

Cross(X)link Mapping and AnalySis (XMAS) – Smooth Integrative Modeling in ChimeraX

Ilse M. Lagerwaard^{1,2,\$}, Pascal Albanese^{1,2,\$}, Andris Jankevics^{1,2}, Rodrigo Vargas Honorato³, Alexandre M.J.J. Bonvin³, Richard A. Scheltema^{1,2,*}

1. Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute of Pharmaceutical Sciences, Utrecht University, 3584 Utrecht, The Netherlands
2. Netherlands Proteomics Center, 3584 Utrecht, The Netherlands
3. Bijvoet Centre for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

\$ equal contribution

* Address correspondence to: r.a.scheltema@uu.nl

Keywords: ChimeraX, XL-MS, XlinkX/PD, PhoX, DSSO, DSS, fibrin.

Abstract

Crosslinking mass spectrometry (XL-MS) has demonstrated its ability to add tremendous value to structural biology investigations. Its main strength lies in uncovering structural information from proteins in the form of distance constraints between neighboring amino acids, even in particularly complex samples containing proteins with regions difficult to access by other techniques. However, interpreting XL-MS data in a structural context has been cumbersome, even though several approaches have been developed. Moreover, with the introduction of the highly flexible and widely used ChimeraX – a software package for visualization of structural data – XL-MS analysis tools previously available in its predecessor, Chimera, were lost. We therefore introduce XMAS, a bundle for ChimeraX that allows users to load results from several XL-MS search engines straight into ChimeraX and directly map the information onto protein structures. Besides automatically locating the distance constraints on the protein structure, XMAS offers the possibility to work with replicate experiments and/or different crosslinkers, and filter on reproducibility of distance constraints across replicates to increase the data quality. We additionally introduce the concept of self-links, which allows modeling of homomeric interactions. Furthermore, we implemented a seamless connection to the structural modeling suite of HADDOCK to ease difficult tasks in the structural modeling pipeline. Here, key functionalities of XMAS are demonstrated on data obtained from human fibrin clots. The software is freely available from the ChimeraX toolshed, with an extensive user manual for quickly grasping the functionality.

Introduction

Proteins are large, macro-molecular machines responsible for complex cellular behavior, such as sensing the external environment, reproduction, immune defense, and much more. Proteins can act alone, but can also cooperate in larger assemblies that enable more complex behavior than the sum of the individual parts would suggest (1). It has been estimated that 80% of proteins engage in these interactions, leading to estimations of 130,000 to 650,000 unique protein to protein interactions (2, 3). The study of protein structure and the protein complexes they form is an important field that aids our understanding of complex diseases driven by failures in the protein machinery, for instance, and requires solid analytical approaches. A powerful technique to analyze proteins is shotgun mass spectrometry, where proteins are digested into peptides before separating the peptide mixture by liquid chromatography and measuring the eluting peptides through mass spectrometry (4). The technique has proved instrumental in uncovering relative quantities of proteins between different cellular states (5-7), location of post-translational modifications (6), melt curves of proteins (7, 8), location in cellular compartments (8-10), and many other details of the life of proteins inside and outside cells. To make shotgun mass spectrometry capable of capturing protein structural information, crosslinking mass spectrometry (XL-MS) was developed (9-11). In XL-MS, agile reagents capable of forming a covalent bond between amino acids in close proximity are supplemented to the protein mixture. Typically, NHS ester chemistry is utilized to capture lysine-lysine pairs, but other chemistries have also been introduced (11). After digestion, the peptide mixture contains three distinct products: normal peptides (unaffected by the used reagent), mono-links (a single peptide where one amino acid is covalently linked by one reactive head on the used reagent, while the other head is quenched on for example water), and cross-linked peptides (two peptides covalently bound by the reagent). The term crosslink is defined as two crosslinker-connected amino acids, independently from the presence of missed cleavages or modifications in the containing peptides (12). Distance constraints can be derived from the peptide identities of the experimentally obtained crosslinks. These distance constraints can aid in delineating the structure of individual proteins and the interaction interface(s) between multiple proteins. In recent years, XL-MS has been made generally applicable by developments

that specifically solved issues with low stoichiometry of the desired products and complexities in data analysis (13-16). The advent of cryo-electron microscopy (EM) in structural biology supported the technique's gain in popularity, since XL-MS was found to be a perfect companion (16); it enables interpretation of cryo-EM and cryo-electron tomography (ET) densities in which the protein structures cannot be solved *ab initio*, or cryo-EM structures in which the identity of interacting subunits cannot be determined at intermediate resolution (17, 18).

More recently, XL-MS has shown great potential by leveraging the spectacular advances in protein prediction solutions *in-silico* like AlphaFold (19) and RoseTTAFold (20) and even the recently introduced ability to use AlphaFold2 to predict protein complexes for subsequent validation using XL-MS data (21) or modeling of larger assemblies (22). Using the distance constraints, the produced models can be validated on correctness. Moreover, structural modeling tools like HADDOCK (23) can be used to build credible structural models of protein complexes. Placing distance constraints in the context of the three-dimensional protein structures is critical in this respect and has received considerable attention in the past. For example, tools have been developed to visualize the distance constraints in plots showing the primary protein structure to gain insights into hotspots of crosslinking, type of folding and many other details (24-28). These approaches have enabled scientists to make use of the structural information derived from XL-MS experiments, but at a price. Data reformatting and additional scripting have been unavoidable for many years, demanding a steep learning curve of researchers in the field and, importantly, making the analysis pipeline error prone.

Here, we introduce XMAS (Cross(X)link Mapping and AnalySis), a plugin that is fully integrated in ChimeraX (29) and allows data from XL-MS experiments – generated from multiple search engines – to be directly loaded into the environment without the need for data reformatting. Additionally, the plugin automatically localizes the crosslinks inside the protein sequence/structure even when there is no perfect match between sequences (*e.g.*, if the structure contains only part of the complete protein sequence). We also introduce the concept of self-links, which assist in building protein complex models for homomers. Since self-links can conveniently be inferred from existing XL-MS datasets, no extra experimental setup is required. Furthermore, XMAS provides means of plotting crosslink distances and overlap between replicate experiments, as well as an automatic search for the shortest possible crosslinks in cases where multiple locations are possible (*e.g.*, for homomers). Lastly, to enable full use of the distance constraints in a structural modeling context, XMAS automates the connection to the HADDOCK structural modeling environment (23, 30). This considerably reduces the number of data manipulation steps and simplifies XL-MS analysis for non-expert users.

Materials and methods

Implementation – XMAS was implemented as a ChimeraX bundle written in Python 3.9. It was packaged as a Python wheel according to the ChimeraX developer guide for Building and Distributing Bundles, which can be found at https://www.cgl.ucsf.edu/chimerax/docs/devel/writing_bundles.html. Of the third-party Python packages preincluded in the ChimeraX environment, XMAS uses PyQt5 (for the graphical user interface (GUI)), NumPy (for fast processing of large datasets), and Matplotlib (for generating plots). Additionally, Pandas (for handling tabular data format files), openpyxl (for importing Excel files), matplotlib-venn (for plotting Venn diagrams), seaborn (for plotting box and strip plots), and QtRangeSlider (to create slider widgets for defining a range of values) are employed. These external libraries are automatically installed in ChimeraX upon installation of XMAS when necessary.

XMAS can be downloaded and installed easily *via* the ChimeraX toolshed (<https://cxtoolshed.rbvi.ucsf.edu/>). Alternatively, instructions for manual installation are provided in the "INSTALL.TXT" file included in the install package. To ensure support of all XMAS functionalities, the minimum ChimeraX version required is 1.3. The bundle source code, including a user manual, is freely available at <https://github.com/ScheltemaLab> under an MIT software license.

