Supporting Information

INCLUDED IN THIS DOCUMENT

- **Supplementary Results**
- **Supplementary Methods**
- **Supplementary Figures 1 – 11.**

PROVIDED AS SEPARATE FILES

Supporting Dataset 1: Unique residue pairs (URPs), cross-linked peptides and cross-link spectral matches (CSMs) produced by the identification of DHSO and DSSO cross-links with XlinkX and DMTMM crosslinks with pLink 2 software.

Supporting Dataset 2: Summary statistics for selected large scale cross-linking studies of the human cell.

Supporting Dataset 3: Results of the mapping cross-links to experimental structures curated in the PDB database.

Supporting Dataset 4: Results of the mapping of intra-protein URPs to AlphaFold 2 monomer structural predictions.

Supporting Dataset 5: Cross-linked protein-protein interactions (PPIs) and their relationship to biological databases.

Supporting Dataset 6: The quality, size and results of mapping of inter-protein URPs to AlphaFold 2 structural predictions for dimeric cross-linked protein-protein interactions (PPIs).

Supporting Dataset 7: List of CORUM-annotated protein complexes with cross-linked protein-protein interactions (PPIs).

SUPPLEMENTARY RESULTS

The MIC60-MIC25-MIC19 hetero-tetramer has a 2:1:1 stoichiometry

We examined the seven-membered MICOS complex, for which we observed 12 URPs between five of the subunits. MICOS comprises MIC60, MIC13, MIC27, MIC25, MIC10, MIC26 and MIC19 (Uniprot: Q16891, Q5XKP0, Q6UXV4, Q9BRQ6, Q5TGZ0, Q9BUR5 and Q9NX63; CORUM #6255) and is essential for the proper formation and maintenance of crista junctions in the mitochondrial inner membrane (1). Despite its importance, the only existing structures are of distant fungal homologues of the coiled-coil domain of MIC60 (PDB: 7PUZ (2)) and of a complex between the C-terminal mitofilin domain of MIC60 and the CHCH domain of MIC19 (PDB: 7PV1 and 7PV0 (2)).

AlphaFold Multimer-v2 models of the full heptameric complex predicted that most of the sequences did not form 'traditional' globular folds **(Supp Figure 10)**. Nevertheless, these models shared a core architecture that place MIC60, MIC25 and MIC19 in the same general spatial arrangement **(Supp Figure 10)**, in agreement with published biochemical data (1). The interaction between the mitofilin domain of MIC60 and the CHCH domain of MIC19 was also recapitulated in two out of the five calculated models. The other four subunits (MIC10, MIC13, MIC26 and MIC27) did not form substantial interactions with this core. MIC26 and MIC27 are apolipoproteins and is it possible that their presence in the complex is lipid-dependent.

Given that (a) MIC60 has been reported to form homo-oligomers (2), (b) a crystal structure corroborates the predicted MIC60-MIC19 interaction (2), and (c) the predicted structures of MIC19 and MIC25 were highly similar, we also used AlphaFold Multimer-v2 to explore the possibility that these proteins form a higher order assembly. Of the many permutations that we tested, the most striking – and the one predicted with highest confidence – was a 2:1:1 MIC60-MIC19-MIC25 complex **(Supp Figure 11)**. This complex displayed uniformly high pLDDT scores across all subunits (average pLDDT = 77) (which were higher than those

observed for other assemblies tested; average pLDDT ≤ 70) and comprised a MIC60 homodimer bound symmetrically to MIC19 and MIC25 subunits via the MIC60 mitofilin domains and MIC19/MIC25 CHCH domains. The MIC60-MIC19 and MIC60-MIC25 interfaces closely resemble the MIC60-MIC19 crystal structure (PDB: 7PV0; RMSDs < 0.7 Å). Furthermore, all six inter-protein URPs and 12 (of 13) intra-protein URPs involving high-confidence residues are satisfied in this architecture **(Supp Figure 11)**, underscoring the orthogonal power provided by XL-MS data.

SUPPLEMENTARY METHODS

Subcellular fractionation

The general procedure for subcellular fractionation were similar as described in (3, 4). For each biological replicate, 2 × 10^8 freshly-harvested suspension HEK Expi293F™ (ThermoFisher Scientific) cells were used. Subcellular fractions were prepared by incubating the cell pellet with 5× the packed cell volume in swelling medium (15 mM KCl, 1.5 mM MgOAc, 10 mM HEPES-KOH, pH 7.5) for 10 min at 4°C. The cells were then centrifuged (300 × *g*, 2 min, 4°C) and the supernatant was discarded. The pellet was then homogenised by vortexing for 10 s, followed by repeatedly flushing the cells through a G-25 syringe. To this homogenate, 0.2 volumes of osmotic balancer buffer (375 mM KCl, 22.5 mM MgOAc, 1.25 M sucrose, 50 mM HEPES-KOH, pH 7.5) was added. The homogenate was then centrifuged (750 × *g*, 10 min, 4°C). This pellet consisted of the crude nuclei and was further treated as described below. The supernatant was transferred to a fresh tube and centrifuged at 3000 × *g*, 10 min, 4°C. This crude mitochondrial pellet was further treated as described below. The supernatant was again transferred to a fresh tube and centrifuged at 20,000 × *g*, 20 min, 4°C. The subsequent pellet, enriched with lysosomes, peroxisomes and Golgi membranes (henceforth described as the Golgi pellet for brevity), was further treated as described below. The supernatant should comprise of cytosolic and microsomal proteins. The crude nuclear pellet was washed in resuspension buffer 1 (10 mM HEPES-NaOH, 0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, pH 7.4) + 0.5% (v/v) IGEPAL CA-630 and re-centrifuged (750 × *g*, 10 min, 4°C), twice. The final pellet was resuspended in resuspension buffer 1 (without IGEPAL CA-630). The crude mitochondrial pellet was washed in resuspension buffer 2 (10 mM HEPES-NaOH, 0.2 mM EDTA, 0.25 M sucrose, pH 7.4) and re-centrifuged (3000 × *g*, 10 min, 4°C), twice. The final pellet was resuspended in resuspension buffer 2. The Golgi pellet was washed in resuspension buffer 2 (10 mM HEPES-NaOH, 0.2 mM EDTA, 0.25 M sucrose, pH 7.4) and re-centrifuged (3000 × *g*, 10 min, 4°C) once. The final pellet was resuspended in resuspension buffer 2. Protein content

of each subcellular fraction was quantified using a BCA protein assay kit (Pierce) and adjusted to 2 mg/mL prior to cross-linking.

Protein cross-linking

For each cross-linking experiment, 2–6 mg of protein from each organellar fraction was used. The concentration of each cross-linker used for cross-linking experiments were pre-determined empirically, as in (5). Disuccinimidyl sulfoxide (DSSO; 100 mM stock solution in anhydrous DMSO) was added to a final concentration of 5 mM for all subcellular fractions except for the cytoplasmic fraction (1 mM instead) and allowed to react for 1 h. Dihydrazide sulfoxide (DHSO (6); 100 mM stock solution in Milli-Q water) in combination with 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM; 200 mM stock solution in Milli-Q water) were added to final concentrations of 8 mM and 16 mM, respectively, and allowed to react for 1.5 h. Note: DMTMM is required for DHSO cross-linking, activating carboxyl groups in acidic residues to enable coupling with the hydrazide reactive groups in DHSO. DMTMM is itself also capable of directly catalysing carboxyl to primary amine coupling in amino acids residues to produce zero-length, but non-cleavable K-D/E cross-links. All reactions were incubated at 37°C.

Post-cross-linking, DSSO reactions were quenched with a final concentration of 100 mM NH_4HCO_3 at 37°C for 15 min, snap-frozen in liquid nitrogen and freeze-dried. As excess DHSO/DMTMM cannot be quenched, DHSO samples were chilled on ice, briefly sonicated to rupture organellar membranes, and 1 volume of ice-cold acetone was added. This mixture was then vortexed, and proteins were allowed to precipitate at -20°C for 2 h. Precipitated proteins were centrifuged at 20,000 × *g*, 15 min, 4°C. The supernatant was discarded, and the pellet was allowed to air-dry.

Protein digestion and peptide fractionation

Cross-linked sample trypsinisation and peptide size exclusion chromatography were performed essentially as described previously (7). Briefly, dried, cross-linked samples were resuspended in 8 M urea to give a final concentration of 5 mg/mL of protein. Sonication was employed to help re-solubilize the protein pellets. Samples were then reduced (10 mM DTT, 37°C, 30 min) and alkylated (15 mM iodoacetamide, 20 min, room temperature in the dark). The samples were then diluted to 4 M urea with 50 mM Tris-HCl pH 8 and Trypsin/Lys-C mix (Promega) was added to an enzyme:substrate ratio of 1:250 (w/w) and incubated at 37°C, 4 h. Following this, the samples were further diluted to 0.8 M urea with 50 mM Tris-HCl pH 8, additional Trypsin (Promega) was added at an enzyme:substrate ratio of 1:200 (w/w), and the sample was further incubated at 37°C overnight (16 h minimum). After the overnight digestion, the samples were acidified with formic acid to 2% (v/v) and centrifuged at 16,000 × *g* for 10 min. The supernatant was then desalted using either 500-mg or 1-g Sep-Pak tC18 cartridges (Waters), and eluted in 60:40:0.1 acetonitrile:water:formic acid (v/v/v), snap-frozen in liquid nitrogen and freeze-dried.

For size exclusion chromatography fractionation (SEC), the dried desalted peptides were resuspended at 2 mg of peptide per 250 µL of SEC mobile phase (acetonitrile:water:trifluoroacetic acid, 30:70:0.1 (v/v/v)) and separated on a Superdex Peptide HR 10/30 column. A maximum of 2 mg of peptide was injected onto the column per SEC run. A flow rate of 0.5 mL/min was used and the separation was monitored by UV absorption at 215, 254 and 280 nm. Fractions were collected as 0.5-mL fractions. Based on the UV absorption traces, fractions of interest (retention volumes ~9–13 mL) were pooled, snap-frozen and freeze-dried.

