predictions for *Tetrahymena* and human IFT-A proteins (**Figure 1—figure supplement 3**). Accordingly, we observed 86% agreement between the human model and the *Tetrahymena* XL/MS data, measured as the percentage of XL/MS amino acid pairs that were predicted to be 35 Å or less apart when mapped onto the AlphaFold2 models of human IFT proteins. Taken together, the high level of agreement between predicted structures and experimental cross-linking data strongly suggests that the predicted structures accurately capture relevant IFT-A subunit conformations in these highly conserved eukaryotic proteins.

Integrative modeling of the IFT-A complex

We next sought to build a molecular model of the assembled monomeric IFT-A complex. To this end, we first mined our XL/MS data for intermolecular cross-links, providing distance restraints between pairs of amino acids located in two different IFT-A protein subunits (**Figure 2, green lines**). With these new data providing experimentally-supported intermolecular contacts, we modeled all possible pairs of IFT-A proteins with AlphaFold-Multimer (55) to test if the algorithm correctly identified the true interaction partners, judged by the crosslinks, and provided 3D models for the interacting domains (**Figure 2—figure supplement 1**). We found near-perfect concordance between AlphaFold-predicted interaction partners and the experimentally confirmed interaction partners, the sole exception being a high-confidence AlphaFold interaction between IFT122 and IFT139 with no supporting cross-links.

Using AlphaFold's predicted aligned error (PAE) confidence scores as a guide (as detailed in (56)), we constructed high-confidence models of interacting domains between IFT121-IFT122, IFT121-IFT139, IFT121-IFT43, IFT122-IFT139, IFT122-IFT140, IFT122-IFT144, and IFT140-IFT144 (**Figure 2—figure supplement 2A**). The combination of cross-linked residue pairs, AlphaFold2 monomer structures, and AlphaFold-Multimer domain-domain interaction models provided sufficient spatial restraints to build an initial model of IFT-A, which we improved by using a four-step integrative modeling approach (57–59) (**Figures 2, Figure 2—figure supplement 3, Figure 2—figure supplement 4, and Methods**).

We began by gathering data to provide spatial restraints for each of the IFT-A proteins in a manner consistent with prior structural information. We modeled most IFT-A proteins (IFT121, IFT122, IFT139, IFT140, and IFT144) as chains of rigid bodies (**Figure 2—figure supplement 2B**), introducing breaks between individual rigid bodies that corresponded to protein loop regions with lower pLDDT AlphaFold2 confidence scores. For IFT43, helices were modeled as rigid bodies, while the remainder of the protein was modeled using flexible beads, as it is thought to be disordered (39). We then further restrained the model by representing protein interaction interfaces derived from the PAE plots as rigid bodies. We used 98 DSSO chemical cross-links for the modeling, 29 of which were intermolecular, to further refine the model.

Next, we performed an optimized sampling through 20 independent modeling runs, with each run beginning from a unique initial configuration for the model. This exploration of the configuration space (10,000 frames) enabled the IFT-A subunits to find positions that best satisfied the structural restraints. In all, we sampled 200,000 possible configurations for the IFT-A complex. The models were initially clustered based on scores to identify those that best satisfied the input restraints. We selected the cluster of models that best agreed with the chemical cross-linking data. This cluster included models from multiple independent runs that converged on similar arrangements of the IFT-A complex, indicating that the model could be independently determined from different starting configurations (see **Figure 2—figure supplement 3** and **Methods**).

To arrive at a final model, we considered the ensemble of 9,121 models associated with the top-scoring cluster. The cluster had a weighted root-mean-square (RMSF) cluster precision of 15 Å (60), which denotes the average fluctuations of the individual residues (or beads) in 3D space across the ensemble. We visualized these fluctuations across the top-scoring models to define the probability density for each subunit; these probability densities and the centroid of the model ensemble provided our current best estimate of the 3D structure of monomeric IFT-A (**Figure 2, Figure 2—figure supplement 3**).

As an initial validation of our IFT-A model, we first examined the agreement with the chemical cross-links. Consistent with the length of the cross-linker and the two coupled lysine side chains, we again considered a cross-link satisfied if the lysine Cα residues were positioned within 35 Å of each other for at least one model in the ensemble. In all, 92% of the cross-links were well-satisfied by our top-scoring ensemble of models (**Figure 2—figure supplement 4**).

We also assessed the quality of the ensemble solution by splitting the models in the final cluster in half and computing the probability densities separately, testing for convergence. We observed a high cross-correlation coefficient between the two samples (.94) confirming a high degree of convergence between models in our final cluster (60). These values indicate that the selected cluster contains convergent modeling runs in which different starting positions yield highly similar end arrangements of IFT-A subunits. Thus, our structural model of the IFT-A monomer was both strongly internally consistent with the input data from XL/MS and structure predictions and the same solution was repeatedly derived by independent modeling runs.

Assembly of IFT-A monomers into a polymeric train

As the model was determined primarily using cross-links and structural constraints derived from monomeric IFT-A, we next sought to determine how the monomeric IFT-A might assemble into the polymeric form found in anterograde trains in cilia, and how it might orient with respect to IFT-B and the ciliary membrane. To address these questions, we determined a 23 Å resolution structure of the IFT-A complex by cryo-ET and subtomogram averaging, as observed *in situ* in the context of a flagellar anterograde IFT transport train within intact *Chlamydomonas* flagella (**Figure 3**). To arrive at this structure, we incorporated 7,900 additional particles into the subtomogram averaging from our previous studies (37,61) (**Table S1** and **Methods**), and this improved resolution served to constrain and inform our modeling.

We first performed a rigid body docking of our IFT-A monomeric complex into the subtomogram average using the ChimeraX fit-in-map tool (62). While IFT43, IFT121, IFT139, and the N-terminus of IFT122 fit well into the train, IFT140, IFT144, and the C-terminus of IFT122 required further fitting using molecular dynamics-based flexible fitting (63) (**Figure 4, Figure 4—figure supplement 1, Figure 4—animation 1**).

Accommodating the monomeric IFT-A into the anterograde train cryo-ET structure required a rotation of the C-terminus of IFT122 into the neighboring volume, where the extended TPR tail could be fit into a clearly delineated tube of density (**Figure 4—figure supplement 1**). By preserving the relative positions of IFT140/144 relative to the IFT122 C-terminal domain, their positions were also clearly evident in the cryo-ET density, in spite of a large movement of both IFT140 and IFT144 relative to IFT43/121/122 (N-terminus)/139 (**Figure 4—animation 1**). Notably, we observed IFT140 rearranged to bridge adjacent IFT-A complexes within the train, with the N-terminus of IFT140 interacting with the C-terminus of an adjacent IFT140 protein. This arrangement was especially interesting in light of AlphaFold's suggestion that IFT140 formed a loop with the N-terminus interacting with its own C-terminus, which may stem in part from evolutionary couplings that reflect the native polymeric state.

Validation of the modeled IFT-A structure

To further validate our structural model, we investigated whether our model was consistent with the extensive prior biochemical literature identifying direct interactions among the IFT-A proteins. Both co-sedimentation assays and visual immunoprecipitation (VIP) experiments in *Chlamydomonas* and mammalian cells (39,64) suggest that IFT122, IFT140, and IFT144 form a core complex within IFT-A (**Figure 5B.I and B.II**). Within this core complex, the C-terminal domain of IFT122 directly interacts with IFT140 and IFT144 (65), with a stable heterodimer being formed between IFT122 and IFT144, and specifically, residues 357-653 of IFT144 are required for the interaction with IFT122 (64,65). Furthermore, visible immunoprecipitation (VIP) assays show the N-terminal domain of IFT122 interacts with IFT121 and IFT43 (65). Our model of the IFT-A structure agrees perfectly with these previous findings (**Figure 5B.III-5B.VI**).

Moreover, several studies have cataloged the extensive interactions of IFT121. VIP experiments were used to observe the robust binding of both the IFT121 C-terminal fragment to all other subunits in the complex and the IFT121 fragment 545-800 to IFT122 (66). Moreover, the interaction between IFT121 and IFT43 was observed for *C. reinhardtii* proteins by yeast 2-hybrid assay and recombinant bacterial co-expression (39). In addition, bacterial co-expression assays demonstrated an interaction between IFT122 and IFT43 (39). Again, our IFT-A model agreed with all of these independent data (**Figure 5B.VII - 5B.X**).

Finally, several studies focus on direct interactions of IFT139 within IFT-A. One study in mammalian cells using the VIP assay suggests that IFT43, IFT121, and IFT122 interact with IFT139. The study implies that IFT122 may be required for the interaction of IFT139 with IFT43 and IFT121 (64). Interestingly, while our data showed no direct cross-links between IFT122 and IFT139, our model suggests that the interaction between the peripheral subunits (IFT43, IFT121, and IFT139) is facilitated by IFT122 (**Figure 5B.XI**). It has also been suggested from IFT139 loss-of-function experiments that IFT139 is the most distal subunit (64). Consistent with these data, our model places IFT139 most distal to the core (**Figure 5B.XII**).

Discussion

As structural biology moves towards tackling more complicated problems *in situ*, AI-predicted structures and chemical cross-links provide a complement to cryo-electron microscopy and tomography studies to illuminate the architecture of challenging, transient, or less abundant protein complexes. Here, we have combined cryo-electron subtomogram averaging of intact cilia, chemical cross-links of highly enriched soluble endogenous IFT-A complexes, and AlphaFold2 predicted structures of individual proteins and protein pairs to build a comprehensive 3D model of the IFT-A ciliary trafficking protein complex. The IFT-A structure is strongly supported by previous biochemical interaction studies and reveals new and conserved packing modes among proteins sharing these domain architectures. Moreover, the model provides testable new hypotheses and sheds new light on the precise mechanisms underlying IFT-related human genetic diseases.

Conserved interactions between structurally similar proteins

We constructed our model using spatial information from organisms that have diverged since the last eukaryotic common ancestor (LECA). Despite this, IFT has proven to be highly conserved (38). In one example, there is a 41% sequence identity shared between the IFT172 protein from human and *T. brucei*, an early branching supergroup (Excavata) of eukaryotes (38).

This structural conservation between species is also evident from IFT-B subtomogram averages collected from mammalian primary cilia and *C. reinhardtii* motile cilia (67), which display similar overall morphologies to the IFT-B monomers.

Furthermore, our new model allows us to compare the protein-protein interactions in IFT-A to those with similar domain architectures to better understand the functions of individual subunits within the complex. Phylogenetic evidence suggests that the IFT complex is a sister structure to COPI and a member of the proto-coatomer family (38). The αβ subunits of the COPI complex interact via their TPR tail domain (38,68), and we observed a similar binding mode in IFT-A, where the subunits IFT121, IFT122, IFT140, and IFT144 interact *via* their TPR tail domains (**Figure 6**). Phylogenetic analyses also suggest that IFT-A and the BBSome likely evolved from a single IFT complex through subunit duplication (38). Interestingly, the BBSome contains β-propeller structures similar to the WD40 domains of IFT-A, in addition to TPR-based subunits similar to IFT139. In the BBSome, the TPR protein, BBS4, interacts with the β-propeller of BBS1 (69). Similarly, in our model of IFT-A, there is an analogous interaction between the WD40 domain of IFT122 and TPR-based IFT139.

The dominant IFT-A subunit domain structure is two WD40 heads and a TPR tail. Because this domain architecture is so prevalent in the IFT-A complex, we were curious about its representation in the human proteome. We investigated the interactions formed between proteins with the same domain architecture. Sequence-based alignment has traditionally been highly successful in identifying proteins with homologous domains; however, AlphaFold2 has recently enabled new approaches to identify more distantly related homologs that nonetheless retain structural similarity. AlphaFold2 has been used to create databases of proteome-wide structure predictions (70), enabling the identification of proteins with similar architectures that are dissimilar in sequence. With this resource now available, an entire proteome can serve as the target of structural similarity searches for discovery of distant homologs, such as by using the program Dali (71,72). We used this approach to compare the four IFT-A proteins sharing the canonical WD40/TPR domain architecture to the set of human proteins with AlphaFold2-predicted structures and found the set of proteins with related structures (**Figure 6—figure supplement 1**). All contained a TPR tail and at least one WD40 head domain.

Analysis of these structurally similar proteins reveals they are generally members of protein complexes involved in transport and trafficking. For example, the elongator complex contains two copies of the ELP1 protein, which interact via their TPR tail domains (73,74) (**Figure 6**). Similar to IFT-A, the elongator complex has a role in trafficking (75). We identified two of the subunits of the BLOC-2 complex–HPS3 and HPS5–in our Dali search (**Figure 6—figure supplement 1**), as being structurally similar to IFT-A proteins. The BLOC-2 complex is essential for trafficking and is specifically required for the transportation of tyrosinase and Tyrp1 from early endosomes to maturing melanosomes (76). Intriguingly, one-third of the mutations leading to Hermansky–Pudlak syndrome occur in BLOC-2 subunits (77) suggesting that a better understanding of the complex architecture could impact the treatment of this rare disease. These complexes all have a shared role in cellular transport and trafficking with IFT-A (78–80).

In addition to overall structural similarity, the interactions among TPR tail domains of these proteins are similar to those of IFT-A (**Figure 6**). This is not surprising given the reported role of TPR domains in stabilizing intracomplex interactions (81). The recurring interaction between TPR domains raises the question of the role of the associated WD40 domains in these protein complexes. In clathrin, another complex identified in our search, the WD40 domains are known to selectively bind unique cargo peptides in vesicular transport (82). A similar function to this was previously suggested for the WD40 domains in IFT-A and the BBSome (83). We hypothesize that the WD40 domains in these TPR-containing proteins may be specialized for participating in transient interactions such as the binding of cargo proteins and that this role may prove useful for explaining the position of the domains towards the exterior of the complex as well as the range of phenotypes associated with IFT-based diseases.

Human disease mutations of the IFT-A complex

Variants in the human genes encoding IFT-A subunits are linked to diverse ciliopathies (84), and we examined our IFT-A structure to shed new light on the molecular basis of IFT-A associated diseases. Both the evolutionary history of the IFT-A protein domains (81) and our new model suggest that the stable formation of the IFT-A complex is mediated by interactions of the TPR domains. While we observe some disease-associated variants occurring within the TPR domains, the large majority of known disease-associated variants (79%) (85,86) are located within the WD40 regions (**Figure 7**). This suggests that the variants located on the WD40 domains act by disrupting the association of important cargo proteins rather than by disrupting overall IFT-A complex formation. This idea is consistent with the proposal that the WD-repeat domains provide large surface areas for potential binding, allowing these domains to simultaneously bind several distinct cargo proteins (87). To explore these ideas, we integrated the known genetics and cell biology of a subset of IFT-A related ciliopathy variants with our new structure of the IFT-A complex (**Figure 8**).

Expression of IFT140 lacking the N-terminal WD repeats is sufficient to disrupt ciliogenesis to some extent but fails to rescue the ciliary localization of GTPases, lipid-anchored proteins, and cell signaling proteins (88). This result suggests that the WD40 domains play an important role in cargo transport (88) and is consistent with the position of the IFT140 TPR domain in the core of the IFT-A complex and the position of the WD40 domains on the surface of the complex (**Figure 8A**). With this in mind, it is interesting that two disease-associated missense variants in the IFT140 WD40 domains (Y311C and E664K)(89,90) (**Figure 8B**) were shown to localize normally to axonemes (90), indicating their normal interaction with other IFT-A proteins. By contrast, the S939P variant of IFT40 lies not just in the TPR domain but also near the interaction interface with both IFT122 and IFT144 (**Figure 8D**) and this variant displays an aberrant cellular localization (91), suggesting failure to assemble into a normal IFT-A complex.

Likewise, loss of IFT122 results in ciliogenesis defects (65), but cilia can be rescued by the ciliopathy-associated variants such as the W7C and G513V alleles (65), which do not map to the TPR domains but instead to the WD40 repeats (**Figure 8C**). Moreover, cilia rescued with these disease-associated alleles still exhibit abnormal localization of the ciliary proteins INPP5E and GPR161 (65), raising the possibility that they specifically disrupt the association of IFT122 with membrane protein cargoes. This is supported by the location of the WD40 domains in our IFT-A structure as sitting under the membrane (**Figure 8A**). These and other ciliopathy-associated missense variants in the WD40 domains of IFT122, including S373F and V553G, also disrupt the association of the IFT-A core with the IFT-A peripheral proteins, IFT121 and IFT43 (65), and the positions of all of these alleles near those interfaces in our model are consistent with this result (**Figure 8C**).

The L795P variant in the peripheral component IFT139 is also interesting, as this allele lies in the TPR domain near the protein's interface with IFT121 (**Figure 8E**). The localization of this variant seems normal (29), yet functional assays in both zebrafish and *C. elegans* demonstrate that the allele is pathogenic suggesting a biochemical perturbation (29,92). Our model suggests that this allele's pathogenicity may be exerted by disrupting interaction with IFT121(**Figure 8E**). Thus, our molecular model of the IFT-A complex provides insights into the molecular basis of IFT-A related human ciliopathies.