Crosslink datasets – Commercial plasminogen-depleted fibrinogen, purified from human plasma (Merck), was dissolved in ice-cold crosslinking buffer at a concentration of ~10 μ M (1mg/mL, 100 μ g per 1.5-mL vial). The crosslinking buffer contained 50 mM Hepes (pH 7.4), 120 mM NaCl, and 2 mM CaCl₂ (all Sigma-Aldrich). Fibrinogen clotting was activated by adding 1 U/mL human thrombin (Sigma-Aldrich). Samples were incubated at 37°C for 30 min in an Eppendorf Thermomixer C orbital shaker to allow further clotting. Two sets of three independent clotting reactions, each on 100 μ g of fibrinogen, were crosslinked at room temperature for 30 min with two different crosslinking reagents: either 1) 2 mM (3,5-bis(((2,5-dioxopyrrolidin-1-yl)oxy) carbonyl)phenyl)phosphonic acid (PhoX) or 2) 1 mM disuccinimidyl suberate (DSS) (Thermo Scientific). The crosslinking reaction was subsequently quenched with 50 mM Tris (Roche). Next, predigestion was performed by incubation with Lys-C (Sigma-Aldrich) at 37°C for 30 min, at a Lys-C/protein ratio of 1:70 (wt/wt). Proteins were denatured with 4 M urea (Merck) at 54°C for 1 h, prior to sonication in a Bioruptor bath sonicator (Diagenode) for 15 cycles of 30 s ON and 30 s OFF. Thereafter, samples were reduced with 8 mM DTT (Sigma-Aldrich) at 54°C for 30 min in an Eppendorf Thermomixer C at 800 rpm, followed by alkylation with 200 mM IAA (GE Healthcare) at room temperature for 1 h while protected from light exposure. Samples were then diluted with 50 mM Tris at a sample/Tris ratio of 1:1.1 (vol/vol) to reduce the urea concentration to <2 M. Trypsin and Lys-C (Sigma-Aldrich) were added at a protease/protein ratio of 1:25 and 1:70 (wt/wt), respectively, for protein digestion overnight at 37°C (~16 h). Prior to peptide cleanup, samples were acidified to pH <2 with 0.5% trifluoroacetic acid (TFA) (Honeywell) and desalted using an Oasis HLB μ Elution 30 μ m plate (Waters). After washing the plate columns, twice with 150 μ L of acetonitrile (ACN) (Biosolve) and twice with 150 μ L of 0.1% TFA, acidified samples were loaded. Columns were then washed twice with 150 μ L of 0.1% TFA again, and peptides were eluted with 50% ACN/0.1% TFA (thrice 35 μ L). The obtained eluates were evaporated by vacuum centrifugation in a Savant SpeedVac Vacuum Concentrator (Thermo Scientific). For subsequent fractionation of crosslinked peptides, two different strategies were adopted, depending on the chemical crosslinker. For PhoX crosslinked samples, the previously optimized enrichment protocol leveraging the immobilized metal affinity chromatography (IMAC) enrichable phosphonate group was used (13). Briefly, crosslinked peptides were enriched with Fe(III)-NTA-5 μ L cartridges using the AssayMAP Bravo Platform (Agilent Technologies). Cartridges were primed with 200 μ L of 0.1% TFA in ACN and equilibrated with 200 μ L of 80% ACN/0.1% TFA. Samples were dissolved in 200 μ L of 80% ACN/0.1% TFA and loaded onto the cartridge. The columns were washed with 250 μ L of 80% ACN/0.1% TFA and the crosslinked peptides were eluted with 35 μ L of 10% ammonia directly into 35 μ L of 10% formic acid (FA) (98-100% purity; Merck). Samples crosslinked with DSS were enriched by strong cation exchange (SCX) chromatography using the same AssayMAP Bravo Platform with a protocol specifically optimized for the purpose of micro-fractionation of crosslinked peptides pairs as follows: dried samples were resuspended in 100 μ L of 20% ACN/0.1% TFA, using SCX cartridges in an AssayMAP Agilent Bravo robot (both Agilent Technologies). Fractions were retrieved by five consecutive elutions of 25 μ L each with buffers containing increasing concentrations (50, 250, 450, 600, and 750 mM) of ammonium acetate (diluted from a 7.5 M stock; Sigma-Aldrich), in addition to 20% ACN, and TFA to acidify buffers to pH <2. Next, the 250-, 450-, 600-, and 750-mM ammonium acetate SCX fractions were desalted on an Oasis HLB μ Elution 30 μ m plate as described previously in this section. All fractions were evaporated again in a Savant SpeedVac Vacuum Concentrator and stored at -20°C prior to MS analysis.

For MS acquisition, samples were resuspended in 2% FA, and peptides were separated using an Ultimate 3000 LC system (Thermo Scientific) coupled online to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific). Firstly, peptides were trapped in reverse-phase solvent A (0.1% FA), by applying a 5- μ L/min column flow rate for 5 min on a 2-cm double-frit trap column (100- μ m inner diameter) packed in-house with Reprosil-Pur C18-AQ 3 μ m beads (Dr. Maisch). Secondly, peptide separation was achieved by a 45-min gradient elution with 9-42% solvent B (80% ACN/0.1% FA), applying a 300-nL/min column flow rate on a 50-cm single-frit analytical column (75- μ m inner diameter) packed in-house with InfinityLab Poroshell 120 EC-C18 2.7 μ m beads (Agilent Technologies).

Survey MS1 Orbitrap scans were acquired at 60,000 resolution, with an automatic gain control (AGC) target of $1e6$ ions, maximum injection time (IT) of 50 ms, and scan range of 375-1500 m/z. Precursors with a 3-8 charge state were selected for fragmentation using a stepped HCD collision energy mode (normalized energies of 24-27-30% for DSS and 19-27-35% for PhoX). Dynamic exclusion properties were set to sort by highest charge state. Full MS2 Orbitrap scans were acquired at 30,000 resolution, with an AGC target of $8e5$ ions and dynamic maximum IT. Spectra were analyzed using Thermo Proteome Discoverer 2.4.1.15 software with incorporated XlinkX nodes, using DSS (31) or PhoX (13) specific settings. Crosslinks with a false discovery rate at 1% and score >40 were kept for further analysis. Unless otherwise specified, only crosslinks present in at least three out of three replicates were used for structural modeling (31).

Structural modeling – As a proof of concept for testing XMAS to uncover novel structural insights, we used the laterally aggregating fibrinogen protofibrils model from Klykov *et al.* (PDBdev ID: 0000030) (32) that was based on an available X-ray crystal structure (PDB ID: 3ghg) as a starting model. The starting model contained two docked regions of the α -chain (α 220-249 and α 412-472), which were removed in this study. The remaining dimer of trimers, containing the structurally resolved part of the fibrinogen α , β and γ chains (hereafter defined as $\alpha\beta\gamma$ -main), was extended with the α 220-265 and α 423-487 regions (hereafter called α 200 and α 400) extracted from the AlphaFold model of the human fibrinogen α chain (downloaded from <https://alphafold.ebi.ac.uk/entry/P02671>) (Supplementary Fig. 1). Using automatic detection of homodimers in XMAS, ten reproducible overlapping peptide pairs from the DSSO and DSS datasets and four from the PhoX dataset, were identified for the α 400 domain only. In addition to these highly reproducible crosslinks, one self-link (α 463- α 463) has been found in all datasets, confirming that the α 400 region is a pivotal region in aggregating fibrinogen protofibrils irrespective of the sample source (plasma vs. commercial purified fibrinogen) and the sample preparation procedures (this work vs. Klykov *et al.* (32)). The detected distance constraints specifically defining the homomeric interaction interface were directly used as ambiguous interaction restraints (AIR) for molecular docking using the HADDOCK 2.4 server (23). The AIR were submitted as a tbl file defining a median distance of 10 Å (lower bound up to 5 Å and upper bound up to 25 Å). The tbl file was generated using the GenTBL online tool (accessible here: <https://alcazar.science.uu.nl/services/GenTBL/>). Representative structures of the top three clusters in the HADDOCK output were evaluated in light of the submitted XL-MS restraints to select the model best supported by experimental evidence (Fig. 2). Then, the α 200 model was docked to the predicted α 400-dimer. Lastly, the resulting α 400-200 module was docked to the $\alpha\beta\gamma$ -main fibril. For docking of α 200 to the α 400-dimer, and of the α 400-200 module to the $\alpha\beta\gamma$ -main fibril, “true positive” restraints to be used as AIR were determined with a rotational sampling step using the DisVis server (30), which was performed as described in the “Integrative modeling assisted by XMAS” section of the Results. Restraints with a z-score lower than 0.5 were included for the α 200 to α 400-dimer docking, while a more stringent cut-off of 0 (only negative z-score values) was selected for the α 400-200 module to $\alpha\beta\gamma$ -main fibril docking (Supplementary Fig. 2). The selected restraints were then used to generate the AIR tbl restraint file with