Following SEC, peptides were further fractionated using high-pH reverse phase liquid chromatography (bRPLC). Dried peptides from the SEC step were resuspended at 2 mg per 4.5 mL of bRPLC buffer A (3% (v/v) acetonitrile, 5 mM ammonium formate, pH 8.5), filtered

through 0.45 µm nylon filters and loaded onto a XBridge Peptide BEH C18 column (4.6 mm × 250 mm, 300 Å, 5 µm; Waters) using high pressure liquid chromatography (GBC LC1150) system. A maximum of 2 mg peptide was injected per run. Peptides were separated at a flow rate of 1 mL/min using a linear gradient of 3–40% bRPLC buffer B (80% (v/v) acetonitrile, 5 mM ammonium formate, pH 8.5) over 84 min, followed by a linear increase to 75% buffer B over 12 min. Peptides were monitored via UV (GBC LC1210) absorption at 215 and 280 nm. Ninety-six 1 mL fractions were collected across the whole bRPLC run. The collected fractions were then concatenated into 12 fractions by combining 8 fractions that are 12 fractions apart (*e.g.*, the first pooled fraction comprised of original fractions 1, 13, 25, 37, 49, 61, 73 and 85) (8). The pooled fractions were then snap-frozen and freeze-dried.

Mass spectrometry

Dried peptides were resuspended in 4% (v/v) acetonitrile, 0.1% (v/v) formic acid and loaded onto a 30 cm × 75 µm inner diameter column packed in-house with 1.9 µm C18AQ particles (Dr Maisch GmbH HPLC) using a Dionex UltiMate 3000 UHPLC (ThermoFisher Scientific). Peptides were separated using a linear gradient of 10–50% buffer B either over 81 min or 107 min at 300 nL/min at 55°C (buffer A consisted of 0.1% (v/v) formic acid, while buffer B was 80% (v/v) acetonitrile and 0.1% (v/v) formic acid).

Mass analyzes were performed using either an Orbitrap Fusion tribrid or a Q-Exactive HF-X mass spectrometer (ThermoFisher Scientific). On the Orbitrap Fusion tribrid mass spectrometer, the 81-min gradient above was employed and the CID-EThcD-MS2-HCD-MS3 protocol (9) was used. Specifically, following each full-scan MS1 at 60,000 resolution at 200 *m/z* (350 – 1400 *m/z*; 50 ms injection time), precursor ions were selected for sequential CID-EThcD-MS2 acquisitions in a data-dependent manner (CID-MS2: R = 30,000, 18 NCE; EThcD-MS2: R = 30,000, calibrated charged dependent ETD parameters enabled, supplemental HCD at 30 NCE, 54 ms injection time; both: 1.6 *m/z* isolation window, 5×10^4

intensity threshold, minimum charge state of +4, dynamic exclusion of 20 s). Subsequently, mass-difference-dependent HCD-MS3 acquisitions were triggered if a mass difference of (Δ = 31.9721 Da) was observed in the CID-MS2 spectrum (HCD, 30 NCE, 2 *m/z* isolation window; 5×10^3 intensity threshold; charge state of +2–4; ion trap scan rate = rapid). Total duty cycle time = 1 s. Additionally, to maximise the detection of DMTMM-mediated cross-linked peptides, DHSO/DMTMM samples were also re-analyzed on the Q-Exactive HF-X. For these analyzes, the 107-min gradient above was employed and the following MS protocol was used: Following each full-scan MS1 at 60,000 resolution at 200 *m/z* (350 – 1400 *m/z*, AGC = 3×10^6, 100 ms max injection time), up to 12 most abundant precursor ions were selected MS2 in a data-dependent manner (HCD, *R* = 15,000, AGC = 2×10^5, stepped NCE = (25, 30, 35), 25 ms max injection time, 1.4 *m/z* isolation window, minimum charge state of +4; dynamic exclusion of 20 s).

Identification of cross-linked peptides

DSSO and DHSO cross-linked peptides were identified using the XlinkX 2.0 (9) nodes as implemented in Proteome Discover v2.3 (ThermoFisher Scientific). The following key parameters were used: peptide mass between 300–10,000 Da, minimum peptide length of 5 residues, precursor mass tolerance ±10 ppm, product-ion mass tolerance of ±20 ppm for Orbitrap data and ±0.5 Da for ion trap data, allowable variable modification = oxidation (M), allowable static modification = carbamidomethyl (C), enzyme specificity of Trypsin with up to two missed cleavages (excluding the site of cross-linking), and FDR control of CSMs at 1%. The search database used was the UniProt human reference proteome (UP000005640; May 2020; 20,286 entries). For DSSO cross-links, the settings were as follows: allowable cross-linking sites were Lys, Ser, Tyr and Thr (or Lys-Lys only in pilot searches), cross-link mass-shift 158.0038 Da, cross-link fragment mass on cleavage = 54.0106 Da (alkene) and 85.9826 Da (thiol). For DHSO cross-links, the settings were as follows: allowable crosslinking sites were Asp and Glu, crosslink mass-shift 186.0575 Da, crosslink fragment

mass on cleavage = 68.0375 Da (alkene) and 100.0095 Da (thiol). XlinkX score titrations were performed using the method described in (10) considering Lys-Lys linkages only, and analysing spectra from a smaller pilot experiment of DSSO cross-links (one biological replicate, four organellar fractions, 56 HPLC fractions – a total of 5,916 CSMs identified at default XlinkX settings). Three score combinations were assessed, where D = delta XlinkX score and S = XlinkX score: D4S40 (default), D10S60 (from (10)) and D20S80. Note that XlinkX FDR is approximated using the equation DD/(TT + DD) as XlinkX 2.3 does not report hybrid TD matches, where D = decoy database match and T = target database match. This can lead to inappropriate FDR control. The XlinkX score titration revealed that default score cut-offs (XlinkX score = 40, delta = 4; abbreviated as D4S40) produced inter-links with an inflated FDR even at the CSM level at which they are controlled, while D20S80 could produce CSMs of equivalent quality for known and novel protein-protein interactions, and control the false discovery rate to <1% at all levels of redundancy **(Supplementary Figure 1A and 1B)**. The D20S80 search setting was used for DSSO and DHSO spectra in the final dataset. Due to computational constraints, searches were done in batches of organelles and biological replicates. For DSSO, only Lys-Lys linkages were allowed in a first pass search against the whole human reference proteome (as described above), whilst in a second pass, all linkages (Lys, Ser, Tyr and Thr) were permitted using a reduced search database consisting of only the proteins identified in the first pass. The identifications resulting from these DSSO searches (D20S80, with Lys, Ser, Tyr and Thr specificity) were used for the final dataset.

DMTMM crosslinked peptides were identified using pLINK2 v2.3.9 (11). Key pLink 2 search parameters were as follows: Peptide mass between 600–10,000 Da and peptide length between 6–100 were considered, precursor mass tolerance ± 15 ppm, product-ion mass tolerance ± 20 ppm, variable modification = oxidation (M), fixed modification = carbamidomethyl (C), enzyme specificity of trypsin with up to two missed cleavages (excluding the site of crosslinking) per chain, and a 1% FDR. DMTMM cross-linker settings:

cross-linking sites used in the final dataset were Asp, Glu and protein C-terminus to Lys and protein N-terminus, crosslink mass-shift –18.0106 Da. The search database used was the UniProt human reference proteome (UP000005640; May 2020; 20,286 entries). Decoy hits for decoy analysis were extracted from the unfiltered identification list table (the .csv output table with no suffix) by filtering for rows with the following attributes: Q.value <= 0.01, Peptide_Type = 3, Target_Decoy != 2 and isFilterIn = 1. We also considered the possibility of misidentified cross-linked peptides as co-eluted linear peptides (12) but did not find significant evidence of this occurring.

Cross-link post-hoc filtering and consolidation

The R programming language (v4.1.3) was used to filter and consolidate DSSO, DHSO and DMTMM cross-link spectral matches (CSMs) across replicates and software outputs (XlinkX and pLink 2), and collapse redundant identifications to the levels of unique residue pairs (URPs) and protein-protein interactions (PPIs). For post-hoc filtering steps, decoy hits from XlinkX identifications of DSSO and DHSO data were removed. For pLink 2 identifications of DMTMM data, CSMs with a pLink score ≤ 0.34 were also filtered out to ensure appropriate FDR control at higher levels of redundancy (URP and PPI levels) **(Supplementary Figure 1C)**.

To consolidate filtered cross-links across software suites and replicates, the identity of each linked peptide's sequence, the (peptide-based) site of cross-linker modification, and the linked amino acid type was parsed for each CSM from each software output. This enabled removing variability introduced across multiple pLink 2 and XlinkX search engine runs, where the assignment of master protein accessions (for peptide sequences redundantly mapped to the protein sequence database) can sometimes differ. Specifically, for each peptide sequence within the cross-link, the first protein accession (alphabetically) which mapped to the peptide was reassigned as the master protein for that sequence. This

therefore standardised the protein(s) described by the CSM. The protein-level position of the cross-link modification in each peptide were then recalculated for use in the URP. If the two peptides could be mapped to the same protein sequence (an intra-protein cross-link), the first shared mapped protein accession was chosen to describe both proteins in the pair, with the protein-level positions recalculated accordingly for use in the URP. Homodimer URPs were defined when the sequences in an intra-protein cross-linked peptide overlapped.