Comparison with two independently-determined IFT-A structures

Concurrently with our initial preprint (93), two additional structures were reported for IFT-A particles from other species and determined using different techniques, one for the reconstituted human IFT-A particle determined by single particle cryo-EM (94) and one of the *Chlamydomonas* IFT-A/IFT-B train reconstructed from an 18 Å cryo-ET subtomogram average (95). Because our structural constraints–the chemical cross-links and the co-evolutionary couplings incorporated into AlphaFold2–likely represent a combination of IFT-A in the anterograde and retrograde forms, we were interested to compare our model to these contemporaneous structures. While a full comparison will have to await release of atomic coordinates, an initial inspection (**Figure 4—figure supplement 2**) indicates that all three structures are highly concordant in the reasonably compact region of IFT-A defined by the WD40 domains of IFT121/122 and their interactions with IFT139. While IFT43 is absent from (95) and generally highly disordered, our model agrees well with that of Hesketh *et al.* for the positions of its more ordered segments. Finally, all three IFT-A polymer models appear to agree substantially with regard to the general orientation and placement of the IFT122 extended TPR domain, and to the IFT140-mediated monomer-monomer interactions, placing the N-terminus of IFT140 from one monomer to interact with the C-terminus of an adjacent IFT140 protein. This concordance is especially notable in light of the different experimental techniques employed by each group and provides some degree of confidence that the structures faithfully capture representative IFT-A conformations. The fact that the combination of cross-linking mass-spectrometry, AlphaFold-Multimer, and integrative modeling produced a highly concordant structure with the significantly higher resolution single particle cryo-EM reconstruction (94) also suggests that such a readily accessible combined modeling approach might be broadly suitable for many other protein complexes.

Limitations arising from integrative modeling and alternative conformations

A consideration of IFT-A's biological role in ciliary cargo transport is essential in interpreting our model. It has been previously observed that when the IFT trains (including IFT-A, IFT-B, and the BBsome) reach the ciliary tip, a structural rearrangement occurs where kinesin dissociates from the complex and the dynein motors power the retrograde movement (96). The extent of the rearrangement for the IFT-A complex is not known. However, because our data were derived from IFT complexes solubilized from intact cilia, they should inform models of IFT in both directions. The subtomogram average that we have used in our polymeric model represents the anterograde arrangement and while the retrograde arrangement is not known, it was shown to differ in morphology from the anterograde arrangement (97). It is important to remember that our results represent an ensemble of models that best satisfy the input data, and that some cross-link violations remain for our model (**Figure 2—figure supplement 4**). These remaining violations may point to a possible structural rearrangement in IFT-A proteins during the transition from anterograde to retrograde IFT.

Similarly, the assembly of the IFT-A complex into oligomeric trains clearly suggests structural rearrangements relative to a monomeric IFT-A complex (**Figure 4, Figure 4—figure supplement 1**), consistent with observations for the human IFT-A monomer (94). Based on known size calibrants for our SEC column chromatography, our chemical cross-links were derived from the monomeric IFT-A assembly (**Figure 1—figure supplement 2**). However, as the samples were subsequently further concentrated for crosslinking, we speculate that concentrating the particles might have induced some degree of oligomerization and interactions with IFT-B, which may explain the small number of cross-links consistent with IFT-A/A and IFT-A/B interactions. It is unclear, however, if these interactions better reflect particular anterograde or retrograde conformations, and it is possible that additional conformational changes may accompany train assembly or cargo binding that are not captured by these data.

For example, in our monomeric integrative model of the IFT-A complex we observe a large cross-link violation between the WD40 domain of IFT121 (K16) and IFT144 (K831). Interestingly, while AlphaFold2 places the WD40 domains of IFT121 and IFT122 as interacting and is supported by low PAE values between the residues within this region, we observe no cross-links between the proteins in this region. We do, however, observe cross-links between the WD40 domains of IFT121 and IFT122 and other proteins. These cross-links raise the possibility of domain rearrangements of the WD40 domains of IFT121 and IFT122 upon interactions with other proteins, or even transient rearrangements reflecting a dynamic and flexible complex. As for any structural biology study, while certain regions of the model are highly confident, we anticipate that additional local structural rearrangements are likely to occur with changes to assembly state and direction of travel.

Methods

Tetrahymena culture

Tetrahymena thermophila SB715 were obtained from the *Tetrahymena* Stock Center (Cornell University, Ithaca, NY) and grown in Modified Neff medium obtained from the stock center. Cells were routinely maintained at room temperature (~21° C) in 10 ml cultures and were expanded to 3 liters at 30°C with shaking (100 rpm) for preparation of cilia.

Tetrahymena membrane and matrix preparation

Cilia extracts were made as outlined in Gaertig *et al.* (98). Briefly, cilia were released by either pH shock or dibucaine treatment and recovered by centrifugation. For cross-linking experiments, cilia were extracted with 1% NP40 in HEPES-Cilia Wash Buffer (H-CWB, where 50 mM HEPES pH 7.4 was used in place of the 50 mM Tris to ensure compatibility with DSSO, and 0.1 mM PMSF was added.) Axonemes were removed by centrifugation at 10,000 x g, 10 minutes 4°C. Protein concentration of the soluble membrane and matrix fraction (M+M) was determined by DC BioRad Assay.

Tetrahymena IFT-A sample preparation

Two types of IFT-A-containing samples were generated: (1) IFT-A enriched fractions from preparative-scale SEC separations, or (2) IFT-A-containing ion exchange chromatography (IEX) fractions. Preparative scale SEC fractionation began with 2.8 mg M+M extract in 2 ml H-CWB. Separation was on a HiLoad 16/600 Superdex 200 PG (preparative grade) column (Cytiva) at a flow rate of 1 ml/min, mobile phase 50 mM HEPES, pH 7.4, 50 mM NaCl, 3 mM $MgSO_4$, 0.1 mM EGTA. 1.5 ml fractions were collected and analyzed by mass spectrometry to confirm the IFT-A elution peak (fractions 16-18, as initially identified from analyses of *Tetrahymena* whole cell extract, corresponding to monomeric IFT-A; see **Figure 1—figure supplement 2**). Molecular mass of eluted IFT-A was estimated using a commercial mixture of molecular weight markers (Sigma-Aldrich #MWGF1000) run under the same conditions (Blue Dextran (approximate molecular mass ~2,000 kDa), bovine thyroglobulin, horse spleen apoferritin, bovine serum albumin, and yeast alcohol dehydrogenase) in combination with 8 well-characterized protein complexes of known or inferred molecular mass that were observed in the separations of *Tetrahymena* whole cell extract run under the same conditions. For sample preparation, IFT-A containing fractions from 2 identical sequential separations were pooled; IFT-A and B were estimated by mass spectrometry to each comprise approximately 3% of the sample. Membrane contamination was removed by centrifugation 100,000 x g 1.25 hours 4°C in an NVT65.2 rotor (Beckman Coulter). The clarified supernatant was concentrated to 50 µl by ultrafiltration (Sartorius Vivaspin Turbo, 100,000 MWCO) according to manufacturer's instructions. The second set of samples was generated from 1.5 mg of cilia M+M subjected to ultrafiltration (Amicon Ultra Ultracel 10k NMWL, #UFC501096) to adjust salt and concentrate protein for fractionation using a mixed-bed ion exchange column (PolyLC Inc., #204CTWX0510) with a Dionex UltiMate3000 HPLC system. The chromatographic method was performed as in McWhite *et al.* (99), but with 10 mM HEPES pH 7.4 replacing Tris in both Buffers A and B.

Chemical cross-linking / mass spectrometry

Crosslinking was performed on both samples described above : (1) IFT-A enriched SEC fractions and (2) IFT-A-containing IEX fractions. The first sample (representing approx. 40 µg of protein) was cross-linked by addition of DSSO (freshly made 50 mM stock in anhydrous DMSO) to 5 mM final concentration. After 1 hour at room temperature crosslinking was quenched by addition of 1 M Tris pH 8.0 to 20 mM for 25 minutes at room temperature. Peptides were reduced, alkylated, digested with trypsin, and desalted using C18 spin tips (Thermo Scientific HyperSep SpinTip P-20 BioBasic # 60109-412) as in Havugimana *et al.* (100) with the exception that reduction was accomplished with 5.0 mM TCEP (Thermo Scientific #77720) instead of DTT. To enrich for cross-linked peptides, the desalted peptides were dried and resuspended in 25 µl 30% Acetonitrile, 0.1% TFA, and separated on a GE Superdex 30 Increase 3.2/300 size exclusion column (Cytiva) at 50 µl/minute flow rate using an ÄKTA Pure 25 FPLC chromatography system (Cytiva). 100 µl fractions were collected, dried, and resuspended in 5% acetonitrile, 0.1% formic acid for mass spectrometry.

The second set of IFT-A samples was crosslinked using 50 mM DSSO stock prepared with anhydrous DMF immediately before use, diluted with 10 mM HEPES pH 7.4, and added to each 0.5 ml column fraction to a final concentration of 0.5 mM. Cross-linking proceeded for 1 hour at room temperature and was quenched by addition of 1 M Tris pH 8.0 to 28 mM. Samples were prepared for mass spectrometry using Method 1 from (99) with no enrichment for crosslinked peptides.

Mass spectra were collected on a Thermo Orbitrap Fusion Lumos tribrid mass spectrometer. Peptides were separated using reverse phase chromatography on a Dionex Ultimate 3000 RSLCnano UHPLC system (Thermo Scientific) with a C18 trap to Acclaim C18 PepMap RSLC column (Dionex; Thermo Scientific) configuration. Data were collected from an aliquot of the cross-linked peptides prior to SEC enrichment using a standard top speed HCD MS1-MS2 method (101) and analyzed using the Proteome Discoverer basic workflow, and the proteins identified served as the reference proteome for subsequent cross-link identification from the cross-link-enriched fractions.

To identify DSSO cross-links, spectra were collected as follows: peptides were resolved using a 115 min 3-42% acetonitrile gradient in 0.1% formic acid. The top speed method collected full precursor ion scans (MS1) in the Orbitrap at 120,000 m/z resolution for peptides of charge 4-8 and with dynamic exclusion of 60 sec after selecting once, and a cycle time of 5 sec. CID dissociation (25% energy 10 msec) of the cross-linker was followed by MS2 scans collected in the orbitrap at 30,000 m/z resolution for charge states 2-6 using an isolation window of 1.6. Peptide pairs with a targeted mass difference of 31.9721 were selected for HCD (30% energy) and collection of rapid scan rate centroid MS3 spectra in the ion trap. Data were analyzed using the XlinkX node of Proteome Discoverer 2.3 and the XlinkX_Cleavable processing and consensus workflows (47) and results exported to xiView (102) for visualization.

Chlamydomonas cell culture

Chlamydomonas wild-type cells (CC-124 mt- and CC-125 mt+) were obtained from the *Chlamydomonas* resource center (https://www.chlamycollection.org); ift46-1::IFT46-YFP was a gift of K. Huang and G. Witman (103). Cells were cultured in TAP (Tris-Acetate-Phosphate) medium as described by the resource center. For long-term storage, cells were grown on TAP plates with 1.5% agar at room-temperature. For sample preparation, fresh liquid cultures of 300 mL were grown for three to four days at 22°C under a light-dark cycle with constant aeration.

Preparation of cryo-TEM grids by plunge freezing

TEM gold grids with Holey Carbon support film (Quantifoil Micro Tools GmbH, Au 200 mesh R3.5/1) were glow-discharged in a plasma cleaner (Diener electronic GmbH, Femto). For plunge freezing, a Leica EM Grid Plunger (GP) was used at 18°C and a humidity of ~80%. 3 μL undiluted *Chlamydomonas* cells were applied to the grid and mixed with 1 μL 10 nm colloidal gold particles (BBI solutions). Blotting was performed from the back for 1 sec and the grid was plunged into liquid ethane at -182°C and stored in liquid nitrogen until data acquisition.

Cryo-ET Data Acquisition

For data acquisition, a Thermo Fisher (former FEI) Titan Halo cryo-TEM was used, equipped with a field emission gun (FEG), operating at 300 kV. Images were recorded on a K2 summit direct electron detector (Gatan) with an energy filter (GIF, Gatan image filter). Digital Micrograph software (Gatan) was used to tune the GIF, and SerialEM software (104) was employed for the automated acquisition of tomographic tilt series in low-dose mode. Tilt series were acquired at a magnification of 30,000X with a pixel size of 0.236 nm in super-resolution mode of the K2 camera. The tilting scheme was bidirectional with a starting angle of -20° and maximal tilt angles of ±64° when possible. Images were acquired every 2 degrees. The defocus range was between -3 and -6 μm, the slit width of the energy filter was 20 eV. Each tomogram had a cumulative dose between 100 and 140 $e^{-}/Å^{2}$ with an image dose of 1.8 to 2.1 $e^{-}/Å^{2}$. Exposure times were between 1.6 and 2.5 sec per image, while each 10 frames were acquired. The sample drift was held well below 1 nm/sec, and 34 grids were imaged in all. Data collection and refinement statistics are summarized in **Table S1**.

Tomographic reconstruction

Frame alignment was done with K2Align (provided by the Baumeister group, MPI for Biochemistry, Munich), which is based on the MotionCorr algorithm (105). Tomograms were reconstructed with Etomo from IMOD (106), using fiducial markers for alignment. CTF curves were estimated with CTFPLOTTER and the data were corrected by phase-flipping with CTFPHASEFLIP, both implemented in IMOD (107). Dose weight filtration was applied (108) and tomograms were reconstructed by weighted back-projection and subsequently binned by 3 and 6, resulting in pixel sizes of 0.708 nm and 1.413 nm.

Subtomogram averaging

IFT-A particles were picked and averaged as described in detail previously (61). The data from wild-type cells (CC-124 and CC-125) was complemented with data from IFT46-YFP cells, which showed no differences in their IFT-A structure compared to wild-type. In short, subtomogram averaging was performed with PEET version 1.11.0 from the IMOD package (109). Particles were picked with 11 nm spacing and pre-aligned to a reference generated from particles of one train on bin6 tomograms. The alignment was then refined on bin3 tomograms. To reduce the influence of IFT-B or the membrane, loose binary masks were applied to the reference. The final average was calculated from 9,350 particles derived from 96 tomograms. The resolution was estimated to 2.3 nm by Fourier shell correlation, using a cut-off criterion of 0.5. A reconstruction of the whole IFT-A polymer was generated in UCSF Chimera by placing several copies of one IFT-A unit along the train polymer with 'Fit in map' and merged with 'vop maximum'.

Modeling of IFT-A protein subunits and pairwise interactions

The IFT-A complex is composed of IFT43, IFT121, IFT122, IFT139, IFT140, and IFT144. Each subunit was modeled independently using the CollabFold notebook (42) with AlphaFold2 (41). Modeling results and statistics show that confident structures were generated for 5 of the 6 protein subunits (**Figure 1—figure supplement 1**). Pairwise interacting protein structures were predicted using the 2.1.2 version/release of AlphaFold-Multimer (55) as implemented on Texas Advanced Computing Center (TACC) Maverick2 and Frontera (110) GPU computer clusters. Predicted aligned error (PAE) plots were used as in (56) to determine interaction interfaces to be represented as rigid bodies in integrative modeling. We use the calibration curve from our previous studies to select pairwise interactions with PAE of less than 3.5 Å.

Domain representation and spatial restraint configuration

We used the Python modeling interface of the Integrative Modeling Platform (57) to model the IFT-A complex, performing the modeling in four stages: 1) gathering data, 2) domain and spatial restraint representation, 3) system restraint and restraint scoring, and 4) model validation (58,59,111). Because there are no available crystal structures for any of the IFT-A subunits, we used AlphaFold2 structural models to construct representations of each of the subunits. IFT139 has a long alpha solenoid domain structure, which was independently supported by 18 cross-links, and it was represented as two rigid bodies. In contrast, IFT43, with its more poorly defined structure, was represented as two alpha helices connected by a flexible string of 1 Å beads. IFT121, IFT122, IFT140, and IFT144 share a similar domain architecture of two WDR domains and an alpha solenoid tail domain. All four of these proteins were represented as chains of rigid bodies that corresponded to regions of high AlphaFold2 confidence scores (high pLDDT scores). For pairs of IFT-A proteins, we used PAE estimated errors (**Figure 2—figure supplement 1**) to determine boundaries of rigid body interactions between pairs of IFT-A proteins predicted by AlphaFold-Multimer to interact. The initial model was built from combining the pairwise AlphaFold-Multimer structure predictions into a consensus model, thus preserving the arrangements of predicted interfaces (**Figure 2—figure supplement 2 & Figure 2—figure supplement 4**).

In total, we considered 98 inter- and intramolecular cross-links, representing the combined cross-linking evidence from two alternative IFT-A enrichment procedures. Cross-links were modeled in IMP using a length of 21 Å as in (51). An excluded volume restraint was incorporated to 10 residue beads ensuring that two volumes do not occupy the same space, and a connectivity restraint was applied between beads to ensure that consecutive protein segments remained nearby in 3D space. The full set of restraints was used in creating the scoring framework for the model, and all relevant data and scripts are available on the Zenodo repository.

