TITLE

Efficient and robust proteome-wide approaches for crosslinking mass spectrometry

AUTHORS

Oleg Klykov[1,2], Barbara Steigenberger[1,2], Sibel Pektaş[1,2], Domenico Fasci[1,2], Albert JR Heck[1,2], Richard A Scheltema[1,2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH, Utrecht, The Netherlands
[2] Netherlands Proteomics Center, Padualaan 8, 3584 CH, Utrecht, The Netherlands

ABSTRACT

Crosslinking mass spectrometry (XL-MS) has received considerable interest due 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 to other methods for determining PPIs, is that it provides direct spatial information for the detected interactions. This is accomplished by use of bi-functional crosslinking 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 mass spectrometry. XL-MS has received the required boost to tackle more complex samples with recent advances in crosslinking chemistry with MS-cleavable or reporter-based crosslinkers and faster, more sensitive and more versatile mass spectrometry platforms. This protocol provides a detailed description of our optimized conditions for a full proteome native protein preparation followed by crosslinking using the gas-phase cleavable crosslinking reagent DSSO. Following crosslinking, we demonstrate extensive sample fractionation and significantly simplified data analysis with XlinkX in Proteome Discoverer and subsequent protein structure investigations with DisVis and HADDOCK. This protocol produces data of high confidence and can be performed within approximately 10 d including structural investigations.

KEYWORDS: Crosslinking Mass Spectrometry / PC9 / protein-protein interactions / XlinkX / cleavable crosslinkers

32 **INTRODUCTION**

33 Structural proteomics, and more specifically crosslinking mass spectrometry (XL-MS), gained a large
34 amount of traction as a supplemental method to protein structure techniques like electron microscopy
35 (EM), nuclear magnetic resonance (NMR) and crystallography in recent years. XL-MS generally provides
36 distance constraints of lower resolution than these other techniques, but is able to pinpoint residues in
37 close proximity to the interaction interfaces between individual subunits for protein complexes of any size
38 in solution. Provided there are sufficient crosslinks, this technique can even allow for the detection and
39 definition of protein interfaces. The information obtained from XL-MS experiments have in many cases
40 been successfully leveraged to produce final protein complex models[1–5]. The distance information in
41 elucidated protein structures is obtained by use of bi-functional crosslinking molecules that link two amino
42 acids in close proximity to one another by a covalent bond. Upon proteolytic digestion of protein-protein
43 complexes, four distinct peptide products are formed. The first consists of single peptides not captured
44 by the crosslinking reagent, which therefore yields no structural information. The second group consists
45 of monolinks or dead-end links[6], constituting single peptides captured by only one side of the crosslinking
46 reagent because the other end of the linker has been quenched. These links yield structural information
47 about the surface exposed regions of the proteins. The third group consists of loop-links, which is
48 comprised of single peptides with two amino acids captured by a single linker. This group of internally
49 crosslinked peptides yield limited structural information due to the close proximity of the linked amino
50 acids. The fourth, and most information rich group consists of two peptides captured by the crosslinking
51 reagent; this yields valuable distance information for protein tertiary structure (if the two peptides
52 originate from the same protein) or for protein quaternary structure (if the two peptides originate from
53 different proteins).

54 XL-MS experiments up to this point have primarily been performed on purified recombinant proteins or
55 protein complexes[7–12]. Working with purified systems has the benefit of sufficient signal for the detection
56 of the generally low abundant crosslinks. In addition, only a handful of protein sequences need to be
57 searched during spectrum matching, allowing for a relatively straightforward data analysis. In contrast to
58 purified proteins, whole cell lysates lead to comprehensive interaction networks which can provide far
59 reaching insights into the relationships between proteins as a result of stimulation. For example, it has
60 been shown that crosslinking complete mitochondria is possible and demonstrates protein reorganization
61 under stimulation[13,14]. However, data analysis together with deep proteome coverage remain major
62 bottlenecks in these more advanced experiments. Several laboratories, including ours, have set out to
63 circumvent these issues by developing user-friendly analysis software, dedicated to unraveling the
64 identities of crosslinked peptides from complex mixtures[15,16]. Adoption of MS-cleavable crosslinkers[17–19]
65 significantly simplifies and even facilitates these identifications due to the presence of characteristic peaks
66 in the recorded spectra. Improvements in sample preparation techniques[20,21] or preserving protein
67 structure with formaldehyde[22] allows researchers to keep protein assemblies intact and capture
68 snapshots of the processes inside the living cell[23].

69 Here we present an optimized protocol for XL-MS that scales from structural analysis of single protein
70 complexes to whole cell lysates. As a showcase study, we demonstrate application of our protocol to non-
71 small cell lung cancer PC9 cells. We selected the MS-cleavable amino-reactive disuccinimidyl sulfoxide
72 (DSSO) crosslinker as it shows high efficiency, is commercially available, and has successfully been applied
73 in a large number of crosslinking experiments[13,22–26]. In addition to an optimized soft lysis and digestion
74 procedure, we present a technique for the efficient enrichment of crosslinked peptides employing offline
75 fractionation based on strong cation exchange (SCX) chromatography[26]. Optimized LC-MS/MS parameters
76 for current, state-of-the-art MS-instrumentation are presented that allow for the efficient detection of
77 DSSO-crosslinked peptides. Detailed data analysis descriptions are available together with an overview of

78 different reagents and research tasks to facilitate researchers aiming to perform these experiments. As a
79 last step, we demonstrate validation of a recorded dataset on a well-resolved crystal structure.

80 **Overview of the procedures**

81 This protocol consists of six major steps and is presented for both low as well as high complexity mixtures.
82 The first step involves gentle lysis of the cells under investigation (**Fig. 1a**). Such a gentle approach
83 liberates the protein complexes from the cellular environment while leaving their interactions intact. This
84 makes the protein complexes amenable for crosslinking, as the reagents used are generally not cell-
85 permeable. After incubating the protein mixture with DSSO, standard shotgun proteomics workflow steps
86 are applied consisting of protein denaturation, reduction, alkylation, and digestion (**Fig. 1b**). Detergents
87 are removed by precipitation or phase extraction approaches and digestion is carried out with a selection
88 of proteases – e.g. LysC/Trypsin or Chymotrypsin – to obtain a unique set of crosslinked peptides normally
89 not visible with a single protease. Further purification of the sample (e.g. removal of artefacts of the soft
90 lysis approach and/or salts from the buffers) is performed using solid-phase extraction with elevated pore
91 size C_{18} material. Following digestion and purification, well-established pre-fractionation techniques are
92 applied which specifically enrich for crosslinked peptides (**Fig. 1c**). This can be achieved with size exclusion
93 chromatography (SEC)[27], or strong cation exchange (SCX)[26]. The individual fractions are analyzed by
94 shotgun mass spectrometry for which an optimized acquisition protocol for the Orbitrap Fusion and
95 Orbitrap Fusion Lumos has been developed (**Fig. 1d**). This protocol makes use of a MS2-MS3 strategy,
96 where each precursor is fragmented and recorded aimed exclusively at breaking the MS-cleavable
97 crosslinker while keeping the individual peptides intact. This provides simultaneous insight into whether
98 the precursor ion represents crosslinked peptides and the masses of the individual peptides. For those
99 cases where it is likely that the precursor represents crosslinked peptides the precursor (MS1) is isolated
100 and fragmented (MS2), followed by selection and fragmentation of fragment peaks resulting originating
101 from the crosslinked peptide pair (MS3), providing insight into their identities. The main advantage of the
102 extra MS3 step is that each peptide can be identified from a single spectrum, resulting in higher quality
103 evidence for each peptide. However, the downside of this approach is that more time needs to be spent
104 per precursor / crosslinked peptide pair leading to a decrease in identification rates. The acquired RAW
105 data files can seamlessly be analyzed with the XlinkX search engine integrated as a node in the well-
106 established Proteome Discoverer software suite (**Fig. 1e**). The data analysis handles spectral processing,
107 potential crosslink detection, and peptide pair identification at low false discovery rates. Integration into
108 the environment of Proteome Discoverer ensures that the results can be navigated in a user-friendly
109 manner. Finally, XlinkX provides direct connections to existing software for visualizing crosslinks like
110 xiNet[28] or Cytoscape[29] (**Fig. 1f**); and generates convenient input for further structural analysis and
111 interpretation of existing crystal structures with Chimera[30] or pyMol[31]. Additionally, XlinkX plays well with
112 the I-TASSER[32–34], DisVis[35,36] and HADDOCK[37,38] software solutions, which make computational predictions
113 for new structures and interaction interfaces based on the detected crosslinks.

114 **Advantages and limitations**

115 The ability of XL-MS to elucidate PPIs complete with structural information is a major advantage over
116 currently widespread techniques like affinity purification mass spectrometry (AP-MS). However,
117 proteome-wide crosslinking applications have so far been limited to laboratories focusing on developing
118 these technologies[39,40] We attempt to facilitate further adoption by providing a detailed protocol to
119 researchers worldwide and utilize commercially available reagents including the lysine-reactive
120 crosslinking reagent DSSO, which links neighboring lysine pairs at a maximum distance of 30 Å. The
121 protocol is set up to provide information for both purified protein samples and complex mixtures. Even
122 though the focus is on DSSO, the steps are also applicable to any reagent with gas-phase cleavable
123 features, such as DSBU[18]. Non-cleavable crosslinking reagents are however not indicated for complex
124 mixtures, as these reagent require more extreme computational approaches to perform database

125 searches. Additional hurdles include complex data analysis, which we have attempted to simplify with the
126 integration of XlinkX into Proteome Discoverer, making a user-friendly XL-MS data analysis approach
127 accessible to a large audience of researchers.

