nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-018-0074-x

Efficient and robust proteome-wide approaches for cross-linking mass spectrometry

Oleg Klykov[1,2], Barbara Steigenberger[1,2], Sibel Pektaş[1,2], Domenico Fasci[1,2], Albert J. R. Heck[1,2], Richard A. Scheltema[1,2*]

Cross-linking mass spectrometry (XL-MS) has received considerable interest, owing to its potential to investigate protein–protein interactions (PPIs) in an unbiased fashion in complex protein mixtures. Recent developments have enabled the detection of thousands of PPIs from a single experiment. A unique strength of XL-MS, in comparison with other methods for determining PPIs, is that it provides direct spatial information for the detected interactions. This is accomplished by the use of bifunctional cross-linking molecules that link two amino acids in close proximity with a covalent bond. Upon proteolytic digestion, this results in two newly linked peptides, which are identifiable by MS. XL-MS has received the required boost to tackle more-complex samples with recent advances in cross-linking chemistry with MS-cleavable or reporter-based cross-linkers and faster, more sensitive and more versatile MS platforms. This protocol provides a detailed description of our optimized conditions for a full-proteome native protein preparation followed by cross-linking using the gas-phase cleavable cross-linking reagent disuccinimidyl sulfoxide (DSSO). Following cross-linking, we demonstrate extensive sample fractionation and substantially simplified data analysis with XlinkX in Proteome Discoverer, as well as subsequent protein structure investigations with DisVis and HADDOCK. This protocol produces data of high confidence and can be performed within ~10 d, including structural investigations.

Introduction

Structural proteomics and, more specifically, XL-MS, has gained a large amount of traction in recent years as a supplemental method to protein structure techniques such as electron microscopy (EM), NMR and crystallography. XL-MS generally provides distance constraints of lower resolution than those of other techniques but is able to pinpoint residues in close proximity to the interaction interfaces between individual subunits of protein complexes of any size in solution. Provided there are sufficient cross-links, this technique can even allow for the detection and definition of protein interfaces. The information obtained from XL-MS experiments has in many cases been successfully leveraged to produce final protein complex models[1–5]. With such experiments, the cross-linking data are combined with crystal structure and cryo-EM data to develop a final model. In the case that the number of cross-links derived from a complex lysate is not sufficient to make reliable conclusions, the protein complex of interest can be enriched to increase the chances of extracting more cross-links.

The distance information in elucidated protein structures is obtained by use of bifunctional cross-linking molecules that link two amino acids in close proximity to one another with a covalent bond. Upon proteolytic digestion of proteins or protein–protein complexes, four distinct peptide products are formed. The first consists of single peptides not captured by the cross-linking reagent, which therefore yields no structural information. The second group consists of monolinks or dead-end links[6], constituting single peptides captured by only one side of the cross-linking reagent, because the other end of the linker has been quenched. These links yield structural information about the surface-exposed regions of the proteins. The third group consists of loop links, which are composed of single peptides with two amino acids captured by a single linker. This group of internally cross-linked peptides yields limited structural information owing to the close proximity of the linked amino acids. The fourth, and most information-rich group, consists of two peptides captured by the cross-linking reagent; this group yields valuable distance information about protein tertiary structure (if the two peptides originate from the same protein) or about protein quaternary structure (if the two peptides originate from different proteins).

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands. [2]Netherlands Proteomics Center, Utrecht, the Netherlands. *e-mail: R.A.Scheltema@uu.nl

XL-MS experiments up to this point have primarily been performed on purified recombinant proteins or protein complexes[7–13]. Working with purified systems has the benefit of sufficient signal for the detection of the generally low-abundant cross-links. In addition, only a handful of protein sequences must be searched during spectrum matching, allowing for a relatively straightforward data analysis. In contrast to purified proteins, whole-cell lysates lead to comprehensive interaction networks, which can provide far-reaching insights into the relationships between proteins as a result of stimulation. For example, it has been shown that cross-linking complete mitochondria and intact nuclei is possible[14–16]. However, data analysis and deep proteome coverage remain major bottlenecks in these more advanced experiments. Several laboratories, including ours, have set out to circumvent these issues by developing user-friendly analysis software dedicated to unraveling the identities of cross-linked peptides from complex mixtures[17,18]. Adoption of MS-cleavable cross-linkers[19–21] substantially simplifies and even facilitates these identifications, owing to the presence of characteristic peaks in the recorded spectra. Improvements in sample preparation techniques[22,23] or preservation of protein structure with formaldehyde[24] allow researchers to keep protein assemblies intact and to capture snapshots of the processes inside the living cell[25].

Here, we present an optimized protocol for XL-MS that scales from structural analysis of single protein complexes to whole-cell lysates. In this protocol, both in vitro–reconstituted complexes and purified complexes are considered to be low-complexity samples, whereas cell lysates are considered to be high-complexity samples. As a showcase study, we demonstrate application of our protocol to reconstituted BSA protein and non-small-cell lung cancer PC9 cells. We selected the MS-cleavable amino-reactive DSSO cross-linker because it shows high efficiency, is commercially available and has successfully been applied in a large number of cross-linking experiments[14,24–28]. In addition to an optimized soft lysis and digestion procedure, we present a technique for the efficient enrichment of cross-linked peptides that uses off-line fractionation based on strong cation exchange (SCX) chromatography[28]. Optimized LC–MS/MS parameters for current, state-of-the-art MS instrumentation are presented that allow for the efficient detection of DSSO-cross-linked peptides. Detailed data analysis descriptions are available, together with an overview of different reagents and research tasks, to assist researchers aiming to perform these experiments. As a last step, we demonstrate validation of a recorded dataset on a well-resolved crystal structure.

Overview of the procedures

This protocol consists of six major steps and is presented for both low- and high-complexity mixtures. The first step involves lysis of the cells under investigation (Fig. 1, top left). General lysis methods often tend to disrupt PPIs or disturb native protein conformations (e.g., SDS, over a period of time, denatures proteins). A more gentle approach uses a syringe for lysis to liberate the protein complexes from the cellular environment while leaving their interactions intact. This makes the protein complexes amenable to cross-linking, as the reagents used are generally not cell-permeable. After incubating the protein mixture with DSSO, standard shotgun proteomics workflow steps are applied, consisting of protein denaturation, reduction, alkylation and digestion (Fig. 1, top right). Detergents are removed by precipitation or phase extraction approaches, and digestion is carried out with a selection of proteases—e.g., LysC/trypsin or chymotrypsin—to obtain a unique set of cross-linked peptides normally not visible with a single protease. Further purification of the sample (e.g., removal of salts from the buffers) is performed using solid-phase extraction (SPE) with larger pore-size C_{18} material. Following digestion and purification, well-established pre-fractionation techniques are applied that specifically enrich for cross-linked peptides (Fig. 1, right). This can be achieved with size-exclusion chromatography (SEC)[29] or SCX[28]. The individual fractions are analyzed by shotgun mass spectrometry, for which an optimized acquisition protocol for the Orbitrap (OT) Fusion and OT Fusion Lumos has been developed (Fig. 1, bottom right).

This acquisition protocol makes use of a MS2–MS3 strategy, in which each precursor is fragmented and recorded; it has been optimized to exclusively break the MS-cleavable cross-linker while keeping the individual peptides intact. This simultaneously provides two pieces of information: whether the precursor ion represents cross-linked peptides, and the masses of the individual peptides. For those cases in which it is likely that the precursor represents cross-linked peptides, the precursor (MS1) is isolated and fragmented (MS2), followed by selection and fragmentation of fragment peaks originating from the cross-linked peptide pair (MS3), providing insight into their identities. The main advantage of the extra MS3 step is that each peptide can be identified from a single spectrum,

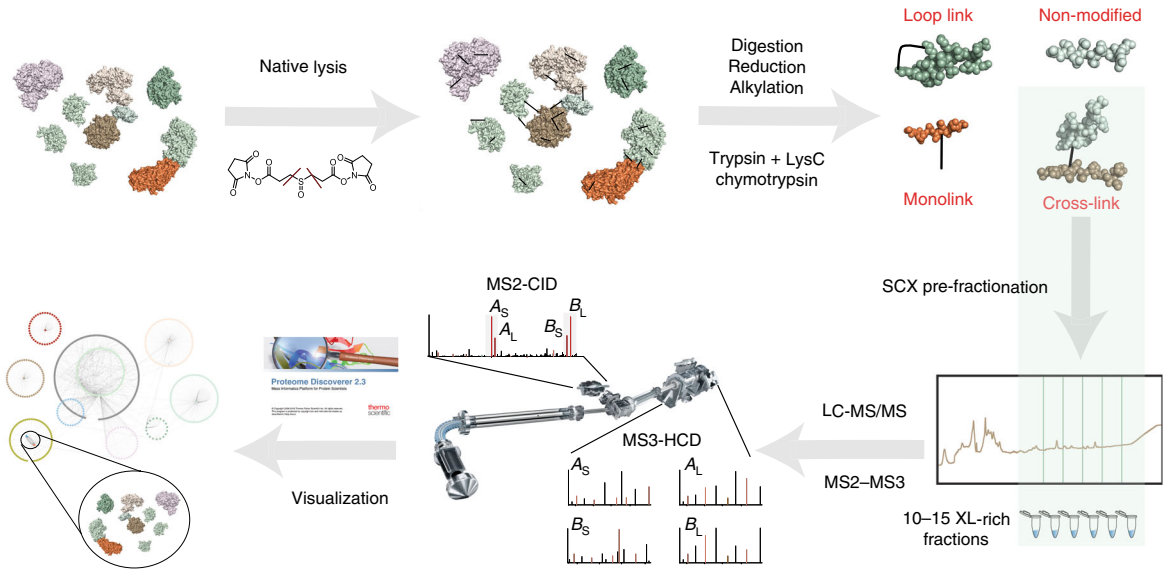

Fig. 1 | Generic workflow for XL-MS experiments. (Top left) Cells or tissue are lysed gently, leaving protein complexes intact. (Top right) After optimized incubation with the cross-linking reagent and proteolytic digestion, four peptide products can be categorized. (Right) XL peptides are enriched and pre-fractionated, using techniques such as SCX. (Bottom middle) Advanced data acquisition techniques utilizing multiple steps of fragmentation techniques are used to identify the peptides. (Bottom left) XlinkX for Proteome Discoverer v. 2.3 is used to identify the cross-linked peptides (Copyright Thermo Fisher Scientific Inc. Used by permission). The resulting data can consequently be visualized (e.g., with xiNET and Cytoscape) and integrated into structural modeling software (e.g., HADDOCK, I-TASSER and DisVis).

resulting in higher-quality evidence for each peptide. However, the downside of this approach is that more time must be spent per precursor/cross-linked peptide pair, leading to a decrease in identification rates.

The acquired RAW data files can be seamlessly analyzed with the XlinkX search engine integrated as a node in the well-established Proteome Discoverer software suite (https://xlinkx.hecklab.com/) (Fig. 1, bottom center). The data analysis handles spectral processing, potential cross-link detection and peptide pair identification at low false-discovery rates (FDRs). Integration into the Proteome Discoverer environment ensures that the results can be navigated in a user-friendly manner. Finally, XlinkX provides direct connections to existing software for visualizing cross-links such as xiNet[30] or Cytoscape[31] (Fig. 1, bottom center), and generates convenient input for further structural analysis and interpretation of existing crystal structures with Chimera[32] or pyMol[33]. In addition, XlinkX plays well with the I-TASSER[34–36], DisVis[37,38] (http://milou.science.uu.nl/cgi/services/DISVIS/disvis/) and HADDOCK[39,40] (http://milou.science.uu.nl/services/HADDOCK2.2/) software solutions, which make computational predictions for new structures and interaction interfaces based on the detected cross-links.

Advantages and limitations

The ability of XL-MS to elucidate PPIs complete with structural information is a major advantage over currently widespread techniques such as affinity purification MS. However, proteome-wide cross-linking applications have so far been limited to laboratories focusing on developing these technologies[41,42]. We attempt to facilitate further adoption by providing a detailed protocol to researchers worldwide that utilizes commercially available reagents, including the lysine-reactive cross-linking reagent DSSO, which links neighboring lysine pairs at a maximum distance of 30 Å. The protocol is set up to provide information for both purified protein samples and complex mixtures. Even though the focus is on DSSO, the steps are applicable to any reagent with gas-phase cleavable features, such as disuccinimidyl dibutyric urea (BuUrBu or DSBU)[20]. Non-cleavable cross-linking reagents are, however, not indicated for complex mixtures, as these reagents require more extreme computational approaches to performing database searches. Additional hurdles include complex data analysis, which we have attempted to simplify with the integration of XlinkX into Proteome Discoverer, making a user-friendly XL-MS data analysis approach accessible to a large audience of researchers.

A major limitation of XL-MS experiments remains: per detected PPI, little information is available in terms of the number of cross-linked peptides. Typically, only the top 20–30% of proteins are detected, and only the top 100–500 proteins have sufficient depth of coverage for successful modeling (see Anticipated results). This can potentially be alleviated through the development of enrichable chemical cross-linkers. The challenges of creating such molecules lie in the ability of the molecules to retain reactivity of the amine reactive groups while keeping their agility in regard to entering the protein structure. Several molecules have to date been reported to achieve these goals and have been successfully applied[43–45]. Another approach to alleviate the relatively low level of detection would be to include cross-linking on acidic groups (e.g., dihydrazide sulfoxide (DHSO)[46]), which could potentially open up new locales of the protein complex for detection. However, these molecules have only recently become commercially available, and therefore were not included in our analysis. Alternatively, in this protocol, we attempt to overcome the limit of detection by extensive pre-fractionation. Such an approach works well and successfully yields thousands of detected cross-linked peptides; however it also has limitations. To illustrate, our selected SCX approach uses charge as the separation mechanism; this is a property that is exuberated for cross-linked peptides, but is not necessarily unique. From the number of detected normal peptides, it is evident that a large background is still present, and therefore a large number of fractions are required to reach sufficient depth.