Biological annotations of proteins and protein-protein interactions

To annotate protein subcellular localisations, data was extracted from annotations of the human proteome hosted on the UniProt database (13). "Residents" of the isolated organellar fractions were those with relevant annotations for the "Subcellular Location" field or the cellular component Gene Ontology (14) term field were assessed for the presence of the following case-insensitive substrings; for the nuclear fraction (post-750 *g* pellet) – "nuclear", "nucleus", "histone", "chromatin", "nucleolus", "nucleolar", "nucleoplasm"; for the mitochondrial fraction (post-3,000 *g* pellet) – "mitochondria", "mitochondrial", "mitochondrion", "mitoribosome"; for the Golgi (lysosome, perioxosome) fraction (post-20,000 *g* pellet) – ribosome", "golgi", "endoplasmic", "reticulum", "lysosome", "peroxisome"; for the cytosol (microsome) fraction (supernatant post-20,000 *g*) – "cytosol", "cytoplasm", "microsome".

To investigate protein abundances in the cross-linked proteome, the PaxDB protein abundance database (15) was used (downloaded August 2021). To annotate protein disorder, the MobiDB protein disorder database (16) (downloaded April 2022) was used. The aggregated majority consensus stringency was used ("prediction-disorder-th_50") to determine whether the residues fell in disordered regions, which means at least 50% of the tested disorder predictor programs agreed on the annotation of the region as disordered.

To annotate the presence and density of PTMs within the sequences of cross-linked proteins, the PhosphoSitePlus database (17) (downloaded April 2022) of experimentally detected protein post-translational modifications (PTMs) was used.

The Agile Protein Interactomes DataServer (APID) database (18) (version 9606_noISI_Q3) of experimentally-derived protein-protein interactions (PPIs) was used to annotate the novelty and types of existing experimental evidence for cross-linked PPIs. The distinction of experimental interaction mapping techniques as either “binary” or “indirect” is used throughout this study and is based on the previous manual curations of experimental evidence codes defined by the original APID study and therefore annotated in their database (18). The resource was also used to determine the number of common interactors between two given proteins. The STRING-db resource (19) (v11.5) was used to determine the degree of predicted functional association between two proteins for random and cross-linked PPIs. Protein pairs without an association in the database were assigned a “combined score” of 0. Only those with a “combined score” of at least 0.4 were considered functionally predicted. To simulate a random population of protein pairs, 300 protein pairs were randomly sampled from the cross-linked proteome. The CORUM database of protein complexes (20) (2018 release) was used to determine shared membership of a cross-linked PPI in a known protein complex. Only the first CORUM entry was considered in cases where the exact same group of proteins (including those that were not cross-linked) were annotated together in multiple complex entries.

Protein structures

The Structure Integration with Function, Taxonomy and Sequence (SIFTS) database (21) (downloaded April 2022) was used to annotate experimental coverage of protein structures curated in the RCSB Protein Data Bank (22). This resource was also used to identify co-crystallized protein-protein interactions (two UniProt accession IDs present in the same PDB

entry), and to determine the relevant PDB entries to download for cross-link mapping exercises. Structure files of interest (.pdb, asymmetric units) were downloaded in batch from the PDB database (22) website. For some large entries, the structures were only available as .cif files (due to the limit in number of atomic coordinates capable of being stored in the .pdb format). For these entries, the two linked chains of interest were extracted from the .cif files and then converted into .pdb format using the functions within the Bio3D R-package (v2.4-3) (23).

All AlphaFold predictions of monomeric protein structures used in this study were precomputed previously (24), and downloaded from the EBI AlphaFold Protein Structure Database (AlphaFold DB) (25) (v1 release). Only proteins which had models where their entire sequence length was considered in a single modelling run were analyzed further (proteins < 2,700 amino acids).

Cross-linked PPI dimers (with at least 2 URPs) and random PPIs (sampled from the cross-linked proteome) were subjected to AlphaFold Multimer v2 modelling analyzes. Only PPIs with a combined amino acid length <2,000 residues were chosen to be modelled, as this approximate cut-off ensured completion of all computational steps on our high-performance computing hardware set-up (Nvidia Tesla Volta V100-SXM2-32GB GPUs on the Gadi supercomputer (National Computational Infrastructure, Australia). Modelling was performed by supplying canonical protein sequences for each complex member as the input for Colabfold (v1.2.0) (26), which uses MMSeqs2 (27) instead of JackHMMER (28) for accelerated MSA creation and implements AlphaFold-Multimer 2 (v2.2.0) (29). Colabfold was executed locally with the options --model-type AlphaFold2-multimer-v2 --recompile-all-models. The modelling with Colabfold was run on Nvidia Tesla Volta V100-SXM2-32GB GPUs on the Gadi supercomputer (National Computational Infrastructure, Australia).

For complexes with more than two members, the Google Collaboratory notebook version of ColabFold (v1.3.0) was used. The complexes were the pentameric tRNA ligase complex

comprising DDX1, FAM98B, RTRAF, RTCB and ASHWIN (Uniprot: Q92499, Q52LJ0, Q9Y224, Q9Y3I0 and Q9BVC5) and the heptameric MICOS complex comprising MIC60, MIC13, MIC27, MIC25, MIC10, MIC26 and MIC19 (Uniprot: Q16891, Q5XKP0, Q6UXV4, Q9BRQ6, Q5TGZ0, Q9BUR5 and Q9NX63). For the tRNA ligase complex, two runs separate runs were performed comprising: (i) one copy each of all five full-length proteins; (ii) one copy each of full-length proteins of DDX1, FAM98B, RTRAF and RTCB. For the MICOS complex, four runs were performed comprising: (i) one copy each of all seven full-length proteins; (ii) four copies of MIC60 residues (181–758); (iii) two copies of MIC60 residues (181–758), one copy of MIC19 residues (65–227) and one copy of MIC25 residues (110–235); (iv) two copies each of MIC60 residues (181–758), MIC19 residues (65–227) and MIC25 residues (110–235).

The command-line implementation of the Protein Interfaces, Surfaces and Assemblies (PISA) tool (v2.1.2) (30), as implemented in the CCP4 software suite (v4-7.1) (31), was used to calculate features of the structural interfaces predicted by the AlphaFold2 modelling. These included the number of detected inter-chain interfaces, the types and strength of chemical bonds, and the size and residues involved in the interface area.

To calculate the degree of atomic clashes in AlphaFold2 structural prediction models, the clashscore was calculated using MolProbity (32) as implemented in Phenix (v1.20.1-4487) (33), using the options o_flips=TRUE, coot=FALSE, probe_dots=FALSE. This was performed using the high-performance computing cluster hosted by the University of New South Wales (Katana). The pLDDT scores of individual residues in AlphaFold monomer models, and additionally the interface residues in dimer models identified by PISA, was accessed by extracting the b-factor variable in the .pdb file format using the Bio3d R-package (v2.4-3) (23). All protein structures were visualized using the PyMOL software (v2.5.2).

Cross-link structure mapping

The Xwalk command line executable program (v0.6) (34) was used (with the -dist, -bb, -f and -euc flag options) to determine the Euclidean distances (alpha-carbon to alpha-carbon) between cross-linked residues on experimental or predicted protein structures previously downloaded (or converted into) .pdb files. For experimental structures, random URPs were generated for each PDB entry by sampling fifteen random but theoretically valid residue pairs (5 per cross-linker; KSTY-KSTY, K-DE, DE-DE) from the protein sequences present in the given PDB. For predicted monomeric structures, these were sampled from the modelled protein sequence. For predicted dimer structures, three random URPs (one for each cross-linker reactivity pair) were sampled where one residue was sampled from the first protein sequence and the other from the second. Xwalk input files were then generated for each structure containing both the experimental and random URPs. For experimental PDB files, this involved accounting for any PDB entry-specific differences in the residue index from that reported in the canonical protein sequences used for cross-link identification. This was achieved by applying a numeric adjustment to the cross-linked residue number by reference to the annotations in the SIFTS resource (21). In the case of chain ambiguity within multimer PDB structures, URPs were redundantly described so that each theoretically possible combination of chain pairs could be considered. For analyzes where the span of distances across multiple PDB entries was calculated, only structures with unique chain mappings (no multimeric chains of the same protein) were used.

Cross-links were visualized on structures in Pymol (version 2.5.2), using either the automated .pml scripts generated by the Xwalk command line tool (version 0.6, using the flags -dist, -bb, -f, -euc, and -pymol) (34) or at a smaller scale, using the PyXlinkViewer PyMOL plugin (35).