128 A major limitation of XL-MS experiments remains that per detected PPI little information is available in
129 terms of number of crosslinked peptides. Typically only the top 20 – 30 % of proteins are detected and
130 only the top 100 to 500 proteins have sufficient depth of coverage for successful modeling (see
131 Anticipated results). This can potentially be alleviated through the development of enrichable chemical
132 crosslinkers. Challenges of creating such molecules lies in the ability of the molecule to retain reactivity of
133 the amine reactive groups, while keeping their agility in entering the protein structure. Several molecules
134 have so far been reported to achieve these goals and with successful application[41–43]. Another approach
135 to alleviate the relatively low level of detection would be to include crosslinking on acidic-groups (e.g.
136 DHSO[44]), which could potentially open up new locales of the protein complex for detection. However,
137 these molecules have only recently become commercially available, and therefore were not included in
138 our analysis. Alternatively, in this protocol we attempt to overcome the limit of detection by extensive
139 pre-fractionation. Such an approach works well and successfully yields thousands of detected crosslinked
140 peptides; however it also has limitations. To illustrate, our selected approach of SCX uses charge as the
141 separation mechanism, which is a property that is exuberated for crosslinked peptides, but is not
142 necessarily unique. From the number of detected normal peptides it is evident that a large background is
143 still present and therefore a large number of fractions are required to reach sufficient depth.

144 For purified complexes the opposite holds true. Mainly driven by recent advances in mass spectrometry,
145 XL-MS has become increasingly capable of generating hundreds[4,24,45,46] and in some cases even
146 thousands[26,47] of crosslinked peptides. Given the limited complexity, the use of non-cleavable crosslinking
147 reagents poses no barrier for successful data analysis. From the generated distance constraints, successful
148 modeling of protein complexes can be achieved and have been reported. The major limitation of such an
149 approach is that the researchers needs to find sufficiently successful routes of purifying the protein-
150 complex prior to investigation. Such a process can take a long time to complete and doubts about the
151 native state of the protein-complex remain. For instance, enrichable handles like His-tags are often
152 expressed on one of the proteins in the complex. Such a handle can be used to successfully enrich for the
153 complex, but can also have repercussions for the protein-complex like allosteric modifications and/or
154 complete loss of interactors.

155 **Potential applications**

156 XL-MS in recent years has received increased attention for its ability to elucidate pseudo-atomic models
157 for protein complexes and has now even seen application in highly complex backgrounds. It has already
158 been demonstrated as a useful tool for identifying purified protein complexes, and together with X-ray
159 crystallography, HDX, NMR and cryo-EM, it has become one of the staples in the structural biologist's
160 toolbox. An area where XL-MS can really take center stage is the application to transient interactions,
161 which are so labile that any sample preparation step can potentially disrupt them and inevitably result in
162 the inability to measure them. Even though PPIs within a cell might range from very stable to transient,
163 for the crosslinking reagent there is no differences, and all interactions in close proximity are potentially
164 caught by the crosslinking reagent. After successful crosslinking residues in close proximity the interaction
165 has been stabilized and normal analysis steps can be executed without losing the transient interaction.
166 Such an approach would require cell permeable crosslinking reagents, for which a number of examples
167 have been posed[21,48–50] but so far have not yet seen wide-spread application.

168 The second area where XL-MS promises to be very useful is the mapping of dynamic behavior of the
169 protein complex. A well-known example is the TRiC or CCT complex, which was reported to have a lid
170 structure that can be closed by application of ATP. Upon quantifying detected crosslinked peptides[51], this

171 lid behavior could be mapped[52]. Further steps to streamline this process have recently been taken by
172 either integrating quantifiable labels directly in the crosslinking reagents[43,53–56] or the application of TMT
173 labeling to the crosslinked peptides[57]. Even though in this particular example, the lysine residues are
174 occupied by the crosslinking reagent, TMT labeling can still successfully be applied to the N-terminus of
175 the peptides.

176 Both of these application areas are well supported by XlinkX and, combined with support for the spatial
177 restraints coming from crosslink identifications in modern protein modelling engines (HADDOCK, I-
178 TASSER, ROSETTA), will allow researchers to immediately postulate structural models from discovery
179 experiments[58,59]. With time, enrichable crosslinking chemistry will evolve to allow this for most protein
180 complexes in complex mixtures.

181 **Experimental design**

182 **Preparation of crosslinked lysate.** In this protocol both *in vitro* reconstituted complexes and purified
183 complexes are considered as low complexity samples, while cell lysates are considered as high complexity
184 samples. We make use of two types of buffers for cleanup and resuspension of the proteins / protein-
185 complexes. PBS is intended for washing samples extensively to remove traces of any buffer compounds
186 harmful to the crosslinking reaction. For example, primary amines must be avoided when performing
187 crosslinking reactions because they interfere with NHS-ester derivatives such as DSSO. In addition, specific
188 crosslinker properties must be taken into account, as for example thiols should not be present in
189 crosslinking reaction mixture in case of thiol-cleavable reagents. The crosslinking buffer (i.e. the buffer in
190 which the crosslinking reaction will be performed, but which does not contain any crosslinking reagent),
191 is intended to re-suspend the proteins prior to further steps. The crosslinking buffer is supplemented
192 directly before the crosslinking reaction with protease inhibitors and minor amounts of a reducing agent.
193 Protease inhibitors are required to prevent protein degradation due to the presence of endogenous
194 proteases, while the role of reducing agents is to prevent oxidative stress and keep proteins in their native
195 state.

196 Low complexity samples can directly be re-suspended in the crosslinking buffer. Soft lysis of the high
197 complexity samples is achieved by pushing the cells through a syringe needle to preserve organelles and
198 protein assemblies. A 27 ¾ G syringe diameter is recommended for most mammalian cell types as this is
199 the smallest diameter available and it is not required to preserve intra-cellular vesicles for this procedure.
200 In case of bacterial, plant or fungal samples harsher lysis conditions may be required. Prior to further
201 experiments, the protein concentration can be estimated with a Bradford of bicinchoninic acid assay
202 (BCA). We recommend a final protein concentration in the range 1 – 3 mg/ml to avoid aggregation of the
203 proteins, and potential precipitation, leading to incorrect results.

204 **Optimization of crosslinking reaction conditions.** Prior to the crosslinking reaction, the optimal
205 concentration of crosslinking reagent for the acquired sample concentration must be determined. To
206 achieve this, different aliquots of the sample can be crosslinked with a range of crosslinking reagent
207 concentrations and visualized by SDS-PAGE (**Fig. 2**). Depending on the sample amounts, the visualization
208 can be done for high amounts of input with a standard Coomassie or for low amounts of input with silver
209 staining. For low complexity samples, the crosslinking reagent concentration range can be built from 0.1
210 mM to 10 – 15 molar excess of crosslinking reagent to protein, while for high complexity samples the
211 range of crosslinking reagent can be from 0.1 mM to 5 mM. Please note that the limited reaction efficiency
212 for the currently available crosslinking reagents will significantly decrease identification rates for low
213 concentrations. In contrast, high concentrations will block lysines, leading to incomplete digestion with
214 lysine-cleaving proteases, and additionally can lead to over-crosslinking where individual proteins are

215 captured by the crosslinking reagent and unspecifically aggregate. It is therefore critical to use the lowest
216 possible crosslinker concentration which still initiates the reaction.

217 **Protein digestion.** The native protein complex state has been captured after the crosslinking experiment,
218 and the still present nucleic acids can be sheared to prevent precipitation after protein denaturation. In
219 our protocol we achieve this by adding benzonase to a final concentration of 1%, but in principle any
220 enzyme with similar properties can be used. Natively folded proteins, supported by disulfide bridges, are
221 generally more difficult to digest as the protease tends to be sterically hindered. To achieve the maximum
222 efficiency for protein reduction and alkylation, we employ tris(2-carboxyethyl)phosphine hydrochloride
223 (TCEP), which can be used simultaneously with 2-chloroacetamide (CAA) at elevated temperatures. Here
224 we perform the reduction/alkylation reaction at 37 ^{o}C with longer incubation times to prevent
225 modification of proteins by Urea artefacts, which is not needed if other more stable chaotropic agents are
226 used. For the digestion, a short pre-digestion with LysC followed by standard overnight trypsin digestion
227 is recommended to increase protein sequence coverage. Alternatively, the ultimate coverage of the
228 investigated proteins can potentially be increased by employing other proteases with cleavage sites
229 different from ones involved in crosslinking reaction, e.g. chymotrypsin[60]. It is beneficial to perform the
230 digestion procedure in solution, as the efficiency of in-gel digestion is lower and leads to fewer
231 identifications. Although LysN and chymotrypsin can easily be adapted for in-gel digestion, other
232 proteases can be less efficient for this approach. We note that for whole proteome approaches the benefit
233 of using alternative proteases is less clear and therefore such an approach is recommended mainly for
234 low complexity samples.

235 **Pre-fractionation of crosslinked peptides.** Fractionation increases the depth of analysis required to reach
236 the relatively low-abundant crosslinked peptides. The specific properties of crosslinked peptides allow for
237 separation of single peptides from crosslinked peptides by either SCX[26], given that crosslinked tryptic
238 peptides tend to carry extra charges compared to single tryptic peptides, or size-exclusion
239 chromatography (SEC)[27], given that crosslinked peptides are generally larger than single peptides. This
240 procedure can be performed in Stage-Tip format for low complexity samples. However, for all sample
241 types, the best coverage for samples with a loading amount of at least 100 µg is achieved either with
242 HPLC- or AKTA-assisted fractionation. In our laboratory, we found 500 µg as the optimal amount. Quality
243 control of the setup prior to separation can be performed with a partial tryptic digest of BSA[60]. When the
244 BSA run meets expectations in terms of signal and reproducibility, the sample can be fractionated
245 (**Supplementary Fig. 1**). With the described HPLC SCX setup, 25 to 50 fractions containing ca. 50 µl each
246 can be collected (**Table 1**). When samples other than whole cell lysates are fractionated 25 fractions are
247 sufficient.