defined distances (5-25 Å for PhoX or 6-33 Å for DSS). To prepare models for molecular docking in HADDOCK, both multi-chain “fixed” models used for DisVis rotational sampling (the α 400-dimer and $\alpha\beta\gamma$ -main fibril) were merged as single chains renamed “A”, and amino acid positions were renumbered from “1” onwards using PDBtools (33). Likewise, both “scanning” chains, (*i.e.*, the structures to be docked to the main “fixed” model; α 200 and the α 400-200 module), were also renamed “A” and their amino acid positions renumbered from “1”. The HADDOCK output was then re-evaluated in XMAS by mapping the XL-MS data on the best scoring HADDOCK clusters and comparing them based on XL-MS distances (top three in the examples for Fig. 2 and Fig. 3).

Results

Human fibrin clots as a model system – To demonstrate the capabilities of XMAS, we used fibrin clots as our model system. Fibrin clots are formed by activation of fibrinogen hexamers, major constituents of plasma that assist in wound healing. Activation releases fibrinopeptides from fibrinogen, forming fibrin. Clot formation then occurs through fibrin oligomerization, followed by lateral aggregation, packing into fibrin fibers, and consequent branching (34). Fibrin(ogen) consists of three chains: α , β , and γ . The α -chain contains unstructured and flexible regions, which are challenging to investigate by conventional structural biology approaches. Many structural details indeed remain out of reach, although fibrin clots formed in human plasma have already been studied successfully with XL-MS by Klykov *et al.*, using DSSO crosslinking in combination with structural modeling of missing regions of the fibrin structure (32). The main goal of the present study was to extend the fibrin model from Klykov *et al.* using XMAS. To achieve this, we performed additional experiments on commercially available fibrinogen purified from human plasma. Compared to the previous study, a significantly lower amount (approximately ten- to twentyfold less) of fibrinogen was used. The protein was clotted in triplicate independent reactions and crosslinked with either PhoX or DSS (see Materials and Methods for details). The resulting dataset consisted of 339 PhoX crosslinks (156 inter- and 183 intralinks) and 273 DSS crosslinks (122 inter- and 151 intralinks).

For our structural modeling approach, we took advantage of the AlphaFold prediction for the α -chain (downloaded from <https://alphafold.ebi.ac.uk/entry/P02671>). From this predicted model, we extracted the α 220-265 and α 423-487 (Supplementary Fig. 1) domains (hereafter α 200 and α 400, respectively) in ChimeraX, and used the XMAS interface to DisVis and HADDOCK to integrate these protein sections into the current fibrin model. This increased the sequence coverage of the α -chain model by 4% with respect to the previous model. Only those structural domains for which an average confidence was achieved (*i.e.*, pLDDT > 50) and for which crosslinks have been identified were considered. In the following paragraphs, we show how this is accomplished with XMAS, including how to visualize, compare and analyze XL-MS data, as well as how to detect homomeric interactions.

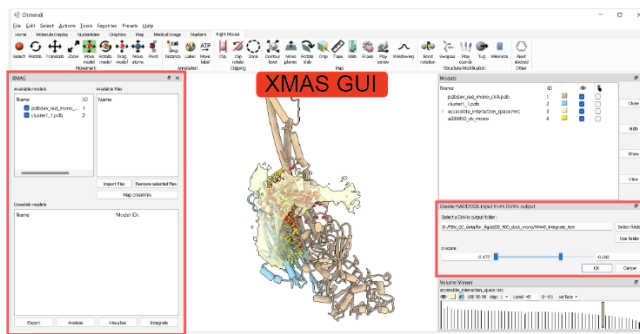
Mapping crosslinks to protein structures, more than a simple visualization step – One of the major bottlenecks in the use of XL-MS derived distance constraints is the need to prepare results for loading into programs like ChimeraX through file reformatting. To prevent the required labor-intensive and error-prone file manipulations, XMAS natively reads the evidence files containing crosslink data from a number of commonly used XL-MS search engines: XlinkX (35), pLink (36), and Xi (37). Given its open architecture, XMAS can easily be modified to support extra output types. Besides these raw outputs, XMAS includes support to the open standard file format mzIdentML, which is evolving to become a community standard for XL-MS data (38). To further improve user-friendliness, XMAS performs automatic sequence alignment of the peptide sequences of the identified crosslinks (extracted from “Evidence files” in Fig. 1A-B) against

the protein sequence(s) of the selected PDB structure(s) (“Molecular models” in Fig. 1A-B). This effectively prevents localization mistakes of the distance constraints in the structure. If a match is found for both crosslinked amino acids, a graphical object connecting their α -carbon atoms named *pseudobond* is generated. The “Crosslink models” panel will then contain all the *pseudobond* files generated with an evidence file for the given structures (Molecular models). Information on this mapping stage is given in the “Log” window, including the location of a report file in tsv format where the *pseudobond* strings and other metadata relative to each crosslink of the initial evidence file are stored. During this mapping step, all possible combinations of residue-to-residue contacts are made if a chain is present multiple times in the selected models or the protein contains multiple stretches with the same sequence. This obviously generates surplus *pseudobonds* mapped within the same structure, an issue which is resolved in a later step, but that is nonetheless needed initially to avoid biases in data interpretation.

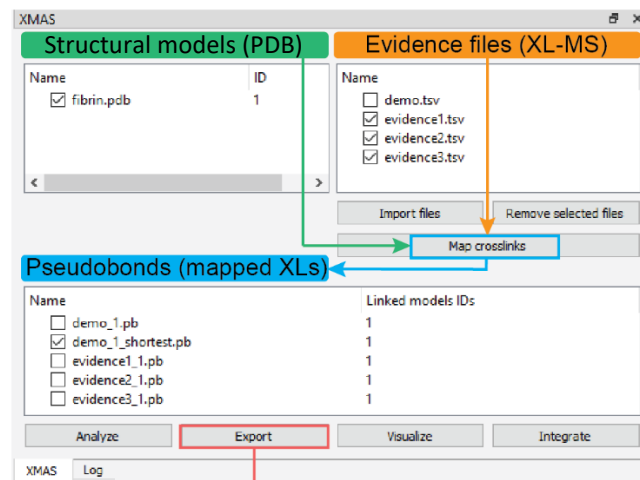
XMAS enables handling and comparing multiple large datasets – XMAS can load multiple datasets, and offers the option to visualize the overlap between these datasets in a Venn diagram. Visualizing the overlap between replicates, for example, provides critical insights into the reproducibility of the experiment (Fig. 1C). Complementary to this, we have found it effective over the years to also filter the datasets on reproducible links, for which a setting to select crosslinks detected in n out of the total number of experiments is incorporated in XMAS (Fig. 1C). This feature can for instance be used to locate all crosslinks that are present in only one out of multiple datasets, and map all these unique crosslinks onto the protein structures, thereby shedding light on differences between datasets.