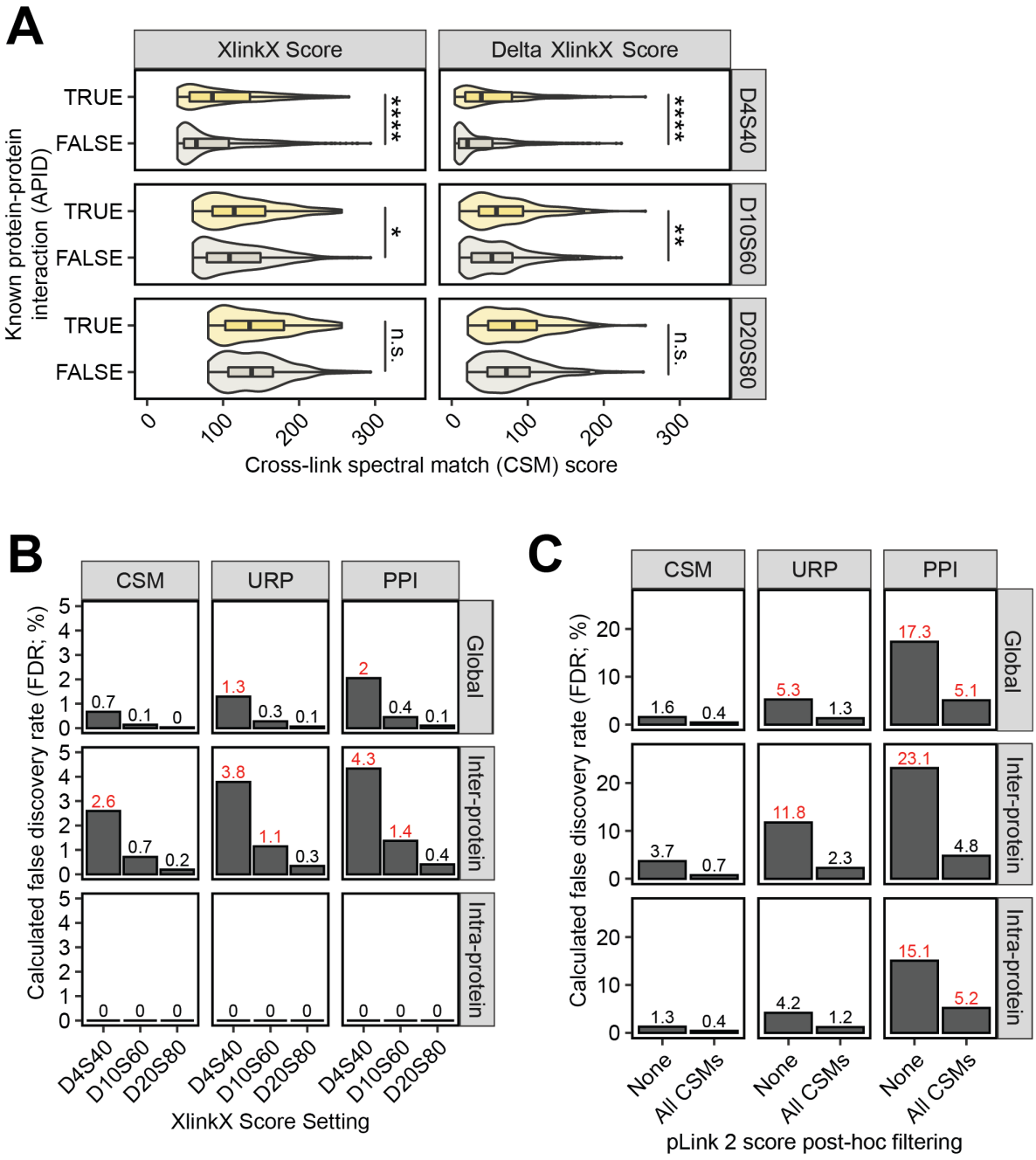
Homology search and structural alignments to PDB structures containing homologous proteins

To identify experimental structures existing for protein sequences homologous to cross-linked proteins, protein sequences were subjected to BLAST analysis (36) to search for homologous sequences. The protein sequences were downloaded in FASTA format by batch retrieval of UniProtKB IDs from the UniProt website. Either the webserver, or a command-line implementation of the NCBI BLAST+ blastp program constructed using the functions within the biopython Python-package (v1.79), was used to query these sequences against a locally downloaded 'pdbaa' database (retrieved from the NCBI FTP server). Protein homologues were defined to be any matches from this BLAST search pertaining to an e-value < 1×10^{-50}. A set of all PDB structure identifiers containing a homologue was obtained for each protein. For protein pairs, the intersection between the individual sets of homologue-containing structures were taken for further analysis.

Structural alignment was performed for each protein pair that had homologue-containing PDB structures. For each pair, the structure with the greatest homology, as determined by the smallest e-values from the BLAST search, was obtained in .cif file format using functions from the biopython Python-package (v1.79) to access the wwPDB API. The relevant chains in the PDB entry, as specified in the BLAST homology search, were chosen and aligned to the AF Multimer structure using functions from the pymol Python-package (v2.5.4). Structural alignments were stored in .pdb format and RMSD values from the alignment were saved.

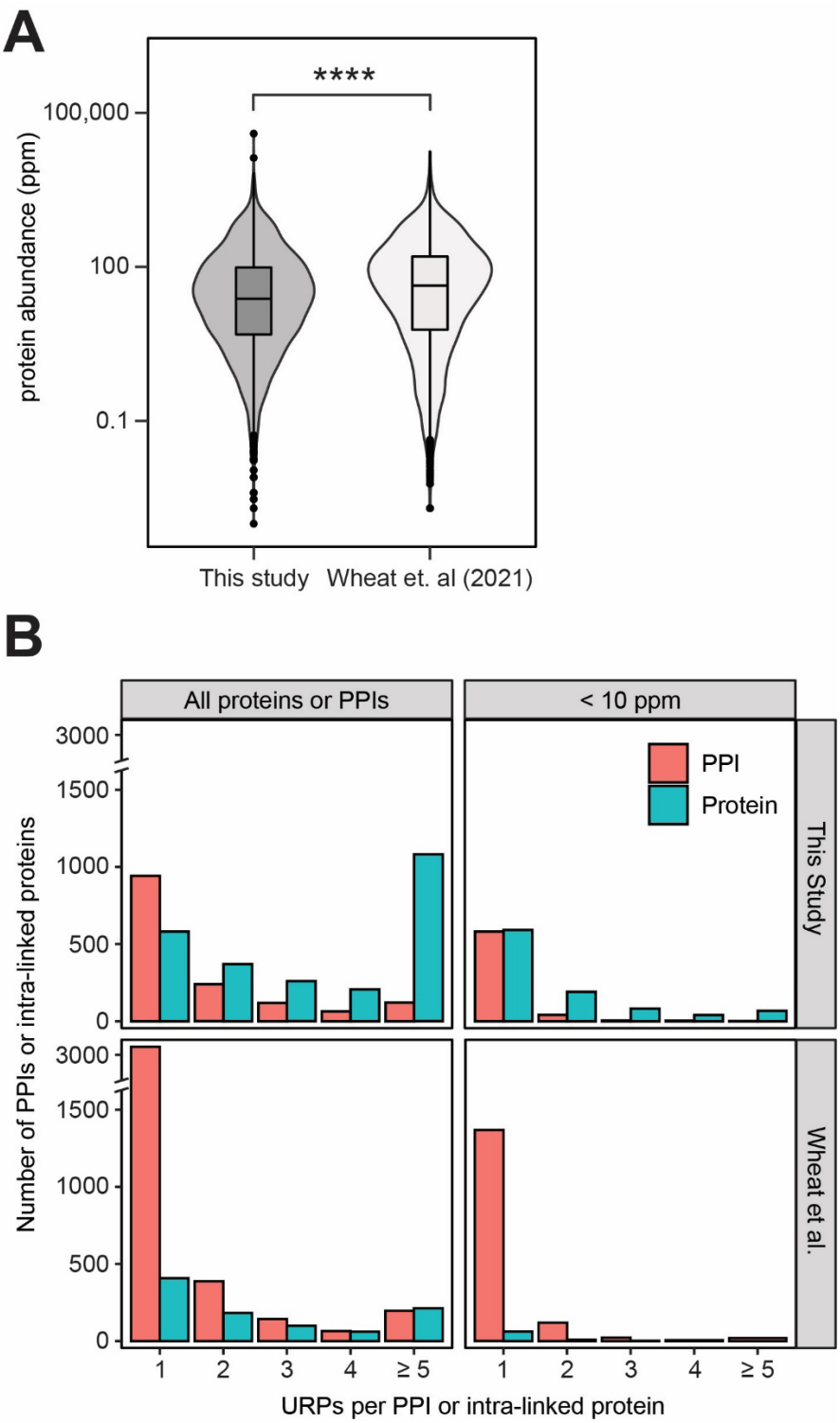
Statistical Analyses

Data manipulation and statistical analyses were performed within R (v4.1.3) using RStudio, using the inbuilt statistical functions such as the Wilcoxon signed rank test with continuity correction "wilcox.test".



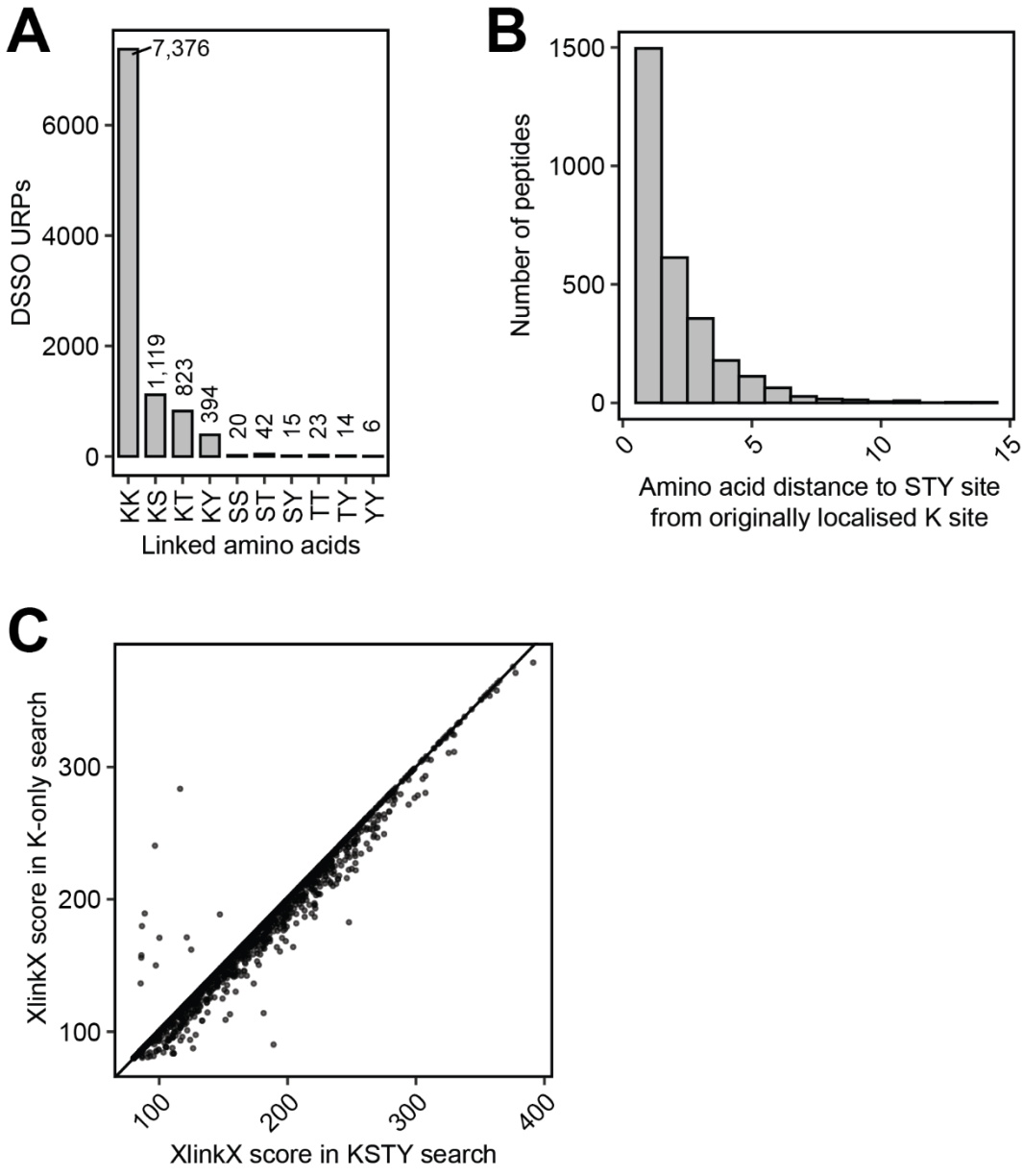
Supplementary Figure 1: False discovery rate control measures to reduce incorrect identifications and propagation of error. (A) and **(B)** show the results of XlinkX search settings titrated on a small pilot dataset, performed by varying XlinkX score cut-off settings in the XlinkX Search node. These search settings enable filtering of matched spectra (and importantly, individual peptide identifications within the cross-link) before FDR control by the XlinkX 'validator' node, which was set to 1% at the cross-link spectral match (CSM) level. The scores used were D4S40 (default settings, minimum delta XlinkX score of 4, minimum XlinkX