System sampling, scoring of restraints, and initial model validation

The 10 rigid bodies were randomized into their initial configurations. A steep gradient descent minimization based on connectivity was used to ensure that neighboring residues were close to each other before Monte Carlo sampling. We then performed 20 independent runs of Monte Carlo sampling, each run starting from a unique initialization configuration and sampling 10,000 frames, thus sampling 200,000 total configurations. Ensembles of models were then clustered first based on cross-linking agreement, sequence connectivity, enclosed volume, and total score. We selected the cluster with models from multiple runs showing high agreement with the cross-linking data and high overall scores for subsequent analyses. The top cluster was assessed against the input data and tested for convergence using the sampling exhaustiveness protocol (60).

Docking and polymer modeling

Our integrative IFT-A monomer model was rigidly docked into the IFT-A train using the ChimeraX fit in map tool (112). Rigidly docking the model placed 54% of the atoms within the map. IFT43, IFT121, IFT139, and the C-terminus of IFT122 subcomplex fit better into the subtomogram average with 67% of the atoms within the map. The remainder of the IFT-A integrative model was fit into the map by successive rounds of breaking the structures and refining their fit using Namdinator (63). The final flexibly fit structure places 71% of all atoms within the map.

Data deposition

Mass spectrometry proteomics data was deposited in the MassIVE/ProteomeXchange database (113) under accession number PXD032818. Cryo-tomography data was deposited in the Electron Microscopy Data Bank (114) under accession number EMD-26791. IFT-A models were deposited in the PDB-Dev database (115) as well as on Zenodo at doi: 10.5281/zenodo.7222413, along with additional supporting materials, including integrative modeling data and code.

Author contributions

Conceptualization: C.L.M., E.M.M., G.P. Methodology: C.L.M., O.P., M.A.J. Formal analysis: C.L.M., O.P., M.A.J., E.M.M. Investigation: C.L.M., O.P., M.A.J., G.H, C.N. Resources: E.M.M., D.W.T., G.P. Writing–original draft: C.L.M., O.P., J.B.W., E.M.M., Writing–review & editing: C.L.M., O.P., M.A.J, J.B.W., E.M.M., D.W.T., G.P.

Acknowledgements

The authors gratefully acknowledge suggestions from Anthony Roberts (Birkbeck/UCL) and the two anonymous referees that substantially improved the paper, as well as Jaime Hibbard for helpful discussion, Agnes Toth-Petroczy (MPI-CBG, Dresden) for early discussions about homology modeling, and the generous support of the *Tetrahymena* stock center (Cornell University) and *Chlamydomonas* resource center that made this project possible. Research was funded by grants from the National Institute of General Medical Sciences R35GM122480 (to E.M.M.) and R35GM138348 (to D.W.T.), National Science Foundation (2019238253 to C.L.M.), National Institute of Child Health and Human Development (HD085901 to J.B.W. and E.M.M.), Army Research Office (W911NF-12-1-0390 to E.M.M.), Welch Foundation (F-1515 to E.M.M., F-1938 to D.W.T.), and Max Planck Society (M.A.J. and G.P.). D.W.T. is a CPRIT Scholar supported by Cancer Prevention and Research Institute of Texas (RR160088). The authors would like to thank the Electron Microscopy Facility of the MPI-CBG. The authors acknowledge the Texas Advanced Computing Center at The University of Texas at Austin for providing high-performance computing resources that have contributed to the research results reported within this paper.

References

1. Gigante ED, Caspary T. Signaling in the primary cilium through the lens of the Hedgehog pathway. WIREs Dev Biol. 2020;9(6):e377.
2. Luck DJ. Genetic and biochemical dissection of the eucaryotic flagellum. J Cell Biol. 1984 Mar;98(3):789–94.
3. Bloodgood RA. From central to rudimentary to primary: the history of an underappreciated organelle whose time has come. The primary cilium. Methods Cell Biol. 2009;94:3–52.
4. Garcia-Gonzalo FR, Reiter JF. Scoring a backstage pass: mechanisms of ciliogenesis and ciliary access. J Cell Biol. 2012 Jun 11;197(6):697–709.
5. Legendre M, Zaragosi LE, Mitchison HM. Motile cilia and airway disease. Semin Cell Dev Biol. 2021 Feb;110:19–33.
6. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev Mol Cell Biol. 2017 Sep;18(9):533–47.
7. Kozminski KG, Johnson KA, Forscher P, Rosenbaum JL. A motility in the eukaryotic flagellum unrelated to flagellar beating. Proc Natl Acad Sci U S A. 1993 Jun 15;90(12):5519–23.
8. Lechtreck KF, Van De Weghe JC, Harris JA, Liu P. Protein transport in growing and steady-state cilia. Traffic Cph Den. 2017 May;18(5):277–86.
9. Taschner M, Lorentzen E. The Intraflagellar Transport Machinery. Cold Spring Harb Perspect Biol. 2016 Oct 3;8(10):a028092.
10. Huang K, Diener DR, Mitchell A, Pazour GJ, Witman GB, Rosenbaum JL. Function and dynamics of PKD2 in Chlamydomonas reinhardtii flagella. J Cell Biol. 2007 Nov 5;179(3):501–14.
11. Qin H, Burnette DT, Bae YK, Forscher P, Barr MM, Rosenbaum JL. Intraflagellar transport is required for the vectorial movement of TRPV channels in the ciliary membrane. Curr Biol CB. 2005 Sep 20;15(18):1695–9.
12. Kobayashi T, Ishida Y, Hirano T, Katoh Y, Nakayama K. Cooperation of the IFT-A complex with the IFT-B complex is required for ciliary retrograde protein trafficking and GPCR import. Kaibuchi K, editor. Mol Biol Cell. 2021 Jan 1;32(1):45–56.
13. Wingfield JL, Lechtreck KF, Lorentzen E. Trafficking of ciliary membrane proteins by the intraflagellar transport/BBSome machinery. Essays Biochem. 2018 Dec 7;62(6):753–63.
14. Craft JM, Harris JA, Hyman S, Kner P, Lechtreck KF. Tubulin transport by IFT is upregulated during ciliary growth by a cilium-autonomous mechanism. J Cell Biol. 2015 Jan 19;208(2):223–37.
15. Hao L, Thein M, Brust-Mascher I, Civelekoglu-Scholey G, Lu Y, Acar S, et al. Intraflagellar transport delivers tubulin isotypes to sensory cilium middle and distal segments. Nat Cell Biol. 2011 Jun 5;13(7):790–8.
16. Bhowmick R, Li M, Sun J, Baker SA, Insinna C, Besharse JC. Photoreceptor IFT complexes containing chaperones, guanylyl cyclase 1 and rhodopsin. Traffic Cph Den. 2009 Jun;10(6):648–63.
17. Ye F, Breslow DK, Koslover EF, Spakowitz AJ, Nelson WJ, Nachury MV. Single molecule imaging reveals a major role for diffusion in the exploration of ciliary space by signaling receptors. eLife. 2013 Aug 6;2:e00654.
18. Belzile O, Hernandez-Lara CI, Wang Q, Snell WJ. Regulated membrane protein entry into flagella is facilitated by cytoplasmic microtubules and does not require IFT. Curr Biol CB. 2013 Aug 5;23(15):1460–5.
19. Harris JA, Liu Y, Yang P, Kner P, Lechtreck KF. Single-particle imaging reveals intraflagellar transport–independent transport and accumulation of EB1 in *Chlamydomonas* flagella. Holzbaur E, editor. Mol Biol Cell. 2016 Jan 15;27(2):295–307.
20. Ishikawa H, Marshall WF. Intraflagellar Transport and Ciliary Dynamics. Cold Spring Harb

Perspect Biol. 2017 Mar 1;9(3):a021998.

21. Pigino G. Intraflagellar transport. Curr Biol. 2021 May;31(10):R530–6.
22. Pazour GJ, Dickert BL, Vucica Y, Seeley ES, Rosenbaum JL, Witman GB, et al. Chlamydomonas IFT88 and its mouse homologue, polycystic kidney disease gene tg737, are required for assembly of cilia and flagella. J Cell Biol. 2000 Oct 30;151(3):709–18.
23. Huangfu D, Liu A, Rakeman AS, Murcia NS, Niswander L, Anderson KV. Hedgehog signalling in the mouse requires intraflagellar transport proteins. Nature. 2003 Nov 6;426(6962):83–7.
24. Haycraft CJ, Zhang Q, Song B, Jackson WS, Detloff PJ, Serra R, et al. Intraflagellar transport is essential for endochondral bone formation. Dev Camb Engl. 2007 Jan;134(2):307–16.
25. Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, et al. Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. Am J Hum Genet. 2010 Sep 10;87(3):418–23.
26. Ashe A, Butterfield NC, Town L, Courtney AD, Cooper AN, Ferguson C, et al. Mutations in mouse Ift144 model the craniofacial, limb and rib defects in skeletal ciliopathies. Hum Mol Genet. 2012 Apr 15;21(8):1808–23.
27. Schmidts M, Vodopiutz J, Christou-Savina S, Cortés CR, McInerney-Leo AM, Emes RD, et al. Mutations in the gene encoding IFT dynein complex component WDR34 cause Jeune asphyxiating thoracic dystrophy. Am J Hum Genet. 2013 Nov 7;93(5):932–44.
28. Taylor SP, Dantas TJ, Duran I, Wu S, Lachman RS, University of Washington Center for Mendelian Genomics Consortium, et al. Mutations in DYNC2LI1 disrupt cilia function and cause short rib polydactyly syndrome. Nat Commun. 2015 Jun 16;6:7092.
29. Davis EE, Zhang Q, Liu Q, Diplas BH, Davey LM, Hartley J, et al. TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. Nat Genet. 2011 Mar;43(3):189–96.
30. Taschner M, Weber K, Mourão A, Vetter M, Awasthi M, Stiegler M, et al. Intraflagellar transport proteins 172, 80, 57, 54, 38, and 20 form a stable tubulin-binding IFT-B2 complex. EMBO J. 2016 Apr 1;35(7):773–90.
31. Taschner M, Kotsis F, Braeuer P, Kuehn EW, Lorentzen E. Crystal structures of IFT70/52 and IFT52/46 provide insight into intraflagellar transport B core complex assembly. J Cell Biol. 2014 Oct 27;207(2):269–82.
32. Bhogaraju S, Taschner M, Morawetz M, Basquin C, Lorentzen E. Crystal structure of the intraflagellar transport complex 25/27. EMBO J. 2011 May 18;30(10):1907–18.
33. Wachter S, Jung J, Shafiq S, Basquin J, Fort C, Bastin P, et al. Binding of IFT22 to the intraflagellar transport complex is essential for flagellum assembly. EMBO J. 2019 May 2;38(9):e101251.
34. Bhogaraju S, Cajanek L, Fort C, Blisnick T, Weber K, Taschner M, et al. Molecular basis of tubulin transport within the cilium by IFT74 and IFT81. Science. 2013 Aug 30;341(6149):1009–12.
35. Taschner M, Lorentzen A, Mourão A, Collins T, Freke GM, Moulding D, et al. Crystal structure of intraflagellar transport protein 80 reveals a homo-dimer required for ciliogenesis. eLife. 2018 Apr 16;7:e33067.
36. Hoek H van den, Klena N, Jordan MA, Viar GA, Schaffer M, Erdmann PS, et al. In situ architecture of the ciliary base reveals the stepwise assembly of IFT trains [Internet]. 2021 [cited 2022 Jan 10]. p. 2021.10.17.464685. Available from: https://www.biorxiv.org/content/10.1101/2021.10.17.464685v2
37. Jordan MA, Diener DR, Stepanek L, Pigino G. The cryo-EM structure of intraflagellar transport trains reveals how dynein is inactivated to ensure unidirectional anterograde movement in cilia. Nat Cell Biol. 2018 Nov;20(11):1250–5.
38. van Dam TJP, Townsend MJ, Turk M, Schlessinger A, Sali A, Field MC, et al. Evolution of

modular intraflagellar transport from a coatomer-like progenitor. Proc Natl Acad Sci. 2013 Apr 23;110(17):6943–8.
39. Behal RH, Miller MS, Qin H, Lucker BF, Jones A, Cole DG. Subunit Interactions and Organization of the Chlamydomonas reinhardtii Intraflagellar Transport Complex A Proteins. J Biol Chem. 2012 Apr;287(15):11689–703.
40. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science [Internet]. 2021 Aug 20 [cited 2022 Jan 11]; Available from: https://www.science.org/doi/abs/10.1126/science.abj8754
41. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.
42. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all [Internet]. In Review; 2021 Nov [cited 2022 Jan 11]. Available from: https://www.researchsquare.com/article/rs-1032816/v1
43. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021 Aug;596(7873):590–6.
44. Leitner A, Faini M, Stengel F, Aebersold R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. Trends Biochem Sci. 2016 Jan;41(1):20–32.
45. O'Reilly FJ, Rappsilber J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. Nat Struct Mol Biol. 2018 Nov;25(11):1000–8.
46. Tang X, Wippel HH, Chavez JD, Bruce JE. Crosslinking mass spectrometry: A link between structural biology and systems biology. Protein Sci Publ Protein Soc. 2021 Apr;30(4):773–84.
47. Liu F, Lössl P, Scheltema R, Viner R, Heck AJR. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. Nat Commun. 2017 May 19;8:15473.
48. Rajagopalan V, Corpuz EO, Hubenschmidt MJ, Townsend CR, Asai DJ, Wilkes DE. Analysis of properties of cilia using Tetrahymena thermophila. Methods Mol Biol Clifton NJ. 2009;586:283–99.
49. Lucker BF, Behal RH, Qin H, Siron LC, Taggart WD, Rosenbaum JL, et al. Characterization of the intraflagellar transport complex B core: direct interaction of the IFT81 and IFT74/72 subunits. J Biol Chem. 2005 Jul 29;280(30):27688–96.
50. Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. Protein Sci Publ Protein Soc. 2014 Jun;23(6):747–59.
51. Erzberger JP, Stengel F, Pellarin R, Zhang S, Schaefer T, Aylett CHS, et al. Molecular architecture of the 40S · eIF1 · eIF3 translation initiation complex. Cell. 2014 Aug 28;158(5):1123–35.
52. Fernandez-Martinez J, Kim SJ, Shi Y, Upla P, Pellarin R, Gagnon M, et al. Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. Cell. 2016 Nov;167(5):1215-1228.e25.
53. LoPiccolo J, Kim SJ, Shi Y, Wu B, Wu H, Chait BT, et al. Assembly and Molecular Architecture of the Phosphoinositide 3-Kinase p85α Homodimer. J Biol Chem. 2015 Dec 18;290(51):30390–405.
54. Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, et al. Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. Mol Cell Proteomics MCP. 2014 Nov;13(11):2927–43.
55. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer [Internet]. Bioinformatics; 2021 Oct [cited 2022 May

6]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.10.04.463034
56. McCafferty CL, Pennington EL, Papoulas O, Taylor DW, Marcotte EM. Does AlphaFold2 model proteins' intracellular conformations? An experimental test using cross-linking mass spectrometry of endogenous ciliary proteins [Internet]. Bioinformatics; 2022 Aug [cited 2022 Oct 14]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.08.25.505345
57. Webb B, Viswanath S, Bonomi M, Pellarin R, Greenberg CH, Saltzberg D, et al. Integrative structure modeling with the integrative modeling platform. Protein Sci. 2018;27(1):245–58.
58. Saltzberg D, Greenberg CH, Viswanath S, Chemmama I, Webb B, Pellarin R, et al. Modeling biological complexes using integrative modeling platform. In: Biomolecular Simulations. Springer; 2019. p. 353–77.
59. Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, et al. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol. 2012;10(1):e1001244.
60. Viswanath S, Chemmama IE, Cimermancic P, Sali A. Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. Biophys J. 2017 Dec;113(11):2344–53.
61. Jordan MA, Pigino G. In situ cryo-electron tomography and subtomogram averaging of intraflagellar transport trains. Methods Cell Biol. 2019;152:179–95.
62. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605–12.
63. Kidmose RT, Juhl J, Nissen P, Boesen T, Karlsen JL, Pedersen BP. Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. IUCrJ. 2019 Jul 1;6(Pt 4):526–31.
64. Hirano T, Katoh Y, Nakayama K. Intraflagellar transport-A complex mediates ciliary entry and retrograde trafficking of ciliary G protein–coupled receptors. Marshall W, editor. Mol Biol Cell. 2017 Feb;28(3):429–39.
65. Takahara M, Katoh Y, Nakamura K, Hirano T, Sugawa M, Tsurumi Y, et al. Ciliopathy-associated mutations of IFT122 impair ciliary protein trafficking but not ciliogenesis. Hum Mol Genet. 2018 Feb 1;27(3):516–28.
66. Fu W, Wang L, Kim S, Li J, Dynlacht BD. Role for the IFT-A Complex in Selective Transport to the Primary Cilium. Cell Rep. 2016 Nov;17(6):1505–17.
67. Kiesel P, Alvarez Viar G, Tsoy N, Maraspini R, Gorilak P, Varga V, et al. The molecular structure of mammalian primary cilia revealed by cryo-electron tomography. Nat Struct Mol Biol. 2020 Dec;27(12):1115–24.
68. Lee C, Goldberg J. Structure of Coatomer Cage Proteins and the Relationship among COPI, COPII, and Clathrin Vesicle Coats. Cell. 2010 Jul;142(1):123–32.
69. Yang S, Bahl K, Chou HT, Woodsmith J, Stelzl U, Walz T, et al. Near-atomic structures of the BBSome reveal the basis for BBSome activation and binding to GPCR cargoes. eLife. 2020 Jun 8;9:e55954.
70. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022 Jan 7;50(D1):D439–44.
71. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010 Jul;38(Web Server issue):W545-549.
72. Bayly-Jones C, Whisstock J. Mining folded proteomes in the era of accurate structure prediction [Internet]. 2021 [cited 2022 Jan 13]. p. 2021.08.24.457439. Available from: https://www.biorxiv.org/content/10.1101/2021.08.24.457439v1