For purified complexes, the opposite holds true. Mainly driven by recent advances in mass spectrometry, XL-MS has become increasingly capable of generating hundreds[4,26,47,48], and in some cases even thousands[28,49], of cross-linked peptides. Given the limited complexity, the use of non-cleavable cross-linking reagents poses no barrier to successful data analysis. From the generated distance constraints, successful modeling of protein complexes can be achieved and has been reported. The major limitation of such an approach is that the researchers need to find sufficiently successful routes for purification of the protein complex before investigation. Such a process can take a long time to complete, and doubts about the native state of the protein complex remain. For instance, enrichable handles such as His-tags are often expressed on one of the proteins in the complex. Such a handle can be used to successfully enrich for the complex, but can also have repercussions for the protein complex, such as allosteric modifications and/or complete loss of interactors.

Potential applications

XL-MS in recent years has received increased attention for its ability to elucidate pseudo-atomic models for protein complexes and has now even seen application in highly complex backgrounds. It has already been demonstrated to be a useful tool for identifying purified protein complexes, and together with X-ray crystallography, hydrogen–deuterium exchange, NMR and cryo-EM, it has become one of the staples in the structural biologist's toolbox. An area in which XL-MS can really take center stage is its application to transient interactions, which are so labile that any sample preparation step can potentially disrupt them and inevitably result in the inability to measure them. Even though PPIs within a cell might range from very stable to transient, for the cross-linking reagent, there are no differences, and all interactions in close proximity are potentially caught by the cross-linking reagent. After successful cross-linking of residues in close proximity, the interaction has been stabilized and normal analysis steps can be executed without losing the transient interaction. Such an approach would require cell-permeable cross-linking reagents, for which a number of compounds have been proposed[23,50–52], but so far none have been widely applied.

The second area in which XL-MS promises to be very useful is the mapping of dynamic behavior of the protein complex. A well-known example is the *TRiC* or *CCT* complex, which was reported to have a lid structure that can be closed by application of ATP. Upon quantifying detected cross-linked peptides[53], this lid behavior could be mapped[54]. Further steps to streamline this process have recently been taken by either integration of quantifiable labels directly into the cross-linking reagents[45,55–58] or the application of TMT labeling to the cross-linked peptides[59]. Even though, in this particular example, the lysine residues are occupied by the cross-linking reagent, tandem mass tag (TMT) labeling can still be successfully applied to the N terminus of the peptides.

Both of these application areas are well supported by XlinkX and, combined with support for the spatial restraints coming from cross-link identifications in modern protein modeling engines (HADDOCK, I-TASSER and ROSETTA), will allow researchers to immediately postulate structural models from discovery experiments[60,61]. With time, enrichable cross-linking chemistry will evolve to allow this for most protein complexes in complex mixtures.

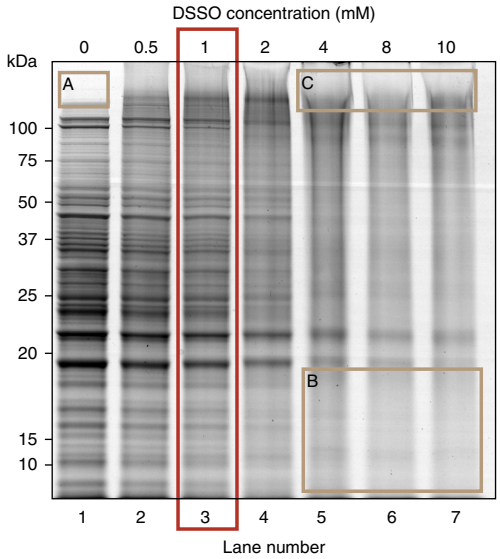

Fig. 2 | Optimization of cross-linking reaction conditions by titration. Lane 1 corresponds to the experiment when no cross-linking agent was added; the high-MW region is empty (box A). For lanes 5–7, the cross-linking agent was added in high concentrations, leading to a less intense low-MW region (box B), whereas the upper part of the gel is oversaturated with high-MW species (box C). Lane 3, with an intermediate 1 mM concentration of cross-linker, shows the presence of proteins in both regions and is selected as the optimal cross-linker concentration.

Experimental design

Preparation of cross-linked lysate

We make use of two types of buffers for cleanup and resuspension of the proteins/protein complexes. PBS is intended for washing samples extensively to remove traces of any buffer compounds harmful to the cross-linking reaction. For example, primary amines must be avoided when performing cross-linking reactions because they interfere with *N*-hydroxysuccinimide (NHS)-ester derivatives such as DSSO. In addition, specific cross-linker properties must be taken into account, as, for example, thiols should not be present in a cross-linking reaction mixture in the case of thiol-cleavable reagents. The cross-linking lysis buffer (i.e., the buffer in which the cross-linking reaction will be performed, but which does not contain any cross-linking reagent), is intended to resuspend the proteins before further steps. The cross-linking lysis buffer is supplemented with protease inhibitors and minor amounts of a reducing agent directly before the cross-linking reaction. Protease inhibitors are required to prevent protein degradation owing to the presence of endogenous proteases, whereas the role of reducing agents is to prevent oxidative stress and keep proteins in their native state.

Low-complexity samples can be directly resuspended in the cross-linking lysis buffer. For cell culture samples, soft lysis is achieved by pushing the cells through a syringe needle to preserve organelles and protein assemblies. A 27¾-gauge syringe is recommended for most mammalian cell types, as this is one of the smallest diameters available and it is not required to preserve intracellular vesicles for this procedure. In the case of bacterial, plant or fungal samples, harsher lysis conditions may be required. Before further experiments, the protein concentration can be estimated with a Bradford or bicinchoninic acid assay (BCA). We recommend a final protein concentration in the range of 1–3 mg/ml to avoid aggregation of the proteins and potential precipitation, leading to detection of cross-linked peptide pairs originating from non-interacting proteins.

Optimization of cross-linking reaction conditions

Before the cross-linking reaction, the optimal concentration of the cross-linking reagent for the acquired sample concentration must be determined. To achieve this, different aliquots of the sample can be cross-linked with a range of cross-linking reagent concentrations and visualized by SDS–PAGE (Fig. 2). Visualization is dependent on sample amount; visualization for high amounts of input can be done with standard Coomassie staining, and visualization for low amounts of input can be done with silver staining. For low-complexity samples, the cross-linking reagent concentration range can be built from 0.1 mM to a 10–15 molar excess of cross-linking reagent to protein, whereas

for high-complexity samples, the range of cross-linking reagent can be from 0.1 to 5 mM. Note that the limited reaction efficiency of the currently available cross-linking reagents will substantially decrease identification rates for low concentrations. By contrast, high concentrations will block lysines, leading to incomplete digestion with lysine-cleaving proteases and, in addition, can lead to over-cross-linking, in which individual proteins are captured by the cross-linking reagent and aggregate. It is therefore critical to use the lowest possible cross-linker concentration that still initiates the reaction.

Protein digestion

The native protein complex state is captured during the cross-linking experiment, and the still-present nucleic acids can be sheared to prevent precipitation after protein denaturation. In our protocol, we achieve this by adding benzonase to a final concentration of 1% (vol/vol), but in principle, any enzyme with similar properties can be used. Natively folded proteins, supported by disulfide bridges, are generally more difficult to digest because the protease tends to be sterically hindered. To achieve the maximum efficiency for protein reduction and alkylation, we use tris(2-carboxyethyl)phosphine hydrochloride (TCEP), which can be used simultaneously with 2-chloroacetamide (CAA) at elevated temperatures. Here, we perform the reduction/alkylation reaction at 37 °C, with longer incubation times to prevent modification of proteins by urea artifacts; these conditions are not needed if other more stable chaotropic agents are used. For the digestion, a short predigestion with LysC, followed by standard overnight trypsin digestion, is recommended to increase protein sequence coverage. Alternatively, the ultimate coverage of the investigated proteins can potentially be increased by using other proteases with cleavage sites different from ones involved in the cross-linking reaction, e.g., chymotrypsin[62]. It is beneficial to perform the digestion procedure in solution, as the efficiency of in-gel digestion is lower and leads to fewer identifications. Although LysN and chymotrypsin can easily be adapted for in-gel digestion, other proteases can be less efficient in this approach. We note that for whole-proteome approaches, the benefit of using alternative proteases is less clear, and therefore such an approach is recommended mainly for low-complexity samples.

Pre-fractionation of cross-linked peptides

Fractionation increases the depth of analysis required to reach the relatively low-abundant cross-linked peptides. The specific properties of cross-linked peptides allow for separation of single peptides from cross-linked peptides by either SCX[28], given that cross-linked tryptic peptides—as compared to single tryptic peptides—tend to carry extra charges, or SEC[29], given that cross-linked peptides are generally larger than single peptides. This procedure can be performed in Stage-Tip format for low-complexity samples. However, for all sample types, the best coverage for samples with a loading amount of ≥100 µg is achieved with either HPLC- or ÄKTA-assisted fractionation. In our laboratory, we found 500 µg to be the optimal amount. Quality control of the setup before separation can be performed with a partial tryptic digest of BSA[62]. When the BSA run meets expectations in terms of signal and reproducibility, the sample can be fractionated (Fig. 3). With the described HPLC SCX setup (Table 1), 25–50 fractions containing ~50 µl each can be collected. When samples other than whole-cell lysates are fractionated, 25 fractions are sufficient.

Desalting

The samples at this point still contain contaminants and reagents that potentially affect peptide ionization and LC separation. This is especially relevant in those cases in which SCX separation with nonvolatile salts is used and/or when combining SCX fractions. In our laboratory, we found the commercially available SPE C_{18} kits to be most useful. Cross-linked peptides tend to be longer than normal tryptic peptides, owing to missed cleavages where the linker has bound to the peptide, requiring materials with larger pore size to prevent sample losses. If the total amount of peptide material does not exceed 20 µg, we recommend applying C_{18} Stage-Tips, for which a detailed protocol is available[63]. In cases in which a large number of samples with peptide amounts ≤50 µg are used, we recommend commercially available Oasis HLB or Oasis WCX 96-well elution plates.

LC–MS/MS analysis

The amount of peptide material in each of the fractions should be estimated before LC/MS analysis. This step is essential to accurate estimation of the loading amounts that provide enough signal for the detection of the mostly low-abundant cross-linked peptide species while preventing overloading of

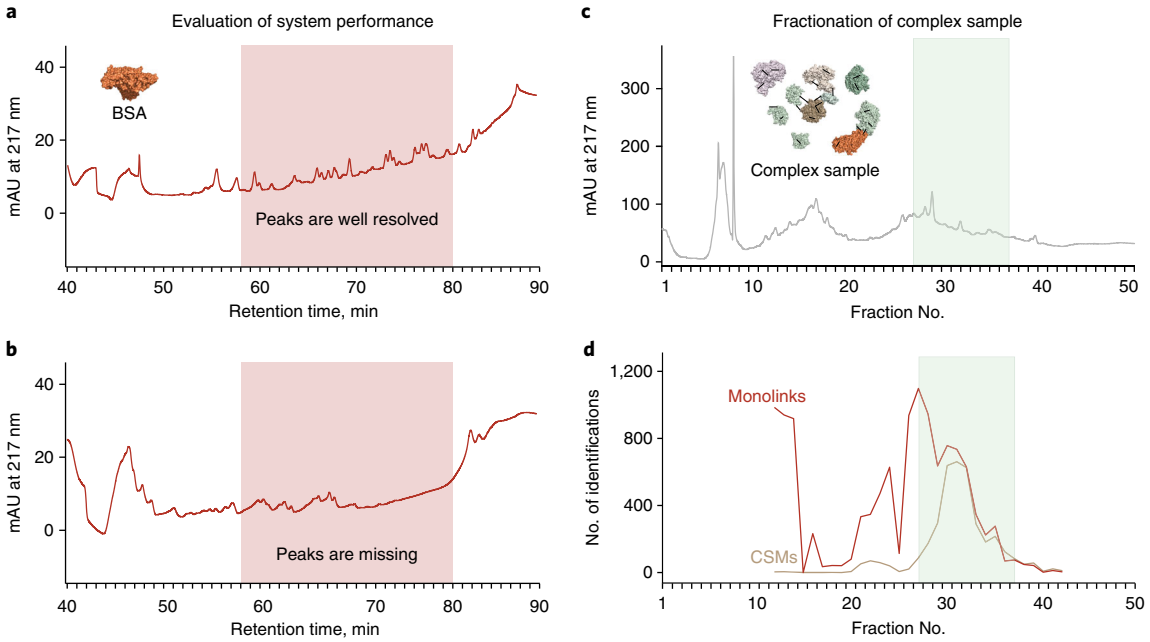

Fig. 3 | SCX fractionation profile for whole PC9 cell lysate samples. a, A SCX chromatogram for a well-performing system recorded at a wavelength of 217 nm. **b**, A SCX chromatogram for separation on an unequilibrated SCX column recorded at a wavelength of 217 nm. **c**, Snapshot of a SCX chromatogram with 50 collected fractions recorded at a wavelength of 217 nm. **d**, Distribution of the number of identified cross-links (number of entries from the cross-link spectral match, or CSM, table) and monolinks (number of entries annotated with water-quenched linker as a modification in the peptide spectral match, or PSM, table) over the analyzed fractions. The most CSM-rich fractions are highlighted.