248 **Desalting.** The samples at this point still contain contaminants and reagents which potentially affect
249 peptide ionization and LC separation. This is especially relevant in those cases where SCX separation with
250 non-volatile salts is used and/or when combining SCX fractions. In our laboratory we found the
251 commercially available Solid Phase Extraction (SPE) C18 kits to be most useful. Crosslinked peptides tend
252 to be longer than normal tryptic peptides due to missed cleavages where the linker has bound to the
253 peptide, requiring materials with elevated pore size to prevent sample losses. If the total amount of
254 peptide material does not exceed 20 µg, it is recommended to apply C18 Stage-Tips, for which a detailed
255 protocol is available[61]. In those cases where a large number of samples with peptide amounts up to 50 µg
256 are used, we recommend commercially available Oasis HLB or Oasis WCX 96-well elution plates.

257 **LC-MS/MS analysis. T**he amount of peptide material in each of the fractions should be estimated prior to
258 LC/MS analysis. This step is essential for the accurate estimation of the loading amounts that provide
259 enough signal for the detection of the mostly low-abundant crosslinked peptides species, while preventing

260 overloading of the LC column. This is especially helpful for those cases where extensive pre-fractionation
261 of peptides is employed. Detection of peptide concentration after digestion can be done with
262 commercially available colorimetric or fluorometric assays[62]. In the HPLC-SCX pre-fractionation setup
263 described here, the amount of sample to inject for each collected fraction can be estimated from the LC-
264 UV trace after determination of a select number of fractions. It is essential to perform nanoflow LC-
265 separation to improve chromatographic separation. MS equipment must be able to perform acquisition
266 up to MS3 level. At the MS1 level, the survey scan is recorded in Orbitrap (OT) at high resolution. For
267 selected precursors collisional-induced dissociation (CID) is applied and signature peaks for the
268 crosslinkers are recorded at middle resolution as the MS2 level. Fragments exhibiting patterns associated
269 to the used cleavable crosslinker are further subjected to a low-resolution MS3 scan in the ion trap (IT).
270 Before starting the acquisition it is recommended to perform quality control of the LC-MS/MS system, for
271 which a detailed description is available from Giansanti et al.[60]

272 **Data analysis with XlinkX.** This step is split into processing and consensus workflows and is exemplified
273 in **Figure 4**. Standard crosslinking workflows can be found as common templates in Proteome Discoverer
274 and detailed settings are described in **Supplementary Table 1**. In the processing workflow (**Fig. 4a**) the
275 first two nodes, called "Spectrum Files" and "Spectrum Selector", are common for most of PD processing
276 workflows and their settings can remain default. The next node is "XlinkX Detect", where the used
277 crosslinker and MS acquisition strategy can be specified. Depending on the crosslink chemistry, a number
278 of analysis options are available. For setting up a new crosslinker as a chemical modification, please refer
279 to **Supplementary Figure 2** and **Supplementary Tutorial**. When using non-MS-cleavable crosslinkers the
280 "NonCleavable" option should be set. For MS-cleavable crosslinkers like DSSO or DSBU, depending on the
281 acquisition strategy "MS2_MS2" (e.g. CID/ETD), "MS2_MS3" (e.g. CID/MS3 HCD) or MS2_MS2_MS3 (e.g.
282 CID/ETD/MS3 HCD) should be set. Then the workflow splits into two independent branches, where the
283 nodes "XlinkX Filter" regulate the flow of fragmentation spectra into two branches. The "Peptides Search"
284 branch executes a standard proteomics search with a peptide search engine and FDR correction method
285 of choice on those fragmentation spectra that do not contain reporter ions indicative to the presence of
286 the crosslinker. In addition to non-crosslinked peptides, partially modified peptides can be identified at
287 this step by setting relevant "Dynamic Modifications" (e.g. 'type 0 crosslinks' or monolinks). The
288 "Crosslinks" branch consists of the "XlinkX Search" and "XlinkX Validation" nodes, which perform the
289 search and validation of the fragmentation spectra that contain reporter ions indicative to the presence
290 of the crosslinker. The consensus workflow (**Fig. 4b**) consists of the following standard nodes: "MSF Files",
291 "PSM Grouper", "Peptide Validator", "Peptide and Protein Filter" connected to "Protein Marker", "Protein
292 Scorer" connected to "Protein FDR Validation" and "Protein Grouper"; their settings can remain default.
293 For the crosslinking workflow two additional nodes are introduced: "XlinkX Crosslinks Grouping" combines
294 detected Crosslink Spectra Matches (CSMs) to Crosslinks, and "XlinkX Crosslink Export" exports the results
295 in the format for the xiNET visualization toolkit[28].

296 **Search output.** The output of the crosslinking search nodes consist of several tables. The first table
297 'Crosslink Spectra Match' (CSM), analogously to the 'Peptide Spectra Match' (PSM) table, contains the
298 identification information of each set of fragmentation spectra (in the case of the 'MS2-MS3' strategy
299 employed here the initial CID scan revealing the presence of the diagnostic ions is combined with the
300 associated MS3 scans in a single set of scans). This table can contain the same peptide pair identity several
301 times due to the presence of the peptide pair with different modifications and/or charge states. The table
302 'Crosslinks' contains the on sequence identity grouped 'CSM' entries for a direct overview of the relevant
303 structural information yielded from the experiment. When relevant crosslinker modifications are set in
304 the peptide search branch (node number 7 in "Peptide Search" path on **Fig. 4a**), monolinks (dead-ends or
305 type 0 crosslinks) are shown as peptide modification in the PSM table. The search results in 2 additional
306 files that can be directly loaded into the xiNET online visualization platfrom[28]. In our laboratory we have

307 found this visualization tool especially useful for illustrating the interaction network of a single protein or
308 protein complex (**Fig. 5a**).

309 **Additional data validation.** As an additional validation step, the identified crosslinks can be mapped on a
310 protein complex with a well-resolved structure. In principle, any resolved protein complex with a high
311 number of identified crosslinks can be used. When preparing whole cell lysates of human cell lines, we
312 found it useful to map identified crosslinks of the human ribosome (**Fig. 6ab**); a complex which is
313 characteristic for many cells, and is highly abundant with well resolved crystallography. With the current
314 protocol hundreds of distance restraints within the DSSO maximum crosslinking distance of 30 Å can be
315 identified and mapped on the human ribosome structure with 3.6 Å resolution[63]. Prior to modelling of the
316 protein complex, the identified inter-subunit crosslinks (interlinks) can be validated with DisVis[35,36]. This
317 free tool provides, based on the set of crosslinks, predictions of the possible interaction interfaces and
318 provides an estimation of how many crosslinks violate all possible conformations. For the DisVis input one
319 can use available PDB structures or, in case none are available, generate a structure for which several
320 online platforms are available that make predictions based on amino acids sequences and detected in the
321 XL-MS experiment intra-link distance constraints (e.g. I-TASSER[58]). Further modelling steps based on the
322 confirmed distance restraints can be done with HADDOCK, ROSETTA or X-MOD[64].
323

324

325 **MATERIALS**

326 **REAGENTS**

327 • 10% Criterion XT Bis-Tris Protein Gel (Bio-Rad, cat. no. 3450111)
328 • 2-Chloroacetamide (CAA; Sigma-Aldrich, cat. no. C0267)
329 • Acetic Acid (Merck, cat. no. 1000632500)
330 **CAUTION** Corrosive and flammable, avoid skin contact and inhalation.
331 • Acetonitrile (ACN; Biosolve, cat. no. 012007)
332 **CAUTION** Highly flammable
333 • Ammonium Acetate ($CH_3CO_2NH_4$; AmAc; Sigma-Aldrich, cat. no. A1542)
334 • Ammonium Bicarbonate (NH_4HCO_3; Sigma-Aldrich, cat. no. 09830)
335 • BenchMark Unstained Protein Ladder (Thermo Fisher Scientific, cat. no. 10747012)
336 • Benzonase (Merck, cat. no. 70664-3)
337 • Bovine Serum Albumin (BSA; Sigma-Aldrich, cat. no. A2153)
338 • Bradford Protein Assay Dye Reagent Concentrate (Bio-Rad, cat. no. 5000006)
339 • Cells to be analyzed. The procedure is optimized for lysis of PC9 cells (formerly known as PC-14,
340 Sigma-Aldrich, cat. no. 90071810-1VL)
341 • Complete mini EDTA-free cocktail (Roche, cat. no. 000000011836170001)
342 • Dimethyl sulfoxide (DMSO; Sigma-Aldrich, cat. no. D8418)
343 • Disuccinimidyl sulfoxide (DSSO, Thermo Fisher Scientific, cat. no. A33545)
344 • DL-Dithiothreitol (DTT; Sigma-Aldrich, cat. no. 43815)
345 • Formic Acid (FA; Fluka, cat. no. 94318)
346 **CAUTION** Corrosive, flammable and toxic, avoid skin contact and inhalation
347 • GelCode Blue Stain Reagent (Coomassie stain; Thermo Fisher Scientific, cat. no. 24592)
348 • HEPES (Sigma-Aldrich, cat. no. H3375)
349 • High-Purity water obtained from a Q-POD MilliQ purification system (Merck, cat. no. ZMQSP0D01)
350 • LysC, MS grade (Wako Chemicals, cat. no. 129-02541)
351 • Magnesium chloride hexahydrate ($MgCl_2x6H_2O$; Sigma-Aldrich, cat. no. M2670)
352 • Methanol Absolute, HPLC supra-gradient (MeOH; Biosolve, cat. no. 0013680602BS)
353 **CAUTION** Flammable and toxic, avoid skin contact and inhalation.
354 • PBS (Lonza, cat. no. BE17-512F)
355 • Precision Plus Protein Dual Color Standards (Bio-Rad, cat. no. 161-0374)
356 • Sodium chloride (NaCl; Merck, cat. no. 1064041000)
357 • Sodium hydroxide (NaOH; Sigma-Aldrich, cat. no. 00000001064621000)
358 • SCX material (cation exchange-SR; Sigma-Aldrich, cat. no. 66889-U)
359 • Tris(2-carboxyethyl)phosphine hydrochloride (TCEP; Sigma-Aldrich, cat. no. C4706)
360 • Trizma, pH = 8.5 (Tris HCl and Tris base; Sigma-Aldrich, cat. no. T8818)
361 • Trypsin, MS-grade (Promega, cat. no. V528A)
362 • Urea (Merck, cat. no. 66612)
363 • XT MOPS running buffer, 20x (Bio-Rad, cat. no. 161-0788)
364 • XT sample buffer, 4x (Bio-Rad, cat. no. 161-0791)