73. Xu H, Lin Z, Li F, Diao W, Dong C, Zhou H, et al. Dimerization of elongator protein 1 is essential for Elongator complex assembly. Proc Natl Acad Sci. 2015 Aug 25;112(34):10697–702.
74. Architecture of the yeast Elongator complex. EMBO Rep. 2017 Feb 1;18(2):264–79.
75. Rahl PB, Chen CZ, Collins RN. Elp1p, the yeast homolog of the FD disease syndrome protein, negatively regulates exocytosis independently of transcriptional elongation. Mol Cell. 2005 Mar 18;17(6):841–53.
76. Bultema JJ, Ambrosio AL, Burek CL, Di Pietro SM. BLOC-2, AP-3, and AP-1 proteins function in concert with Rab38 and Rab32 proteins to mediate protein trafficking to lysosome-related organelles. J Biol Chem. 2012 Jun 1;287(23):19550–63.
77. Dennis MK, Mantegazza AR, Snir OL, Tenza D, Acosta-Ruiz A, Delevoye C, et al. BLOC-2 targets recycling endosomal tubules to melanosomes for cargo delivery. J Cell Biol. 2015 May 25;209(4):563–77.
78. Zanetti G, Pahuja KB, Studer S, Shim S, Schekman R. COPII and the regulation of protein sorting in mammals. Nat Cell Biol. 2012 Jan;14(1):20–8.
79. Popoff V, Adolf F, Brügger B, Wieland F. COPI budding within the Golgi stack. Cold Spring Harb Perspect Biol. 2011 Nov 1;3(11):a005231.
80. Bröcker C, Kuhlee A, Gatsogiannis C, Balderhaar HJ kleine, Hönscher C, Engelbrecht-Vandré S, et al. Molecular architecture of the multisubunit homotypic fusion and vacuole protein sorting (HOPS) tethering complex. Proc Natl Acad Sci. 2012 Feb 7;109(6):1991–6.
81. Grove TZ, Cortajarena AL, Regan L. Ligand binding by repeat proteins: natural and designed. Curr Opin Struct Biol. 2008 Aug;18(4):507–15.
82. ter Haar E, Harrison SC, Kirchhausen T. Peptide-in-groove interactions link target proteins to the beta-propeller of clathrin. Proc Natl Acad Sci U S A. 2000 Feb 1;97(3):1096–100.
83. Bhogaraju S, Engel BD, Lorentzen E. Intraflagellar transport complex structure and cargo interactions. Cilia. 2013 Aug 14;2(1):10.
84. Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. Pediatr Nephrol Berl Ger. 2011 Jul;26(7):1039–56.
85. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D514-517.
86. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014 Jan;42(Database issue):D980-985.
87. Li D, Roberts R. WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell Mol Life Sci CMLS. 2001 Dec;58(14):2085–97.
88. Picariello T, Brown JM, Hou Y, Swank G, Cochran DA, King OD, et al. A global analysis of IFT-A function reveals specialization for transport of membrane-associated proteins into cilia. J Cell Sci. 2019 Feb 11;132(3):jcs220749.
89. Schmidts M, Frank V, Eisenberger T, Al Turki S, Bizet AA, Antony D, et al. Combined NGS approaches identify mutations in the intraflagellar transport gene IFT140 in skeletal ciliopathies with early progressive kidney Disease. Hum Mutat. 2013 May;34(5):714–24.
90. Perrault I, Saunier S, Hanein S, Filhol E, Bizet AA, Collins F, et al. Mainzer-Saldino syndrome is a ciliopathy caused by IFT140 mutations. Am J Hum Genet. 2012 May 4;90(5):864–70.
91. Hull S, Owen N, Islam F, Tracey-White D, Plagnol V, Holder GE, et al. Nonsyndromic Retinal Dystrophy due to Bi-Allelic Mutations in the Ciliary Transport Gene IFT140. Invest Ophthalmol Vis Sci. 2016 Mar;57(3):1053–62.
92. Niwa S. The nephronophthisis-related gene ift-139 is required for ciliogenesis in

Caenorhabditis elegans. Sci Rep. 2016 Aug 12;6:31544.
93. McCafferty CL, Papoulas O, Jordan MA, Hoogerbrugge G, Nichols C, Pigino G, et al. Integrative modeling reveals the molecular architecture of the Intraflagellar Transport A (IFT-A) complex [Internet]. Biochemistry; 2022 Jul [cited 2022 Jul 14]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.07.05.498886
94. Hesketh SJ, Mukhopadhyay AG, Nakamura D, Toropova K, Roberts AJ. IFT-A Structure Reveals Carriages for Membrane Protein Transport into Cilia [Internet]. Cell Biology; 2022 Aug [cited 2022 Oct 14]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.08.09.503213
95. Lacey SE, Foster HE, Pigino G. The Molecular Structure of Anterograde Intraflagellar transport trains [Internet]. Molecular Biology; 2022 Aug [cited 2022 Oct 14]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.08.01.502329
96. Chien A, Shih SM, Bower R, Tritschler D, Porter ME, Yildiz A. Dynamics of the IFT machinery at the ciliary tip. eLife. 2017 Sep 20;6:e28606.
97. Stepanek L, Pigino G. Microtubule doublets are double-track railways for intraflagellar transport trains. Science. 2016 May 6;352(6286):721–4.
98. Gaertig J, Wloga D, Vasudevan KK, Guha M, Dentler W. Discovery and Functional Evaluation of Ciliary Proteins in Tetrahymena thermophila. In: Methods in Enzymology [Internet]. Elsevier; 2013 [cited 2022 Jan 10]. p. 265–84. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780123979445000134
99. McWhite CD, Papoulas O, Drew K, Dang V, Leggere JC, Sae-Lee W, et al. Co-fractionation/mass spectrometry to identify protein complexes. STAR Protoc. 2021 Mar 19;2(1):100370.
100. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, et al. A census of human soluble protein complexes. Cell. 2012 Aug 31;150(5):1068–81.
101. McWhite CD, Papoulas O, Drew K, Cox RM, June V, Dong OX, et al. A Pan-plant Protein Complex Map Reveals Deep Conservation and Novel Assemblies. Cell. 2020 Apr 16;181(2):460-474.e14.
102. Graham M, Combe C, Kolbowski L, Rappsilber J. xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data [Internet]. Molecular Biology; 2019 Feb [cited 2022 Jan 10]. Available from: http://biorxiv.org/lookup/doi/10.1101/561829
103. Lv B, Wan L, Taschner M, Cheng X, Lorentzen E, Huang K. Intraflagellar transport protein IFT52 recruits IFT46 to the basal body and flagella. J Cell Sci. 2017 May 1;130(9):1662–74.
104. Mastronarde DN. Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol. 2005 Oct;152(1):36–51.
105. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods. 2013 Jun;10(6):584–90.
106. Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using IMOD. J Struct Biol. 1996 Feb;116(1):71–6.
107. Xiong Q, Morphew MK, Schwartz CL, Hoenger AH, Mastronarde DN. CTF determination and correction for low dose tomographic tilt series. J Struct Biol. 2009 Dec;168(3):378–87.
108. Grant T, Grigorieff N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. eLife. 2015 May 29;4:e06980.
109. Heumann JM, Hoenger A, Mastronarde DN. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. J Struct Biol. 2011 Sep;175(3):288–99.
110. Stanzione D, West J, Evans RT, Minyard T, Ghattas O, Panda DK. Frontera: The

Evolution of Leadership Computing at the National Science Foundation. In: Practice and Experience in Advanced Research Computing [Internet]. Portland OR USA: ACM; 2020 [cited 2022 Jun 7]. p. 106–11. Available from: https://dl.acm.org/doi/10.1145/3311790.3396656
111. Ganesan SJ, Feyder MJ, Chemmama IE, Fang F, Rout MP, Chait BT, et al. Integrative structure and function of the yeast exocyst complex. Protein Sci. 2020;
112. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX Visualization System. Protein Sci. 2018 Jan;27(1):14–25.
113. Deutsch EW, Bandeira N, Sharma V, Perez-Riverol Y, Carver JJ, Kundu DJ, et al. The ProteomeXchange consortium in 2020: enabling “big data” approaches in proteomics. Nucleic Acids Res. 2020 Jan 8;48(D1):D1145–52.
114. Patwardhan A. Trends in the Electron Microscopy Data Bank (EMDB). Acta Crystallogr Sect Struct Biol. 2017 Jun 1;73(Pt 6):503–8.
115. Burley SK, Kurisu G, Markley JL, Nakamura H, Velankar S, Berman HM, et al. PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models. Struct Lond Engl 1993. 2017 Sep 5;25(9):1317–8.
116. Dodonova SO, Diestelkoetter-Bachert P, von Appen A, Hagen WJH, Beck R, Beck M, et al. VESICULAR TRANSPORT. A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. Science. 2015 Jul 10;349(6244):195–8.
117. Hutchings J, Stancheva VG, Brown NR, Cheung ACM, Miller EA, Zanetti G. Structure of the complete, membrane-assembled COPII coat reveals a complex interaction network. Nat Commun. 2021 Apr 1;12(1):2034.
118. Cheng TC, Hong C, Akey IV, Yuan S, Akey CW. A near atomic structure of the active human apoptosome. eLife. 2016 Oct 4;5:e17755.
119. Zhang S, Chang L, Alfieri C, Zhang Z, Yang J, Maslen S, et al. Molecular mechanism of APC/C activation by mitotic phosphorylation. Nature. 2016 May;533(7602):260–4.
120. Huang G, Zhan X, Zeng C, Zhu X, Liang K, Zhao Y, et al. Cryo-EM structure of the nuclear ring from Xenopus laevis nuclear pore complex. Cell Res. 2022 Apr;32(4):349–58.

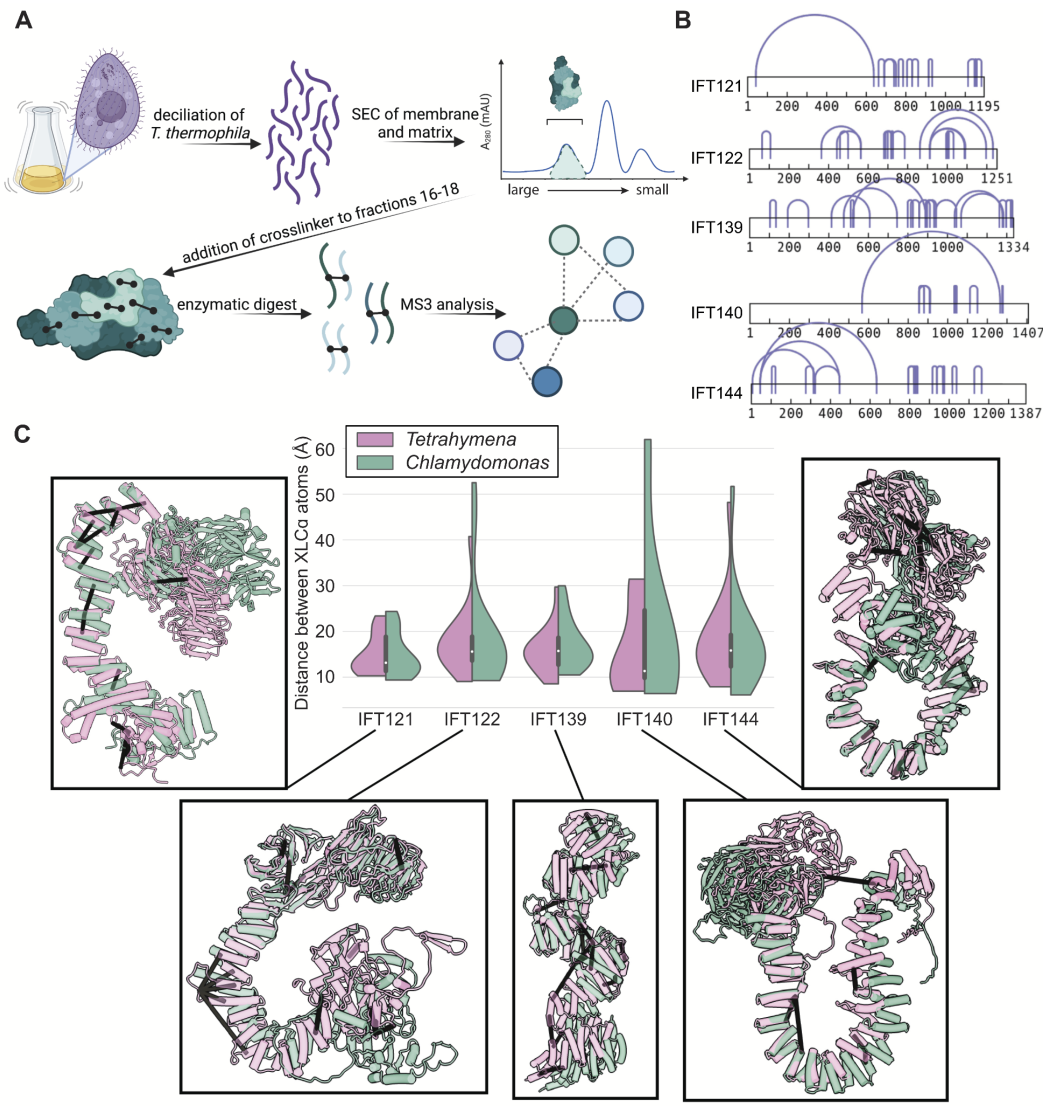


Figure 1. AlphaFold2 structures of the IFT-A subunits are supported by experimentally-determined intramolecular cross-links. (**A**) Sample preparation protocol for obtaining an enriched sample of endogenous IFT-A from *T. thermophila*. The sample preparation was followed by chemical cross-linking using DSSO and tandem (MS^2/MS^3) mass spectrometry to identify cross-linked peptides. Image created with BioRender.com. (**B**) Bar diagrams highlight the extensive intramolecular DSSO cross-links (purple arcs connecting cross-linked amino acid pairs) within each of the IFT-A subunits (bars, numbers denote amino acid positions). (**C**) Violin plots of the distance between Cα atoms of chemically cross-linked residues. Surrounding images show locations of intramolecular crosslinks (black bars) on aligned AlphaFold2 predicted models of IFT-A subunits from *Chlamydomonas reinhardtii* (green) and *Tetrahymena thermophila* (pink). A maximum distance of 35 Å between Cα atoms is expected for DSSO cross-links. 97% of intramolecular cross-links are satisfied for *T. thermophila* and 94% for C. *reinhardtii.*

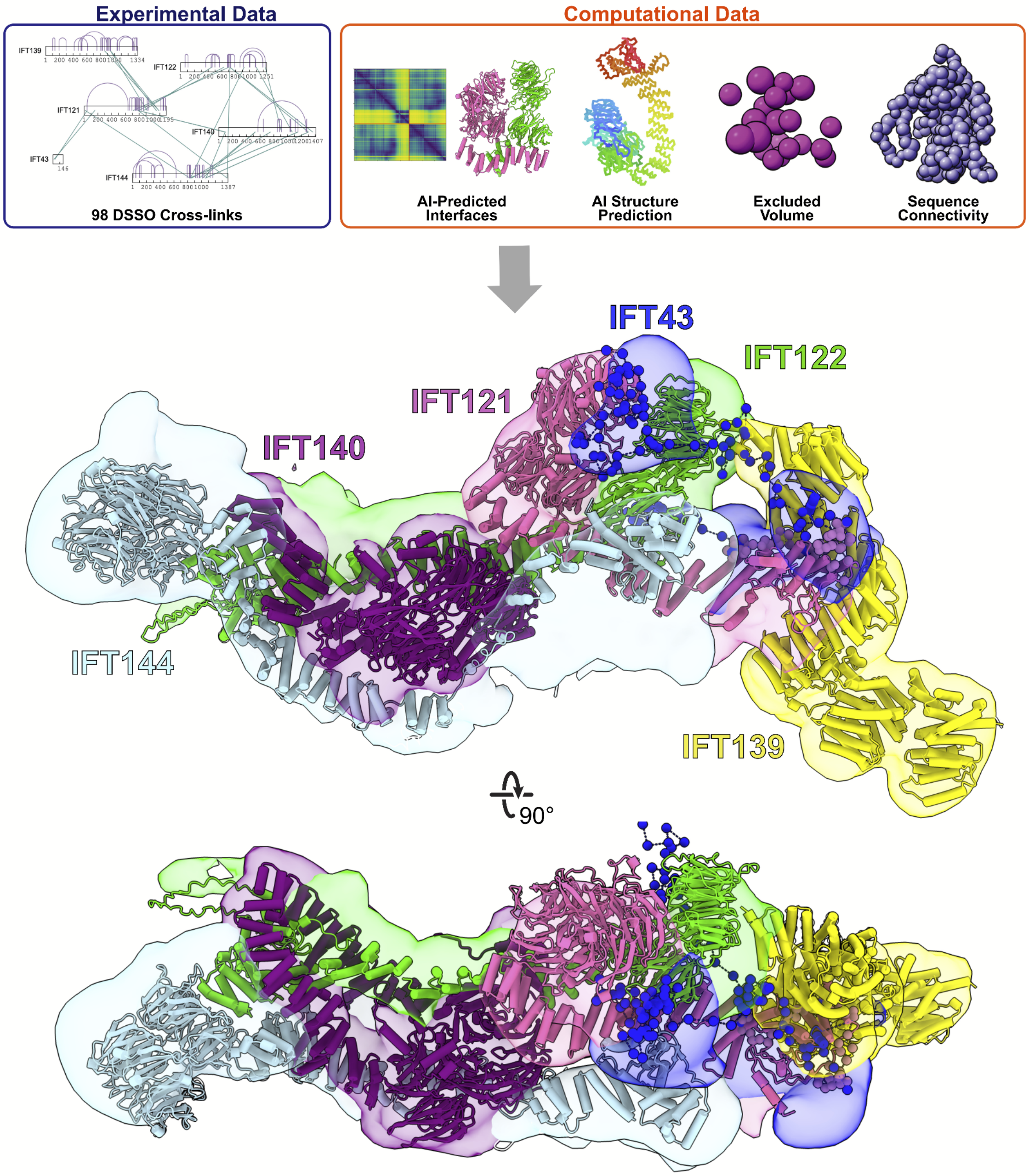


Figure 2. Integrative modeling of the IFT-A monomer identifies subunit locations and interactions. Using spatial restraints (**top panels**) based on chemical cross-links (intramolecular, purple arcs; intermolecular, green lines), AlphaFold2 protein models, and AlphaFold2 high-confidence predicted interfaces between proteins (see **Methods** for more details), an integrative model (**bottom panel**) of monomer IFT-A was determined that best satisfied these spatial restraints. The centroid model of the top-scoring cluster of 9,121 models is shown (ribbon diagram) in the localized probability density (colored volumes) for each subunit.