score 40), D10S60 (as used in (10), minimum delta XlinkX score of 10, minimum XlinkX score 60) and D20S80 (as used for our final dataset, minimum delta XlinkX score of 20, minimum XlinkX score 80). **(A)** The distribution of inter-protein CSM XlinkX scores stratified by protein-interaction novelty as annotated by the APID PPI meta-database. This revealed that D20S80 produced CSMs of equivalent quality between known and novel protein-protein interactions (PPIs). Asterisks show results of Wilcoxon rank sum tests with continuity corrections, where **** is $p < 0.0001$, ** is $p < 0.01$, * is $p \leq 0.05$ and 'n.s.' is $p > 0.05$. **(B)**, **(C)** Calculated false discovery rate (FDR) at different levels of biological interest (URP-level for intra-protein, and PPI-level for inter-protein links). Decoy cross-linked peptide sequences were compared to determine whether they originated from the same reverse protein sequence (intra) or two different sequences (inter) and used to calculate global, inter-link and intra-link false discovery rates at the CSM, URP and protein-pair level. **(B)** XlinkX score titration from the pilot datasets show that the D20S80 search setting was effective in controlling the FDR to < 1% at all levels of redundancy. **(C)** Post-hoc filtered pLink 2 DMTMM identifications from the full dataset. Note: although pLink 2 is reportedly quite effective at controlling FDR, we performed post-hoc score filtering to further control for the propagation of error at the URP and PPI-levels. CSMs were left as reported by pLink 2 (and after in-built FDR control) "None", or filtered for having pLink scores ≤ 0.34 in inter-protein CSMs only or across all CSMs. The filtering strategy for pLink 2 CSMs chosen for our final dataset was to post-hoc filter only the inter-protein CSMs for scores ≤ 0.34, as this filtered intra-protein URPs and inter-protein protein-pairs to < 5% FDR.



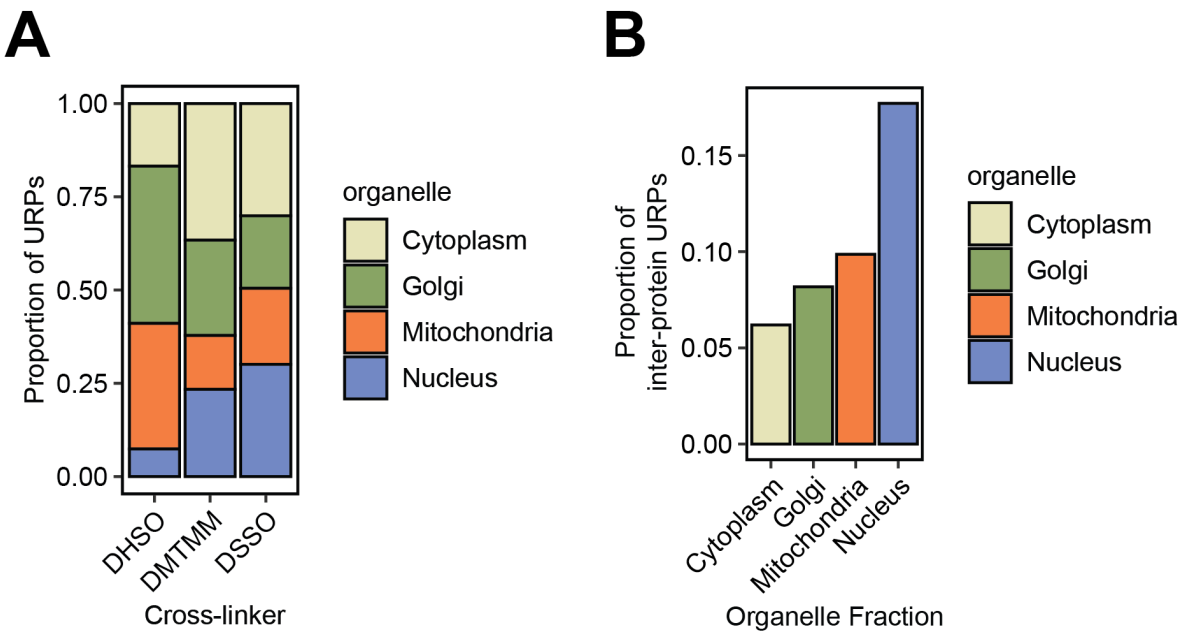
Supplementary Figure 2. High density and deep coverage of the cross-linked proteome. **(A)** PaxDB-annotated abundances for cross-linked proteins in this study and those identified in (37). **** = $p < 2.2 \times 10^{-16}$ in a Wilcoxon rank sum test with continuity correction. **(B)** Histogram of URPs summarised per protein-protein interaction (inter-protein or homo-oligomeric URPs; *orange bars*), or per unique intra-linked protein (intra-protein URPs;

turquoise bars). Histograms of cross-linking densities are shown for all cross-linked proteins and protein-protein interactions, and also for the subset with at least one low-abundance protein (annotated to have a PaxDB abundance lower than 10 ppm). Density statistics for cross-links reported in Wheat et al. (37) are also shown.