Table 1 | HPLC–SCX fractionation settings used during a 65- or 120-min gradient

Time interval (min)	LC gradient (% B)
Pump 1 (trap column)	
0–5 (0–15)	0–0
5–15 (15–35)	100–100
15–17 (35–40)	25–25
17–65 (40–120)	0–0
Pump 2 (separation column)	
0–17 (0–42)	0–2
17–22 (42–50)	2–3
22–32 (50–60)	3–8
32–40 (60–70)	8–20
40–48 (70–80)	20–40
48–53 (80–86)	40–90
53–56 (86–90)	90–90
56–56.5 (90–91)	90–0
56.5–65 (91–120)	0–0

For the methods of 65 and 120 min, the indicated times refer to the total analysis times rather than actual gradient time, which are 39 and 58 min, respectively. LC and MS parameters are provided in parentheses for the 120-min method. Fractions are collected every 2 min from 15 to 60 min for the 65-min method and every 1 min from 40 to 90 min for the 120-min method.

the LC column. This is especially helpful for cases in which extensive pre-fractionation of peptides is used. Detection of peptide concentration after digestion can be done with commercially available colorimetric or fluorometric assays[64]. In the HPLC–SCX pre-fractionation setup described here, the amount of sample to inject for each collected fraction can be estimated from the LC–UV trace after determination of a select number of fractions. It is essential to perform nanoflow LC separation to

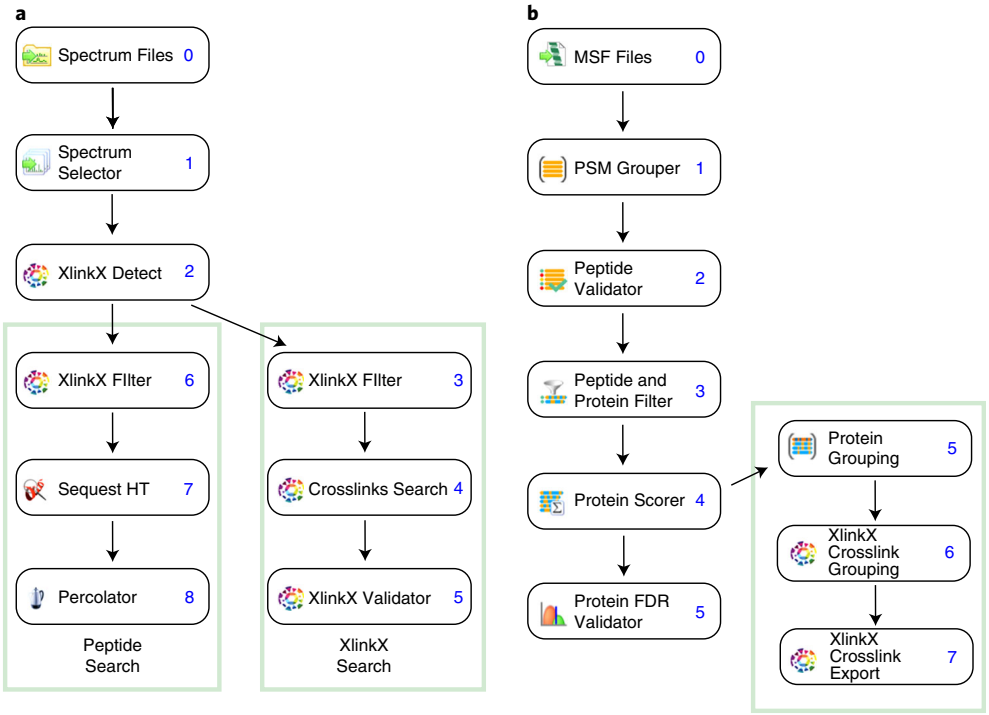

Fig. 4 | Schematic representation of Proteome Discoverer workflows for identification of cross-linked peptides. **a**, Processing workflow. **b**, Consensus workflow. Copyright Thermo Fisher Scientific Inc. Used by permission.

improve chromatographic separation. MS equipment must be able to perform acquisition up to the MS3 level. At the MS1 level, the survey scan is recorded in the OT instrument at high resolution. For selected precursors, collisional-induced dissociation (CID) is applied, and signature peaks for the cross-linkers are recorded at middle resolution as the MS2 level. Fragments exhibiting patterns associated with the used cleavable cross-linker are further subjected to a low-resolution MS3 scan in the ion trap (IT). Before starting the acquisition, we recommend performing quality control of the LC–MS/MS system, for which a detailed description is available from Giansanti et al.[62].

Data analysis with XlinkX

This step is split into processing and consensus workflows and is exemplified in Fig. 4. Standard cross-linking workflows can be found as common templates in Proteome Discoverer, and detailed settings are described in Supplementary Table 1. In the processing workflow (Fig. 4a), the first two nodes, called 'Spectrum Files' and 'Spectrum Selector', are commonly used in most Proteome Discoverer processing workflows, and their settings can remain default. The next node is 'XlinkX Detect', where the used cross-linker and MS acquisition strategy can be specified that work with different fragmentation strategies such as collision induced dissociation (CID), higher-energy collision-induced dissociation (HCD), or electron transfer dissociation (ETD). Depending on the cross-link chemistry, a number of analysis options are available. For setting up a new cross-linker as a chemical modification, refer to Supplementary Fig. 1 and the Supplementary Tutorial. When using non-MS-cleavable cross-linkers, the 'NonCleavable' option should be set. For MS-cleavable cross-linkers such as DSSO or DSBU, depending on the acquisition strategy 'MS2_MS2' (e.g., CID/ETD), 'MS2_MS3' (e.g., CID/MS3 HCD) or 'MS2_MS2_MS3' (e.g., CID/ETD/MS3 HCD) should be set. Then the workflow splits into two independent branches, where the node 'XlinkX Filter' regulates the flow of fragmentation spectra into two branches. The 'Peptide Search' branch executes a standard proteomics search with a peptide search engine and the FDR correction method of choice on those fragmentation spectra that do not contain reporter ions indicative of the presence of a cross-linker. In addition to non-cross-linked peptides, partially modified peptides can be identified at this step by setting relevant 'Dynamic Modifications' (e.g., 'type 0 cross-links' or monolinks). The 'Crosslinks' branch consists of the 'XlinkX Search' and 'XlinkX Validator' nodes, which perform the search and validation of the fragmentation spectra that contain reporter ions indicative of the presence of a cross-linker. The consensus workflow (Fig. 4b) consists of the following standard nodes: 'MSF Files', 'PSM Grouper',

'Peptide Validator', and 'Peptide and Protein Filter', which is connected to 'Protein Scorer', which, in turn, is connected to 'Protein FDR Validator' and 'Protein Grouping'; their settings can remain default. For the cross-linking workflow, two additional nodes are introduced: 'XlinkX Crosslinks Grouping' combines detected cross-link spectra matches (CSMs) with cross-links, and 'XlinkX Crosslink Export' exports the results in the format for the xiNET visualization toolkit[30].

Search output

The output of the cross-linking search nodes consists of several tables. The first table, 'Crosslink Spectra Match' (CSM), analogous to the 'Peptide Spectra Match' (PSM) table, contains the identification information for each set of fragmentation spectra (in the case of the 'MS2–MS3' strategy used here, the initial CID scan revealing the presence of the diagnostic ions is combined with the associated MS3 scans in a single set of scans). This table can contain the same peptide pair identity several times, owing to the presence of the peptide pair with different modifications and/or charge states. The 'Crosslinks' table contains the on-sequence identity-grouped 'CSM' entries for a direct overview of the relevant structural information yielded by the experiment. When relevant cross-linker modifications are set in the peptide search branch (node number 7 in the 'Peptide Search' path in Fig. 4a), monolinks (dead ends or type 0 cross-links) are shown as peptide modifications in the PSM table. The search results in two additional files that can be directly loaded into the xiNET online visualization platform[30]. In our laboratory, we have found this visualization tool especially useful for illustrating the interaction network of a single protein or protein complex after all initial analysis steps involved (Fig. 5).

Additional data validation

As an additional validation step, the identified cross-links can be mapped onto a protein complex with a well-resolved structure. In principle, any resolved protein complex with a high number of identified cross-links can be used. When preparing whole-cell lysates of human cell lines, we found it useful to map identified cross-links of the human ribosome (Fig. 6a,b), a complex that is highly abundant in many cells and has a well-resolved structure. With the current protocol, hundreds of distance restraints within the DSSO maximum cross-linking distance of 30 Å can be identified and mapped onto the human ribosome structure with 3.6-Å resolution[65]. Before modeling the protein complex, the identified inter-subunit cross-links (interlinks) can be validated with DisVis[37,38]. This free tool provides—based on the set of cross-links—predictions of the possible interaction interfaces and provides an estimation of how many cross-links violate all possible conformations. For the DisVis input, one can use available PDB structures or, in the case that none are available, one can generate a structure by using one of several online platforms that make predictions based on amino acid sequences and detected XL-MS experiment intralink distance constraints (e.g., I-TASSER[60]). Further modeling steps based on the confirmed distance restraints can be performed with HADDOCK, ROSETTA or X-MOD[66].

Materials

Biological materials

- Cells to be analyzed. The procedure is optimized for lysis of PC9 cells (formerly known as PC-14 cells; Sigma-Aldrich, cat. no. 90071810-1VL) **! CAUTION** The cell lines used in your research should be regularly checked to ensure that they are authentic and that they are not infected with mycoplasma.

Reagents

! CAUTION General laboratory safety precautions should be taken while you are working with flammable, corrosive and toxic chemicals. Work in a fume hood and wear appropriate gloves and safety goggles. Do not inhale toxic vapors **▲ CRITICAL** To reduce possible cross-linking artifacts, perform optimization of the cross-linker/protein ratio (Steps 2–6, Fig. 2).

- 10% (wt/vol) Criterion XT Bis–Tris protein gel (Bio-Rad, cat. no. 3450111)
- 2-Chloroacetamide (CAA; Sigma-Aldrich, cat. no. C0267)
- Acetic acid (Merck, cat. no. 1000632500) **! CAUTION** Acetic acid is corrosive and flammable; avoid skin contact and inhalation.
- Acetonitrile (ACN; Biosolve, cat. no. 012007) **! CAUTION** Acetonitrile is highly flammable.
- Ammonium acetate ($CH_3CO_2NH_4$; AmAc; Sigma-Aldrich, cat. no. A1542)
- Ammonium bicarbonate (NH_4HCO_3; Sigma-Aldrich, cat. no. 09830)
- BenchMark Unstained Protein Ladder (Thermo Fisher Scientific, cat. no. 10747012)

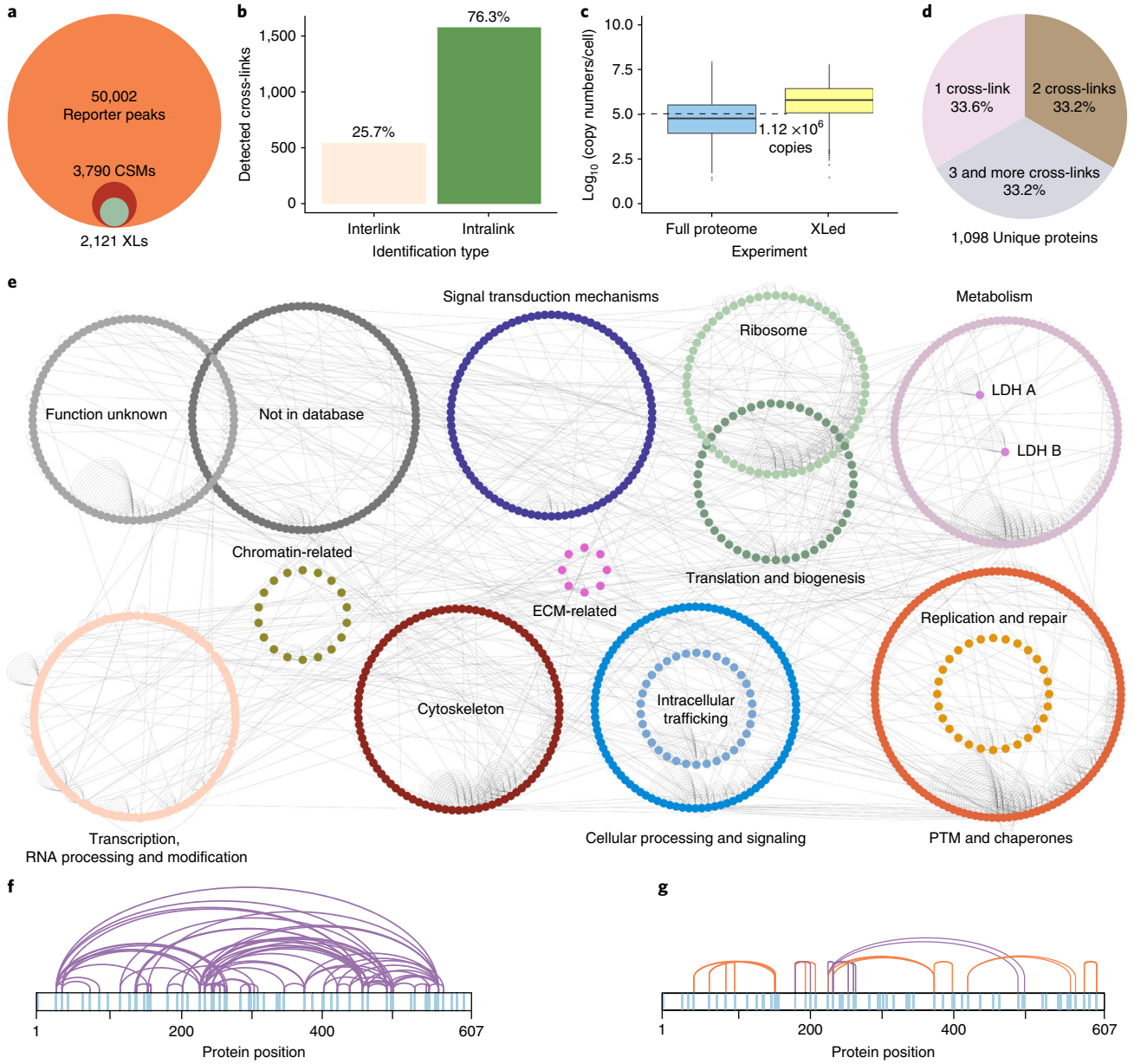

Fig. 5 | Data overview. a, Distribution of detected reporter peaks, CSMs and cross-links. **b**, Types of detected cross-links. **c**, Estimation of sensitivity by correlating detected cross-linked proteins to protein copy numbers per HeLa cell. **d**, Distribution of detected proteins by number of detected cross-links. **e**, Interaction network generated with Cytoscape for the whole-cell lysate, where proteins are grouped by their function according to the EggNOG database. **f**,**g**, xiNET visualization of detected cross-links for the low-complexity BSA sample digested with trypsin (**f**) and chymotrypsin (**g**). Cross-links exclusively detected with chymotrypsin are represented in orange.