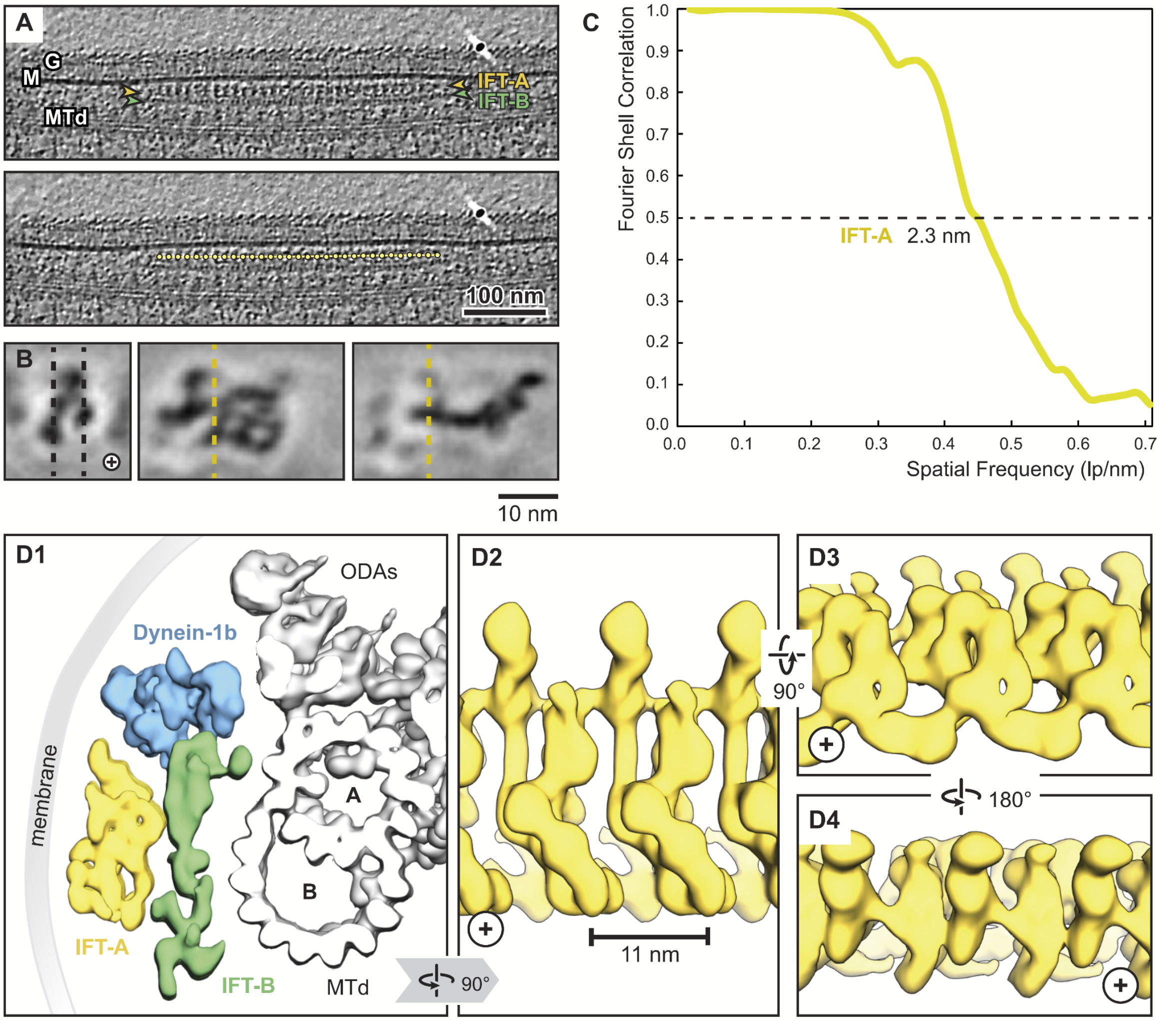


Figure 3. Overview of the IFT-A cryo-ET structure. (**A**) IFT train in a raw tomogram of a *C. reinhardtii* cilium with the section showing the repeats of IFT-A and IFT-B. The picking of IFT-A particles is shown in the lower panel (MTd, microtubule doublet; G, glycocalyx; M, membrane; plus, direction of the ciliary tip)**.** (**B**) Subtomogram average of IFT-A; the left panel shows the same orientation as in A**.** (**C**) The Fourier shell correlation of the subtomogram average indicates a resolution of 2.3 nm at a cut-off criterion of 0.5. The pixel size is 0.71 nm. (**D1**) 3D isosurface of a reconstructed IFT train in between the membrane and the microtubule doublet as seen from the ciliary base towards the tip (ODAs, outer dynein arms; A, B, A- and B-tubule) and composed of three averages: IFT-A (yellow), IFT-B (green) and dynein-1b (blue). (**D2**) The IFT-A polymer as seen from the membrane towards the microtubule doublet. (**D3**) and (**D4**) Views of IFT-A as indicated.

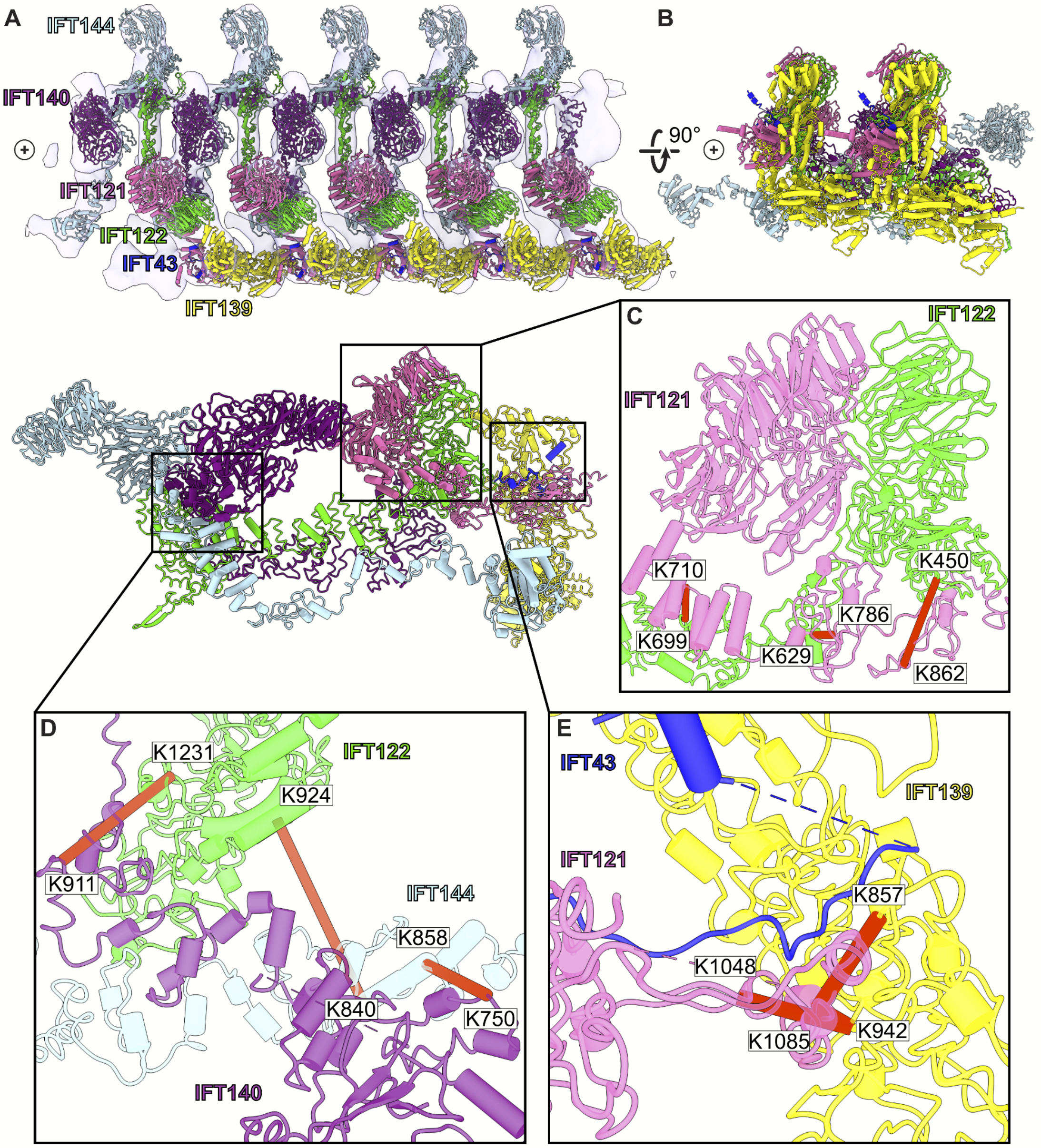


Figure 4. A model of the IFT-A train based on docking the monomeric model into the subtomogram average. (**A**) The final cluster centroid model fit into the five IFT-A polymeric repeat subtomogram average map using molecular dynamics-based flexible fitting (63). (**B**) An alternative side view of the IFT-A trains to show interactions between adjacent monomers. Plus signs in A,B indicate the direction of the ciliary tip. (**C**) The IFT121-IFT122 interaction in the train-docked model with satisfied intermolecular cross-linked pairs shown in red. (**D**) The interaction between IFT122-IFT140-IFT144 is shown with satisfied intermolecular cross-links in red. (**E**) The interaction between IFT43-IFT121-IFT139 with satisfied intermolecular cross-links between IFT121 and IFT139 shown.

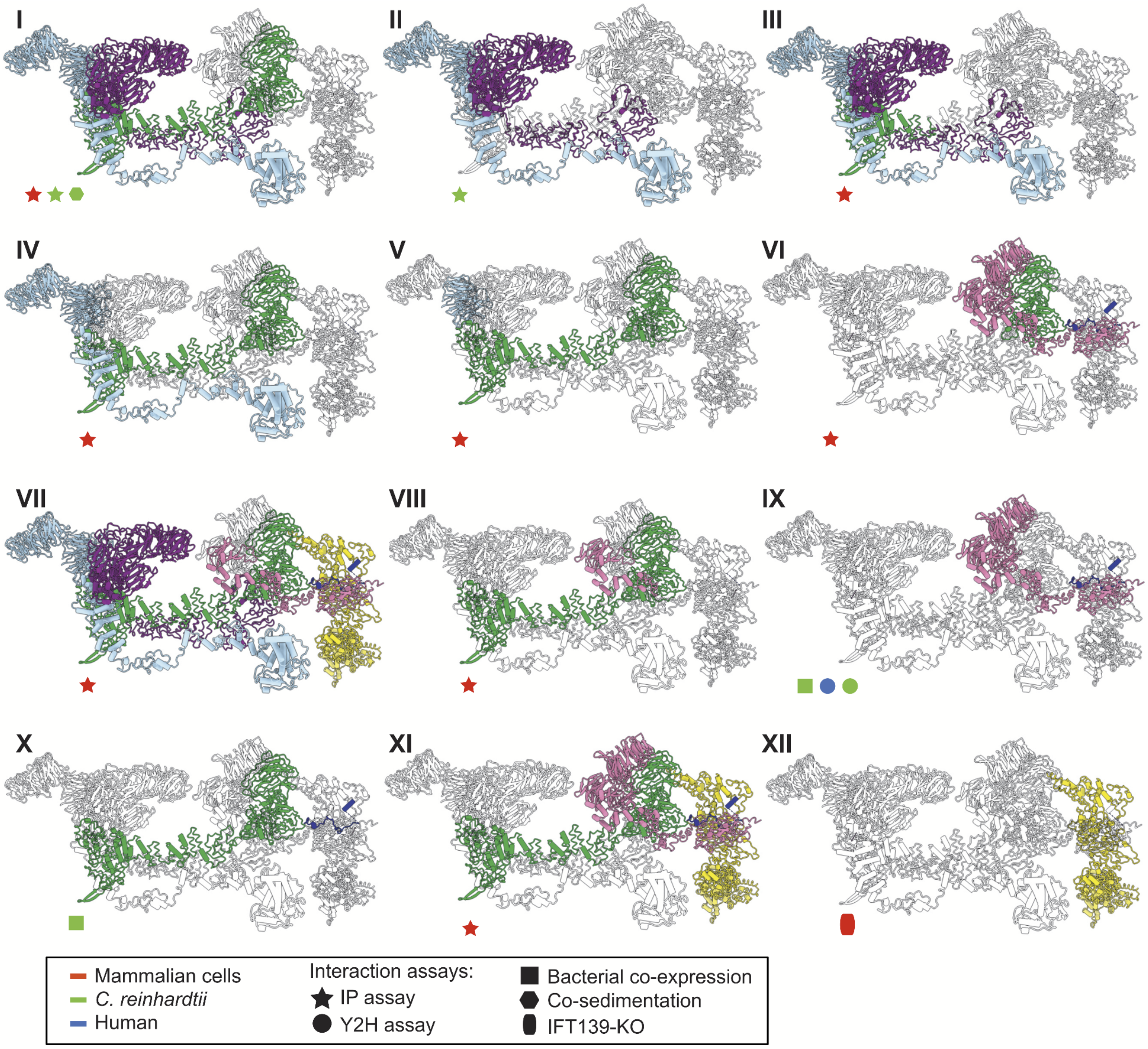


Figure 5. The IFT-A 3D model is highly concordant with literature evidence not used in modeling. Panels **I-XII** highlight specific IFT-A intermolecular interactions, both at the protein and domain level, known from the literature (see the text for detailed discussion and citations). The interacting proteins or domains form direct contacts within the IFT-A model in all cases examined. Support for the interactions was observed across organisms, consistent with high conservation of the IFT-A complex across species. IP, immunoprecipitation; Y2H, yeast 2-hybrid; KO, knockout mutant. Proteins are colored as in Figure 4.