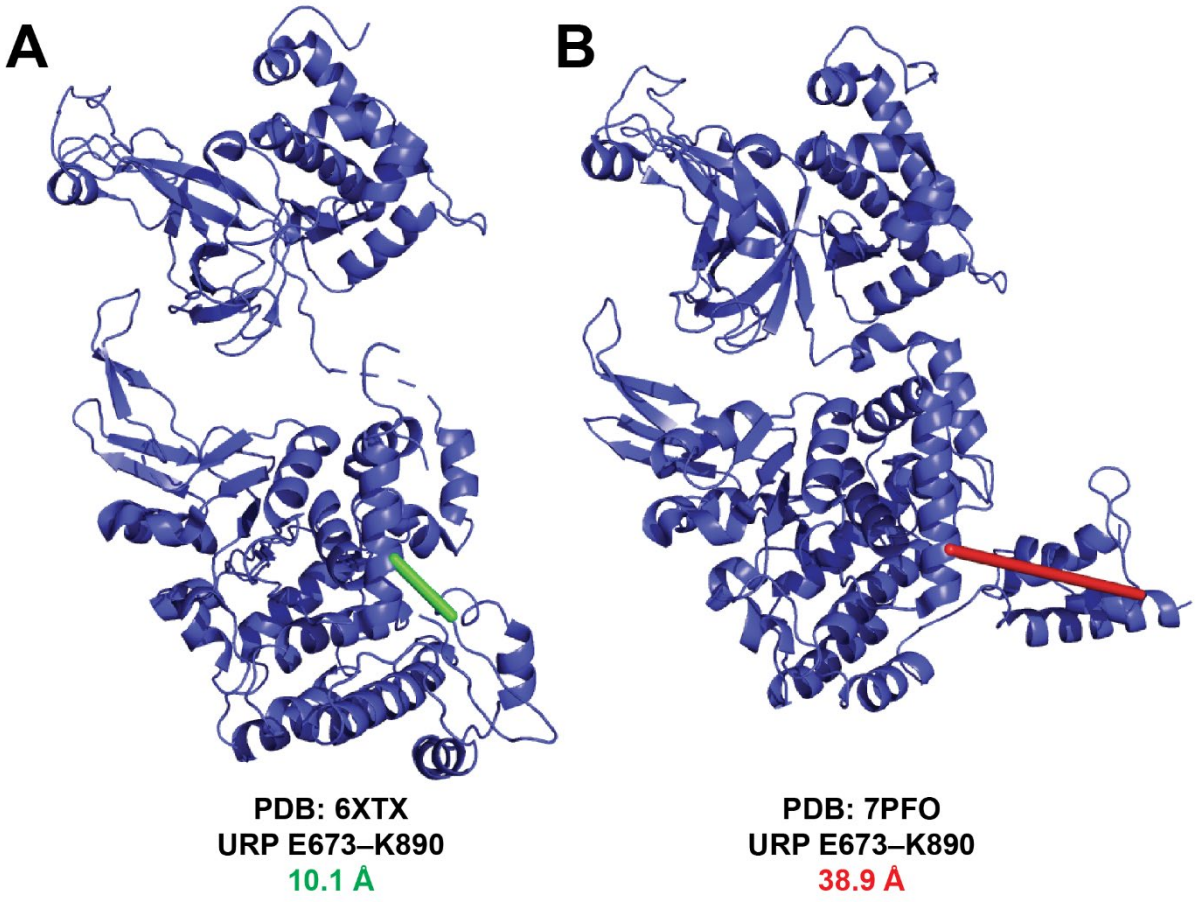


Supp

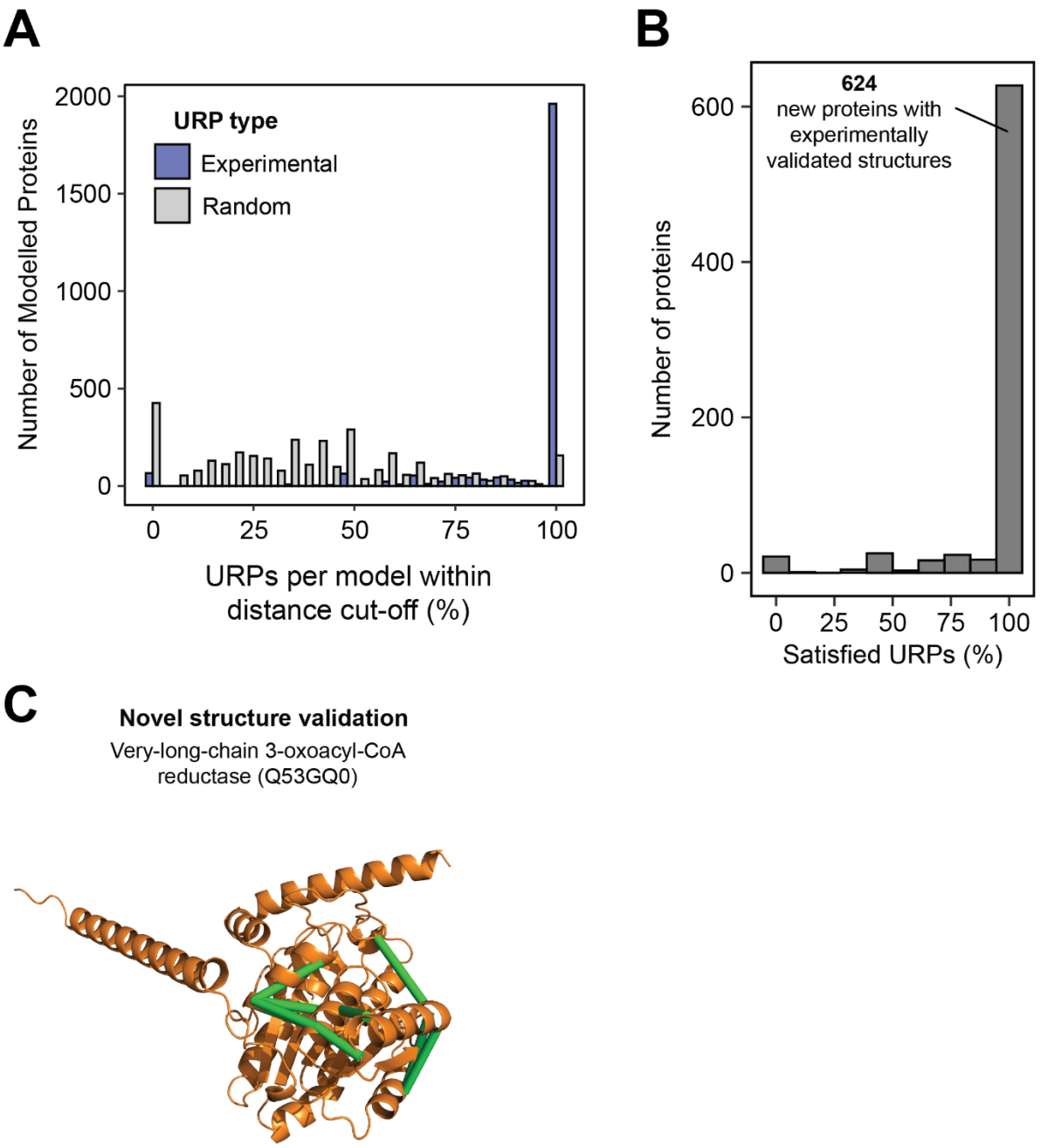
Supplementary Figure 3. Consideration of DSSO S/T/Y reactivity improves quality of cross-links. (A) The breakdown of amino-acid pairs in the final DSSO dataset. **(B)** Distance to nearest lysine-localized residue from K-K searches for peptides where the site of cross-link modification was reassigned as a S/T/Y residue. **(C)** The comparison of cross-link spectral quality scores for matched spectra containing peptides with differently localized cross-linker modified residues between searches are shown for DSSO cross-links identified with XlinkX.



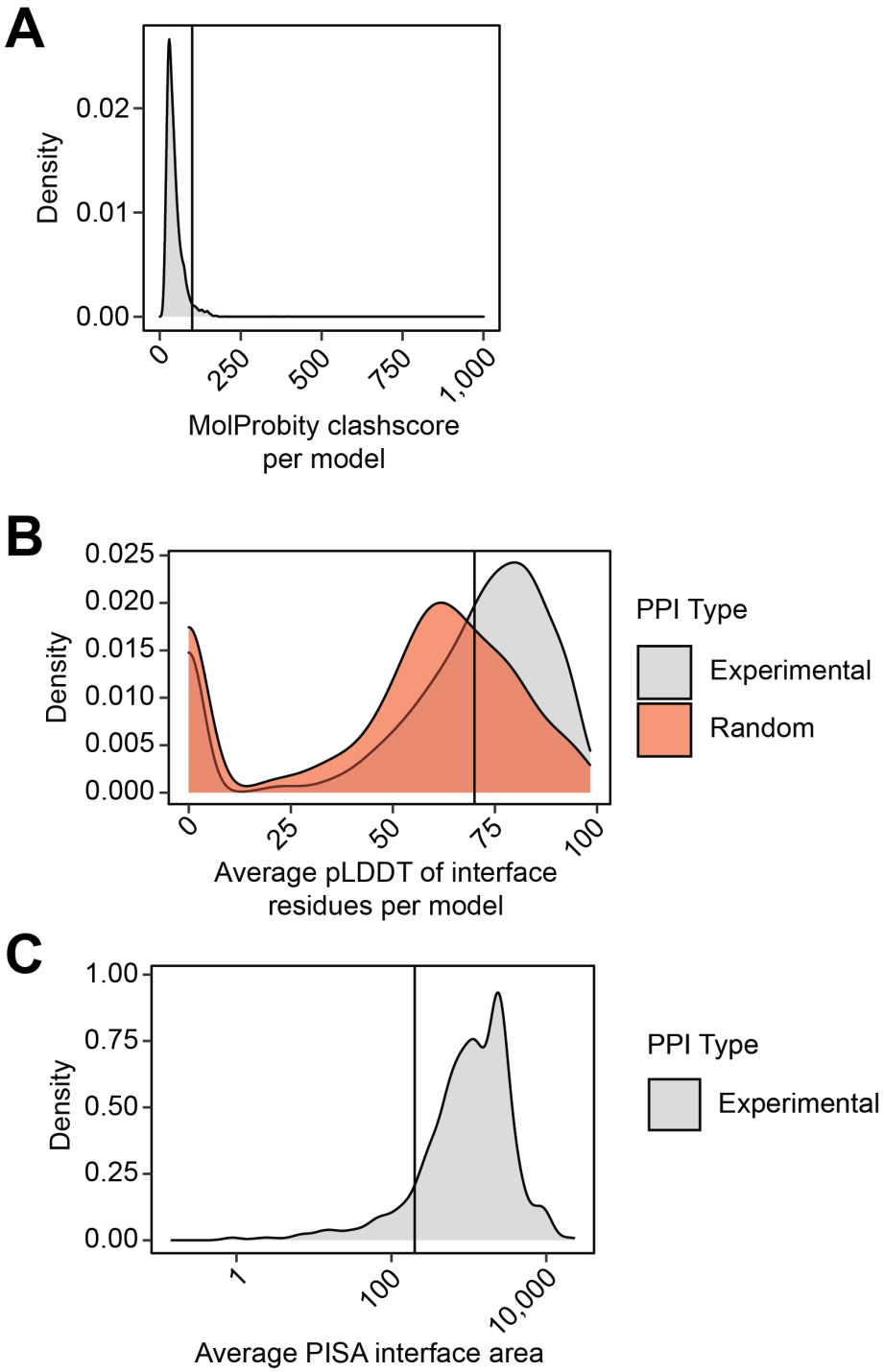
Supplementary Figure 4. Effectiveness of each cross-linker in profiling different subcellular niches. (A) The proportions of URPs identified from each organellar fraction differ between cross-linkers types. **(B)** The organelle fractions themselves differ in their proportion of identified inter-protein URPs. The definition of inter-protein in this graph includes unambiguous homo-dimer URPs.