- Benzonase (Merck, cat. no. 70664-3)
- BSA (Sigma-Aldrich, cat. no. A2153)
- Bradford Protein Assay Dye Reagent Concentrate (Bio-Rad, cat. no. 5000006)
- cOmplete, Mini EDTA-free Protease Inhibitor Cocktail (Roche, cat. no. 000000011836170001)
- Dimethyl sulfoxide (DMSO; Sigma-Aldrich, cat. no. D8418) **! CAUTION** DMSO is toxic.
- Disuccinimidyl sulfoxide (DSSO; Thermo Fisher Scientific, cat. no. A33545)
- DL-Dithiothreitol (DTT; Sigma-Aldrich, cat. no. 43815) **! CAUTION** DTT is toxic.
- Formic acid (FA; Fluka, cat. no. 94318) **! CAUTION** Formic acid is corrosive, flammable and toxic; avoid skin contact and inhalation.
- GelCode Blue Stain Reagent (Coomassie stain; Thermo Fisher Scientific, cat. no. 24592)
- HEPES (Sigma-Aldrich, cat. no. H3375)
- High-purity water obtained from a Q-POD Milli-Q purification system (Merck, cat. no. ZMQSP0D01)

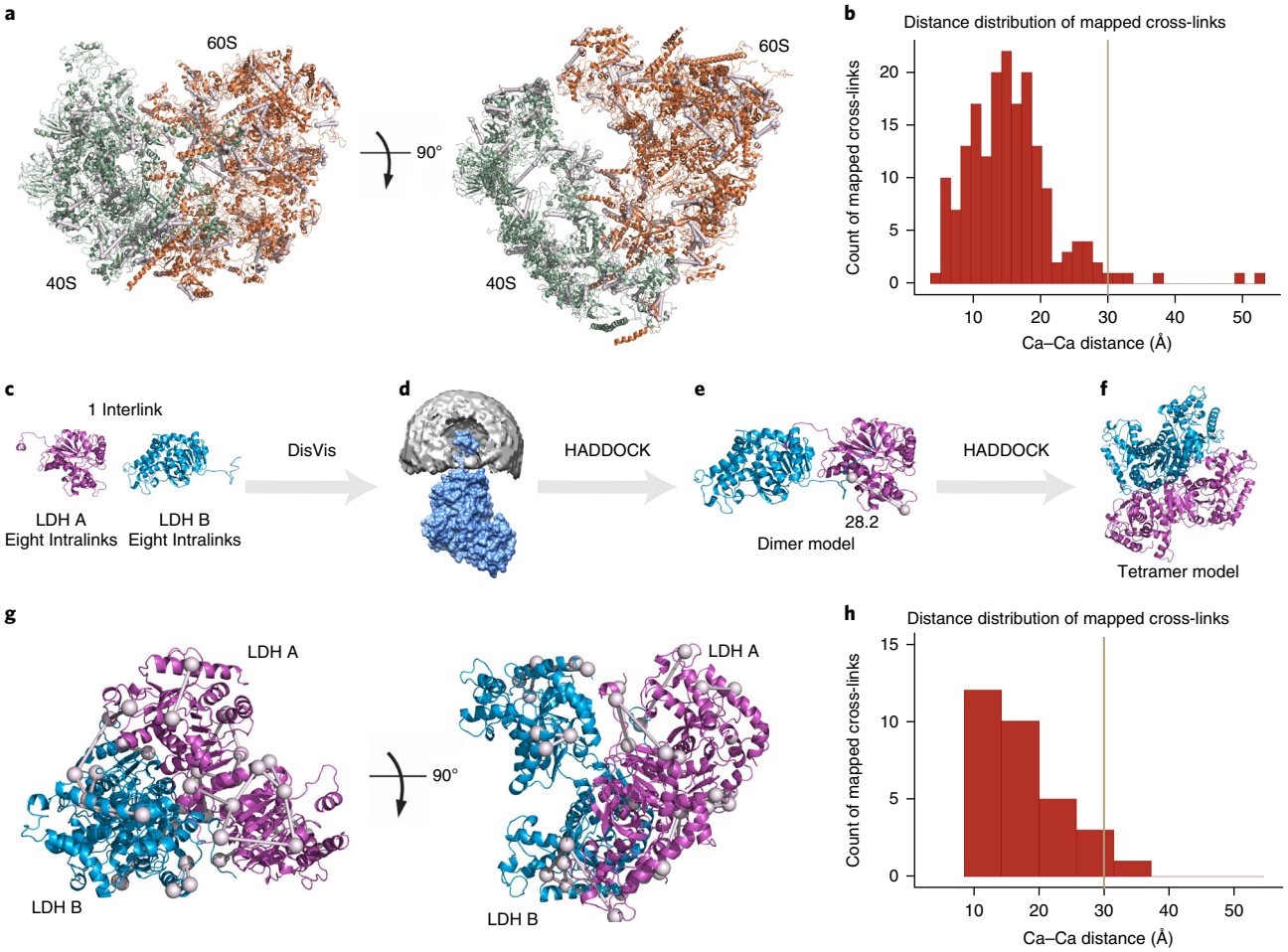

Fig. 6 | Mapping and application of identified crosslinks. a, Validation on the existing ribosome structure. **b**, Distance distribution of cross-links mapped on a ribosome. **c**, DisVis input. **d**, Predicted interaction interface generated by DisVis. **e**, Dimer model obtained using HADDOCK with mapped confirmed interlink. **f**, HADDOCK output with a final tetramer model. **g**, LDH-3 complex consisting of two LDH A and two LDH B subunits with mapped cross-links. **h**, Distance distribution of cross-links mapped on a modeled LDH-3 complex expressed as the distance between the α-carbons (Ca).

- Hydrochloric acid, 1 M solution (HCl; Merck, cat. no. 1.09057) **! CAUTION** HCl is corrosive and toxic.
- LysC (MS grade; Wako Chemicals, cat. no. 129-02541)
- Magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$; Sigma-Aldrich, cat. no. M2670)
- Methanol absolute (HPLC supra-gradient, MeOH; Biosolve, cat. no. 0013680602BS) **! CAUTION** Methanol absolute is flammable and toxic; avoid skin contact and inhalation.
- PBS (Lonza, cat. no. BE17-512F)
- Precision Plus Protein Dual Color Standards (Bio-Rad, cat. no. 161-0374)
- Sodium chloride (NaCl; Merck, cat. no. 1064041000)
- Sodium hydroxide (NaOH; Sigma-Aldrich, cat. no. 00000001064621000)
- SCX material (cation exchange; Sigma-Aldrich, cat. no. 66889-U)
- Tris(2-carboxyethyl)phosphine hydrochloride (TCEP; Sigma-Aldrich, cat. no. C4706)
- Trizma (pH 8.5, Tris-HCl and Tris base; Sigma-Aldrich, cat. no. T8818)
- Trypsin (MS grade; Promega, cat. no. V528A)
- Urea (Merck, cat. no. 66612)
- XT MOPS running buffer (20×; Bio-Rad, cat. no. 161-0788)
- XT sample buffer (4×; Bio-Rad, cat. no. 161-0791)

Equipment

- Gel-running chamber (Bio-Rad, model no. Mini-Protean III Cell)
- Power supply (PowerPac; Bio-Rad, product no. 1645050)

- Molecular imager (Amersham Imager 600; GE Healthcare, product code 29083461)
- Plastic syringe (5 ml; Sigma-Aldrich, product no. Z116866)
- Syringe needle (27¾ gauge; BD Precisionglide; Sigma-Aldrich, product no. Z192384)
- Eppendorf ThermoMixer (Thermo Fisher Scientific, cat. no. 5382000023)
- Eppendorf refrigerated centrifuge (VWR, model no. 5417R)
- Milli-Q Purification System (Millipore)
- pH meter (Meter Lab, model no. PHM210)
- Solid-phase extraction columns (C_{18}, 300-Å pore size; Grace Vydac, product no. 218SPE1000)
- Vacuum concentrator (Savant SpeedVac; Thermo Fisher Scientific, cat. no. SC210-A)
- SCX HPLC Pump A and SCX HPLC Pump B (Agilent Technologies, model no. Agilent 1200)
- 4.6 × 5.0-mm Opti-LYNX trap column (C_{18}, 49-µm particle size; Optimize Technologies, product code 11-02874-TB)
- 50.0 × 1.0-mm PolyLC SCX-separation columns (polysulfoethyl A, 3-µm particle size; PolyLC, item no. 051SE0303)
- Agilent 1290 ultraperformance liquid chromatography (UHPLC) system (Agilent Technologies, model no. Agilent 1290)
- 2 cm × 100-µm, double-frit C_{18} trap column, packed in-house (ReproSil-Pur C18-AQ, 3-µm particle size; Dr. Maisch, mat. no. r13.aq.0001)
- 50 cm × 75-µm C_{18} analytical single-frit column, packed in-house (Poroshell 120 EC-C18, 2.7-µm particle size; Agilent, custom order)
- Orbitrap (OT) Fusion Tribrid or OT Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific)
- Distal coated fused silica emitter (PicoTip emitter; New Objective, part no. FS3602010D20)

Software

- Proteome Discoverer v.2.3 or higher (Thermo Fisher Scientific, cat. no. OPTON-30795) with XlinkX node (Thermo Fisher Scientific, cat. no. OPTON-30799)
- DisVis (http://milou.science.uu.nl/cgi/services/DISVIS/disvis/)
- HADDOCK (http://milou.science.uu.nl/services/HADDOCK2.2/)

Reagent setup

Cross-linking lysis buffer

Prepare the cross-linking lysis buffer from 20 mM HEPES (47.6 mg in 10 ml), 150 mM NaCl (87.6 mg in 10 ml) and 1.5 mM $MgCl_2$ (3 mg of $MgCl_2{\cdot}6H_2O$ in 10 ml) in Milli-Q water. Adjust the pH to 7.8 with a 1 M NaOH solution (400.0 mg in 10 ml). Add one tablet of cOmplete, Mini EDTA-free Protease Inhibitor Cocktail per 10 ml of cross-linking lysis buffer and 0.5 mM DTT (0.8 mg in 10 ml) right before use. ▲ **CRITICAL** Freshly prepare all the solutions. First, dissolve HEPES, NaCl and $MgCl_2{\cdot}6H_2O$ in 5 ml of water, then adjust the pH with NaOH to an optimal range for each cross-linker (e.g., 7.6–8.0 for DSSO) and then bring the solution to a final volume of 10 ml. Keep the buffer on ice. ▲ **CRITICAL** Avoid any reagents that may possibly interfere with the cross-linking reaction. In the case of amino-reactive cross-linkers, buffers containing primary amines must be avoided. Here, DTT is used to prevent oxidative stress, but it may interfere with the cross-linking reaction and must be used at low concentration.

DSSO stock solution

Prepare a 50 mM stock DSSO solution (1 mg in 52 µl) in DMSO. ▲ **CRITICAL** DSSO may undergo hydrolysis. The solution must be prepared directly before use.

Gel-fixing solution

Gel-fixing solution is 50% MeOH (vol/vol) and 10% (vol/vol) acetic acid in Milli-Q water. The solution can be stored at room temperature (20–25 °C) for several months before use.

AmBIC solution

For a 50 mM solution, dissolve 197.5 mg of NH_4HCO_3 in 50 ml of Milli-Q water. The solution can be stored at room temperature for several months before use.

Protein reduction agent: TCEP solution
A 100 mM TCEP stock solution is prepared by dissolving 25 mg of TCEP in 1 ml of AmBiC solution; it can be stored at −20 °C for several months before use. ▲ CRITICAL AmBiC maintains pH and the basic environment of the reduction agent solution.