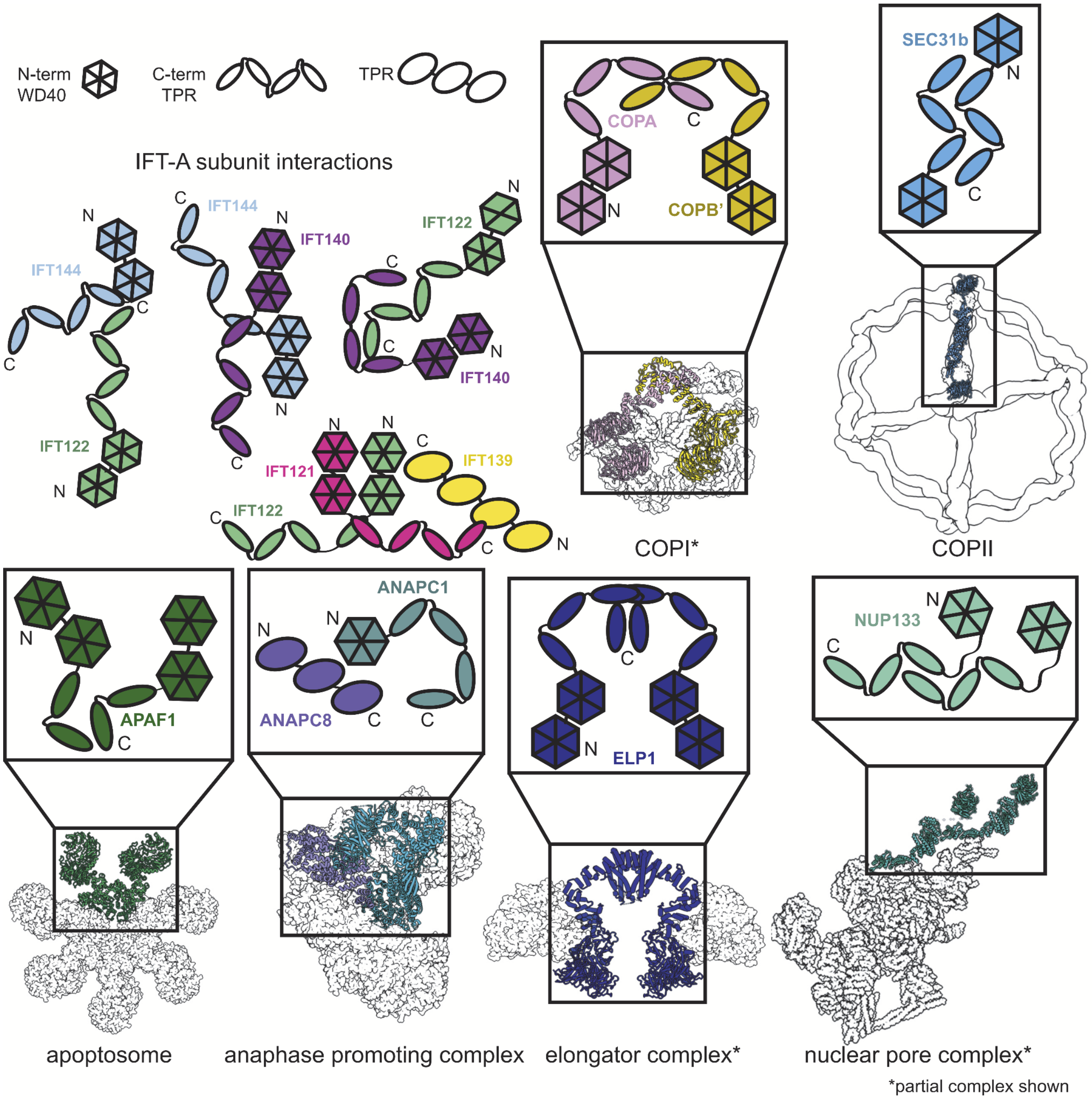


Figure 6. Proteins sharing IFT-A subunit domain architectures vary markedly in their quaternary assemblies. By performing structure-based searches (71) of the set of AlphaFold2-predicted human protein structures (43), we identified proteins that have a similar domain architecture to IFT121, IFT122, IFT140, and IFT144 (as in **Figure 6—figure supplement 1**). Colors highlight the role of these domains in depictions of native complexes containing proteins with these domain architectures. As in our model of the IFT-A complex, stable interactions between TPR domains underlie diverse macromolecular assemblies. COPI PDB: 5A1W (116), COPII PDB: 6ZL0 (117), apoptosome PDB: 5JUY (118), anaphase promoting complex PDB: 5G05 (119), elongator complex PDB: 5CQR (73), nuclear pore complex PDB: 7WB4 (120).

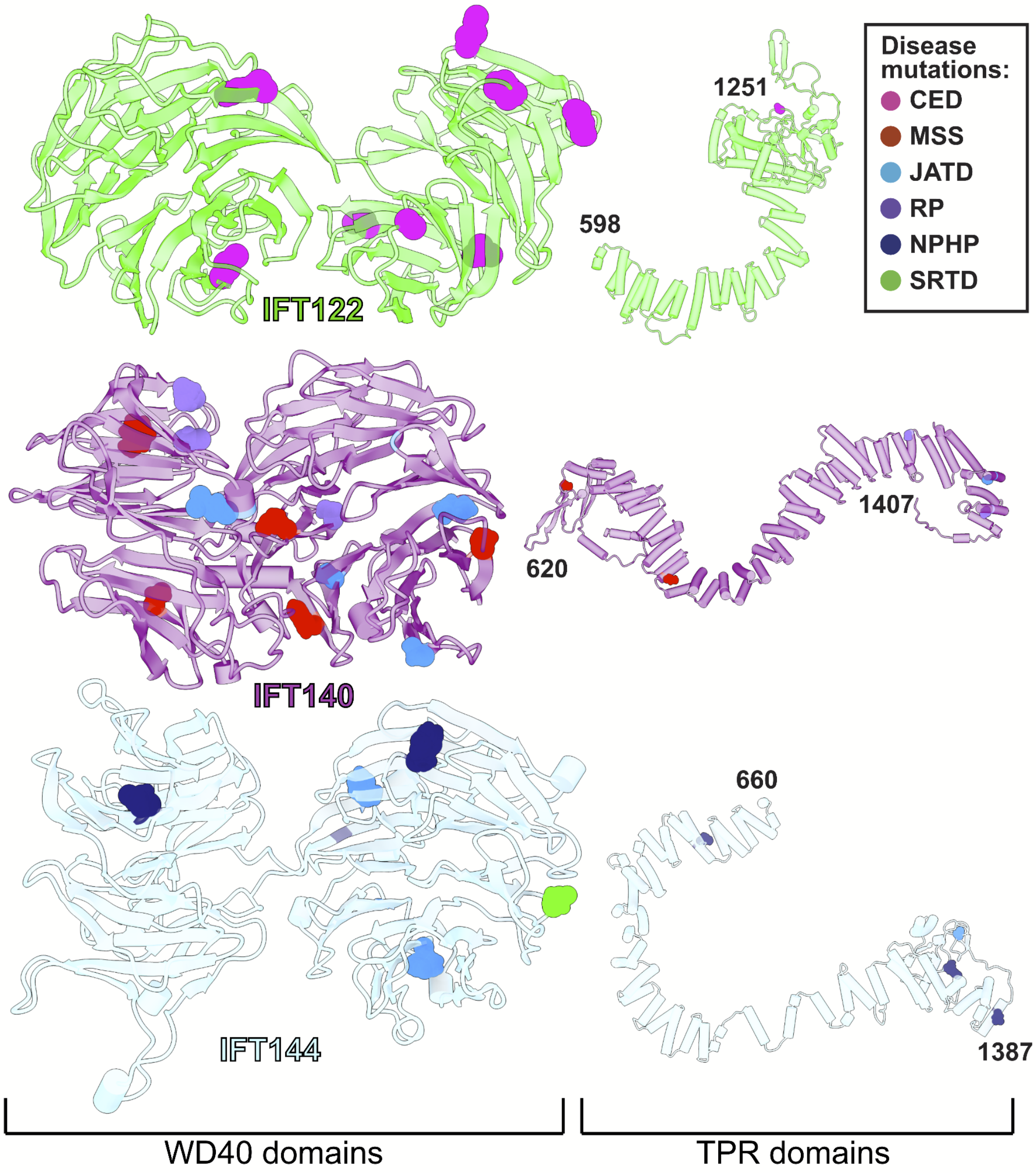


Figure 7. WD40 domains of IFT-A core proteins are hotspots for disease-causing missense mutations. Disease-causing missense mutations are displayed on the structures of IFT122, IFT140, and IFT144 (IFT-A core) and colored by disease. 79% of the mutations on the IFT-A core proteins are concentrated in the WD40 domains of these proteins. The mutations are located in the exposed regions of the domains and do not interfere with other IFT-A interactions suggesting they may disrupt more transient interactions formed between IFT-A and its cargos. TPR domains are shown smaller in scale relative to WD40 domains for display purposes. CED, cranioectodermal dysplasia; JATD, Jeune asphyxiating thoracic dystrophy; NPHP, nephronophthisis; MSS, Mainzer-Saldino syndrome; RP, retinitis pigmentosa; SRTD, short-rib thoracic dysplasia.

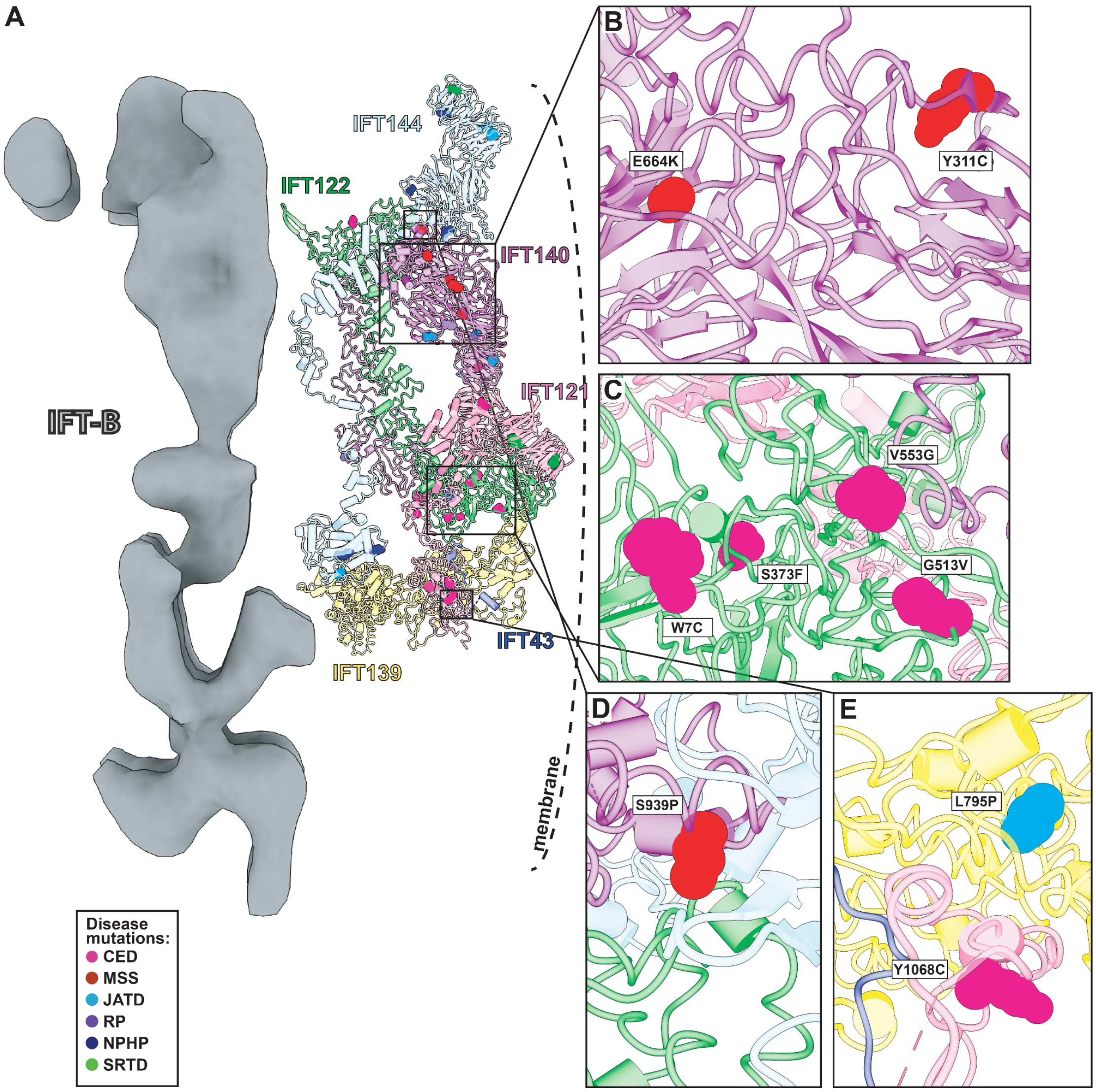


Figure 8. Human disease mutations form clusters on exposed WD40 domains and interaction interfaces of the IFT-A complex. (**A**) Human disease mutations are mapped onto our IFT-A structure with missense mutations colored by disease. The dotted line indicates the ciliary membrane. (**B**) The closeup of MSS variants E664K and Y311C, which are located among many other disease-causing variants on the WD40 domains of IFT140. (**C**) A closeup of CED alleles on the WD40 domain of IFT122 are positioned under the membrane and nearby IFT121. (**D**) IFT140 MSS variant S939P lies at the interaction interface with proteins IFT122 and IFT144. (**E**) A closeup of CED variant Y1068C of IFT121 and its proximity to neighboring proteins, IFT43 and IFT139. The closeup also captures JATD variant L795P of IFT139 and its location near the interaction interface with IFT121. Abbreviations are provided in the **Figure 7** legend.

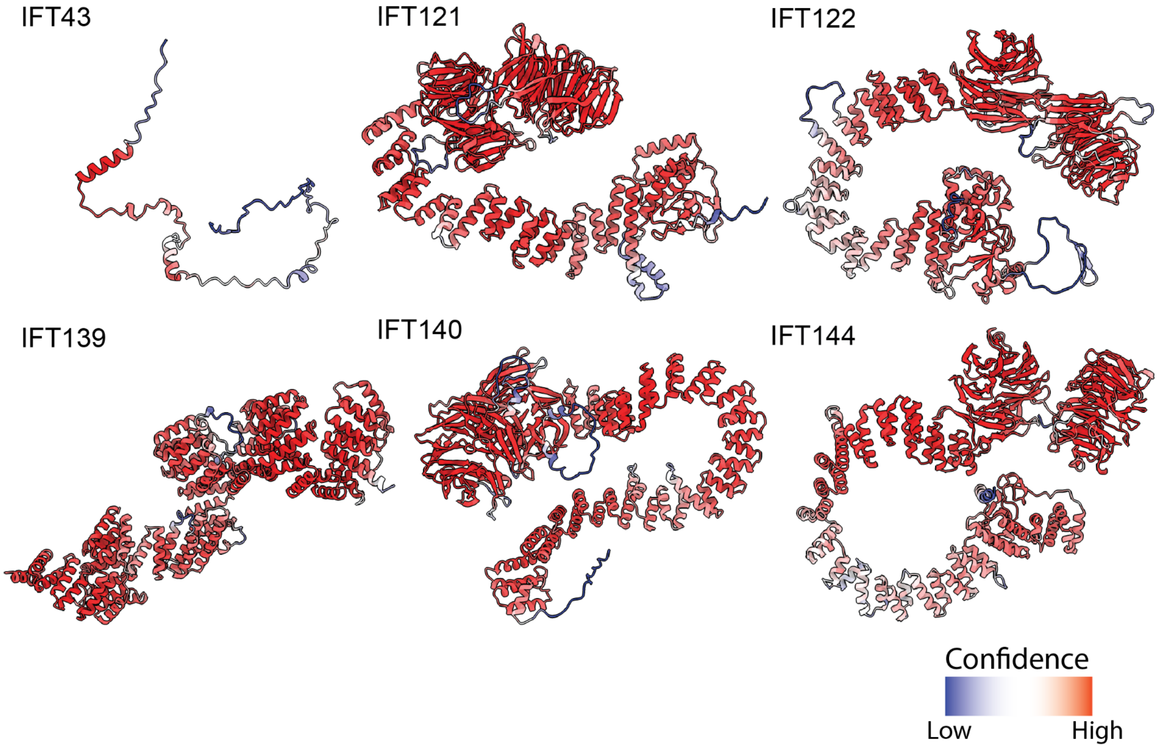


Figure 1—figure supplement 1. AlphaFold2 structural predictions for *T. thermophila* proteins of the IFT-A complex. Structures are colored based on pLDDT (43) where red represents higher confidence predictions and blue lower confidence predictions.