Besides applications of XL-MS on purified or recombinantly expressed protein complexes, one of the strengths of this technique is its applicability on a system-wide scale (17, 39, 40). The integration of XMAS in the ChimeraX environment additionally eases the handling of large structures and/or the many structures uncovered by systems-wide XL-MS datasets. This is demonstrated here for a publicly available dataset (PXD008418) generated from a crosslinking experiment on PC9 cell lysate yielding >2000 crosslinks (25). In this example, we uploaded the PDB codes of the top 18 proteins with >1 crosslink directly from <https://www.rcsb.org/> and mapped the respective crosslinks with the “find shortest” option. This approach allows a rational rapid evaluation of the proteome-wide experiment quality (Supplementary Fig. 3).

A



B



C

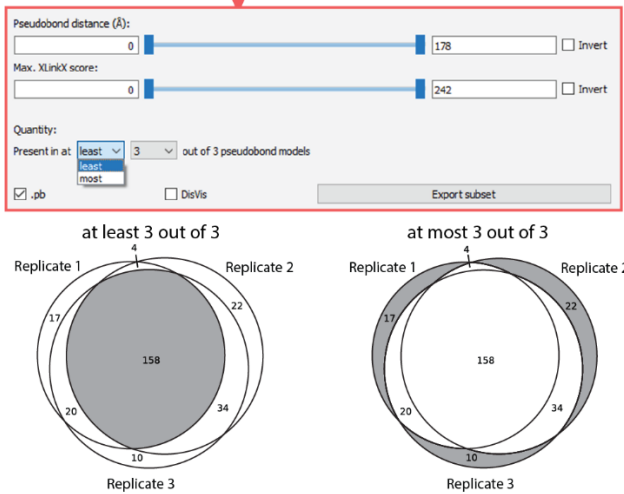


Figure 1. XMAS GUI integrated in ChimeraX 1.3. (A) Main window of ChimeraX 1.3 with localization of XMAS GUI. The main XMAS interface listing molecular models (PDB) and evidence files (crosslinks) that can be matched together with the “Map crosslinks” command. (B) One or multiple evidence file(s) can be mapped onto one or multiple molecular model(s) in any combination. Thereby, a unique *pseudobond* model and corresponding *pseudobond* file is generated for each evidence file. *Pseudobonds* are created for each combination of perfect alignments of both peptides of a crosslinked peptide pair onto the structure sequence. (C) As an example of the “Export” GUI functionalities, the selection of reproducible crosslinks comparing *pseudobond* models in at “least” (common to all datasets) or at “most” (uniquely represented in one or more datasets) n models is shown.

Detecting homomeric interactions – During the mapping stage, XMAS also analyzes the identified peptide pairs to detect possibly occurring homomeric interactions. To achieve this, XMAS highlights those identifications where either the crosslinked residue (or position within the protein sequence) is the same for both peptides of the crosslinked peptide pair, or where the two peptides of the pair share part of their sequence (Fig. 2A). As these situations are physically impossible on the same protein (as would be the case for intralinks), such identifications can only arise from interacting homomers, and this information can be used to drive the homomer modeling and use of XL-MS data to validate the predictions. We name such theoretical crosslinks between overlapping peptides “self-links”. Advantageously, these types of links can also be detected in existing datasets as they do not require additional, more elaborate experimental setups. From reanalyzing and comparing the crosslinked fibrin clots from human plasma (DSSO in Fig. 2B) and the newly generated XL-MS data on *in-vitro* clotted fibrinogen (PhoX and DSS in Fig. 2B), it is worth noting that these homomeric interactions of the α 400 domain are very reproducibly detected, suggesting a significant role in inter-fibril interactions. Although we do not further investigate this role here, we do supply the modeling steps in the following paragraphs that lead to a credible model.

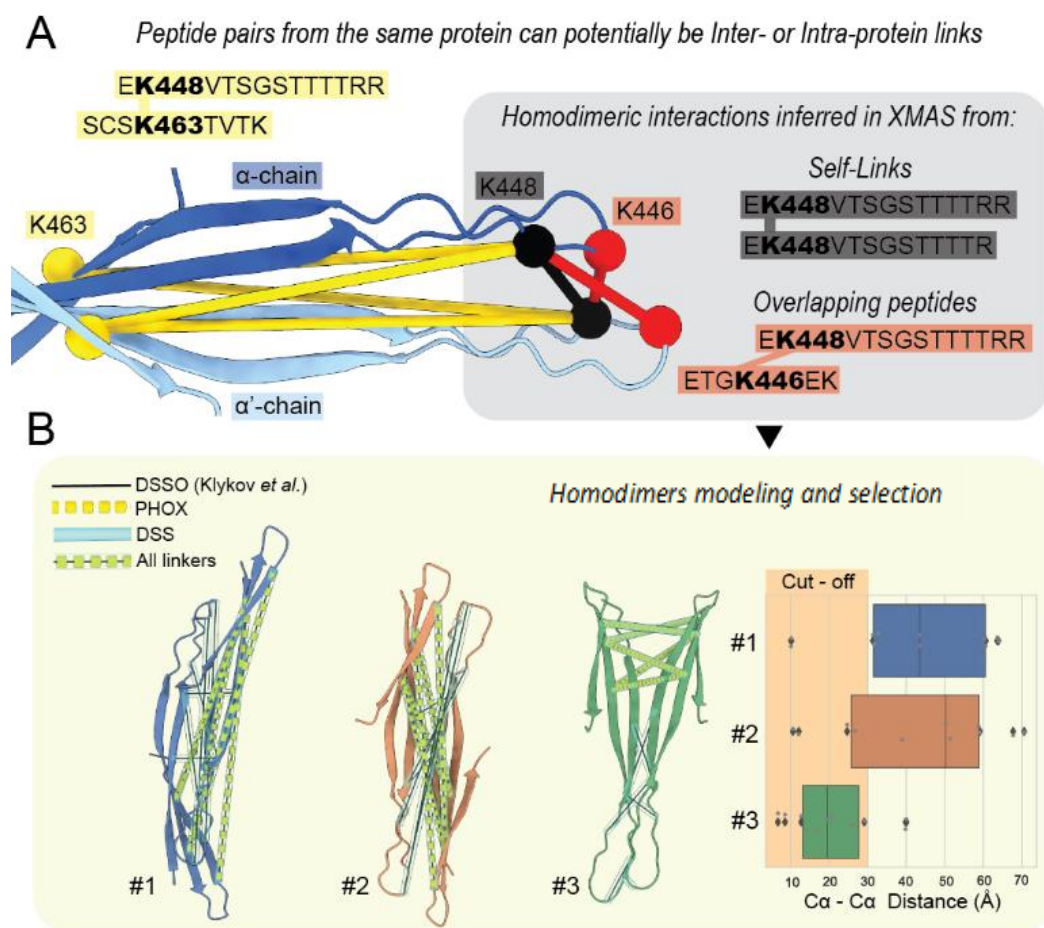


Figure 2. Unbiased mapping and detection of homomeric interactions. (A) Self-links and overlapping peptide pairs are automatically detected from the datasets and indicate positions where homomeric interaction interfaces are located. (B) After extracting these self-links through the XMAS user interface, protein-protein docking software like HADDOCK can be used to assemble the homodimer, resulting in multiple models that can be validated and filtered with the crosslink constraint distances.