Protein alkylation agent: CAA solution
A 400 mM CAA stock solution is prepared by dissolving 38 mg of CAA in 1 ml of AmBiC solution; it can be stored at −20 °C for several months before use. ▲ CRITICAL AmBiC maintains pH and the basic environment of the alkylation agent solution.

HCl solution
A 1 mM HCl solution is prepared by diluting of 10 µl of 1 M stock in 9.99 ml of Milli-Q water. The solution can be stored at room temperature for several months before use.

LysC
Dissolve at least 20 µg of lyophilized LysC in 1 mM HCl, prepare aliquots and store at −80 °C. The solution is stable and can be stored at −80 °C at least until the expiration date provided by the manufacturer. ▲ CRITICAL Each aliquot can be used only once after defrosting.

Trypsin
Dissolve at least 15 µg of lyophilized trypsin in 1 mM HCl, prepare aliquots and store at −80 °C. The solution is stable and can be stored at −80 °C at least until the expiration date provided by the manufacturer. ▲ CRITICAL Each aliquot can be used only once after defrosting.

Sep-Pak washing solution 1
Sep-Pak washing solution 1 is 100% (vol/vol) ACN.

Sep-Pak washing solution 2
Sep-Pak washing solution 2 is 0.1% (vol/vol) formic acid in water. Freshly prepare this solution on the day of use and keep it at ambient temperature.

Sep-Pak elution solution
Sep-Pak elution solution is 80% (vol/vol) ACN and 0.1% (vol/vol) formic acid in water. Freshly prepare this solution on the day of use and keep at ambient temperature.

SCX Stage-Tip washing solution 1
Stage-Tip washing solution 1 is 100% (vol/vol) MeOH.

SCX Stage-Tip washing solution 2
Stage-Tip washing solution 2 is 20% (vol/vol) ACN and 0.4% (vol/vol) formic acid in water. Freshly prepare this solution on the day of use and keep at ambient temperature.

SCX Stage-Tip elution buffers
SCX Stage-Tip elution buffers are based on 20% (vol/vol) ACN, 0.4% (vol/vol) formic acid and 1 M AmAc (770 mg in 10 ml) in water. This buffer is diluted with Stage-Tip washing solution 2 to obtain elution buffers with a range of AmAc concentrations: 25, 50, 200 and 500 mM. Freshly prepare these buffers on the day of use and keep at ambient temperature.

SCX HPLC solvent A
SCX HPLC solvent A consists of 20% (vol/vol) ACN and 0.05% (vol/vol) formic acid in water. The mobile phase should be stored at ambient temperature and should be replaced every 2 months.

SCX HPLC solvent B
SCX solvent B consists of 20% (vol/vol) ACN and 0.05% (vol/vol) formic acid in a 0.5 M NaCl (29.2 g in 1 liter) in water solution. The mobile phase should be stored at ambient temperature and should be replaced every 2 months.

Reverse-phase UHPLC solvent A
Reverse-phase UHPLC solvent A is 0.1% (vol/vol) formic acid in water. The mobile phase should be stored at ambient temperature and should be replaced every 2 months.

Reverse-phase UHPLC solvent B
Reverse-phase UHPLC solvent B is 80% (vol/vol) ACN and 0.1% (vol/vol) formic acid in water. The mobile phase should be stored at ambient temperature and should be replaced every 2 months.

Equipment setup

▲ **CRITICAL** Our list of equipment for the SDS–PAGE gel-running chamber, HPLC pumps and columns for SCX fractionation and LC setups are example setups. The experiments can also be performed with similar equipment.

SCX HPLC fractionation
SCX chromatography is performed on an Agilent 1200 HPLC system. The setup is built with an Opti-LYNX trap column connected to a PolyLC SCX-separation column; for more details, see Hennrich et al.[67]. A detailed description of LC separation parameters can be found in Table 1.

LC–MS/MS analysis
The standard LC–MS/MS setup in our laboratory features an Agilent 1290 Infinity UHPLC system connected to an OT Fusion Lumos. A double-frit C_{18} trap column (Dr Maisch Reprosil C18, 3 µm, 2 cm × 100 µm) is used for rapid sample desalting and is followed by a single-frit analytical C_{18} column. Both columns are packed in-house and configured in a vented column setup[62]. First, the injected samples are loaded onto the trapping column with a flow of 5 µl/min for 5 min of reverse-phase solvent A, and then gradient elution is performed at a column flow rate of ~300 nl/min (split flow from 0.2 ml/min). An effluent from the column is directly introduced into the nano-spray-ionization (NSI) source via a coated fused silica emitter, forcing chromatographic separation of the peptides, using 85-, 115- or 175-min methods with 65-, 95- or 155-min LC gradients, respectively (Table 2).

SCX and C_{18} Stage-Tip preparation
Place three layers of SCX or C_{18} material plugs of 1.2-mm diameter each into a 200-ml pipette tip as described in Rappsilber et al.[63]. ▲ **CRITICAL** In case of input <10 µg, use plugs of 0.5-mm diameter and place them in a gel-loader pipette tip. ▲ **CRITICAL** Do not load >20 µg of peptide material in order to achieve better enrichment of cross-linked peptides and prevent loss of the material, owing to exceeding the maximum material capacity.

Procedure

Preparation of protein sample

1 To prepare low-complexity samples for cross-linking, refer to option A. To prepare whole-cell lysates, refer to option B.

(A) **Low-complexity samples** ● **Timing ~0.5 h**

(i) Resuspend lyophilized protein or dilute a solution of purified protein complex in the required amount of cross-linking lysis buffer to achieve a final concentration of ~1–3 mg/ml.
▲ **CRITICAL STEP** Mix the sample with a pipette to preserve the protein (complex) native state and avoid vortexing.
▲ **CRITICAL STEP** If the sample was frozen to −80 °C, first defrost it on ice and only then bring it to room temperature.

(B) **Whole-cell lysate preparation** ● **Timing ~1 h**

(i) Resuspend the collected cell pellet (10^7 cells) in 1 ml of ice-cold cross-linking lysis buffer and perform soft lysis to keep the organelles and protein assemblies intact. Perform 30–40 quick pushes through a 27¾-gauge syringe.
▲ **CRITICAL STEP** Avoid foaming, as this leads to protein denaturation.
▲ **CRITICAL STEP** The needle width must be chosen according to the size of each cell type.

(ii) Remove the cell debris via centrifugation at 13,800*g* for 10 min at 4 °C.
▲ **CRITICAL STEP** To pellet only the organelles for further cross-linking experiments, apply a lower centrifugation speed. Centrifugation at, e.g., 3,200*g* for 10 min is sufficient for nuclei.

Table 2 | LC and MS parameters used during 85-, 115- or 175-min methods for the Orbitrap Fusion Lumos

LC parameters	
Time interval (min)	**LC gradient (% B)**
0–5	0–7
5–30 (5–100, 5–160)	7–40
30–33 (100–103, 160–163)	40–100
33–34 (103–104, 163–164)	100–100
34–35 (104–105, 164–165)	100–0
35–45 (105–115, 165–175)	0–0
MS parameters	
Polarity	**Positive**
MS1 OT	**Tune Page 3.0, Lumos**
OT resolution	60,000
Scan range, *m/z*	375–1,500
RF lens, %	30
AGC target	4.0×10^5
Maximum IIT, ms	50
dd settings MS2	
dd mode	TOP10
MIPS	Peptide
Intensity	Intensity threshold, 2.0×10^4
Charge state	3–8
Dynamic exclusion	
Exclude after *n* times	
Exclusion duration, s	12 s (16 and 24 s)
Mass tolerance	10-p.p.m. low, 10-p.p.m. high
Sort by charge	Highest charge state
dd-MS2 OT CID	
Isolation window, *m/z*	1.6
CID collision energy, %	30
OT resolution	30,000
AGC target	5.0×10^4
Maximum IIT, ms	100
AAPT	ON
dd-MS2 OT ETD (MS2–MS2–MS3 strategy)	
Isolation window, *m/z*	1.6
Calibrated charge-dependent ETD parameters	ON
OT resolution	30,000
First mass, *m/z*	120
AGC target	1×10^5
Maximum IIT, ms	120
AAPT	ON
dd settings MS3	
dd mode	Four scans
MIPS	Peptide
Charge state	1–6
Targeted mass difference	
Number of precursors in the group	2
Delta M1	31.9721
Partner intensity range	10–100%
Perform scan	Both ions of the same charge
Intensity	Intensity threshold, 5.0×10^3
Precursor ion exclusion, *m/z*	Low 18, high 5

Table continued

Table 2 (continued)

LC parameters	
Time interval (min)	**LC gradient (% B)**
dd-MS3 IT	**CID**
MS isolation window, *m/z*	1.6
MS2 isolation window, *m/z*	2
Collision energy, %	35
Ion trap scan rate	Rapid
First mass, *m/z*	120
AGC target	3.0×10^4
AAPT	ON
Maximum IIT, ms	90

For the methods of 85, 115 and 175 min, the indicated times refer to the total analysis times rather than actual gradient times, which are 68, 98 and 148 min, respectively. In parentheses, LC and MS parameters are provided for the 115- and 175-min methods. OT, Orbitrap.

(iii) Estimate the protein concentration via a BCA or Bradford protein assay. For most mammalian cell lines, one can expect a concentration of ~2 mg/ml from a cell pellet containing 10^7 cells.
▲ **CRITICAL STEP** A minimum protein concentration of 1 mg/ml is recommended.
? TROUBLESHOOTING

Optimization of cross-linker/protein ratio ● Timing ~5 h

2 Prepare seven to ten aliquots of sample containing small amounts of protein material.
▲ **CRITICAL STEP** Per reaction mixture, 1–2 µg of a single protein or complex is sufficient, and 10–15 µg of protein dry weight from a whole-cell lysate is sufficient.

3 Add the required amount of cross-linker stock solution to each aliquot to obtain solutions with a range of cross-linker reagent concentrations. Mix the sample with a pipette.
▲ **CRITICAL STEP** For low-complexity samples, set up a cross-linker concentration range that extends ten times below and ten times above the molar amount of protein. In the case of whole-cell lysates, we recommend starting with concentrations from 0.1 to 5 mM of cross-linker. Note that this range will suffice for most proteins in a proteome-wide context, but might vary for specific protein complexes, owing to differences in the accessibility of the lysines as well as the reactivity of the lysines within their specific microenvironment. The optimization step performed here aims to generate an overall high cross-linking yield without creating artifacts from over-cross-linking.
▲ **CRITICAL STEP** Slowly add DSSO stock to the protein solution to avoid precipitation and keep the solution clear.
? TROUBLESHOOTING

4 Leave the reaction mixture for 30 min at room temperature.
▲ **CRITICAL STEP** The cross-linking reaction is performed at room temperature for a maximum of 30 min. After this time, ~90% of the cross-linking reagent will have hydrolyzed, and no further cross-links between lysines in close proximity will be formed. The temperature can be altered in the case of specific protein complexes; however, for most protein complexes, room temperature is optimal.
▲ **CRITICAL STEP** When preparing replicates, ensure that the protein concentration, cross-linker concentration and cross-linking temperature and time are kept constant.

5 Quench the reaction with Tris solution to a final concentration of 20 mM. Keep at room temperature for 10 min.

6 Run an SDS–PAGE gel using standard laboratory equipement with the range of cross-linker concentrations and select the optimal cross-linker/protein ratio from the moderate concentrations of cross-linker; generally 1 or 2 mM is sufficient, although this should be tested for each sample (see Fig. 2 for further details).
▲ **CRITICAL STEP** To ensure that the correct cross-linker/protein ratio is detected, Steps 2–6 should be performed in replicates.
▲ **CRITICAL STEP** For the presented data with PC9 cells, 1 mM DSSO solution was used.
? TROUBLESHOOTING

Protein cross-linking ● Timing ~1.5 h

7 Add the required amount of cross-linker stock solution, as determined in Step 6, to the protein sample and leave the reaction mixture for 30 min at room temperature.
▲ CRITICAL STEP Slowly add DSSO stock to protein solution to avoid precipitation and keep the solution clear.
? TROUBLESHOOTING

8 Quench with Tris solution to a final concentration of 20 mM. Keep at room temperature for 10 min.

Digestion and desalting of cross-linked proteins ● Timing ~18 h

9 For complex samples, add the benzonase solution to a final concentration of 1% (vol/vol) and keep the solution at 37 °C for 10 min. (Optional) In the case of low-complexity samples, consider adding an appropriate enzyme to a nucleic acid–rich sample to shear relevant artifacts (e.g., RNAse A in the case of 60S ribosome).

10 Add urea to a final concentration of 8 M.
▲ CRITICAL STEP Choose an optimal concentration for each of the chaotropic agents.
? TROUBLESHOOTING

11 Perform the reduction and alkylation reaction simultaneously by adding TCEP and CAA to a final concentration of 10 and 40 mM, respectively. Incubate the solution at 37 °C for at least 1 h.

12 Dilute the sample with AmBIC to decrease the urea concentration.
▲ CRITICAL STEP Dilution is necessary to prevent inhibition of the proteases. LysC is active in <6 M urea, and trypsin is active in <2 M urea.

13 Add LysC for predigestion step in a ratio 1:75 of protease/proteins (wt/wt) and incubate the solution at 37 °C for 2–4 h. Then dilute the reaction mixture 3× with AmBiC and add trypsin in a ratio 1:50 of protease/proteins (wt/wt) and incubate the solution at 37 °C for at least 12 h.
▲ CRITICAL STEP The predigestion step with LysC must be done for at least 2, and should be done for 4 h for maximum efficiency.
▲ CRITICAL STEP When using alternative proteases, e.g., chymotrypsin, adjust the digestion conditions accordingly[62].