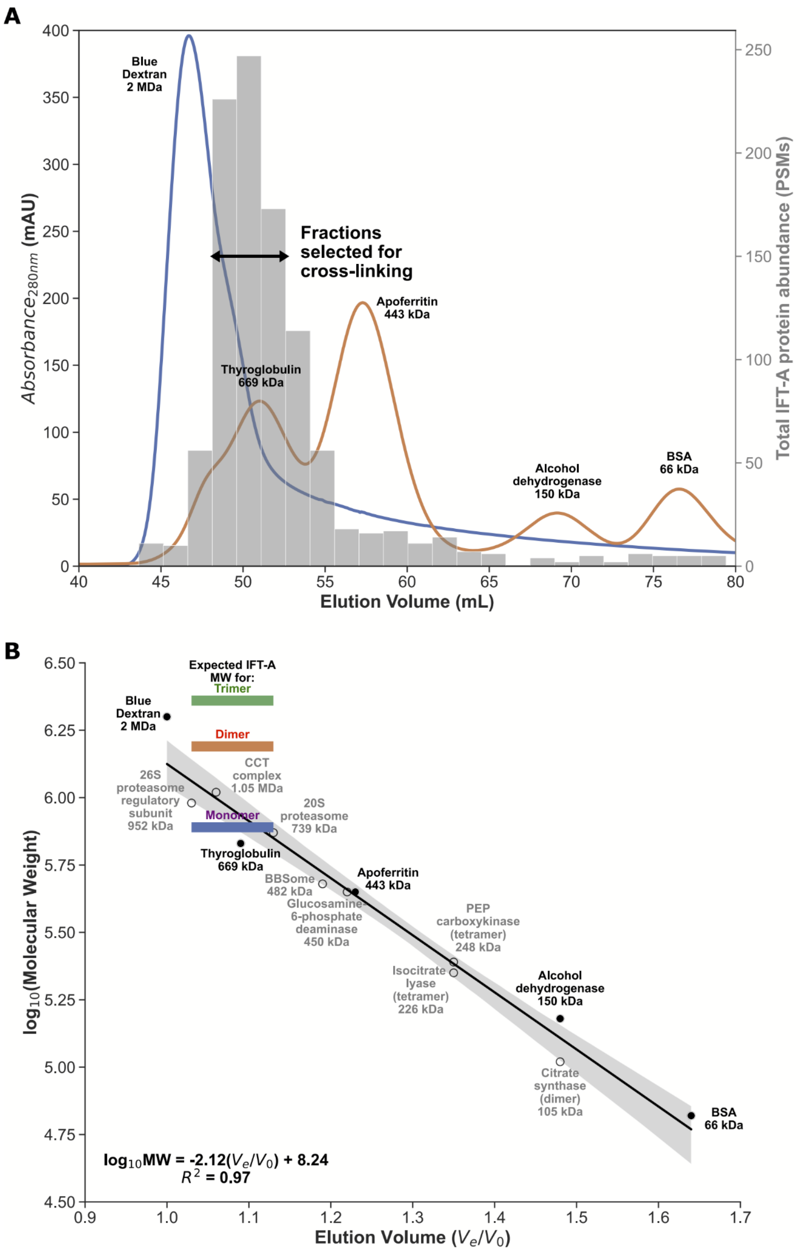


Figure 1—figure supplement 2. Enrichment for monomeric IFT-A from *Tetrahymena* cilia. (**A**) FPLC size exclusion chromatography elution traces for molecular weight calibrants (lines, left axis), overlaid with mass-spectrometry determined IFT-A protein abundances (in protein spectral matches, PSMs; bars, right axis) measured from *Tetrahymena* whole cell extract. Using these data, we identified IFT-A fractions to be subsequently collected from ciliary M+M preparations for cross-linking. (**B**) Linear regression between chromatography elution volumes (V_e, peak elution volume in mL; V_0, void volume from Blue Dextran) and molecular weight calibrants (known size calibrants, filled circles; protein complexes of known composition, open circles) provided a molecular weight estimate for the selected IFT-A fractions ranging between 720 kDa and 1.1 MDa, consistent with an IFT-A monomer, whose expected molecular weight is 772 kDa.

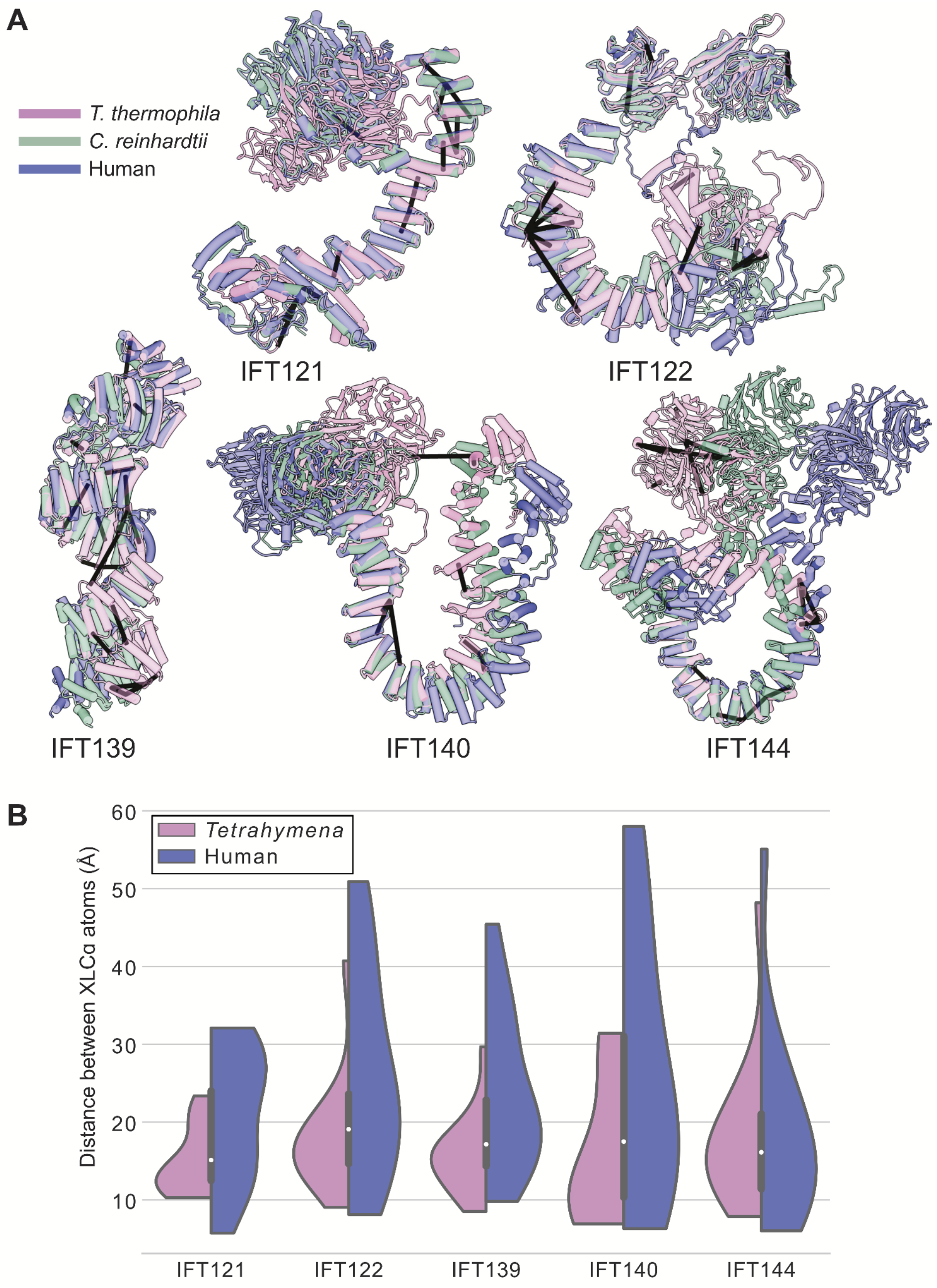
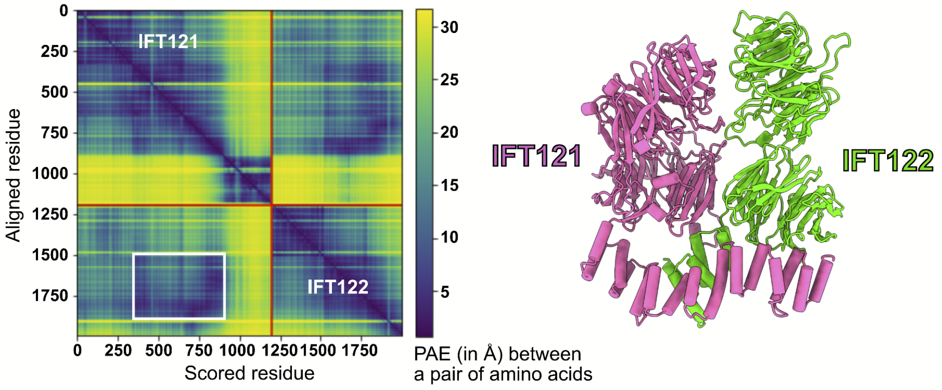


Figure 1—figure supplement 3. *Tetrahymena* IFT-A cross-links mapped onto AlphaFold-predicted structures of human IFT-A proteins. (**A**) Locations of intramolecular crosslinks (black bars) are visualized on structural alignments of AlphaFold2 predicted models of IFT-A subunits from *Tetrahymena thermophila* (pink), *Chlamydomonas reinhardtii* (green), and humans (blue). (**B**) Violin plots show the distance between Cα atoms of chemically cross-linked residues by DSSO. A maximum distance of 35 Å between Cα atoms is expected for DSSO cross-links. 86% of intramolecular cross-links are satisfied for the human structures.

Large interaction predicted, modeled as rigid body



No interaction predicted

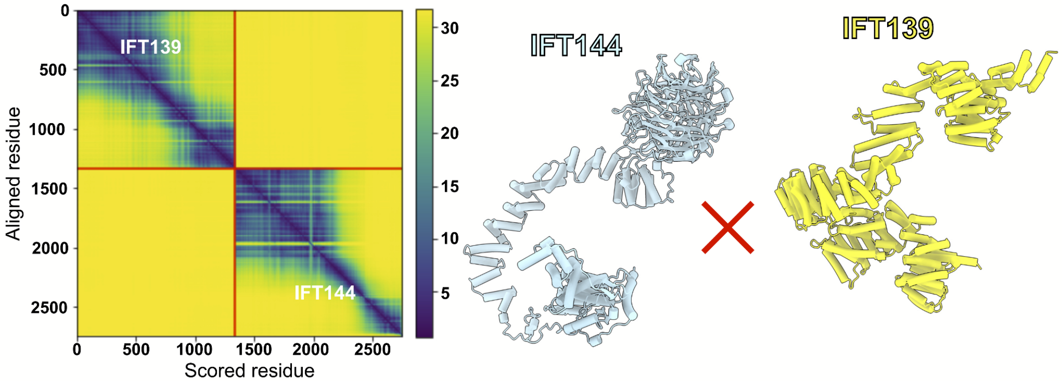


Figure 2—figure supplement 1. AlphaFold PAE analysis was used to determine the boundaries for rigid-bodies in integrative modeling. AlphaFold-Multimer (55) was used to model pairs of IFT-A proteins and compute the predicted aligned error (PAE, reported in Å, where darker colors indicate higher confidence in the relative spatial positions of two amino acid residues) between all pairs of amino acid residues in both structures. Using these PAE scores and thresholds determined by comparing PAE scores to independent structural evidence (56), we defined boundaries of well-predicted interacting domains to be used as rigid bodies for subsequent integrative modeling (see **Methods**), as for the case shown for IFT121-IFT122 (**top panels**). The white box indicates the extracted interacting region, which in this case lies within larger well-predicted monomeric structures; the complete IFT121-122 interaction model is drawn at right. No such information was used from the PAE plots of the proteins that were not predicted to interact, as for the case of IFT139-IFT144 (**bottom panels**), for which neither AlphaFold nor chemical cross-links showed any evidence for interacting.

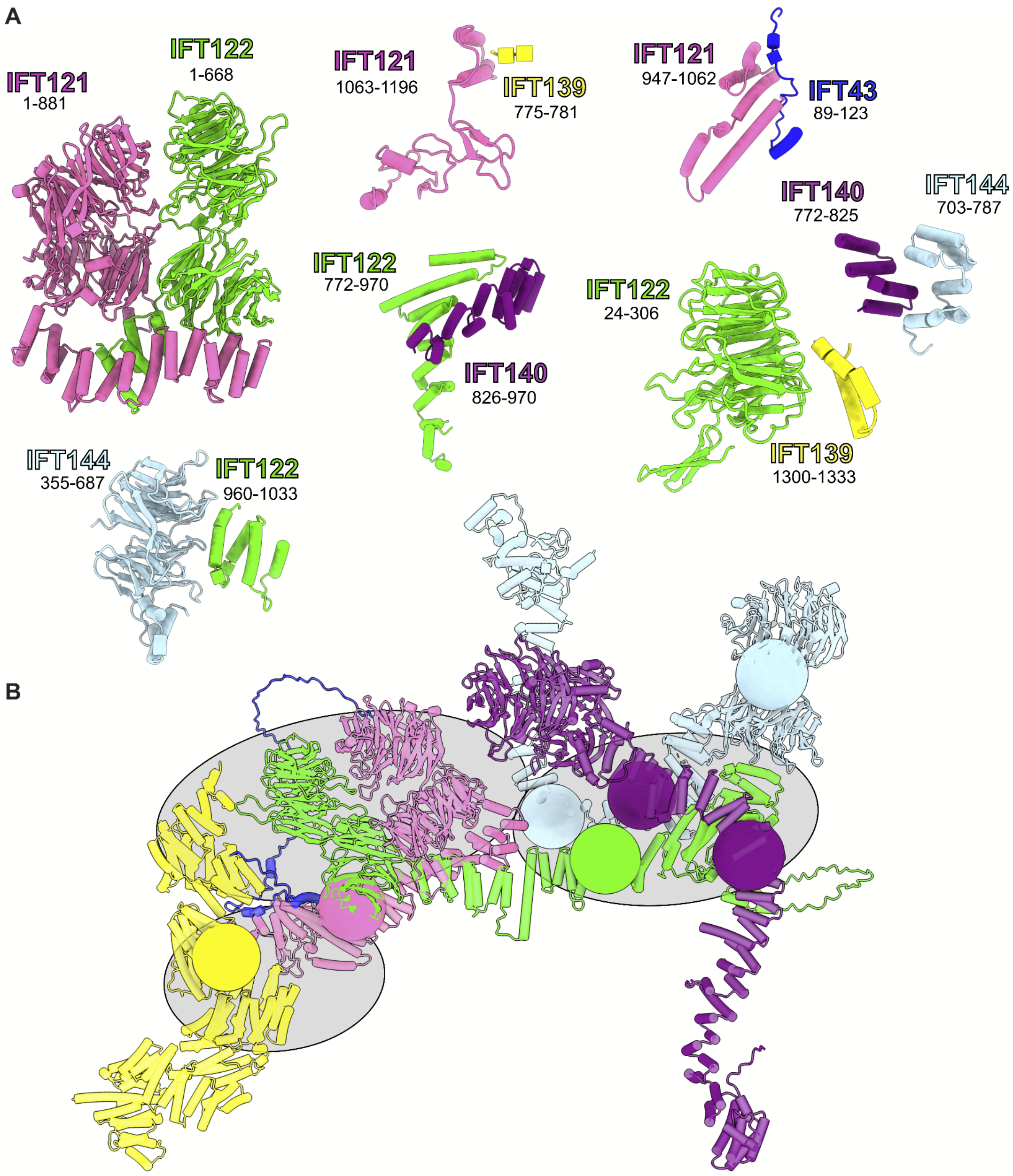


Figure 2—figure supplement 2. Summary of all AlphaFold-Multimer-predicted intermolecular interfaces used as rigid bodies in the model. (**A**) Rigid body protein segment pairs determined from AlphaFold-Multimer on the basis of high-confidence model PAE scores. A starting model for integrative modeling was assembled from these rigid bodies, as shown in (**B**). We restricted the number of degrees of freedom explored during integrative modeling by defining regions that were modeled as groups of rigid bodies (gray ellipses) as well as restricting flexibility to “pivot points” of low AlphaFold confidence (colored spheres) within each protein.