365 **EQUIPMENT**

366 • Gel Running Chamber (Bio-Rad, model no. Mini-Protean III Cell)
367 • Power Supply (PowerPac; Bio-Rad, product no. 1645050)
368 • Molecular Imager (GS-800; Bio-Rad, product code 1707983)

369 • Plastic Syringe 5 mL, with luer slip tip (Sigma-Aldrich, product. no. Z116866)
370 • Syringe needle 27 ¾ G (BD Precisionglide; Sigma-Aldrich, product. no Z192384)
371 • Eppendorf Thermomixer (ThermoFischer Scientific, cat. no. 5382000023)
372 • Eppendorf Refrigerated Centrifuge (VWR, model no. 5417R)
373 • MilliQ Purification System (Millipore)
374 • pH meter (Meter Lab, model no. PHM210)
375 • Solid Phase Extraction columns (C18; 300 Å pore size; Grace Vydac, product no. 218SPE1000)
376 • Vacuum Concentrator (Savant SpeedVac; ThermoFisher Scientific, cat. no. SC210-A)
377 • SCX HPLC Pump A and SCX HPLC Pump B (Agilent Technologies, model no. Agilent 1200)
378 • 4.6 mm x 5.0 mm Opti-lynx trap column (C18; 49 µm particle size; Optimize Technologies, product
379 code 11-02874-TB)
380 • 50.0 mm x 1.0 mm PolyLC SCX-separation columns (PolySULFOETHYL A; 3 µm particle size; PolyLC
381 Inc., item no. 051SE0303)
382 • Agilent 1290 UPLC system (Agilent Technologies, model no. Agilent 1290)
383 • 2 cm x 100 µm double-frit C18 trap column, packed in-house (ReproSil-Pur C18-AQ, 3 µm particle
384 size; Dr Maish, mat. no. r13.aq.0001)
385 • 50 cm x 75 µm C18 analytical single-frit column, packed in-house (Poroshell 120 EC-C18, 2.7 µm
386 particle size; Agilent, lot no. B12087)
387 • Orbitrap Fusion Tribrid or Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher
388 Scientific)
389 • Distal coated fused silica emitter (PicoTip emitter; New Objective, part no. FS3602010D20)
390 • Proteome Discoverer version 2.3 or higher (Thermo Fisher Scientific, cat. no. OPTON-30795) with
391 XlinkX node (Thermo Fisher Scientific, cat. no. OPTON-30799)

392 **REAGENT SETUP**

393 **Crosslinking Lysis Buffer.** Prepare the Crosslinking Lysis Buffer from 20 mM HEPES (47.6 mg in 10 mL), 150
394 mM NaCl (87.6 mg in 10 mL) and 1.5 mM $MgCl_2$ (3 mg of $MgCl_2$ x 6 H_2O in 10 mL) in MilliQ water. Adjust
395 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
396 protease inhibitor cocktail per 10 mL of Crosslinking Lysis Buffer and 0.5 mM DTT (0.8 mg in 10 mL) right
397 before use. **CRITICAL** Freshly prepare all the solutions. First dissolve HEPES, NaCl and $MgCl_2$x6 H_2O in 5 ml
398 of water, then adjust the pH with NaOH to an optimal range for each crosslinker (e.g. 7.6-8.0 for DSSO)
399 and then bring to final volume of 10 ml. Keep buffer on ice. **CRITICAL** Avoid any possible reagents which
400 may interfere with the crosslinking reaction. In case of amino-reactive crosslinkers, buffers containing
401 primary amines must be avoided. Here DTT is used to prevent oxidative stress, but it may interfere with
402 crosslinking reaction and must be used at low concentration.
403 **DSSO stock solution.** Prepare 50 mM stock DSSO solution (1 mg in 52 µl) in DMSO **CRITICAL** DSSO may
404 undergo hydrolysis. The solution must be prepared directly before use.
405 **Gel fixing solution** Gel fixing solution is 50% of MeOH and 10% Acetic Acid (vol/vol) in MilliQ water.
406 **AmBIC solution** For 50 mM solution dissolve 197.5 mg of NH_4HCO_3 in 50 mL of MilliQ water.
407 **Protein Reduction Agent: TCEP solution** 100 mM TCEP stock solution is prepared by dissolving 25 mg of
408 TCEP in 1 mL of AmBiC solution and can be stored at -20 ^oC for several months before use. **CRITICAL** AmBiC
409 maintains pH and keeps basic environment of the reduction agent solution.
410 **Protein Alkylation Agent: CAA solution** 400 mM CAA stock solution is prepared by dissolving 38 mg of
411 CAA in 1 mL of AmBiC solution and can be stored at -20 ^oC for several months before use. **CRITICAL** AmBiC
412 maintains pH and keeps basic environment of the alkylation agent solution.

413 **LysC** Dissolve lyophilized LysC in MilliQ water, aliquot and store at -80 ^{o}C. Solution is stable and can be
414 stored at -80 ^{o}C at least until the expiration date provided by manufacturer. **CRITICAL** Each aliquot can be
415 used only once after defrosting.
416 **Trypsin** Dissolve lyophilized trypsin in MilliQ water, aliquot and store at -80 ^{o}C. Solution is stable and can
417 be stored at -80 ^{o}C at least until the expiration date provided by manufacturer. **CRITICAL** Each aliquot can
418 be used only once after defrosting.
419 **Sep-Pak washing solution 1** Sep-Pak conditioning solution 1 is 100% ACN.
420 **Sep-Pak washing solution 2** Sep-Pak washing solution 2 is 0.1% (vol/vol) formic acid in water. Freshly
421 prepare this solution on the day of use and keep at ambient temperature.
422 **Sep-Pak elution solution** Sep-Pak elution solution is 80% (vol/vol) ACN and 0.1% (vol/vol) formic acid in
423 water. Freshly prepare this solution on the day of use and keep at ambient temperature.
424 **SCX Stage-Tip washing solution 1** Stage-Tip washing buffer 1 is 100% MeOH.
425 **SCX Stage-Tip washing solution 2** Stage-Tip washing solution 2 is 20% (vol/vol) ACN and 0.4% (vol/vol)
426 formic acid in water. Freshly prepare this solution on the day of use and keep at ambient temperature.
427 **SCX Stage-Tip elution solution** SCX Stage-Tip elution solutions are based on 20% (vol/vol) ACN, 0.4%
428 (vol/vol) formic acid and 1 M AmAc (770 mg in 10 ml) in water. This buffer is diluted with Stage-Tip washing
429 solution 2 to obtain elution buffers with range of AmAc concentrations 25 mM, 50 mM, 200 mM and 500
430 mM. Freshly prepare this buffer on the day of use and keep at ambient temperature.
431 **SCX HPLC solvent A** SCX HPLC solvent A consists of 20% ACN (vol/vol) and 0.05% (vol/vol) formic acid in
432 water. Mobile phase should be stored at ambient temperature and be replaced every 2 months.
433 **SCX HPLC solvent B** SCX solvent B consists of 20% ACN (vol/vol) and 0.05% (vol/vol) formic acid in 0.5 M
434 NaCl (29.2 g in 1 L) water solution. Mobile phase should be stored at ambient temperature and be replaced
435 every 2 months.
436 **Reverse-phase UPLC solvent A** Reverse-Phase UPLC solvent A is 0.1% (vol/vol) formic acid in water. Mobile
437 phase should be stored at ambient temperature and be replaced every 2 months.
438 **Reverse-Phase UPLC solvent B** Reverse-Phase UPLC solvent B is 80% (vol/vol) ACN and 0.1% (vol/vol)
439 formic acid in water. Mobile phase should be stored at ambient temperature and be replaced every 2
440 months

441 **EQUIPMENT SETUP**
442 **CRITICAL** Our list of equipment for the SDS-PAGE gel running chamber, HPLC pumps and columns for SCX
443 fractionation and LC setups are example setups. The experiments can also be performed with similar
444 equipment.
445 **SCX HPLC Fractionation** Strong-cation exchange chromatography is performed on an Agilent 1200 HPLC
446 system. The setup is built with an Opti-lynx trap column connected to a PolyLC SCX-separation column;
447 for more details see Hennrich et al[65]. A detailed description of LC separation parameters can be found in
448 **Table 1**.
449 **LC-MS/MS analysis** The standard LC-MS/MS setup in our laboratory features an Agilent 1290 Infinity
450 UHPLC system connected to an Orbitrap Fusion Lumos. A double-frit C18 trap column (Dr Maish Reprosil
451 C18, 3 µm, 2 cm x 100 µm) is used for rapid sample desalting and followed by a single-frit analytical C18
452 column. Both columns are packed in-house and configured in a vented column setup[60]. First, the injected
453 samples are loaded onto the trapping column with a flow of 5 µl/min for 5 min of RP solvent A, and then
454 gradient elution is performed at a column flow rate of approximately 300 nl/min (split flow from 0.2
455 ml/min). Effluent from column is directly introduced into the NSI source via a coated fused silica emitter,
456 forcing chromatographic separation of the peptides using 85, 115 or 175-min methods with 65, 95 or 155-
457 min LC gradients respectively (

458 **Table 2**).