14 Quench the reaction with FA (100% (vol/vol)) to a 2–4% (vol/vol) final concentration.
! CAUTION Experiments with FA must be done in a fume hood.

15 Centrifuge the peptide mixture at 4,500*g* for 10 min at room temperature and take the supernatant.
? TROUBLESHOOTING

16 Condition a 300-Å pore-size C_{18} cartridge with 2 × 1 ml of Sep-Pak washing solution 1, and then equilibrate the column with 2 × 1 ml of Sep-Pak washing solution 2. Refer to Villen et al.[68] and Udeshi et al.[69] for more detailed instructions. (Optional) To prevent unnecessary sample loss, consider avoiding the desalting procedure in the case of a low-complexity sample and directly performing SCX Stage-Tip fractionation, as the sample generally does not need to be concentrated before the fractionation (Step 21A).

17 Load acidified cross-linked peptide digests onto the SPE cartridge. Place a new collection tube below the SPE cartridge and collect the flow-through solution.
▲ CRITICAL STEP To provide optimal loading, maintain a slow flow rate and low pressure to the vacuum scaffold. High pressure may collapse the collection tubing. The cartridge size is selected based on the sample amount[68,69]. Do not allow the cartridge to run dry.
▲ CRITICAL STEP Collection of the flow-through is recommended to allow the possibility of repeating the desalting procedure or for future troubleshooting.

18 Wash the columns with 2 × 1 ml of Sep-Pak washing solution 2 to desalt the peptides.

19 Elute the desalted peptides into new collection tubes with 2 × 200 µl of Sep-Pak elution solution.
▲ CRITICAL STEP To ensure full elution of the peptides from a cartridge, maintain a slow flow rate and low pressure to the vacuum scaffold.

20 Evaporate the obtained fractions with vacuum centrifugation to almost dryness.
▲ CRITICAL STEP Avoid complete dryness to prevent sample loss.
■ PAUSE POINT The sample can be stored at −80 °C for several months before fractionation.

Fractionation of cross-linked peptides by SCX

21 For low-complexity samples, refer to option (A). For whole-cell lysates, refer to option (B).

(A) **Low-complexity sample** ● Timing ~3 h

(i) Prepare the SCX Stage-Tip column as described in the Equipment setup section.

(ii) Condition the SCX Stage-Tips by addition of 100 µl of Stage-Tip washing solution 1. Place the tip in the Eppendorf tube at room temperature and centrifuge at 1,200*g* until the solution passes through the tip.

(iii) Repeat the previous step with 100 µl of 1M Stage-Tip elution buffer and 2 × 100 µl of SCX Stage-Tip washing solution 2.

(iv) Reconstitute the sample at room temperature in 100 µl of 5% DMSO/10% FA/85% HOH (vol/vol/vol) and load onto SCX Stage-Tip. Place a new collection tube to collect the flow-through solution. Centrifuge at a maximum of 1,000*g* until the solution passes through the tip.

▲ CRITICAL STEP To provide optimal loading, maintain a slow flow rate by using lower centrifugation speeds but longer times.

▲ CRITICAL STEP The flow-through should be collected and used again for the loading step to achieve maximum loading efficiency or stored for troubleshooting.

(v) Wash Stage-Tips at room temperature with 200 µl of Stage-Tip washing solution 2 by centrifugation at 1,200*g* until the solution passes through the tip.

(vi) Elute the peptides at room temperature with 2 × 100 µl of each Stage-Tip elution buffer with increasing concentration of AmAc (from 25 mM to 1 M) by centrifugation at a maximum of 1,000*g* until the solution passes through the tip.

▲ CRITICAL STEP To ensure full elution of peptides from a column, maintain a slow flow rate by using lower centrifugation speeds but longer times.

▲ CRITICAL STEP For example, in the presented data on BSA, fractionated with the described SCX Stage-Tip setup, five elution solutions were used: 25 mM, 50 mM, 200 mM, 500 mM and 1 M AmaC. However, only the last three elutions were subjected to LC–MS/MS analysis, as these are most cross-link rich.

(vii) Dry the obtained fractions by vacuum centrifugation to almost dryness.

▲ CRITICAL STEP Avoid complete dryness to prevent sample loss.

(B) **Fractionation of complex samples by SCX HPLC** ● Timing ~6 h

(i) Perform quality control of the SCX fractionation system with tryptic BSA digests (for examples, see Fig. 3a,b).

(ii) Reconstitute the samples in 5% DMSO/10% FA/85% H_2O (vol/vol/vol) and inject into the described SCX HPLC fractionation setup.

▲ CRITICAL STEP A minimum of 100 µg of peptide material is required for this step. However, for isolation of low-abundant cross-linked peptides, the optimal injection amount is 500 µg. For the described HPLC–SCX setup, 20 µl of solvent is sufficient for 500 µg.

(iii) Perform the SCX fractionation with a 65- or 120-min gradient. For a detailed description of the HPLC gradient, refer to Table 1.

▲ CRITICAL STEP The length of the gradient should be adjusted according to the amount of peptides injected.

? TROUBLESHOOTING

(iv) Pool early and late fractions together based on the UV trace and their elution time.

▲ CRITICAL STEP For example, in the presented data on PC9 cell lysates fractionated with the described HPLC–SCX setup and a 120-min gradient, most of the identified cross-links are localized in ten fractions (Fig. 3). Fractions 12–26 are considered to be early, whereas fractions 37–42 are considered to be late. For pilot studies, we advise running LC–MS/MS experiments for at least 20 samples. This number can be decreased to five to ten after the cross-link-rich fractions have been located for the used setup.

(v) An additional desalting step is required for the late and pooled SCX fractions.

▲ CRITICAL STEP The current protocol is optimized for an LC–MS/MS setup incorporating a trapping column. For setups without a trapping column, all fractions must be desalted.

(vi) For pooled fractions, for which the peptide dry weight exceeds 20 µg (e.g., pooled first fractions), repeat Steps 16–19.

(vii) For pooled fractions, for which the peptide dry weight is <20 µg (e.g., pooled last fractions), C_{18} Stage-Tip columns must be used as described below.
▲ **CRITICAL STEP** The same solutions as used for Sep-Pak (Steps 16–19) can be used in cases in which C_{18} material is applied.

(viii) Condition C_{18} Stage-Tips at room temperature by adding 100 µl of Sep-Pak washing solution 1. Place the tip in the Eppendorf tube and centrifuge at 1,200*g* until the solution passes through the tip.

(ix) Repeat the previous step with 2 × 100 µl of Sep-Pak washing solution 2.

(x) Load the samples at room temperature onto the C_{18} Stage-Tip. Place new collection tubes below the Stage-Tips to collect the flow-through solution. Centrifuge at a maximum of 1,000*g* until the solution passes through the tip.
▲ **CRITICAL STEP** For an optimal loading, maintain a slow flow rate by using lower centrifugation speeds but longer times.
▲ **CRITICAL STEP** The flow-through should be collected and either should be used again at the loading step to achieve maximum loading efficiency or should used for future troubleshooting.

(xi) Wash C_{18} Stage-Tips at room temperature with 2 × 100 µl of Sep-Pak washing solution 2 by centrifugation at 1,200*g* until the solution passes through the tip.

(xii) Elute the peptides at room temperature with 2 × 100 µl of Sep-Pak washing solution by centrifugation at a maximum of 1,000*g* until the solution passes through the tip.
▲ **CRITICAL STEP** To ensure full elution of peptides from the Stage-Tips, maintain a slow flow rate by using lower centrifugation speeds but longer times.

(xiii) Dry the obtained fractions by vacuum centrifugation to almost dryness.
▲ **CRITICAL STEP** Avoid complete dryness to prevent sample loss.
■ **PAUSE POINT** Dried fractions can be stored at −80 °C for several months before analysis.

LC–MS/MS analysis ● Timing ~1 d

22 Estimate the amount of peptide in each fraction after SCX fractionation by injecting 0.1–5% of several fractions into a LC–MS/MS system and analyzing with short gradients. The required injection volume for all fractions can then be estimated from the UV trace.
? TROUBLESHOOTING

23 Resuspend the dried peptide pellet in 5% DMSO/10% FA/85% H_2O (vol/vol/vol) and inject an appropriate amount into the LC–MS/MS system. Longer LC gradients and running replicates in parallel tends to increase the number of identified cross-linked peptides. For samples of low complexity, 85-min runs are sufficient. For low amounts (e.g., late pooled SCX fractions), shorter gradients are recommended to increase the sensitivity. The MS acquisition method for MS-cleavable cross-linkers must consist of an MS1 high-resolution survey scan, followed by an MS2 scan, which again is followed, for fragments of potential cross-linked peptides, by MS3 fragmentation. For a review of alternative fragmentation strategies that support XLinkX, refer to the Supplementary Tutorial. In brief, for MS-cleavable cross-linkers such as DSSO or DSBU, depending on the acquisition strategy, 'MS2_MS2' (e.g., CID/ETD), 'MS2_MS3' (e.g., CID/MS3 HCD) or 'MS2_MS2_MS3' (e.g., CID/ETD/MS3 HCD) should be set. Refer to Table 2 for a detailed description of LC and MS parameters; however, the most appropriate settings can also be found as templates in the Xcalibur method editor.

Data analysis with Proteome Discoverer v.2.3 ● Timing ~17 h

24 Perform the data analysis with the XLinkX node incorporated into Proteome Discoverer (v.2.3 or higher). General workflow schemes can be found in Fig. 4 and are also installed as common templates in Proteome Discoverer. The main settings are described below, and a complete list of parameters is available in Supplementary Table 1. A detailed description of the cross-linker modification settings is provided in the Supplementary Tutorial and Supplementary Fig. 1.

25 Set the processing workflow. In the 'XlinkX Detect' node, specify the acquisition strategy as 'MS2_MS3'. Define the cross-link modification as DSSO on lysines.

26 In the peptide search path, set 'XlinkX Filter' to 'Peptides'. In our laboratory, we use the 'Sequest' search node, although other search nodes are equally applicable. Filtered spectra were matched against the *Homo sapiens* database from SwissProt (v.2017_10; 20,230 sequences) or the single BSA

sequence (UniProt ID P02769). Regardless of the search node used, set the enzyme to 'Trypsin' and the maximum number of missed cleavages as 2. Set the precursor mass tolerance to 20 p.p.m. and the fragment mass tolerance to 0.05 Da for IT readout or 20 p.p.m. for OT readout. Set carbamidomethylation of cysteines as a fixed modification and oxidation of methionine and protein N-terminal acetylation as variable modifications. For detection of monolinks, set DSSO-hydrolyzed and DSSO-tris as variable modifications.

▲ **CRITICAL STEP** Adjust enzyme specificity and the number of missed cleavages according to the protease used.

27 In the XlinkX search path, set 'XlinkX Filter' to 'Crosslinks'. In the 'XlinkX Search' node, filtered spectra are matched against the *Homo sapiens* database from SwissProt (v.2017_10, 20,230 sequences) or combined BSA sequence (UniProt ID P02769) and the *E. coli* strain K12 in the SwissProt database (v.2017_10, 4,306 sequences). Set the enzyme name to 'Trypsin (Full)' and the maximum number of missed cleavages to 2. Set the minimum peptide length to 5 and minimum peptide mass to 300; the maximum peptide mass has to be set to 7,000. Set the precursor mass tolerance to 10 p.p.m., the Fourier transform mass spectrometer (FTMS) fragment mass to 20 p.p.m. and the ion trap mass spectrometer (ITMS) fragment mass to 0.5 Da. Set carbamidomethylation of cysteines as a fixed modification and oxidation of methionines and protein N-terminal acetylation as variable modifications. In the 'XlinkX Validation' node, define the FDR threshold as 0.01 and the FDR strategy for low-complexity samples as 'Simple' and for the whole-cell lysates as 'Percolator'.

▲ **CRITICAL STEP** Adjust the enzyme specificity and the number of missed cleavages according to the protease used.

▲ **CRITICAL STEP** The search database must contain at least 100 other protein sequences that are used as support for FDR control.

? TROUBLESHOOTING

28 Set the consensus workflow. In 'XlinkX Crosslink Export', specify the file path where the xiNET input files will be stored.

Data representation and additional validation ● Timing ~1 d

29 In our laboratory, visualization of detected cross-links is done on multiple levels with xiNET, pyMol or Chimera and Cytoscape. Proteome Discoverer generates all required input files for xiNET, including the .fasta file with protein sequences and the detected cross-linked residues, which is suitable for single proteins and protein complexes. If a protein of interest already has an existing crystal structure, visualization in pyMOL or Chimera can be done. In the case of whole-cell lysates, we apply Cytoscape to visualize all detected cross-linked peptide pairs. For the analysis of thousands of detected cross-links, export Proteome Discoverer output tables to Excel format and upload to Cytoscape; then cluster the proteins by biological function (Supplementary Methods 1), pathway or relevant protein complexes.

▲ **CRITICAL STEP** In line with normal shotgun proteomics experiments, when, e.g., triplicates are analyzed, we recommend that only cross-links detected in at least two out of three samples be accepted.