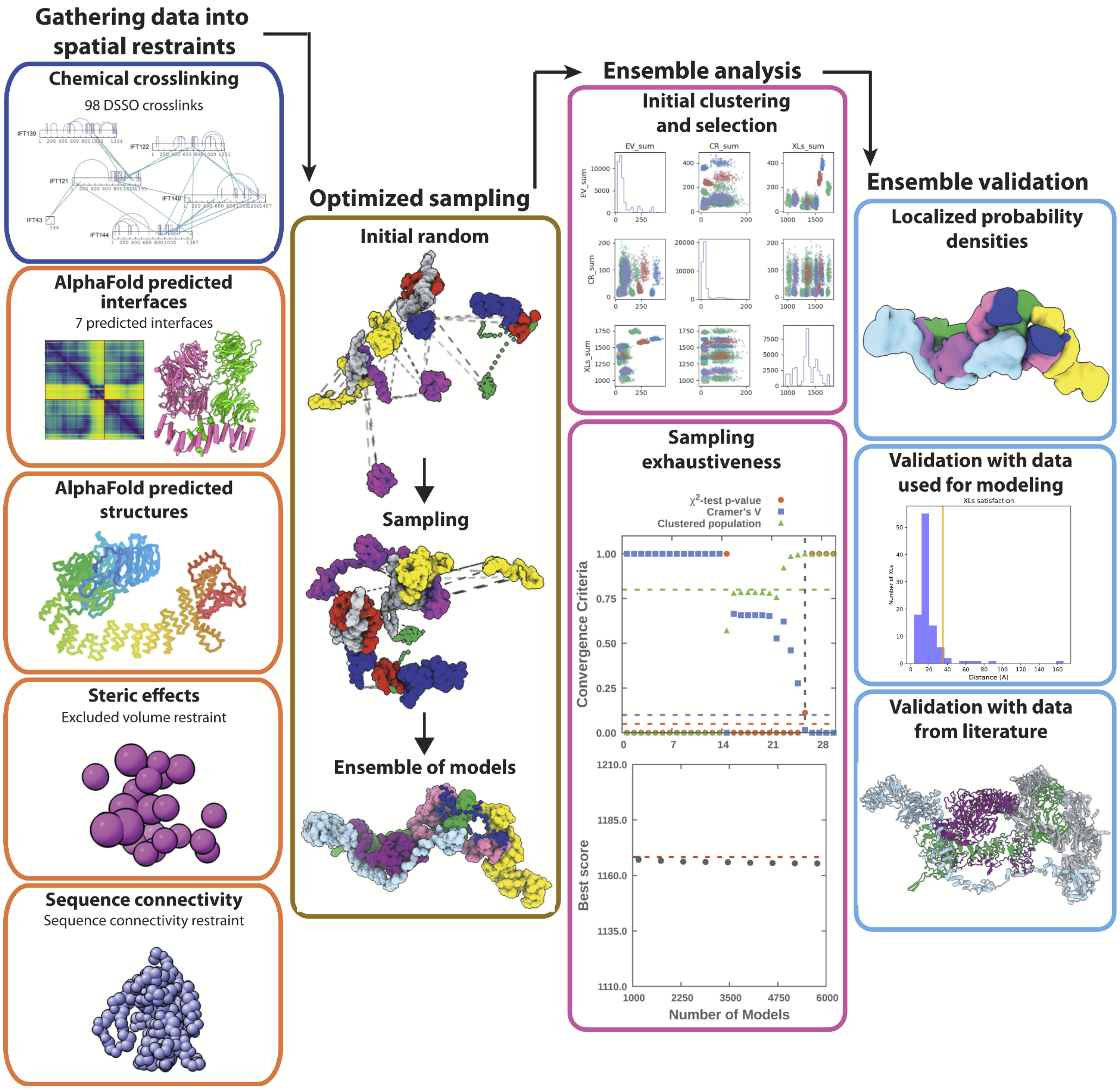


Figure 2—figure supplement 3. Integrative modeling scheme. The four step protocol of IMP (shown as consecutive columns of panels) includes 1) gathering data into spatial restraints, 2) optimized sampling, 3) ensemble analysis, and 4) ensemble validation. In 1) we show the restraints that were used in the model. In 2) we show the initial random positions followed by a Monte Carlo sampling. In 3) we show the ensemble analysis, which includes clustering models by score and testing for convergence of high-scoring models. Finally, in step 4) we validate the model both against data that was used in the model and using protein interactions previously reported in the literature.

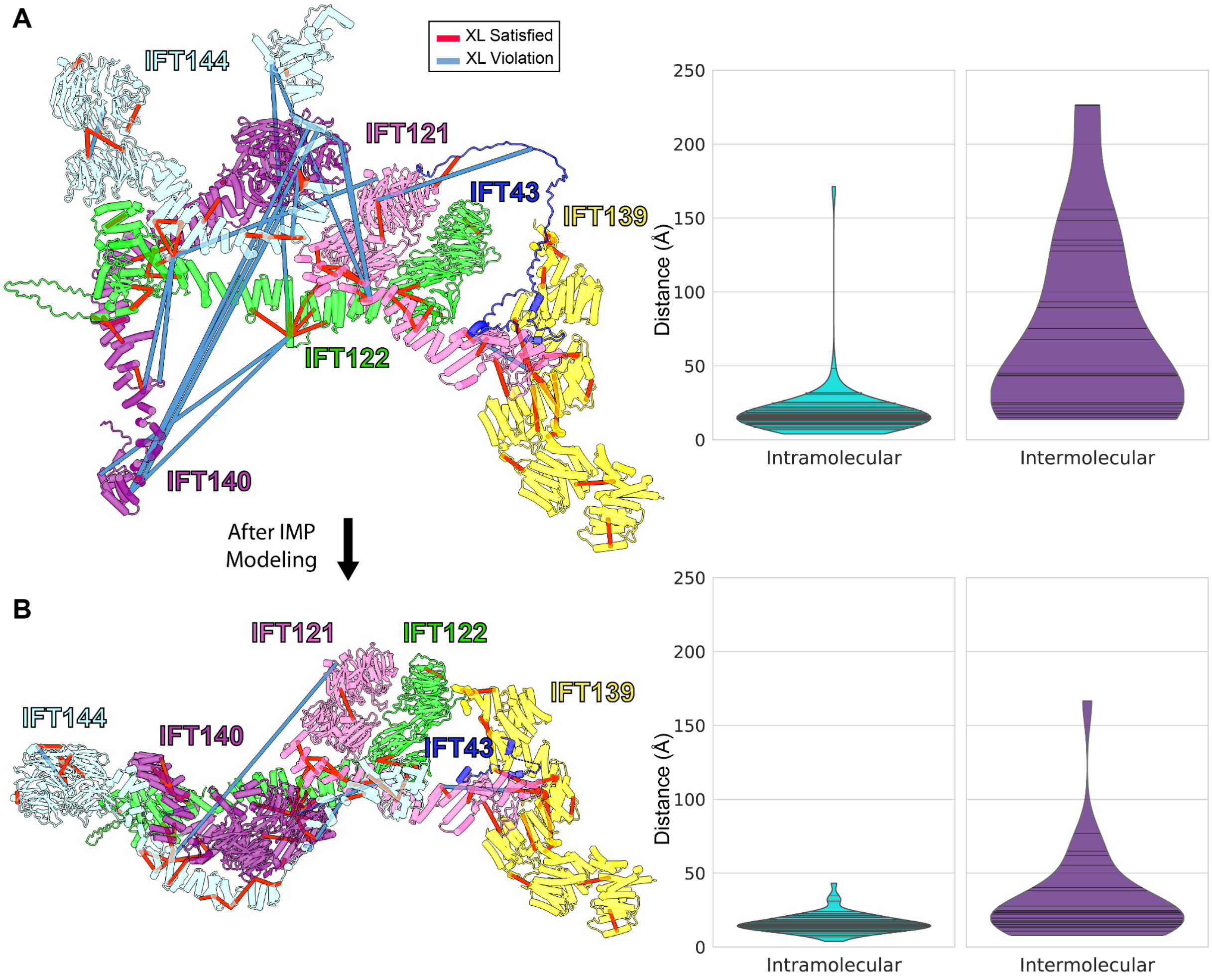


Figure 2—figure supplement 4. Initial (A) and final (B) configurations of the IFT-A monomer model. (**A**) The initial model built based on pairwise AlphaFold-Multimer predictions. The starting model showed numerous violations of intermolecular cross-linker distances (violin plots at right, showing distances between cross-linked amino acid residues in Ångstroms). (**B**) Integrative modeling using IMP markedly reduced the number of violations, producing a final model consistent with 98.6% of intramolecular and 72% of intermolecular cross-links, assessed as the percentage of cross-linked residue pair alpha carbons positioned within 35 Å of each other.

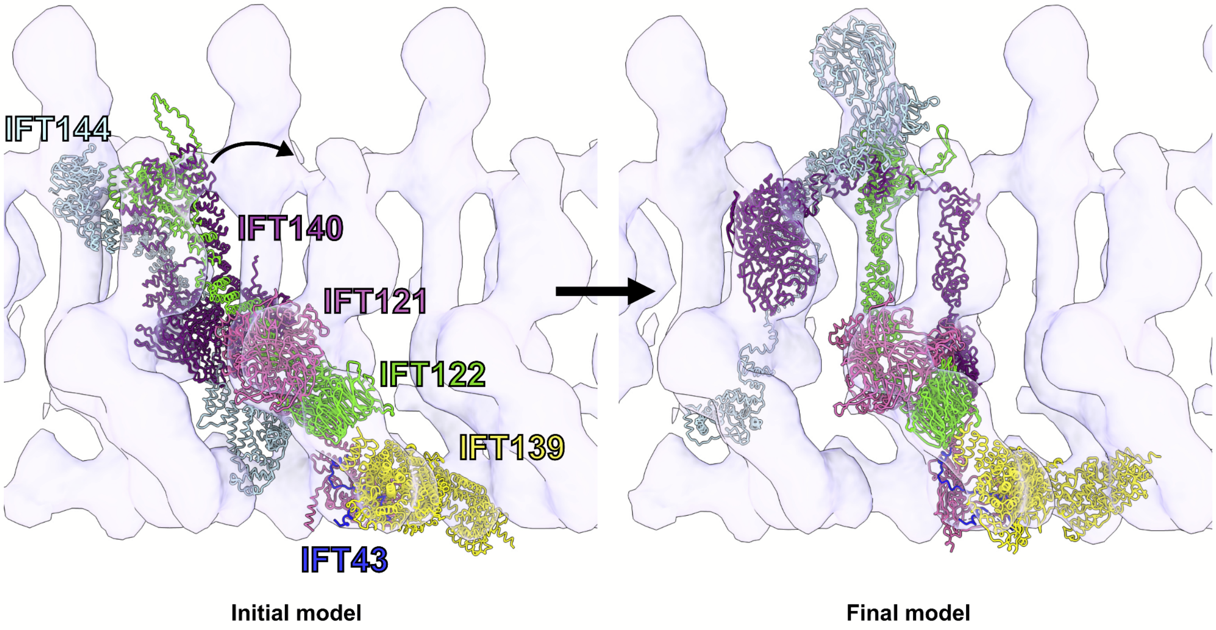
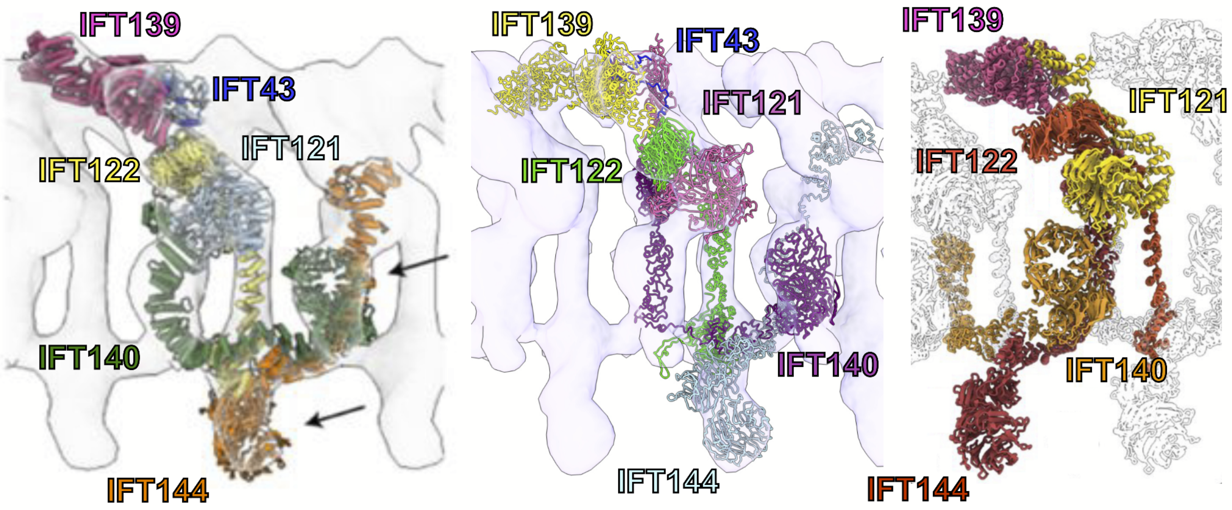


Figure 4—figure supplement 1. Flexible fitting of the IFT-A integrative model into the 23 Å subtomogram average of an anterograde IFT-A train. The starting position of the IFT-A integrative model (**left panel)** was determined by rigidly docking the model into the subtomogram average cryo-ET density using the ChimeraX fit-in-map tool (112). The final structure (**right panel**) was obtained by flexibly fitting into the density using Namdinator (63). **Figure 4—animation 1** provides a visualization of the IFT-A conformational changes required to fit the monomer into the polymeric train.



Hesketh, Mukhopadhyay *et al*., Fig. 5C **This study** **Lacey *et al*., Fig. 4C**

Figure 4—figure supplement 2. Independent 3D structural models of IFT-A derived by distinct workflows nonetheless compare favorably. A comparison of our model (**center panel**), based on determining a monomeric structure using cross-linking mass spectrometry, AlphaFold2, and integrative modeling, followed by docking into a 23 Å subtomogram average of polymeric IFT-A, compares favorably with (**left panel**, reprinting Fig. 5C of (94)) a monomeric structure determined by single particle cryo-EM then docked into a 30 Å subtomogram average of polymeric IFT-A , as well as with (**right panel**, reprinting Fig. 4C of (95)) a polymeric IFT-A model built into an 18 Å subtomogram average. Note the latter model omits IFT43 and the repeating IFT-A monomer adopts an alternative choice of neighboring IFT140/144 subunits, but otherwise shows similar packing of subunits within the cryo-ET density map. (Reprinted images are used with permission under CC-BY-NC-ND 4.0 International license and are unchanged apart from adding labels for clarity.)

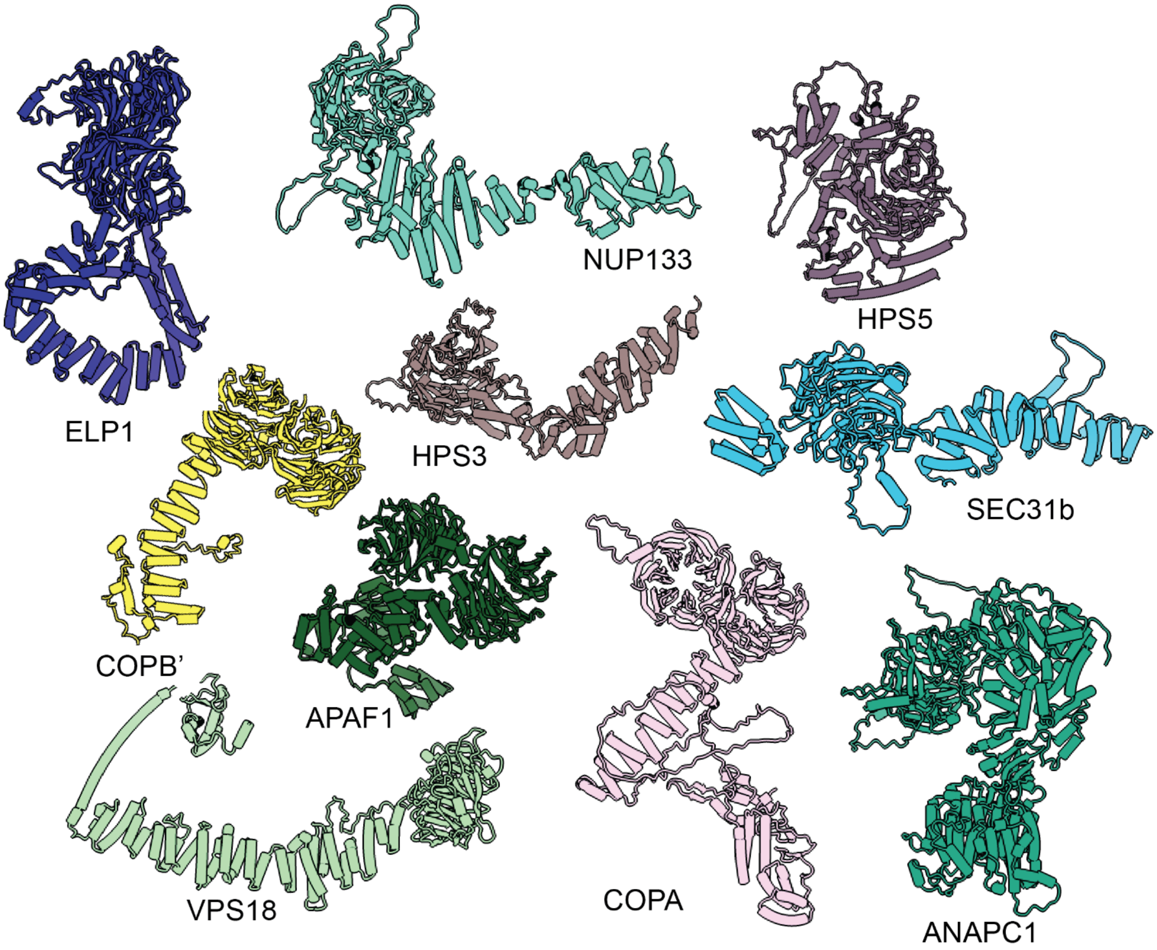


Figure 6—figure supplement 1. Proteins from the human proteome that are structurally similar to IFT121, 122, 140, and 144. IFT121, IFT122, IFT140, and IFT144 were individually searched against AlphaFold-predicted human protein structures (43) using DALI (71) to identify other proteins that are structurally similar and share domain architectures.

Table S1. Cryo-ET data collection and processing statistics

	IFT-A complex (EMD-26791)
Data collection	
Magnification	30,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	100 to 140
Defocus range (μm)	-3 to -6
Pixel size (Å)	14.13 (bin6), 7.08 (bin3)
Tilt-range/step (°)	±64° / 2
Processing	
Symmetry imposed	C1
Final particle images (no.)	9,350
Map resolution (Å)	23
FSC threshold	0.5