459 **SCX and C_{18} Stage-Tip preparation** Place 3 layers of SCX or C_{18} material plugs with 1.2 mm diameter each
460 in 200 ml pipette tip as described in Rappsilber et al[61]. **CRITICAL** In case of input less than 10 µg use plugs
461 of 0.5 mm diameter and place in Gel-Loader pipette tip. **CRITICAL** Do not load more than 20 µg of peptide
462 material to achieve better enrichment of crosslinked peptides and prevent loss of the material due to
463 exceeding maximum material capacity

464 **PROCEDURE**

465 **Preparation of protein sample**

466 **1|** To prepare low-complexity samples for crosslinking, refer to option A. To prepare whole cell
467 lysates, refer to option B.

468 **(A) Low-complexity samples TIMING ~0.5 h**

469 (i) Re-suspend lyophilized protein or dilute solution of purified protein complex in required amount
470 of crosslinking buffer to achieve final concentration of about 1-3 mg/ml

471 **CRITICAL STEP** Mix sample by pipette to preserve the protein (-complex) native state, avoid vortexing.

472 **CRITICAL STEP** If sample was frozen to -80 ^{0}C, first defrost on ice and only then bring to r.t.

473 **(B) Whole cell lysate preparation TIMING ~1 h**

474 (i) Re-suspend the collected cell pellet (10^7 cells) in 1 mL ice-cold crosslinking buffer and perform soft
475 lysis to keep the organelles and protein assemblies intact. Perform 30-40 quick pushes through a
476 27 ¾ G syringe.

477 **CRITICAL STEP** Avoid foaming as this leads to protein denaturation.

478 **CRITICAL STEP** Needle width must be chosen according to the size of the each cell type.

479 (ii) Remove the cell debris via centrifugation at 13,800 g for 10 min at 4 ^{0}C.

480 **CRITICAL STEP** To pellet only the organelles for further crosslinking experiments apply lower
481 centrifugation speed. Centrifugation at e.g. 3,200 g for 10 min is sufficient for nuclei.

482 (iii) Estimate protein concentration via a BCA or Bradford protein assay. For most mammalian cell
483 lines one can expect a concentration of ~2 mg/mL from a cell pellet containing 10^7 cells.

484 **CRITICAL STEP** Minimum protein concentration of 1 mg/ml is recommended.

485 **TROUBLESHOOTING**

486 **Optimization of crosslinker to protein ratio TIMING ~5 h**

487 **2|** Prepare 7-10 aliquots of sample containing small amounts of protein material.

488 **CRITICAL STEP** Per reaction mixture, 1-2 µg of single protein or complex is sufficient and 10-15 µg of
489 protein dry weight from a whole cell lysate is sufficient.

490 **3|** Add required amount of crosslinker stock solution to each aliquot to get solutions with a range of
491 crosslinker reagent concentrations. Mix sample by pipette.

492 **CRITICAL STEP** For the low complexity samples, build concentration range around 10-15 molar excess of
493 the crosslinker. In case of whole cell lysates, concentrations from 0.1 mM to 5 mM of crosslinker are
494 recommended. Please note that this range will suffice for most proteins in a proteome-wide context, but
495 might vary for specific protein complexes due to differences in the accessibility of the lysines as well as
496 the reactivity of the lysines within their specific micro-environment. The optimization step performed
497 here aims to generate an overall high crosslinking yield without creating artefacts from over-crosslinking.

498 **CRITICAL STEP** Slowly add DSSO stock to protein solution to avoid precipitation and keep solution clear.

499 **TROUBLESHOOTING**

500 **4|** Leave reaction mixture for 30 min at r.t.

501 **CRITICAL STEP** The crosslinking reaction is performed at r.t. for a maximum of 30 min. After this time,
502 approximately 90% of the crosslinking reagent will have hydrolyzed and no further crosslinks between
503 lysines in close proximity will be formed. The temperature can be altered in case of specific protein-
504 complexes; however, for most protein complexes room temperature is optimal.

505 **CRITICAL STEP** When preparing replicates, please ensure that the protein concentration, crosslinker
506 concentration and crosslinking temperature and time are kept constant.

507

5| Quench reaction with Tris solution to final concentration of 20 mM. Keep at the r.t. for 10 min.

6| Run an SDS-PAGE gel with the range of crosslinker concentrations and select the optimal crosslinker to protein ratio from the moderate concentrations of crosslinker; 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 crosslinker to 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 is used.
TROUBLESHOOTING

Protein crosslinking TIMING ~1.5 h
7| Add required amount of crosslinker stock solution, as determined in Step 6, to the protein sample and leave reaction mixture for 30 min at r.t.
CRITICAL STEP Slowly add DSSO stock to protein solution to avoid precipitation and keep solution clear.
TROUBLESHOOTING

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

Digestion and desalting of crosslinked proteins TIMING ~18 h
9| For complex samples, add the benzonase solution to a final concentration of 1% (vol/vol) and keep solution for 10 min at 37 ^{o}C. (Optional) In case of low-complexity samples, consider to add an appropriate enzyme to a nucleic acids-rich sample to shear relevant artefacts (e.g. RNAse A in case of 60S ribosome).

10| Add Urea to 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 ^{o}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 proteases. LysC is active in Urea < 6 M, trypsin is active in Urea < 2 M.

13| Add LysC for pre-digestion step in a ratio 1:75 of protease to proteins (w/w) and incubate the solution at 37 ^{o}C for 2-4 h. Then dilute the reaction mixture 3 x with AmBiC and add trypsin in ratio 1:50 of protease to proteins (w/w) and incubate the solution at 37 ^{o}C for at least 12 h.
CRITICAL STEP The pre-digestion step with LysC must be done for at least 2 hours, 4 hours for maximum efficiency.
CRITICAL STEP When using alternative proteases, e.g. chymotrypsin, adjust the digestion conditions accordingly[60].

14| Quench the reaction with FA (100 %) to 2-4% final concentration.
CAUTION Experiments with FA must be done in the fume hood.

15| Centrifuge the peptide mixture at 4,500 g for 10 min at r.t. and take the supernatant.
TROUBLESHOOTING

16| Condition a 300 Å pore size C_{18} cartridge with 2 x 1 mL of Sep-Pak washing solution 1, then equilibrate column with 2 x 1 mL of Sep-Pak washing solution 2. Please refer to Villen et al.[66] and Udeshi

549 et al.[67] for more detailed instructions. (Optional) To prevent unnecessary sample loss, consider avoiding
550 the desalting procedure in case of low-complexity sample and directly perform SCX Stage-Tip
551 fractionation, as the sample generally does not require to be concentrated prior to the fractionation (step
552 21A).

553 **17|** Load acidified crosslinked peptide digests onto the SPE cartridge. Place new collection tube and
554 collect flow-through solution.
555 **CRITICAL STEP** To provide optimal loading, maintain a slow flow rate and low pressure to the vacuum
556 scaffold. High pressure may collapse the collection tubing. The cartridge size is selected based on the
557 sample amount[66,67]. Do not allow cartridge to run dry.
558 **CRITICAL STEP** Collecting flow-through is recommended for the possibility to repeat the desalting
559 procedure or for future troubleshooting.

560 **18|** Wash the columns with 2 x 1 ml of Sep-Pak washing solution 2 to desalt the peptides.

561 **19|** Elute the desalted peptides with 2 x 200 µl of Sep-Pak elution solution into new collection tubes.
562 **CRITICAL STEP** To ensure full elution of peptides from a cartridge, maintain a slow flow rate and low
563 pressure to the vacuum scaffold.

564 **20|** Evaporate obtained fractions with vacuum centrifugation to almost dryness.
565 **CRITICAL STEP** Avoid complete dryness to prevent sample loss.
566 **PAUSE POINT** Sample can be stored at -80 ^{0}C for several months before fractionation.

567 **Fractionation of Crosslinked Peptides by SCX**
568 **21|** For low-complexity samples, please refer to option **(A).** For whole cell lysates, refer to option **(B)**

569 **(A) Low complexity sample TIMING ~3 h**
570 (i) Prepare SCX Stage-Tip column as described in the Equipment Setup section.
571 (ii) Condition SCX Stage-Tips by addition of 100 µl of Stage-Tip washing solution 1. Place the tip
572 in the Eppendorf tube and centrifuge at 1,200 g until solution passes through the tip.
573 (iii) Repeat previous step with 100 µL of Stage-Tip elution buffer and 2 x 100 µl SCX Stage-Tip
574 washing solution 2.
575 (iv) Reconstitute sample in 100 µl of 5% DMSO/10% FA/85% HOH (vol/vol/vol) and load onto SCX
576 Stage-Tip. Place new collection tube to collect flow-through solution. Centrifuge until solution
577 passes through the tip at maximum 1,000 g.
578 **CRITICAL STEP** To provide optimal loading, maintain a slow flow rate by using lower centrifugation speeds
579 but longer times.
580 **CRITICAL STEP** The flow-through should be collected and used again for the loading step to achieve
581 maximum loading efficiency or stored for troubleshooting.
582 (v) Wash Stage-Tips with 200 µl Stage-Tip washing solution 2 by centrifugation at 1,200 g until
583 solution passes through the tip.
584 (vi) Elute peptides with 2 x 100 µl of each Stage-Tip elution buffers with increasing concentration
585 of AmAc (from 25 mM to 1 M) by centrifugation until solution passes through the tip at
586 maximum 1,000 g.
587 **CRITICAL STEP** To ensure full elution of peptides from a column, maintain a slow flow rate by using lower
588 centrifugation speeds but longer times.
589 (vii) Dry obtained fractions by vacuum centrifugation to almost dryness.
590 **CRITICAL STEP** Avoid complete dryness to prevent sample loss.