Integrative modeling assisted by XMAS – With credible structures in hand, for which we here use the existing $\alpha\beta\gamma$ -main fibril model and the AlphaFold predicted versions for the α 400 dimer and α 200, the current fibrin model can be extended to cover a larger part of the amino acid sequence. To achieve this, all the structures can be loaded into a single ChimeraX session and our XL-MS distance constraints can be mapped, after which we aim to uncover the “inter-model” crosslinks. As the next step, a selection is to be made of the subset of XL-MS derived distance constraints that are consistent with a certain position of the α 400-200 module in relation to the $\alpha\beta\gamma$ -main fibril. For this, we rely on the DisVis package (30), where we view the $\alpha\beta\gamma$ -main fibril model as “fixed” and the α 400-200 module as “scanning”. The required input files are interactively generated in XMAS from the “Export” menu (Fig. 1B), where the “fixed” and “scanning” chains, the crosslinks datasets to include, and the minimum/maximum distance specific for the crosslinker in use (*e.g.*, 5-25 Å for PhoX, and 6-33 Å for DSS) can be set. The resulting txt file containing the distance constraints together with the two PDB models are then ready to be uploaded to the freely accessible DisVis webserver (<https://wenmr.science.uu.nl/disvis/>). When the run has completed, the DisVis output folder can directly be opened through the “Integrate” menu. XMAS will load both structures, open the MRC density showing the maximum number of consistent constraints at every position in space, and all the input constraints mapped as a new set of *pseudobonds* (Fig. 3). XMAS color-codes the *pseudobonds* that either violate (red) or agree with (yellow) the distance defined in the DisVis input file. The colors are updated in real-time if the user rotates or moves the “scanning” model in ChimeraX (Fig. 3). For the selection of consistent constraints to keep for a subsequent molecular docking step, an interactive slider selector for the z-scores is supplied, where higher z-scores denote a more likely “false positive” distance constraint. It should be noted, however, that in some cases more than one interaction interface might be defined in the occupancy analysis (*e.g.*, if a protein has multiple binding sites) and therefore constraints excluded merely based on a z-score cut-off may not be “false positives”. The selected distance constraints defining the interaction are then exported as a *pseudobond* file and integrated into a json file containing both the “fixed” and “scanning” chains, as well as the standard parameters to be directly uploaded in the HADDOCK webserver (<https://wenmr.science.uu.nl/haddock2.4>) for the actual docking of the two molecules. After a successful docking run, the clustered solutions can be evaluated in light of the XL-MS experimental evidence, and the better predicted models identified (Fig. 3B).

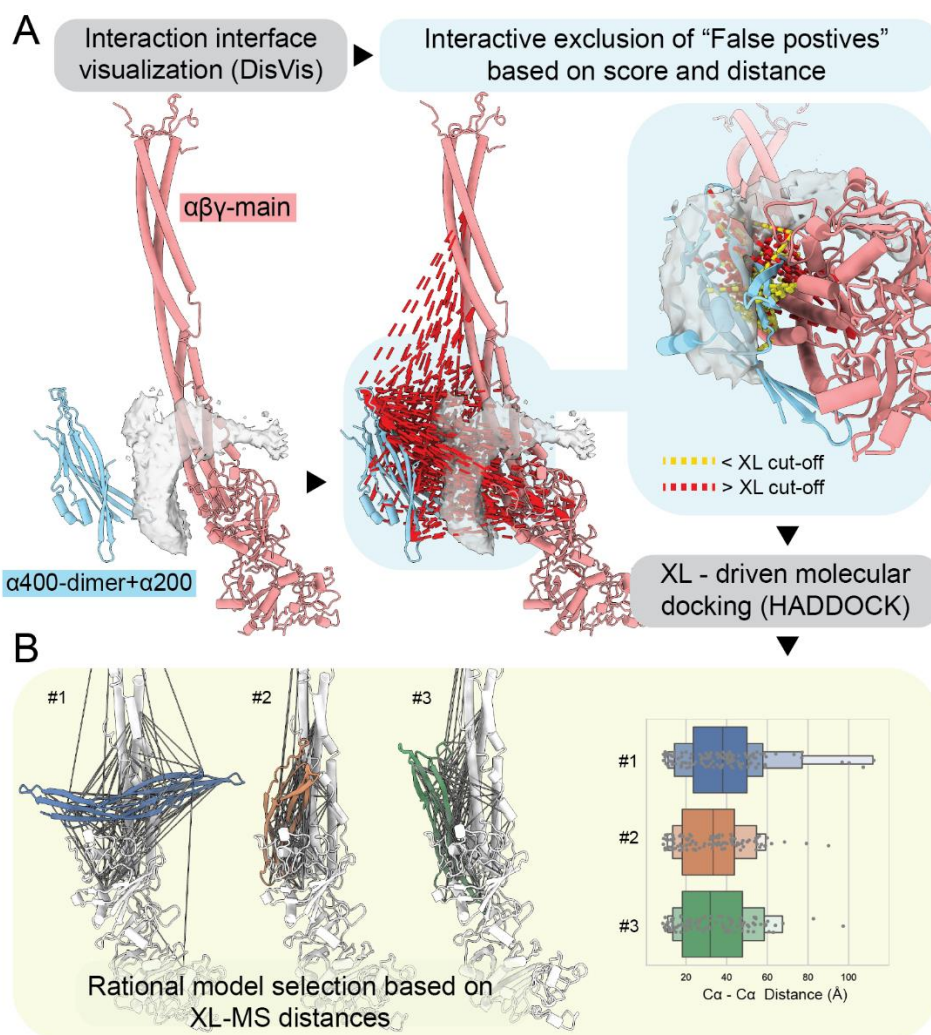


Figure 3. XMAS-assisted integrative modeling. (A) The location of novel protein stretches or interacting proteins can be located based on the distance constraints and DisVis – a process assisted by XMAS in obtaining the desired distance constraints from the complete set and producing the required input files for DisVis. (B) After further filtering the distance constraints in XMAS based on the DisVis results, the input files for HADDOCK are generated from XMAS. The docking run can then be started, and outputs evaluated in the context of experimentally derived distance constraints.

XMAS visualization capabilities allow deep data analysis – Once the integrative model is generated, XMAS offers advanced options for high-quality rendering of images and animations. This is accomplished thanks to the integration into ChimeraX, allowing the visual handling of new and large data types gracefully. We added the “Visualize” menu, providing capabilities to store and display useful data for each crosslink (*i.e.*, *pseudobond*) (Fig. 4). Intuitively, the length of the *pseudobonds* represents the most valuable piece of information. In the XMAS GUI, it is possible to change the color and style of the *pseudobonds* according to an arbitrary crosslinker-related cut-off, which can eventually be combined with a color gradient visualization for the distance (Fig. 4B).