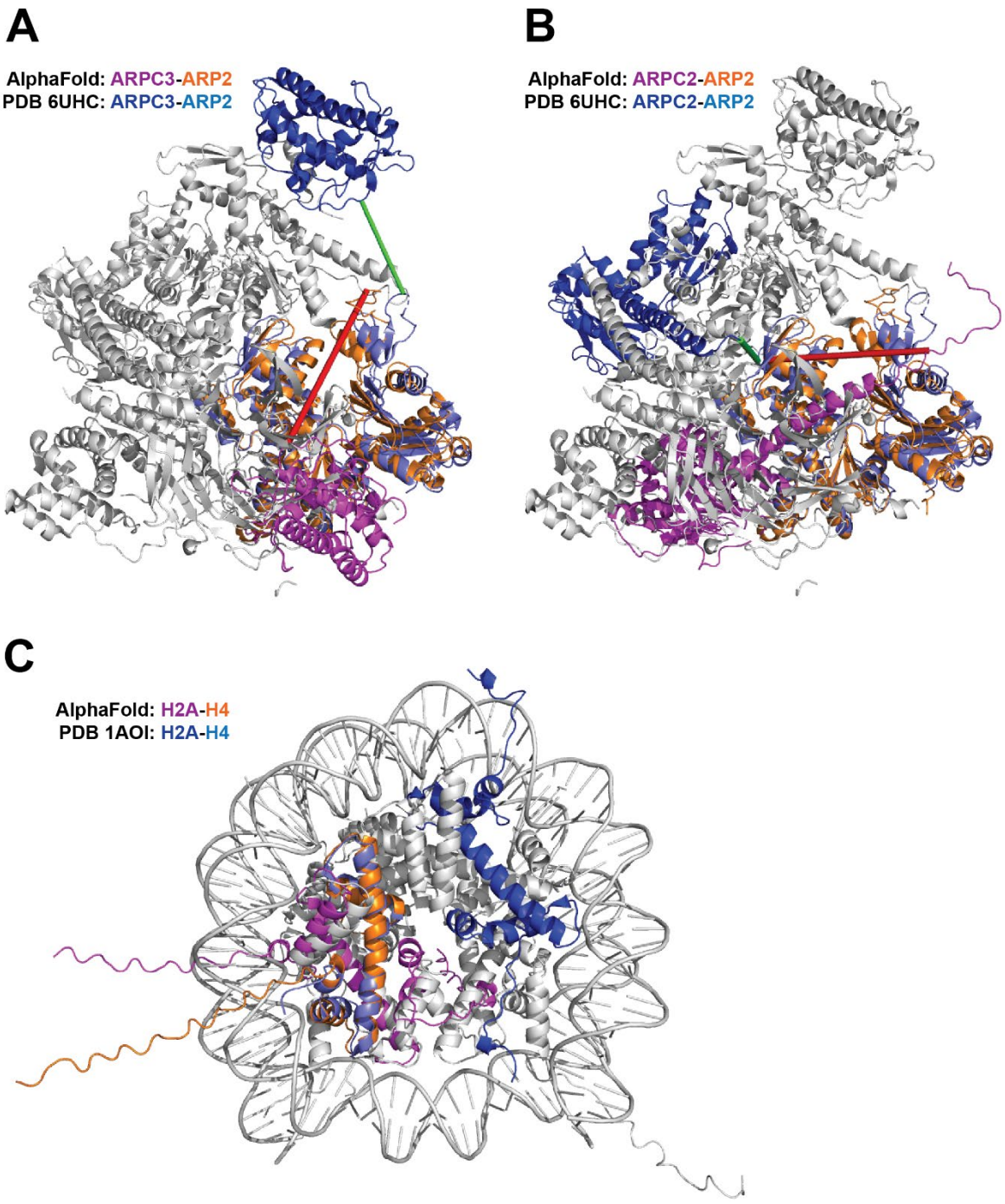
Supplementary Figure 5. The URP E673-K890 in the DNA replication licensing factor MCM2 reports different cross-linker distances in different contexts. MCM2 (P49736) is a member of the CDC45-MCM-GINS (CMG) helicase. **(A)** When the CMG helicase is not engaged in the replisome, the URP E673-K890 reports a distance of ~10 Å. **(B)** When the CMG helicase is engaged in the replisome, the URP E673-K890 reports a distance of ~39 Å.



Supplementary Figure 6. Cross-links validate AlphaFold2 monomer models of proteins absent from the PDB. (A) The satisfaction rate for experimental and random URPs per AlphaFold2 model. For each model, five random URPs were generated for each cross-linker with appropriate sidechain reactivities. **(B)** 624 out of the 737 proteins without corresponding PDB entries satisfied all their high-confidence URPs in their corresponding AF2 models. **(C)** AlphaFold2 model of the very-long-chain 3-oxoacyl-CoA reductase enzyme (Q53GQ0) that does not have any PDB entries for itself or of homologous proteins. Nine URPs support the AF2 model.

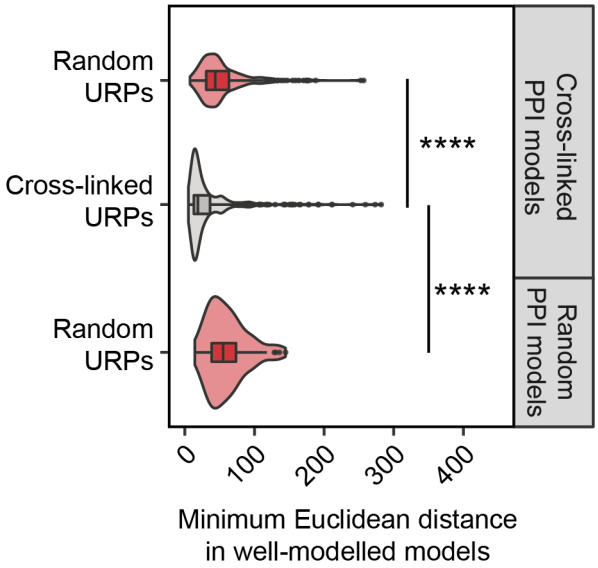


Supplementary Figure 7. AlphaFold Multimer-v2 modelling quality statistics. The distributions of **(A)** MolProbity clashscores (clashing atoms per 1000 atoms); and **(B)** the average interface residue pLDDT scores for all models generated in this analysis; and **(C)** the average interface area as defined by PISA. The line shows the cut-off used to filter for "well-modelled" interfaces.

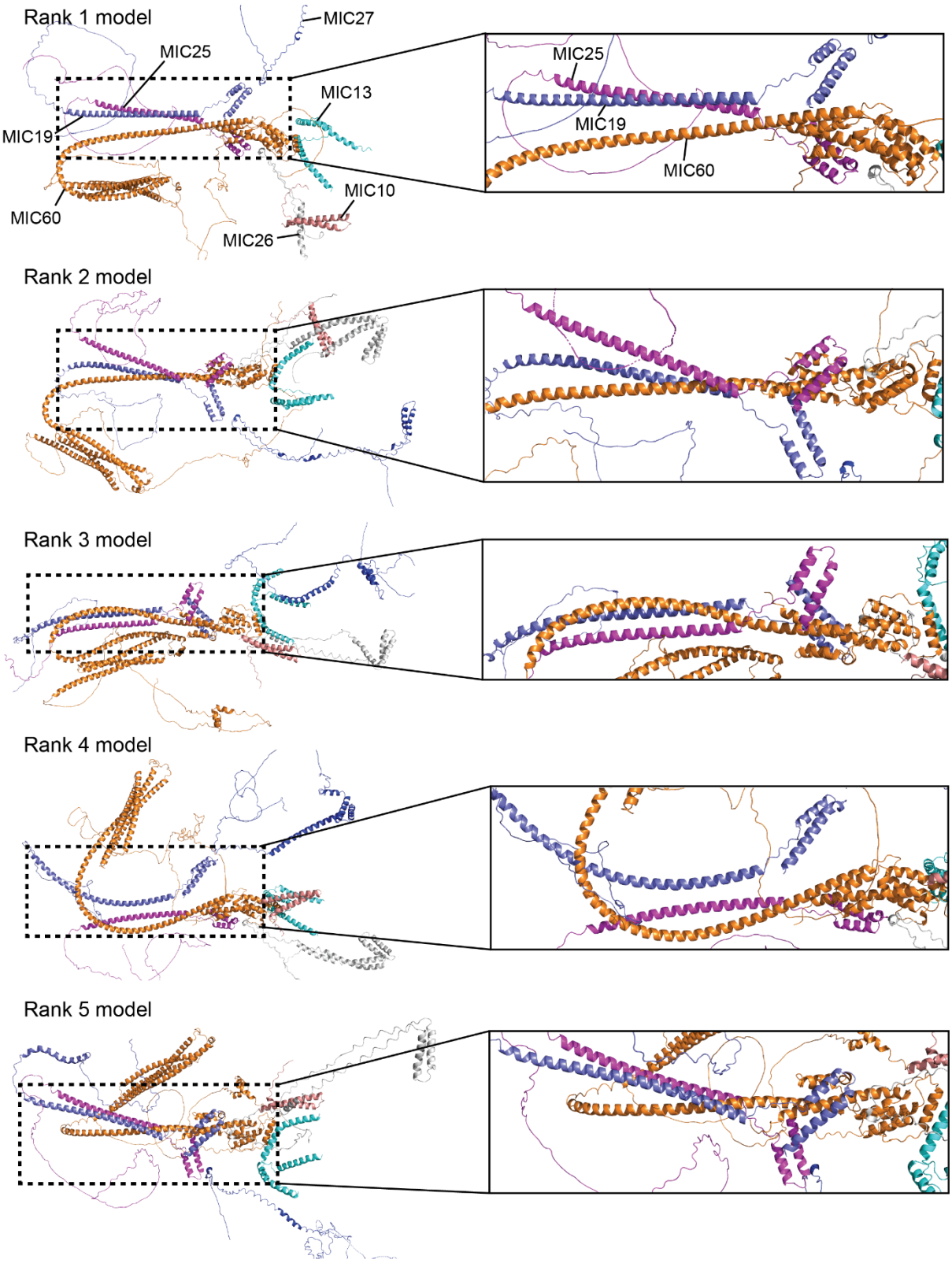


Supplementary Figure 8. Some AF Multimer models do not align well with PDB structures. (A) The predicted AF Multimer-v2 model of ARPC3-ARP2 does not align with the known PDB: 6UHC structure of the full 7-subunit Arp2/3 complex. Two URPs are available and both do not fit the AF model but are satisfied in the PDB structure. **(B)** The predicted AF Multimer-v2 model of ARPC2-ARP2 does not align with the known PDB: 6UHC structure of the full 7-subunit Arp2/3 complex. One URP is available and does not fit the AF model but is

satisfied in the PDB structure. **(C)** The predicted AF Multimer-v2 model of histones H2A-H4 does not align with the known PDB: 1AOI structure of the nucleosome.