30 Check the quality of obtained data by additional validation of detected cross-links on a well-resolved structure of the protein complex. Highlight the protein or protein complex for which a high number of cross-links is available. Filter these complexes on the basis of the availability of structures in existing databases (e.g., human ribosome, PDB ID 4UG0). For quick mapping of a large number of cross-links, use pyMol and run the relevant scripts (Supplementary Methods 2).

▲ **CRITICAL STEP** The structure picked for additional validation must be chosen according to cross-linking experimental conditions.

? TROUBLESHOOTING

Structural investigations ● Timing ~1–10 d

31 Download or generate the PDB structure for each subunit of the protein complex of interest on the basis of available amino acid sequences and obtained intra-protein distance restraints. Submit the PDB structures, together with the interprotein cross-links (interlinks), to the DisVis online platform for additional validation and perform further structural investigations of protein complexes with confirmed cross-links. After cross-links are confirmed, submit a HADDOCK docking run for a protein complex of interest.

? TROUBLESHOOTING

Troubleshooting

Troubleshooting advice can be found in Table 3.

Table 3 | Troubleshooting table

Step	Problem	Possible reason	Solution
1B(iii)	Low protein concentration	Insufficient amount of cells	Use more cells as input material
		Ineffective lysis	Make more pushes through the syringe
3, 7	Precipitate is formed	Cross-linker stock solution is too concentrated	Dilute cross-linker solution
6	Low-MW region is still present on a gel in the case of high cross-linker concentrations	Low activity of the cross-linker	This could be due to hydrolysis of the cross-linker. Therefore, prevent exposure of the stock solution to multiple freeze–thaw cycles. Make sure that the cross-linker stock solution is prepared immediately before the cross-linking reaction or use aliquots of stock solutions only once
			If the cross-linking reagent requires activation (in the case of, e.g., an acidic cross-linker or a zero-length cross-linker), verify whether the correct activation conditions were applied
		Presence of interfering components in the buffer	Define all possible substances that may interfere with the cross-linking reaction. Avoid these in the buffer. In the case of NHS cross-linkers, e.g., DSSO, avoid buffers containing amines and primary alcohols
		Concentration of the cross-linking reagent is too low	Increase the concentration of the cross-linking reagent
10	A precipitate is formed	Insufficient enzyme activity to shear all nucleic acid artifacts	Increase the concentration of the enzyme
		Protein concentration is too high	Dilute the reaction mixture to a final protein concentration of ~1 mg/ml
15	A large pellet is formed after centrifugation of acidified digests	Insufficient digestion	Perform MS measurements and, after an unspecific database search, check the number of missed cleavages and compare this to the maximum allowed number for the protease used, thus checking protease activity. Follow the instructions for optimal enzymatic digestion[62]
		Too little enzyme used	Incubate the reaction for a longer time or increase the amount of enzyme
		Over-cross-linking	Ensure that the lowest possible cross-linking reagent concentration is used
21B(iii)	No or low UV signal	HPLC–SCX system is not performing correctly	Check BSA quality control runs (Fig. 3a,b)
		Peptide concentration is too low	Increase the amount of injected peptide. If necessary, increase the amount of initial cell material and repeat the cross-linking and digestion steps
		Loss of peptides during desalting	Analyze flow-through solutions. If necessary, repeat the loading procedure with collected flow-through solutions and perform further washing and elution steps. Ensure that the procedure is done according to published protocols[62,63]
		Insufficient elution during desalting	Repeat elution of the desalting cartridges. Ensure that the procedure is done according to published protocols[62,63]
22	Weak MS signal intensity	Low amount of injected peptide material	Increase the loading amount to a maximum 10% of a total fraction volume. If the signal is still low, pool a number of SCX fractions
		Loss of peptides during desalting	Analyze flow-through solutions. If necessary, repeat the loading procedure with collected flow-through solutions and perform further washing and elution steps. Ensure that the procedure is done according to published protocols[62,63]

Table continued

Table 3 (continued)

Step	Problem	Possible reason	Solution
		Insufficient elution during desalting	Repeat the elution step on the desalting cartridges. Ensure that the procedure is done according to published protocols[62,63]
		Problems with the LC–MS/MS setup	Refer to detailed evaluation and troubleshooting of the LC–MS/MS system[62]
27	Fewer cross-link identifications than expected	Instrument is out of calibration	Broaden mass window tolerances to, e.g., 50 p.p.m. on all levels. Calibrate the instrument
		Suboptimal software search settings	Ensure that all the settings are filled according to Supplementary Table 1
			Compare the MS acquisition strategy with the setting in the 'XlinkX Detection' node. If necessary, search the data with the correct MS strategy (Supplementary Tutorial)
			Check whether the cross-linker modification in the 'XlinkX Detection' node is set to the reacted version of the linker. We also recommend checking that cross-linker signature peaks are set correctly under the 'Extended Properties' tab in the chemical modifications list (Supplementary Fig. 1)
		Low efficiency of cross-linking reaction	Too few amino acids can be captured by the cross-linking reagent used. Perform the experiment using a cross-linking reagent with alternative chemistry and involving other amino acids
		Reactive residues are outside of the cross-linker's maximum linking distance	Perform the experiment using a cross-linking reagent with alternative chemistry and involving other amino acids
		Problems with LC–MS/MS setup	Refer to detailed evaluation and troubleshooting of the LC–MS/MS system[62]
	Most of the cross-links are detected within the early SCX fractions	Leaking pump B during HPLC–SCX fractionation	Substitute the leaking pump
30	Validation of <90% of cross-links	Wrong PDB structure	Choose an appropriate PDB structure according to your experimental conditions
		Over-cross-linking	Choose a lower concentration of reagent for the cross-linking experiment
31	Most of the cross-links are not confirmed	PDB structures of the subunits were not chosen correctly	Check whether the submitted structures were chosen according to your experimental conditions
		Wrong sequence numbering in the PDB file	Adjust the numbering of the amino acids according to their sequence in the search database

Timing

Step 1A, preparation of low-complexity protein samples: ~0.5 h
Step 1B, preparation of whole-cell lysates: ~1 h
Steps 2–6, optimization of cross-linker/protein ratio: ~5 h
Steps 7 and 8, protein cross-linking: ~1.5 h
Steps 9–20, digestion and desalting of cross-linked proteins: ~18 h
Step 21A, fractionation of low-complexity samples: ~3 h
Step 21B, fractionation of complex samples by HPLC–SCX: ~6 h
Steps 22 and 23, LC–MS/MS analysis: ~1 d
Steps 24–28, data analysis with Proteome Discoverer v.2.3: ~17 h
Steps 29 and 30, data representation and additional validation: ~1 d
Step 31, structural investigations: 1–10 d

Anticipated results

Here, we describe the recommended procedures for performing cross-linking experiments for samples of both low and high complexity. BSA protein digested with two sets of proteases is shown as an example of low-complexity samples, whereas a PC9 cell lysate is taken as a sample of high

complexity. The protocol includes descriptions of both data analysis and data interpretation, and makes use of the MS-cleavable cross-linker DSSO, as this substantially simplifies data analysis for complex samples. Nevertheless, MS-cleavable linkers with alternative chemistry can be used, as the software is fully configurable. In addition, a validation procedure for the human ribosome complex is described in detail.

Low-complexity example

As a proof-of-principle example of single-protein cross-linking, we show BSA digested with a combination of the protease LysC with trypsin and chymotrypsin. The detected cross-links were mapped onto the BSA structure (PDB ID 4F5S with a resolution of 2.5 Å). For the LysC/trypsin digest, 75 unique distance restraints were mapped onto BSA, out of which 59 (79%) were within the DSSO cross-linking range of 30 Å (Fig. 5f). With chymotrypsin, 20 BSA cross-links were detected, out of which 100% were within the DSSO cross-linking range (Fig. 5g). Out of all detected cross-links, 65% are unique for chymotrypsin and were not detected with standard proteases.

High-complexity example

For our high-complexity example, the optimal cross-linker/protein ratio was established before the main cross-linking experiment. Small amounts of sample were cross-linked with a range of linker reagent concentration, after which the optimal ratio could be established by SDS–PAGE (Fig. 2). For the lane in which no cross-linking reaction was performed (Fig. 2, lane 1), the high-MW region is empty, indicating that the available protein complexes disintegrated under the denaturing conditions and demonstrating that none of the available protein complexes are held together by the cross-linking reagent (Fig. 2, box A). For the lanes in which the cross-linking reagent was added in high concentrations (Fig. 2, lanes 5–7), the lower-MW region is less intense (Fig. 2, box B), indicating that non-interacting small species form unspecific agglomerates and move to the upper part of the gel (Fig. 2, box C). The intermittent lanes, with a concentration ~1 mM, show a good compromise between the two (Fig. 2, lanes 2–4). After optimization of the cross-linking conditions and application of the protocol to the sample, the final peptide mixtures were fractionated by SCX-based separation to specifically enrich for cross-linked peptides (Fig. 3a). Based on the UV trace, intensity and fraction retention time, 31 fractions potentially containing the cross-linked peptides were selected for further analysis (Fig. 3b). As can be noted from the distribution of the detected CSMs, the early fractions (e.g., 12–26) and late fractions (e.g., 37–50) can be pooled. This results in a total of 12 samples for analysis, substantially decreasing LC–MS/MS measurement time with no dramatic loss in terms of detected cross-links. For the whole-cell lysate, the total analysis time was 17 h (~10 h for the XlinkX search alone) on a 3.47-GHz, eight-core computer with 48 Gb of RAM. We identified ~50,000 reporter ions, resulting in 3,790 CSMs grouped into 2,121 unique peptide pairs (Fig. 5a). Out of these, 545 cross-links or 25.7% are inter-protein and 1,695 are intra-protein cross-links (Fig. 5b). To estimate the sensitivity of our approach, we correlated the detected cross-linked proteins with protein copy numbers per HeLa cell[70] (Fig. 5c). The lower border for 50% of the identifications is at ~1.12×10^6 copies per cell. In total, it was possible to identify 1,098 unique cross-linked proteins (Fig. 5d). ~66% of the detected PPIs are mapped by one or two cross-links, whereas for the rest, three or more unique peptide pairs are identified. The interaction network is visualized with Cytoscape with grouping of detected identifications by protein function extracted from the EggNOG database v.4.5.1 (Fig. 5e) or with xiNET for the low-complexity samples (Fig. 5f,g).

The detected distance constraints for the ribosome can be mapped onto the existing crystal structure (PDB ID 4UG0 with a resolution of 3.6 Å). A total of 225 unique peptide pairs were detected for this protein complex, out of which 182 could be visualized on the crystal structure. Considering that the maximum DSSO cross-linking distance is limited to 30 Å, 177 cross-links (97%) were valid (Fig. 6a,b). As a further example of cross-link validation, we utilize DisVis to map detected peptide pairs for the lactate-dehydrogenase (LDH) complex, consisting of LDH A (LDH M) and LDH B (LDH H). PC9 cells are non-small-cell lung cancer cells for which LDH-5, a protein complex that is built from four LDH A subunits, is overexpressed[71]. Previously, LDH-3, a complex that is built up from two LDH A and two LDH B subunits, was reported to be found in cells within the lung[72]. In total, we identified 17 unique distance restraints, including 16 intralinks and 1 interlink between LDH A and LDH B, making LDH-3 a potential target for structural modeling (Fig. 6c–f). The PDB structure of each subunit (PDB ID 4OJN for LDH A, with a resolution of 2.4 Å; and PDB ID 1I0Z for LDH B, with a resolution of 2.1 Å) is available, and 15 out of 16 intralinks were judged as correct on

these initial LDH A and LDH B structures. Both structures have been uploaded to the DisVis online platform with one detected inter-subunit distance restraint (Fig. 6c), and this link was confirmed by DisVis as a true positive (Supplementary Results 7). The invalid intralink belongs to the LDH B subunit and connects opposite pieces of the structure (residue 7 to residue 308), implying that for the correct distance representation, this link has to be mapped on a full tetramer complex model. Therefore, we submitted the initial PDB structures of each protein and predicted a model using the HADDOCK online docking suite with cross-link distance constraints specified. It was possible to map the interlink on the obtained dimer model with a distance of 28.2 Å (Fig. 6e). Consequently, both dimers were submitted to HADDOCK with no distance restraints specified, resulting in a tetrameric complex structure (Fig. 6f). On this final model, it was possible to map the last intralink between the antiparallel LDH B subunits with distances of 24.2 and 40.2 Å, of which we select the shortest as the correct interpretation (Fig. 6g). As we did not specify any distance constraints, one of the dimers modestly changed conformation, leading the 28.2-Å cross-link to become 31.2 Å; this would have been resolved by specifying distance constraints, but we show here with an independent method that the intralink can be confirmed. The distance distribution of the described cross-links is shown in Fig. 6h. Even though the number of cross-links is very limited, it was possible to postulate a model for the LDH-3 structures (Supplementary Results 8 and 9) that explains all detected cross-links. From such a model, no hard conclusions can be derived, but it can serve as an excellent starting point for further structural studies.

Reporting Summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Software availability

The XlinkX software is available at https://www.hecklab.com/software/xlinkx/.