591 **(B) Fractionation of complex samples by SCX HPLC TIMING ~6 h**

592 (i) Perform quality control of the SCX fractionation system with tryptic BSA digests (for examples
593 see **Supplementary Figure 1**).
594 (ii) Reconstitute samples in 5% DMSO/10% FA/85% H_2O (vol/vol/vol) and inject into the
595 described SCX HPLC fractionation setup.
596 **CRITICAL STEP** A minimum of 100 µg peptide material is required for this step. However, for isolation of
597 low-abundant crosslinked peptides the optimal injection amount is 500 µg. For the described HPLC-SCX
598 setup, 20 µl of solvent is sufficient for 500 µg.
599 (iii) Perform the SCX fractionation with a 65 or 120 min gradient. For a detailed description of the
600 HPLC gradient please refer to **Table 1.**
601 **CRITICAL STEP** The length of the gradient should be adjusted according to the amount of peptides
602 injected.
603 **TROUBLESHOOTING**

604 (iv) Pool early and late fractions together based on the UV-trace and their elution time.
605 **CRITICAL STEP** For example in the presented data on PC9 cell lysates, fractionated with described HPLC-
606 SCX setup and 120 min gradient, most of the identified crosslinks are localized in 10 fractions (**Fig. 3**).
607 Fractions 12-26 are considered as early, while fractions 37-42 as late. For pilot studies it is advised to run
608 LC-MS/MS experiments for at least 20 samples. This number can be decreased to 5-10 after the
609 crosslink-rich fractions have been located for the used setup.
610 (v) An additional desalting step is required for the late and pooled SCX fractions.
611 **CRITICAL STEP** The current protocol is optimized for an LC-MS/MS setup incorporating a trapping column.
612 For setups without trapping column, all fractions must be desalted.
613 (vi) For pooled fractions, where the peptide dry weight exceeds 20 µg (e.g. pooled first fractions),
614 repeat steps 16-19.
615 (vii) For pooled fractions, where the peptide dry weight is less than 20 µg (e.g. pooled last
616 fractions), C_{18} Stage-Tip columns must be used as described below.
617 **CRITICAL STEP** The same solutions as for Sep-Pak (step 16-19) can be used in those cases where C_{18}
618 material is applied.
619 (viii) Condition C_{18} Stage-Tips by adding 100 µl of Sep-Pak washing solution 1. Place the tip in the
620 Eppendorf tube and centrifuge at 1,200 g until solution passes through the tip.
621 (ix) Repeat previous step with 2 x 100 µl Sep-Pak washing solution 2.
622 (x) Load the samples onto the C_{18} Stage-Tip. Place new collection tubes to collect flow-through
623 solution. Centrifuge until solution passes through the tip at maximum 1000 g.
624 **CRITICAL STEP** For an optimal loading, maintain a slow flow rate by using lower centrifugation speeds but
625 longer times.
626 **CRITICAL STEP** The flow-through should be collected and either be used again at the loading step to
627 achieve maximum loading efficiency or used for future troubleshooting.
628 (xi) Wash C_{18} Stage-Tips with 2 x 100 µl Sep-Pak washing solution 2 by centrifugation until solution
629 passes through the tip at 1,200 g.
630 (xii) Elute peptides with 2 x 100 µl of Sep-Pak washing solution by centrifugation at maximum 1000
631 g until solution passes through the tip.
632 **CRITICAL STEP** To ensure full elution of peptides from the Stage-Tips, maintain a slow flow rate by using
633 lower centrifugation speeds but longer times.
634 (xiii) Dry obtained fractions by vacuum centrifugation to almost dryness.
635 **CRITICAL STEP** Avoid complete dryness to prevent sample loss.
636 **PAUSE POINT** Dried fractions can be store at -80 ^{o}C for several months before analysis.

637 **LC-MS/MS Analysis TIMING ~1 d**

638 **22|** Estimate the peptide amount in each fraction after SCX fractionation by injecting 0.1-5% of several
639 fractions on a LC-MS/MS system and analyze at short gradients. The required injection volume for all
640 fractions can then be estimated from the UV-trace.
641 **TROUBLESHOOTING**

642 **23|** Re-suspend the dried peptide pellet in 5% DMSO/10% FA/85% H_2O (vol/vol/vol) and inject an
643 appropriate amount into the LC-MS/MS system. Longer LC gradients and running replicates in parallel
644 tends to increase the number of identified crosslinked peptides. For samples of low-complexity, 85-min
645 runs are sufficient. For low amounts (e.g. late pooled SCX fractions), shorter gradients are recommended
646 to increase the sensitivity. The MS acquisition method for MS-cleavable crosslinkers must consist of a MS1
647 high-resolution survey scan, followed by an MS2 scan, which again is followed for fragments of potential
648 crosslinked peptides by MS3 fragmentation. For a review of alternative fragmentation strategies
649 supported XLinkX, please refer to **Supplementary Tutorial**. Shortly, For MS-cleavable crosslinkers like
650 DSSO or DSBU, depending on the acquisition strategy "MS2_MS2" (e.g. CID/ETD), "MS2_MS3" (e.g.
651 CID/MS3 HCD) or MS2_MS2_MS3 (e.g. CID/ETD/MS3 HCD) should be set. Please refer to

652 **Table 2** for a detailed description of LC and MS parameters; however, the most appropriate settings can
653 also be found as templates in the Xcalibur method editor.

654 **Data Analysis with Proteome Discoverer 2.3 TIMING ~17 h**
655 **24|** Perform the data analysis with XLinkX node incorporated in Proteome Discoverer (version 2.3 or
656 higher). General workflow schemes can be found in **Figure 4** and are also provided with the Proteome
657 Discoverer installed as common templates. Main settings are described below and a complete list of
658 parameters is available in **Supplementary Table 1.** Detailed description of the crosslinker modification
659 settings are provided in **Supplementary Tutorial** and **Supplementary Figure 2.**

660 **25|** Set processing workflow. In the node "XlinkX Detect", specify acquisition strategy as "MS2_MS3".
661 Define crosslink modification as DSSO on lysines.

662 **26|** In the peptide search path, set "XlinkX Filter" to "Peptides". In our laboratory we use the
663 "Sequest" search node, although other search nodes are equally applicable. Filtered spectra were
664 matched against the *Homo sapiens* database from *SwissProt* (version 2017_10, 20,230 sequences) or the
665 single BSA sequence (Uniprot ID P02679). Regardless of the search node used, set the enzyme to "Trypsin"
666 and maximum number of missed cleavages as 2. Set the precursor mass tolerance at 20 ppm and fragment
667 mass tolerance at 0.05 Da for ion trap readout or 20 ppm for the Orbitrap readout. Set
668 carbamidomethylation of cysteines as a fixed modification and oxidation of methionine and protein N-
669 terminal acetylation as variable modifications. For detection of monolinks, set DSSO-hydrolyzed and
670 DSSO-tris as variable modifications.
671 **CRITICAL STEP** Adjust enzyme specificity and number of missed cleavages according to the used protease.

672 **27|** In the XlinkX search path, set "XlinkX Filter" to "Crosslinks". In the "XlinkX Search" node, filtered
673 spectra are matched against *Homo sapiens* database from *SwissProt* (version 2017_10, 20,230 sequences)
674 or combined BSA sequence (Uniprot ID P02679) and *E. coli* strain K12 SwissProt database (version
675 2017_10, 4306 sequences). Set the enzyme name to "Trypsin (Full)" and maximum number of missed
676 cleavages to 2. Set minimum peptide length to 5 and minimum peptide mass to 300, while maximum
677 peptide mass has to be set to 7000. Search the precursor mass tolerance at 10 ppm, FTMS fragment mass
678 at 20 ppm and ITMS fragment mass at 0.5 Da. Set carbamidomethylation of cysteines as a fixed
679 modification and oxidation of methionines and protein N-terminal acetylation as a variable modifications.
680 In the "XlinkX Validation" node define FDR threshold as 0.01 and FDR strategy for low complexity samples
681 as "Simple" and for the whole cell lysates as "Percolator".
682 **CRITICAL STEP** Adjust enzyme specificity and number of missed cleavages according to the used protease.
683 **CRITICAL STEP** Search database must contain at least 100 other protein sequences which are used as
684 support for FDR control.
685 **TROUBLESHOOTING**

686 **28|** Set consensus workflow. In "XlinkX Crosslink Export" specify the file path, where the xiNET input
687 files will be stored.

688 **Data representation and additional validation TIMING ~1 d**
689 **29|** In our laboratory, visualization of detected crosslinks is done on multiple levels with xiNET, pyMol
690 or Chimera and Cytoscape. PD generates all required input files for xiNET including the fasta file with
691 protein sequences and the detected crosslinked residues, which suits for single proteins and protein
692 complexes. If a protein of interest already has an existing crystal structure, visualization in pyMOL or
693 Chimera can be done. In case of whole cell lysates, we apply Cytoscape to visualize all detected crosslinked
694 peptide pairs. For the analysis of thousands of detected crosslinks, export PD output tables to Excel format

695 and upload to Cytoscape and then cluster proteins by biological function (**Supplementary Methods 1**),
696 pathway or relevant protein complexes.
697 **CRITICAL STEP** In line with normal shotgun proteomics experiments, when e.g. triplicates are analyzed,
698 we recommend that only crosslinks detected in at least 2 out of 3 samples are accepted.