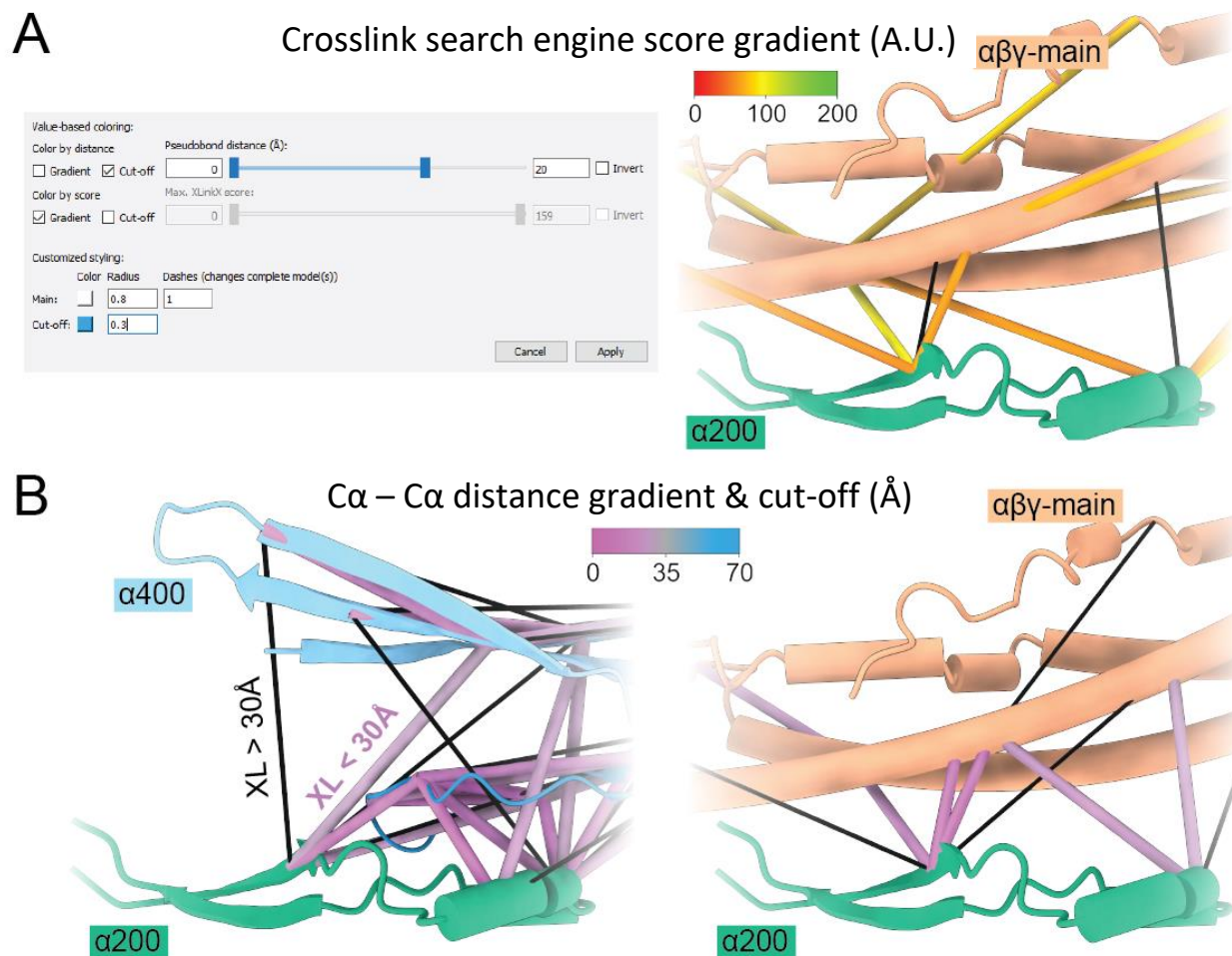


Figure 4. Advanced visualization options supported by XMAS. (A) Various settings pertaining to the distance constraints can be adopted (left panel), with the identification score as example (right panel). **(B)** The constraints can easily be color-coded according to the distance they span.

Future developments, usability, and community support – The open architecture of XMAS, its distribution as open-access software hosted at GitHub and the possibility to install it from the ChimeraX toolshed will allow non-expert users to use the tool and expert users to build up novel and tailored functionalities in future releases. From our side, future features will include: 1) dedicated pipelines for the visualization and analysis of quantitative data, 2) direct visualization of annotated MS spectral data, and 3) integration with molecular dynamics simulation trajectories to analyze protein conformational changes. With the ease XMAS brings to structural modeling using XL-MS data, we anticipate more projects to leverage this type of data. The combination of cryo-ET with XL-MS and protein computational modeling will also likely represent a step toward the visualization of the molecular sociology of the cell (41), for which XMAS already provides ample tools.

Conclusions

The field of XL-MS is rapidly expanding and gaining popularity among structural and cell biologists. The visualization and interpretation of these data in a structural context are, however, relying on piecemeal strategies, mainly implemented in specialized mass spectrometry or molecular modeling-oriented laboratories. To standardize XL-MS data processing and to popularize the use (and reuse) of XL-MS data in an open fashion, we designed XMAS to fulfill the needs of both proficient and inexperienced users. The key features provided are 1) a robust mapping and visualization procedure relying on sequence alignment directly onto PDB structures and detecting homomeric interactions based on peptide sequence overlap, 2) analytical and plotting capabilities resulting in publication-ready figures and plots, and 3) a seamless interface with integrative modeling tools. Thanks to its integration into the fantastic ChimeraX environment, with its many user-friendly and stunning visualization options, we think XMAS provides a future-proof option for the integration of XL-MS data in structural modeling projects.

Acknowledgements – The Dutch Research Council (NWO) supported this research through funding of the large-scale proteomics facility Proteins@Work (project 184.032.201) embedded in the Netherlands Proteomics Centre. We further acknowledge funding through the European Union Horizon 2020 program INFRAIA project Epic-XS (Project 823839). Finally, this work is part of the research program NWO TA with project number 741.018.201, which is partly financed by the NWO.

Author contributions – IL: implementation, software design, writing; PA: conceptualization, software design, supervision, writing; AJ: implementation; RVH: implementation; AB: funding acquisition, supervision; RAS: conceptualization, funding acquisition, supervision, writing.

References

1. Luck,K., Kim,D.K., Lambourne,L., Spirohn,K., Begg,B.E., Bian,W., Brignall,R., Cafarelli,T., Campos-Laborie,F.J., Charlotheaux,B., *et al.* (2020) A reference map of the human binary protein interactome. *Nature*, **580**, 402–408.
2. Stumpf,M.P.H., Thorne,T., De Silva,E., Stewart,R., Hyeong,J.A., Lappe,M. and Wiuf,C. (2008) Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 6959–6964.
3. Venkatesan,K., Rual,J.F., Vazquez,A., Stelzl,U., Lemmens,I., Hirozane-Kishikawa,T., Hao,T., Zenkner,M., Xin,X., Goh,K. II, *et al.* (2009) An empirical framework for binary interactome mapping. *Nat. Methods*, **6**, 83–90.
4. Aebersold,R. and Mann,M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198–207.
5. Altelaar,A.F.M., Munoz,J. and Heck,A.J.R. (2012) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* 2013 **141**, **14**, 35–48.
6. Aebersold,R., Agar,J.N., Amster,I.J., Baker,M.S., Bertozzi,C.R., Boja,E.S., Costello,C.E., Cravatt,B.F., Fenselau,C., Garcia,B.A., *et al.* (2018) How many human proteoforms are there? *Nat. Chem. Biol.*, **14**, 206–214.
7. Savitski,M.M., Reinhard,F.B.M., Franken,H., Werner,T., Savitski,M.F., Eberhard,D., Molina,D.M., Jafari,R., Dovega,R.B., Klaeger,S., *et al.* (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* , **346**, 1255784.
8. Itzhak,D.N., Tyanova,S., Cox,J. and Borner,G.H.H. (2016) Global, quantitative and dynamic mapping of protein subcellular localization. *Elife*, **5**, e16950.
9. Clegg,C. and Hayes,D. (1974) Identification of neighbouring proteins in the ribosomes of *Escherichia coli*. A topographical study with the cross-linking reagent dimethyl suberimidate. *Eur. J. Biochem.*, **42**, 21–28.
10. Sun,T.T., Bollen,A., Kahan,L. and Traut,R.R. (1974) Topography of ribosomal proteins of the *Escherichia coli* 30S subunit as studied with the reversible cross-linking reagent methyl 4-mercaptobutyrimidate. *Biochemistry*, **13**, 2334–2340.
11. Steigenberger,B., Albanese,P., Heck,A.J.R. and Scheltema,R.A. (2020) To cleave or not to cleave in XL-MS? *J. Am. Soc. Mass Spectrom.*, **31**, 196–206.
12. Iacobucci,C., Piotrowski,C., Aebersold,R., Amaral,B.C., Andrews,P., Bernfur,K., Borchers,C., Brodie,N.I., Bruce,J.E., Cao,Y., *et al.* (2019) First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study. *Anal. Chem.*, **91**, 6953–6961.
13. Steigenberger,B., Pieters,R.J., Heck,A.J.R. and Scheltema,R.A. (2019) PhoX: An IMAC-Enrichable Cross-Linking Reagent. *ACS Cent. Sci.*, **5**, 1514–1522.
14. Liu,F., Lössl,P., Scheltema,R., Viner,R. and Heck,A.J.R. (2017) Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.*, **8**, 15473.
15. Lenz,S., Sinn,L.R., O’Reilly,F.J., Fischer,L., Wegner,F. and Rappsilber,J. (2021) Reliable identification of protein-protein interactions by crosslinking mass spectrometry. *Nat. Commun.*, **12**, 3564.
16. Rappsilber,J. (2011) The beginning of a beautiful friendship: Cross-linking/mass spectrometry and