References

1. Holzer, S. et al. Crystal structure of the N-terminal domain of human Timeless and its interaction with Tipin. *Nucleic Acids Res.* **45**, 5555–5563 (2017).
2. Komolov, K. E. et al. Structural and functional analysis of a β2-adrenergic receptor complex with GRK5. *Cell* **169**, 407–421 (2017).
3. Kim, S. J. et al. Molecular architecture of the Nup82 complex, the cytoplasmic mRNA export platform in the nuclear pore complex. *Cell* **167**, 1215–1228 (2016).
4. Plaschka, C. et al. Architecture of the RNA polymerase II–mediator core initiation complex. *Nature* **518**, 376–380 (2015).
5. Snijder, J. et al. Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state. *Science* **355**, 1181–1184 (2017).
6. Schilling, B., Row, R. H., Gibson, B. W., Guo, X. & Young, M. M. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **14**, 834–850 (2003).
7. Arlt, C., Ihling, C. H. & Sinz, A. Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. *Proteomics* **15**, 2746–2755 (2015).
8. Legal, T., Zou, J., Sochaj, A., Rappsilber, J. & Welburn, J. P. I. Molecular architecture of the Dam1 complex – microtubule interaction. *Open Biol.* **6**, 150327–150336 (2016).
9. Mosalaganti, S. et al. Structure of the RZZ complex and molecular basis of its interaction with Spindly. *J. Cell Biol.* **216**, 961–981 (2017).
10. Ciferri, C. et al. Molecular architecture of human polycomb repressive complex 2. *Elife* **1**, e00005 (2012).
11. Gupta, K., Watson, A. A., Baptista, T., Scheer, E. & Anna, L. Architecture of TAF11/TAF13/TBP complex suggests novel regulatory state in general transcription factor TFIID function. *eLife* **6**, e30395 (2017).
12. Scott, H. et al. Spatial organization and molecular interactions of the *Schizosaccharomyces pombe* Ccq1–Tpz1–Poz1 Shelterin Complex. *J. Mol. Biol.* **429**, 2863–2872 (2017).
13. Stucchi, R. et al. Regulation of KIF1A-driven dense core vesicle transport: Ca^{2+}/CaM controls DCV binding and Liprin-α/TANC2 recruits DCVs to postsynaptic sites. *Cell Reports* **24**, 685–700 (2018).
14. Liu, F., Lössl, P., Rabbitts, B. M., Balaban, R. S. & Heck, A. J. R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for co-existing respiratory supercomplexes. *Mol. Cell. Proteomics* **16**, RA117.000470 (2017).
15. Schweppe, D. K. et al. Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc. Natl Acad. Sci. USA* **114**, 1732–1737 (2017).
16. Fasci, D. et al. Histone interaction landscapes visualized by crosslinking mass spectrometry in intact cell nuclei. *Mol. Cell. Proteomics* **17**, 2018–2033 (2018).

17. Götze, M. et al. Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. *J. Am. Soc. Mass Spectrom.* **26**, 83–97 (2014).
18. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
19. Kao, A. et al. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol. Cell. Proteomics* **10**, M110.002212 (2011).
20. Müller, M. Q., Dreiocker, F., Ihling, C. H., Schäfer, M. & Sinz, A. Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal. Chem.* **82**, 6958–6968 (2010).
21. Hage, C., Iacobucci, C., Rehkamp, A., Arlt, C. & Sinz, A. The first 'zero-length' mass spectrometry-cleavable cross-linker for protein structure analysis. *Angew. Chem. Int. Ed. Engl.* **56**, 14551–14555 (2017).
22. Weisbrod, C. R. et al. In vivo protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *J. Proteome Res.* **12**, 1569–1579 (2013).
23. De Jong, L. et al. In-culture cross-linking of bacterial cells reveals large-scale dynamic protein-protein interactions at the peptide level. *J. Proteome Res.* **16**, 2457–2471 (2017).
24. Wang, X. et al. The proteasome-interacting Ecm29 protein disassembles the 26S proteasome in response to oxidative stress. *J. Biol. Chem.* **292**, 16310–16320 (2017).
25. Liu, F., Lössl, P., Scheltema, R., Viner, R. & Heck, A. J. R. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **8**, 15473.1–15473.8 (2017).
26. Benda, C. et al. Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. *Mol. Cell* **56**, 43–54 (2014).
27. Fagerlund, R. D. et al. Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex. *Proc. Natl Acad. Sci. USA* **114**, 5122–5128 (2017).
28. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12**, 1179–1184 (2015).
29. Leitner, A. et al. Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteomics* **11**, M111.014126 (2012).
30. Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14**, 1137–1147 (2015).
31. Cline, S. M. et al. Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2**, 2366–2382 (2007).
32. Pettersen, E. F. et al. UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
33. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8 https://sourceforge.net/projects/pymol/files/pymol/1.8/ (2015).
34. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.1–40.8 (2008).
35. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
36. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2014).
37. van Zundert, G. C. P. & Bonvin, A. M. J. J. DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. *Bioinformatics* **31**, 3222–3224 (2015).
38. van Zundert, G. C. P. et al. The DisVis and PowerFit Web Servers: explorative and integrative modeling of biomolecular complexes. *J. Mol. Biol.* **429**, 399–407 (2016).
39. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
40. van Zundert, G. C. P. et al. The HADDOCK2.2 Web Server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
41. Zhong, X. et al. Large-scale and targeted quantitative cross-linking MS using isotope-labeled protein interaction reporter (PIR) cross-linkers. *J. Proteome Res.* **16**, 720–727 (2016).
42. Navare, A. T. et al. Probing the protein interaction network of *Pseudomonas aeruginosa* cells by chemical cross-linking mass spectrometry. *Structure* **23**, 762–773 (2015).
43. Tang, X., Munske, G. R., Siems, W. F. & Bruce, J. E. Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Anal. Chem.* **77**, 311–318 (2005).
44. Burke, A. M. et al. Synthesis of two new enrichable and MS-cleavable cross-linkers to define protein–protein interactions by mass spectrometry. *Org. Biomol. Chem.* **13**, 5030–5037 (2015).
45. Tan, D. et al. Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. *eLife* **5**, 1–52 (2016).
46. Gutierrez, C. B. et al. Developing a novel acidic residue reactive and sulfoxide-containing MS-cleavable homobifunctional cross-linker for probing protein-protein interactions. *Anal. Chem.* **88**, 8315–8322 (2016).
47. Wang, X. et al. Molecular details underlying dynamic structures and regulation of the human 26S proteasome. *Mol. Cell. Proteomics* **16**, 840–854 (2017).
48. Herzog, F. et al. Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337**, 1348–1352 (2012).
49. Chavez, J. D. et al. Chemical crosslinking mass spectrometry analysis of protein conformations and supercomplexes in heart tissue. *Cell Syst.* **6**, 1–6 (2017).
50. Nowak, D. E., Tian, B. & Brasier, A. R. Two-step cross-linking method for identification of NF-kB gene network by chromatin immunoprecipitation. *Biotechniques* **39**, 715–724 (2005).

51. Yu, C. et al. Characterization of dynamic UbR-proteasome subcomplexes by in vivo cross-linking (X) assisted bimolecular tandem affinity purification (XBAP) and label-free quantitation. *Mol. Cell. Proteomics* **15**, 2279–2292 (2016).
52. Walker-Gray, R., Stengel, F. & Gold, M. G. Mechanisms for restraining cAMP-dependent protein kinase revealed by subunit quantitation and novel crosslinking approaches. *Proc. Natl Acad. Sci. USA* **114**, 10414–10419 (2017).
53. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
54. Walzthoeni, T. et al. xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat. Methods* **12**, 1185–1190 (2015).
55. Fischer, L., Chen, Z. A. & Rappsilber, J. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics* **88**, 120–128 (2013).
56. Chen, Z. A., Fischer, L., Cox, J. & Rappsilber, J. Quantitative cross-linking/mass spectrometry using isotope-labeled cross-linkers and MaxQuant. *Mol. Cell. Proteomics* **15**, 2769–2778 (2016).
57. Schmidt, C. et al. Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nat. Commun.* **4**, 1–11 (2013).
58. Chavez, J. D. et al. Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat. Commun.* **6**, 7928–7940 (2015).
59. Yu, C. et al. Developing a multiplexed quantitative cross-linking mass spectrometry platform for comparative structural analysis of protein complexes. *Anal. Chem.* **88**, 10301–10308 (2016).
60. Rampler, E. et al. Comprehensive cross-linking mass spectrometry reveals parallel orientation and flexible conformations of plant HOP2-MND1. *J. Proteome Res.* **14**, 5048–5062 (2015).
61. Kahraman, A. et al. Cross-link guided molecular modeling with ROSETTA. *PLoS ONE* **8**, e73411 (2013).
62. Giansanti, P., Tsiatsiani, L., Low, T. Y. & Heck, A. J. R. Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nat. Protoc.* **11**, 993–1006 (2016).
63. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
64. Jones, L. J., Haugland, R. P. & Singer, V. L. Development and characterization of the NanoOrange protein quantitation assay: a fluorescence-based assay of proteins in solution. *Biotechniques* **34**, 850–861 (2003).
65. Hennrich, M. L., Groenewold, V., Kops, G. J. P. L., Heck, A. J. R. & Mohammed, S. Improving depth in phosphoproteomics by using a strong cation exchange-weak anion exchange-reversed phase multi-dimensional separation approach. *Anal. Chem.* **83**, 7137–7143 (2011).
66. Ferber, M. et al. Automated structure modeling of large protein assemblies using crosslinks as distance restraints. *Nat. Methods* **13**, 515–523 (2016).
67. Meiring, H. D., van der Heeft, E., ten Hove, G. J. & de Jong, A. Nanoscale LC-MS(n): technical design and applications to peptide and protein analysis. *J. Sep. Sci.* **25**, 557–568 (2002).
68. Villén, J. & Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **3**, 1630–1638 (2008).
69. Udeshi, N. D., Mertins, P., Svinkina, T. & Carr, S. A. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc.* **8**, 1950–1960 (2013).
70. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A 'Proteomic Ruler' for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **13**, 3497–3506 (2014).
71. Koukourakis, M. I. et al. Lactate dehydrogenase-5 (LDH-5) overexpression in non-small-cell lung cancer tissues is linked to tumour hypoxia, angiogenic factor production and poor prognosis. *Br. J. Cancer* **89**, 877–885 (2003).
72. Zimmerman, H. J., Pincus, M. R. & Henry, J. B. in *Clinical Diagnosis and Management by Laboratory Methods* (ed. Henry, J. B.) 281–284 (WB Saunders, Philadelphia, 1996).

Acknowledgements

We thank all Heck-group members for their helpful contributions, and, in particular, F. Liu for her initial development in proteome-wide cross-linking, M. Damen and D. Hagemans for SCX support, and K. Dingess and K. Stecker for their help with the manuscript. From Thermo Fisher Scientific, we thank B. Delanghe, K. Fitzemeier and F. Berg for their collaboration on incorporating the XLinkX cross-link search engine into the Proteome Discoverer software and R. Viner for her collaborative work on DSSO cross-linking and support in mass spectrometry method development. We acknowledge financial support from the large-scale proteomics facility Proteins@Work (Project 184.032.201), embedded in the Netherlands Proteomics Centre and supported by the Netherlands Organization for Scientific Research (NWO). Additional support came from the European Union Horizon 2020 program FET-OPEN project MSmed, Project 686547.

Author contributions

O.K., A.J.R.H. and R.A.S. conceived the study. O.K. prepared the cell extracts and acquired the MS data. O.K. and R.A.S. analyzed the data. B.S., S.P. and D.F. provided various optimizations of the protocol. All authors critically read and commented on the manuscript.

Data availability

Data are publicly available through PRIDE repository PXD008418.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-018-0074-x.

Correspondence and requests for materials should be addressed to R.A..S.

Published online: 16 November 2018

Related links

Key references using this protocol

Fagerlund, R. D. et al. *Proc. Natl Acad. Sci. USA* **114**, E5122–E5128 (2017): http://www.pnas.org/content/114/26/E5122.long

Liu, F. et al. *Nat. Commun.* **8**, 15473 (2017): https://www.nature.com/articles/ncomms15473

Benda, C. et al. *Mol. Cell* **56**, 43–54 (2014): https://www.sciencedirect.com/science/article/pii/S1097276514007096?via%3Dihub

Fasci, D., van Ingen, H., Scheltema, R. A. & Heck, J. R. *Mol. Cell Proteomics* 17, 2018–2033 (2018). http://www.mcponline.org/content/17/2/216.long

nature research

Corresponding author(s): Richard A. Scheltema

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Orbitrap Fusion Lumos Tune Applciation 3.0.2022.16
Data analysis	Thermo Scientific Proteome Discoverer 2.3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data are publicly available at PRIDE repository with accession PXD008418

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	There are millions of different cells are harvested from the plate, assuming that population contains all the cell types/ Same for recombinant proteins or protein complexes, they undergo conformational changes within nanoseconds and therefore, we capture the most stable and relevant ones.
Data exclusions	Score cutoff is applied as described in the manuscript
Replication	Data can be replicated and reproduced if necessary
Randomization	Whole cell population and recombinant proteins were investigated, therefore randomization is not applicable.
Blinding	Cells and recombinant proteins were investigated, therefore blinding is not applicable.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☐ Unique biological materials
☐	☐ Antibodies
☐	☒ Eukaryotic cell lines
☐	☐ Palaeontology
☐	☐ Animals and other organisms
☐	☐ Human research participants

Methods

n/a	Involved in the study
☐	☐ ChIP-seq
☐	☐ Flow cytometry
☐	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	Cell lines and recombinant proteins are available commercially.

Antibodies

Antibodies used	n/a
Validation	n/a

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Sigma-Aldrich cat no 90071810
Authentication	None of the cell lines were authenticated
Mycoplasma contamination	Tested mycoplasma negative

Commonly misidentified lines
(See ICLAC register)

n/a

Palaeontology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session
(e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.