699 **30|** Check the quality of obtained data by additional validation of detected crosslinks on a well-
700 resolved structure of the protein complex. Highlight the protein or protein complex for which a high
701 number of crosslinks is available. Filter these complexes based on availability of structures in existing
702 databases (e.g. human ribosome, PDB entry 4UG0). For quick mapping of a large number of crosslinks,
703 use pyMol and run relevant scripts (**Supplementary Methods 2).**
704 **CRITICAL STEP** Structure for additional validation must be chosen according to crosslinking experimental
705 conditions.
706 **TROUBLESHOOTING**

707 **31|** Download or generate the PDB structure for each subunit of the protein complex of interest based
708 on available amino acid sequence and obtained intra-protein distance restraints. Submit PDB structures
709 together with the inter-protein crosslinks (interlinks) to the DisVis online platform for additional validation
710 and perform further structural investigations of protein complex with confirmed crosslinks. After
711 crosslinks are confirmed, submit a HADDOCK docking run for a protein complex of interest.
712 **TROUBLESHOOTING**

713 ## TROUBLESHOOTING

714 Troubleshooting advice can be found in

715 **Table 3**

716 **TIMING**

717 Step 1A, preparation of low-complexity protein samples: 0.5 h

718 Step 1B, preparation of whole cell lysates: 1 h

719 Steps 2-6, optimization of crosslinker to protein ratio: 5 h

720 Steps 7-8, protein crosslinking: 1.5 h

721 Steps 9-20, digestion and desalting of crosslinked proteins: 18 h

722 Step 21A, fractionation of low-complexity samples: 3 h

723 Step 21B, fractionation of complex samples by HPLC-SCX: 6 h

724 Steps 22-23, LC-MS/MS analysis: 1 d

725 Steps 24-28, data analysis with Proteome Discoverer 2.3: 17 h

726 Steps 29-30, data representation and additional validation: 1 d

727 Step 31, structural investigations: 1-10 d

728

729

730

731 **ANTICIPATED RESULTS**
732 Here we describe the recommended procedures to perform crosslinking experiments for samples of both
733 low and high complexity. The protocol includes descriptions of both data analysis as well as data
734 interpretation and makes use of the MS-cleavable crosslinker DSSO, as this significantly simplifies data
735 analysis for complex samples. Nevertheless, MS-cleavable linkers with alternative chemistry can be used
736 as the software is fully configurable. Additionally, a validation procedure is described in detail for the
737 human ribosome complex. As an proof-of-principle example of single protein crosslinking we show BSA
738 digested with a combination of the proteases LysC with trypsin and chymotrypsin. The detected crosslinks
739 were mapped on BSA PDB structure ID 4F5S (resolution of 2.5 Å). For the LysC/trypsin digest 75 unique
740 distance restraints were mapped on BSA, out of which 59 or 79% were within the DSSO crosslinking range
741 of 30 Å (**Fig. 5f**). With chymotrypsin 20 BSA crosslinks were detected out of which 100% were within the
742 DSSO crosslinking range (**Fig. 5g**). Out of all detected crosslinks 65% are unique for chymotrypsin and were
743 not detected with standard proteases.

744 For our high complexity example, the optimal crosslinker-to-protein ratio was established prior to the
745 main crosslinking experiment. Small amounts of sample were crosslinked with a range of linker reagent
746 concentration, after which the optimal ratio could be established by SDS-PAGE (**Fig. 2**). For the lane where
747 no crosslinking reaction was performed (**Fig. 2**, lane 1), the high MW region is empty indicating that the
748 available protein complexes disintegrated under the denaturing conditions; demonstrating that none of
749 the available protein complexes are held together by the crosslinking reagent (**Fig. 2,** box A). For the lanes
750 where the crosslinking reagent was added in high concentrations (**Fig. 2**, lanes 5-7), the lower MW region
751 is less intense (**Fig. 2,** box B) indicating that non-interacting small species form unspecific agglomerates
752 and move to the upper part of the gel (**Fig. 2,** box C). The intermittent lanes with a concentration around
753 1 mM show a good compromise between the two (**Fig. 2**, lanes 2-4). After optimization of the crosslinking
754 conditions and application to the final sample, the final peptide mixtures were fractionated by SCX-based
755 separation to specifically enrich for crosslinked peptides (**Fig. 3a**). Based on the UV trace, intensity and
756 fraction retention time, 31 fractions potentially containing the crosslinked peptides were selected for
757 further analysis (**Fig. 3b)**. As can be noted from the distribution of detected CSMs, the early fractions (e.g.
758 12-26) and late fractions (e.g. 37-50) can be pooled. This results in a total of 12 samples for analysis,
759 significantly decreasing LC-MS/MS measurement time with no dramatic loss in terms of detected
760 crosslinks. For the whole cell lysate, the total analysis time was 17 hours (approximately 10 hours for the
761 XlinkX search alone) on a 3.47 GHz 8-core computer with 48 Gb of RAM. We identified approximately
762 50,000 reporter ions, resulting in 3,790 CSMs grouped into 2,121 unique peptide pairs (**Fig 5a**). Out of this
763 545 crosslinks or 25.7% are inter-protein and 1695 are intra-protein crosslinks (**Fig. 5b**). To estimate the
764 sensitivity of our approach, we correlated the detected crosslinked proteins to protein copy numbers per
765 HeLa cell[68] (**Fig. 5c**). The lower border for 95% of the identifications is roughly at 1.12e6 copies per cell. In
766 total it was possible to identify 1098 unique crosslinked proteins (**Fig. 5d**). Approximately 66% of the
767 detected protein-protein interactions are mapped by 1 or 2 crosslinks, while for the rest 3 or more unique
768 peptide pairs are identified. The interaction network is visualized either with Cytoscape with grouping of
769 detected identifications by protein function extracted from the EggNOG database v. 4.5.1 (**Fig. 5e**) or with
770 xiNET for the low complexity samples (**Fig. 5fg**).

771 The detected distance constraints for the Ribosome can be mapped on the existing crystal structure (3.6
772 Å resolution; PDB ID 4UG0). A total of 225 unique peptide pairs for this protein-complex were detected,
773 out of which 182 could be visualized on the crystal structure. Considering that the maximum DSSO
774 crosslinking distance is limited by 30 Å, 177 crosslinks or 97% were valid (**Fig. 6ab**). As a further example
775 of crosslink validation we utilize DisVis to map detected peptide-pairs for the lactate-dehydrogenase (LDH)
776 complex, consisting of LDH A (LDH M) and LDH B (LDH H). PC9 cells are non-small cell lung cancer cells for
777 which LDH-5 is overexpressed, a protein complex which is built from four LDH A subunits[69]. Previously,

778 LDH-3 was reported for this cell line, a complex which is built up from two LDH A and two LDH B subunits[70].
779 In total we identified 17 unique distance restraints, including 16 intralinks and 1 interlink between LDH A
780 and LDH B, making LDH-3 a potential target for structural modelling (**Fig. 6c-f**). The PDB structure of each
781 subunit (PDB ID 4OJN for LDH A with resolution 2.4 Å, and LDH B 1I0Z with resolution 2.1 Å) is available
782 and 15 out of 16 intralinks were judged as correct on these initial LDH A and LDH B structures. Both
783 structures have been uploaded to the DisVis online platform with 1 detected intersubunit distance
784 restraint (**Fig. 6c**) and this link was confirmed by DisVis as a true positive (**Supplementary Results 7**). The
785 invalid intralink belongs to the LDH B subunit and connects opposite pieces of the structure (residue 7 to
786 residue 308), implying that for the correct distance representation this link has to be mapped on a full
787 tetramer complex model. Therefore we submitted the initial PDB structures of each protein and predicted
788 a model with the HADDOCK online docking suite with crosslink distance constraints specified. It was
789 possible to map the interlink on the obtained dimer model with a distance of 28.2 Å (**Fig. 6e**).
790 Consequently, both dimers were submitted to HADDOCK with no distance restraints specified, resulting
791 in a tetrameric complex structure (**Fig. 6f**). On this final model it was possible to map the last intralink
792 between the antiparallel LDH B subunits with distances of 24.2 Å and 40.2 Å, of which we select the
793 shortest as the correct interpretation. As we did not specify any distance constraints, one of the dimers
794 modestly changed conformation leading the 28.2 Å crosslink to become 31.2 Å; this would have been
795 resolved by specifying distance constraints, but we show here with an independent method that the
796 intralink can be confirmed. The distance distribution of the described crosslinks is shown on **Figure 6h**.
797 Even though the amount of crosslinks is very limited, it was possible to postulate a model for the LDH-3
798 structures (**Supplementary Results 8-9**) that explains all detected crosslinks. From such a model no hard
799 conclusions can be derived, but it can serve as an excellent starting point for further structural studies.
800

801 **ACKNOWLEDGEMENTS** We thank all Heck-group members for their helpful contributions, and in
802 particular Fan Liu for her initial development in proteome wide crosslinking, Mirjam Damen and
803 Dominique Hagemans for the SCX support. From Thermo Fisher Scientific we thank Bernard Delanghe, Kai
804 Fitzemeier and Frank Berg for their collaboration on incorporating the XLinkX crosslink search engine into
805 the Proteome Discoverer software and Rosa Viner for her collaborative work on DSSO crosslinking and
806 support in mass spectrometry method development.

807 We acknowledge financial support by the large-scale proteomics facility Proteins@Work (Project
808 184.032.201) embedded in the Netherlands Proteomics Centre and supported by the Netherlands
809 Organization for Scientific Research (NWO). Additional support came through the European Union
810 Horizon 2020 program FET-OPEN project MSmed, Project 686547.