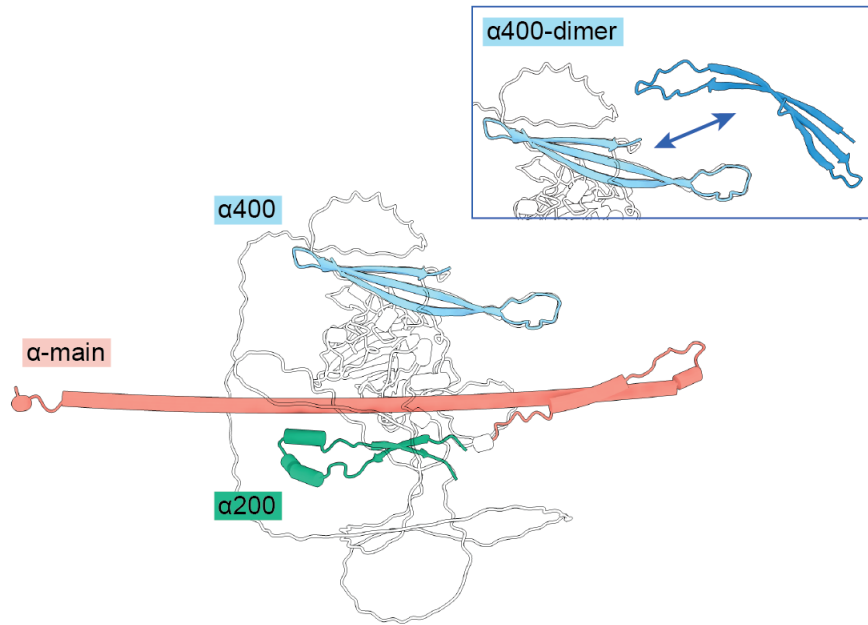
modelling of proteins and multi-protein complexes. *J. Struct. Biol.*, **173**, 530–540.

17. O'Reilly,F.J., Xue,L., Graziadei,A., Sinn,L., Lenz,S., Tegunov,D., Blötz,C., Singh,N., Hagen,W.J.H., Cramer,P., *et al.* (2020) In-cell architecture of an actively transcribing-translating expressome. *Scienc.*, **369**, 554–557.
18. Kosinski,J., Mosalaganti,S., von Appen,A., Teimer,R., DiGuilio,A.L., Wan,W., Bui,K.H., Hagen,W.J.H.H., Briggs,J.A.G.G., Glavy,J.S., *et al.* (2016) Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science*, **352**, 363–365.
19. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
20. Baek,M., DiMaio,F., Anishchenko,I., Dauparas,J., Ovchinnikov,S., Lee,G.R., Wang,J., Cong,Q., Kinch,L.N., Dustin Schaeffer,R., *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, **373**, 871–876.
21. Burke,D.F., Bryant,P., Barrio-Hernandez,I., Memon,D., Pozzati,G., Shenoy,A., Zhu,W., Dunham,A.S., Albanese,P., Keller,A., *et al.* (2021) Towards a structurally resolved human protein interaction network. *bioRxiv*, 10.1101/2021.11.08.467664.
22. Rantos,V., Karius,K. and Kosinski,J. (2021) Integrative structural modeling of macromolecular complexes using Assemblin. *Nat. Protoc.*, **17**, 152–176.
23. Honorato,R. V., Koukos,P.I., Jiménez-García,B., Tsaregorodtsev,A., Verlato,M., Giachetti,A., Rosato,A. and Bonvin,A.M.J.J. (2021) Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.*, **8**, 708.
24. Graham,M., Combe,C., Kolbowski,L. and Rappsilber,J. (2019) xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data. *bioRxiv*, 10.1101/561829.
25. de Graaf,S.C., Klykov,O., van den Toorn,H. and Scheltema,R.A. (2019) Cross-ID: Analysis and Visualization of Complex XL–MS-Driven Protein Interaction Networks. *J. Proteome Res.*, **18**, 642–651.
26. Bullock,J.M.A., Thalassinos,K. and Topf,M. (2018) Jwalk and MNXL web server: model validation using restraints from crosslinking mass spectrometry. *Bioinformatics*, **34**, 3584–3585.
27. Schiffrin,B., Radford,S.E., Brockwell,D.J. and Calabrese,A.N. (2020) PyXlinkViewer: A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. *Protein Sci.*, **29**, 1851–1857.
28. Kosinski,J., von Appen,A., Ori,A., Karius,K., Müller,C.W. and Beck,M. (2015) Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.*, **189**, 177–83.
29. Goddard,T.D., Huang,C.C., Meng,E.C., Pettersen,E.F., Couch,G.S., Morris,J.H. and Ferrin,T.E. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.*, **27**, 14–25.
30. Van Zundert,G.C.P. and Bonvin,A.M.J.J. (2015) DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. *Bioinformatics*, **31**, 3222–3224.
31. Hevler,J.F., Lukassen,M. V, Cabrera-Orefice,A., Arnold,S., Pronker,M.F., Franc,V. and Heck,A.J.R. (2021) Selective cross-linking of coinciding protein assemblies by in-gel cross-linking mass spectrometry.

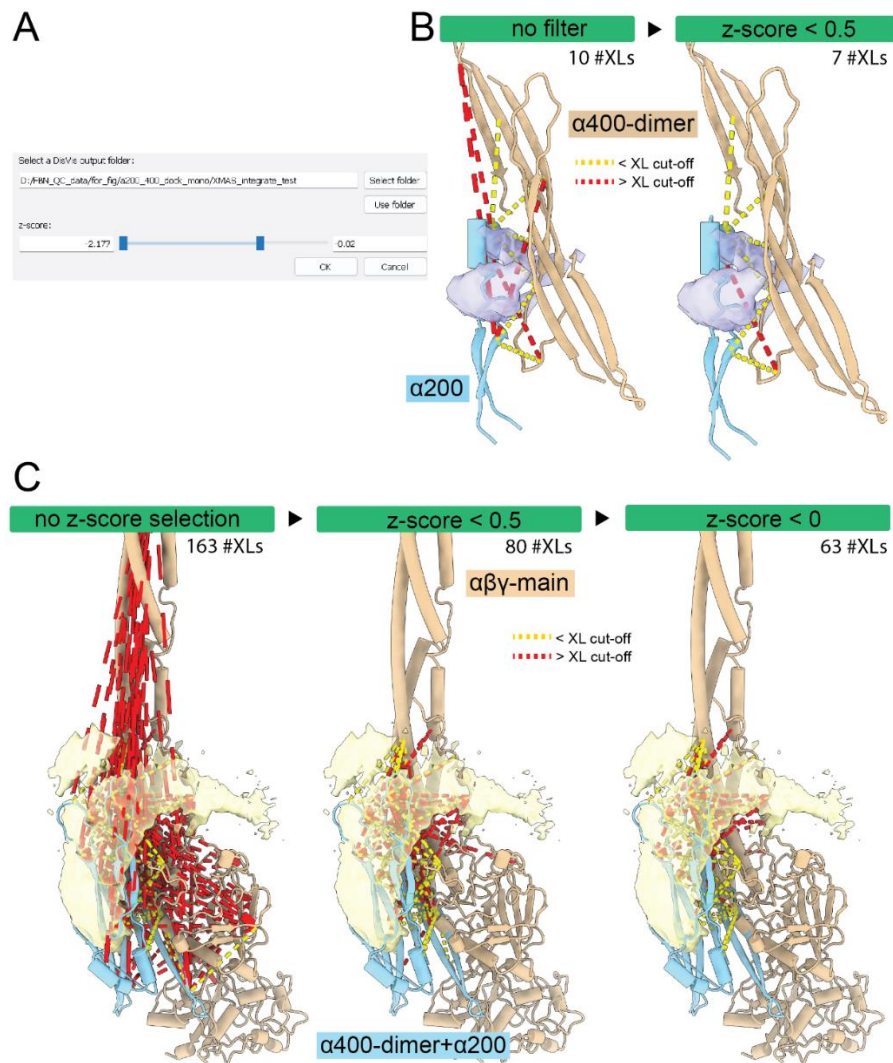
EMBO J., **40**, e106174.