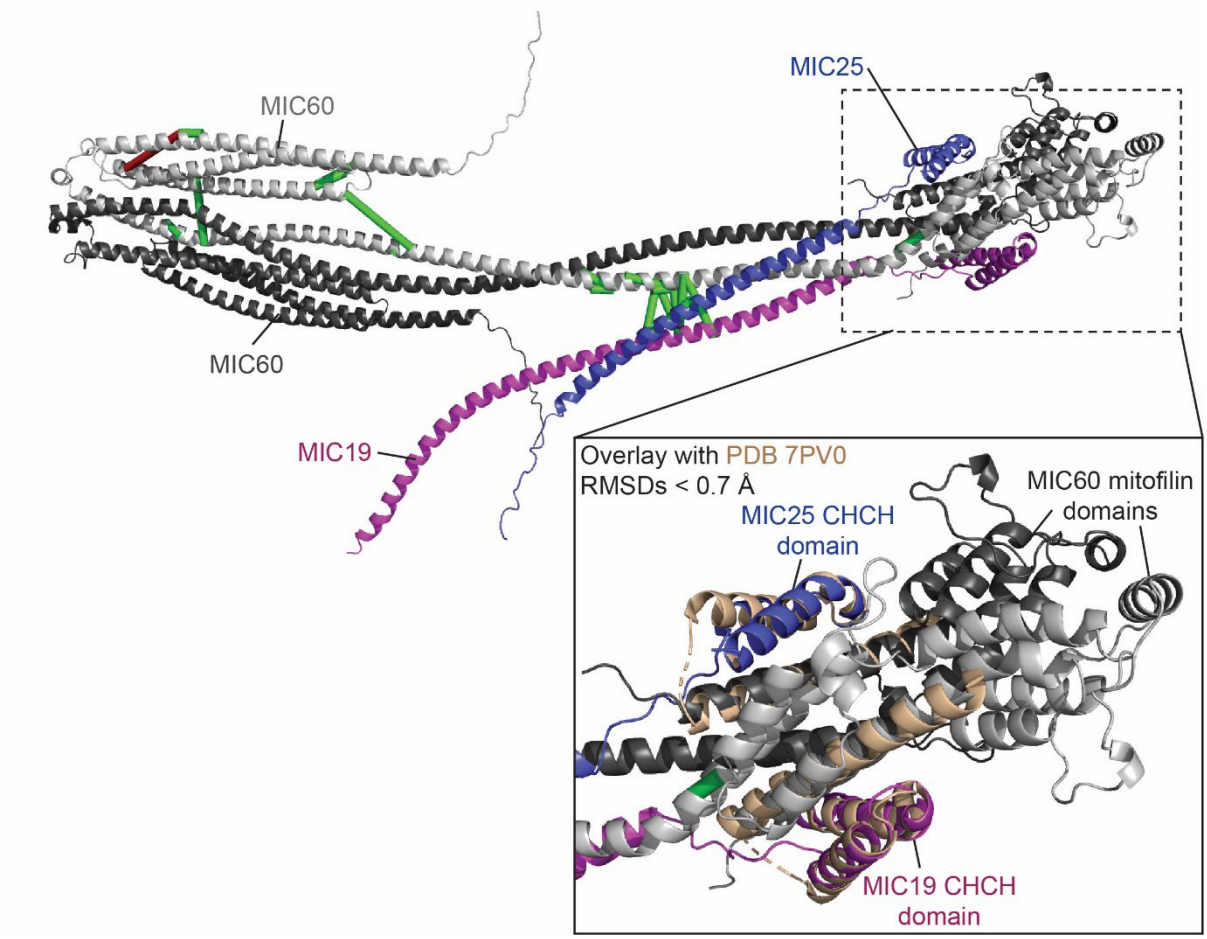


Supplementary Figure 9. AlphaFold Multimer-v2 modelling crosslinking statistics. (A) The distribution of measured Euclidean distances (Å, Cα-Cα) of experimentally derived and randomly generated URPs. Only URPs involving high-confidence residues with pLDDT ≥ 70 were used for model assessment. Experimentally derived URPs were found to be significantly shorter than randomly generated URPs. **** is $p = 2.2 \times 10^{-16}$.



Supplementary Figure 10. AlphaFold Multimer-v2 models of the MICOS complex have generally similar core structures comprised of MIC60, MIC25 and MIC19. While the five ranked AF Multimer-v2 models do not have an overall consensus structure for the

heptameric complex, the spatial arrangement of MIC60 (*orange*), MIC25 (*magenta*) and MIC19 (*slate*) were generally similar. Enlarged images of the MIC60-MIC25-MIC19 core (*dashed boxes*) are shown on the right of each model.



Supplementary Figure 11. MIC60-MIC25-MIC19 is predicted to have a 2:1:1 stoichiometry. The representative rank 1 AF Multimer model of the 2:1:1 MIC60-MIC25-MIC19 subcomplex is shown. This complex displays uniformly high pLDDT score (average ~77) and comprises a MIC60 homodimer with its mitofilin domains bound symmetrically to the CHCH domains of MIC19 and MIC25. This MIC60-MIC19/MIC25 interaction overlays well (RMSDs < 0.7 Å) with the distant fungal MIC60-MIC19 structure (PDB: 7PV0; *beige*). MIC60 is in *cyan* and *orange*, MIC25 is in *blue*, MIC19 is in *magenta*.

Supplemental References:

1. R. Anand, A. S. Reichert, A. K. Kondadi, Emerging Roles of the MICOS Complex in Cristae Dynamics and Biogenesis. *Biology (Basel)* **10** (2021).
2. A. Kroupova *et al.*, Molecular architecture of the human tRNA ligase complex. *Elife* **10** (2021).
3. J. M. Graham, Homogenization of mammalian cultured cells. *ScientificWorldJournal* **2**, 1630-1633 (2002).
4. J. M. Graham, Preparation of crude subcellular fractions by differential centrifugation. *ScientificWorldJournal* **2**, 1638-1642 (2002).
5. O. Klykov *et al.*, Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. *Nat Protoc* **13**, 2964-2990 (2018).
6. A. Leitner *et al.*, Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc Natl Acad Sci U S A* **111**, 9455-9460 (2014).
7. J. K. K. Low *et al.*, The Nucleosome Remodeling and Deacetylase Complex Has an Asymmetric, Dynamic, and Modular Architecture. *Cell Rep* **33**, 108450 (2020).
8. Y. Wang *et al.*, Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019-2026 (2011).
9. F. Liu, P. Lossl, R. Scheltema, R. Viner, A. J. R. Heck, Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat Commun* **8**, 15473 (2017).
10. T. K. Bartolec *et al.*, Cross-linking Mass Spectrometry Analysis of the Yeast Nucleus Reveals Extensive Protein-Protein Interactions Not Detected by Systematic Two-Hybrid or Affinity Purification-Mass Spectrometry. *Anal Chem* **92**, 1874-1882 (2020).
11. Z. L. Chen *et al.*, A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun* **10**, 3404 (2019).
12. S. H. Giese, A. Belsom, L. Sinn, L. Fischer, J. Rappsilber, Noncovalently Associated Peptides Observed during Liquid Chromatography-Mass Spectrometry and Their Effect on Cross-Link Analyses. *Anal Chem* **91**, 2678-2685 (2019).
13. C. UniProt, UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480-D489 (2021).
14. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
15. M. Wang, C. J. Herrmann, M. Simonovic, D. Szklarczyk, C. von Mering, Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163-3168 (2015).
16. D. Piovesan *et al.*, MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Res* **49**, D361-D367 (2021).
17. P. V. Hornbeck *et al.*, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512-520 (2015).
18. D. Alonso-Lopez *et al.*, APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database (Oxford)* **2019** (2019).
19. D. Szklarczyk *et al.*, The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605-D612 (2021).
20. M. Giurgiu *et al.*, CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res* **47**, D559-D563 (2019).
21. J. M. Dana *et al.*, SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Res* **47**, D482-D489 (2019).
22. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).
23. B. J. Grant, L. Skjaerven, X. Q. Yao, The Bio3D packages for structural bioinformatics. *Protein Sci* **30**, 20-30 (2021).

24. K. Tunyasuvunakool *et al.*, Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596 (2021).
25. M. Varadi *et al.*, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-D444 (2022).
26. M. Mirdita *et al.*, ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022).
27. M. Steinegger, J. Soding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).
28. L. S. Johnson, S. R. Eddy, E. Portugaly, Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).
29. R. Evans *et al.*, Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 10.1101/2021.10.04.463034, 2021.2010.2004.463034 (2022).
30. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797 (2007).
31. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242 (2011).
32. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
33. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
34. A. Kahraman, L. Malmstrom, R. Aebersold, Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* **27**, 2163-2164 (2011).
35. B. Schiffrin, S. E. Radford, D. J. Brockwell, A. N. Calabrese, PyXlinkViewer: A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. *Protein Sci* **29**, 1851-1857 (2020).
36. S. F. Altschul *et al.*, Protein database searches using compositionally adjusted substitution matrices. *FEBS J* **272**, 5101-5109 (2005).
37. A. Wheat *et al.*, Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. *Proc Natl Acad Sci U S A* **118** (2021).