32. Klykov,O., Van Der Zwaan,C., Heck,A.J.R., Meijer,A.B. and Scheltema,R.A. (2020) Missing regions within the molecular architecture of human fibrin clots structurally resolved by XL-MS and integrative structural modeling. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 1976–1987.
33. Rodrigues,J.P.G.L.M., Teixeira,J.M.C., Trellet,M. and Bonvin,A.M.J.J. (2018) pdb-tools: a swiss army knife for molecular structures. *F1000Research*, **7**, 1961.
34. Weisel,J.W. and Litvinov,R.I. (2017) Fibrin Formation, Structure and Properties. *Subcell. Biochem.*, **82**, 405–456.
35. Klykov,O., Steigenberger,B., Pektaş,S., Fasci,D., Heck,A.J.R. and Scheltema,R.A. (2018) Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. *Nat. Protoc.*, **13**, 2964–2990.
36. Chen,Z.-L., Meng,J.-M., Cao,Y., Yin,J.-L., Fang,R.-Q., Fan,S.-B., Liu,C., Zeng,W.-F., Ding,Y.-H., Tan,D., *et al.* (2019) A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.*, **10**, 3404.
37. Fischer,L., Chen,Z.A. and Rappsilber,J. (2013) Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics*, **88**, 120–128.
38. Leitner,A., Bonvin,A.M.J.J., Borchers,C.H., Chalkley,R.J., Chamot-Rooke,J., Combe,C.W., Cox,J., Dong,M.Q., Fischer,L., Götze,M., *et al.* (2020) Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry. *Structure*, **28**, 1259–1268.
39. Fasci,D., van Ingen,H., Scheltema,R.A. and Heck,A.J.R. (2018) Histone Interaction Landscapes Visualized by Crosslinking Mass Spectrometry in Intact Cell Nuclei. *Mol. Cell. Proteomics*, **17**, 2018–2033.
40. Chavez,J.D., Mohr,J.P., Mathay,M., Zhong,X., Keller,A. and Bruce,J.E. (2019) Systems structural biology measurements by in vivo cross-linking with mass spectrometry. *Nat. Protoc.*, **14**, 2318–2343.
41. Klykov,O., Kopylov,M., Carragher,B., Heck,A.J.R., Noble,A.J. and Scheltema,R.A. (2022) Label-free visual proteomics: Coupling MS- and EM-based approaches in structural biology. *Mol. Cell*, **82**, 285–303.

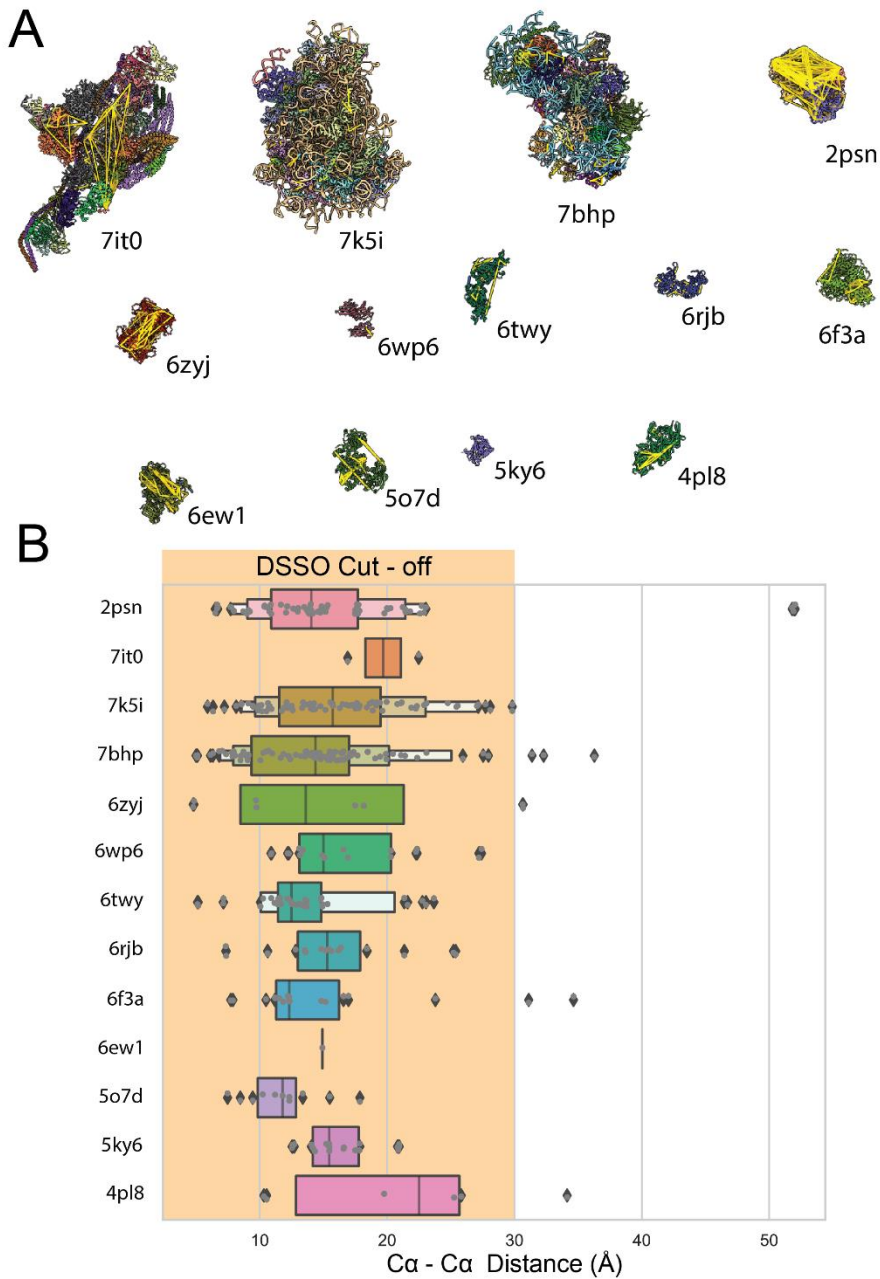
Supplementary figures



Supplementary Figure 1. Individual structures used for integrative modeling of the fibrin chain. The α 200 (α 220-265) and α 400 (α 423-487) regions extracted from the AlphaFold model (downloaded from <https://alphafold.ebi.ac.uk/entry/P02671>) of the human fibrinogen α chain (Uniprot ID: P02671). These models were used in this work in combination with newly acquired XL-MS data and XMAS to extend the previously available structural model (PDBdev ID: 0000030).



Supplementary Figure 2. Examples of the integrative modeling pipeline through XMAS implemented in this work. (A) The “Integrate” XMAS module was used to filter out putative false positive restraints determined by DisVis using the z-score selector. The restraints automatically (dis)appear according to the selected cut-off while the slider moves, allowing the user to evaluate the data in real-time. In addition, color-coding based on the distance cut-off set in the DisVis restraints file is automatically read and applied (in this case *pseudobonds* < 5 Å or > 25 Å are colored in red). In these two examples, the number of selected crosslinks (#XLs) further used for docking the α200 to α400 dimer (B), and the resulting α400-200 module to the αβγ-main chain (C) are indicated for 2 and 3 steps of z-score cut-offs, respectively.



Supplementary Figure 3. Evaluation of proteome-wide XL-MS experiment quality. XMAS integrated into ChimeraX allows the smooth handling of multiple large structures and datasets. In this example, the thirteen PDB structures (PDB IDs indicated) with the most identified crosslinks were selected from a proteome-wide XL-MS experiment on human cell lysate (25). **(A)** Crosslinks can be mapped in all structures within seconds and **(B)** the distances of the “shortest” *pseudobonds* can be plotted for each peptide pair, to avoid biases due to homomeric complexes, allowing for a rapid and thorough evaluation of the data quality.