811 **AUTHOR CONTRIBUTIONS** OK, AJRH and RAS conceived of the study. OK prepared the cell extracts and
812 acquired the MS data. OK and RAS analyzed the data. BS, SP and DF provided various optimizations to the
813 protocol. All authors critically read and commented on the manuscript.

814 **COMPETING FINANCIAL INTERESTS** The authors declare no competing financial interests.

815 **REFERENCES**

816 1. Holzer, S. *et al.* Crystal structure of the N-terminal domain of human Timeless and its interaction
817 with Tipin. *Nucleic Acids Res.* **45,** 5555–5563 (2017).

818 2. Komolov, K. E. *et al.* Structural and Functional Analysis of a β2-Adrenergic Receptor Complex with
819 GRK5. *Cell* **169,** 407–421 (2017).

820 3. Kim, S. J. *et al.* Molecular Architecture of the Nup82 Complex, the Cytoplasmic mRNA Export
821 Platform in the Nuclear Pore Complex. *Cell* **167,** 1215–1228 (2016).

822 4. Plaschka, C. *et al.* Architecture of the RNA polymerase II–Mediator core initiation complex.
823 *Nature* **518,** 376–380 (2015).

824 5. Snijder, J. *et al.* Structures of the cyanobacterial circadian oscillator frozen in a fully assembled
825 state. **355,** 1181–1184 (2017).

826 6. Schilling, B., Row, R. H., Gibson, B. W., Guo, X. & Young, M. M. MS2Assign, automated
827 assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am.*
828 *Soc. Mass Spectrom.* **14,** 834–850 (2003).

829 7. Arlt, C., Ihling, C. H. & Sinz, A. Structure of full-length p53 tumor suppressor probed by chemical
830 cross-linking and mass spectrometry. *Proteomics* **15,** 2746–2755 (2015).

831 8. Legal, T., Zou, J., Sochaj, A., Rappsilber, J. & Welburn, J. P. I. Molecular architecture of the Dam1
832 complex – microtubule interaction. *Open Biol.* **6,** 150327–150336 (2016).

833 9. Mosalaganti, S. *et al.* Structure of the RZZ complex and molecular basis of its interaction with
834 Spindly. *J. Cell Biol.* **216,** 961–981 (2017).

835 10. Ciferri, C. *et al.* Molecular architecture of human polycomb repressive complex 2. *Elife* **1,** e00005
836 (2012).

837 11. Gupta, K., Watson, A. A., Baptista, T., Scheer, E. & Anna, L. Architecture of TAF11 / TAF13 / TBP
838 complex suggests novel regulatory state in General Transcription Factor TFIID function. **6,** e30395
839 (2017).

840 12. Scott, H. *et al.* Spatial Organization and Molecular Interactions of the Schizosaccharomyces
841 pombe Ccq1–Tpz1–Poz1 Shelterin Complex. *J. Mol. Biol.* **429,** 2863–2872 (2017).

842 13. Liu, F., Lössl, P., Rabbitts, B. M., Balaban, R. S. & Heck, A. J. R. The interactome of intact
843 mitochondria by cross-linking mass spectrometry provides evidence for co-existing respiratory
844 supercomplexes. *Mol. Cell. Proteomics* **16,** RA117.000470 (2017).

845 14. Schweppe, D. K. *et al.* Mitochondrial protein interactome elucidated by chemical cross-linking
846 mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 1732–1737 (2017).

847 15. Götze, M. *et al.* Automated assignment of MS/MS cleavable cross-links in protein 3d-structure
848 analysis. *J. Am. Soc. Mass Spectrom.* **26,** 83–97 (2014).

849 16. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9,**
850 904–906 (2012).

851 17. Kao, A. *et al.* Development of a Novel Cross-linking Strategy for Fast and Accurate Identification
852 of Cross-linked Peptides of Protein Complexes. *Mol. Cell. Proteomics* **10,** M110.002212 (2011).

853 18. Müller, M. Q., Dreiocker, F., Ihling, C. H., Schäfer, M. & Sinz, A. Cleavable cross-linker for protein
854 structure analysis: Reliable identification of cross-linking products by tandem MS. *Anal. Chem.* **82,**
855 6958–6968 (2010).

856 19. Hage, C., Iacobucci, C., Rehkamp, A., Arlt, C. & Sinz, A. The First 'Zero-Length' Mass Spectrometry-
857 Cleavable Cross-Linker for Protein Structure Analysis. *Angew. Chemie Int. Ed.* **56,** 14551–14555
858 (2017).

859 20. Weisbrod, C. R. *et al.* In Vivo Protein Interaction Network Identified with a Novel Real-Time Cross-
860 Linked Peptide Identification Strategy. *J. Proteome Res.* **12,** 1569–1579 (2013).

861 21. De Jong, L. *et al.* In-Culture Cross-Linking of Bacterial Cells Reveals Large-Scale Dynamic Protein-
862 Protein Interactions at the Peptide Level. *J. Proteome Res.* **16,** 2457–2471 (2017).

863 22. Wang, X. *et al. The Proteasome-Interacting Ecm29 Protein Disassembles the 26S Proteasome in*
864 *Response to Oxidative Stress*. *J. of Biol. Chem.* **292,** (2017).

865 23. Liu, F., Lössl, P., Scheltema, R., Viner, R. & Heck, A. J. R. Optimized fragmentation schemes and
866 data analysis strategies for proteome-wide cross-link identification. **8,** 15473.1-15473.8 (2017).

867 24. Benda, C. *et al.* Structural Model of a CRISPR RNA-Silencing Complex Reveals the RNA-Target
868 Cleavage Activity in Cmr4. *Mol. Cell* **56,** 43–54 (2014).

869 25. Fagerlund, R. D. *et al.* Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR
870 adaptation complex. *Proc. Natl. Acad. Sci.* **114,** 5122–5128 (2017).

871 26. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by
872 cross-linking mass spectrometry. *Nat Meth* **12,** 1179–1184 (2015).

873 27. Leitner, A. *et al.* Expanding the Chemical Cross-Linking Toolbox by the Use of Multiple Proteases
874 and Enrichment by Size Exclusion Chromatography. *Mol. Cell. Proteomics* **11,** M111.014126
875 (2012).

876 28. Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution.
877 *Mol. Cell. Proteomics* **14,** 1137–1147 (2015).

878 29. Cline, S. M. *et al.* Integration of biological networks and gene expression data using Cytoscape.
879 *Nat. Protoc.* **2,** 2366–2382 (2007).

880 30. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis.
881 *J. Comput. Chem.* **25,** 1605–1612 (2004).

882 31. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

883 32. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9,** 40.1-40.8
884 (2008).

885 33. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure
886 and function prediction. *Nat. Protoc.* **5,** 725–738 (2010).

887 34. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12,** 7–8
888 (2014).

889 35. van Zundert, G. C. P. & Bonvin, A. M. J. J. DisVis: Quantifying and visualizing accessible interaction
890 space of distance-restrained biomolecular complexes. *Bioinformatics* **31,** 3222–3224 (2015).

891 36. van Zundert, G. C. P. *et al.* The DisVis and PowerFit Web Servers: Explorative and Integrative
892 Modeling of Biomolecular Complexes. *J. Mol. Biol.* **429,** 399–407 (2016).

893 37. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A protein-protein docking approach
894 based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125,** 1731–1737 (2003).

895 38. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of
896 Biomolecular Complexes. *J. Mol. Biol.* **428,** 720–725 (2016).

897 39. Zhong, X. *et al.* Large-Scale and Targeted Quantitative Cross-Linking MS Using Isotope-Labeled
898 Protein Interaction Reporter (PIR) Cross-Linkers. *J. Proteome Res.* **16,** 720–727 (2016).

899 40. Navare, A. T. *et al.* Probing the Protein Interaction Network of Pseudomonas aeruginosa Cells by
900 Chemical Cross-Linking Mass Spectrometry. *Struct. Des.* **23,** 762–773 (2015).

901 41. Tang, X., Munske, G. R., Siems, W. F. & Bruce, J. E. Mass spectrometry identifiable cross-linking
902 strategy for studying protein-protein interactions. *Anal. Chem.* **77,** 311–318 (2005).

903 42. Burke, A. M. *et al.* Synthesis of two new enrichable and MS-cleavable cross-linkers to define
904 protein–protein interactions by mass spectrometry. *Org. Biomol. Chem.* **13,** 5030–5037 (2015).

905 43. Tan, D. *et al.* Trifunctional cross-linker for mapping protein-protein interaction networks and
906 comparing protein conformational states. *Elife* **5,** 1–52 (2016).

907 44. Gutierrez, C. B. *et al.* Developing a Novel Acidic Residue Reactive and Sulfoxide-containing MS-
908 cleavable Homobifunctional Cross-linker for Probing Protein-Protein Interactions. *Anal. Chem.* **88,**
909 8315–8322 (2016).

910 45. Wang, X. *et al.* Molecular Details Underlying Dynamic Structures and Regulation of the Human
911 26S Proteasome. *MCP Pap. Press.* M116.065326 (2017).

912 46. Herzog, F. *et al.* Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-
913 Linking and Mass Spectrometry. *Science (80-.).* **337,** 1348–1352 (2012).

914 47. Chavez, J. D. *et al.* Chemical Crosslinking Mass Spectrometry Analysis of Protein Conformations
915 and Supercomplexes in Heart Tissue. *Cell Syst.* **6,** 1–6 (2017).

916 48. Nowak, D. E., Tian, B. & Brasier, A. R. Two-step cross-linking method for identification of NF-kB
917 gene network by chromatin immunoprecipitation. *Biotechniques* **39,** 715–724 (2005).

918 49. Yu, C. *et al.* Characterization of Dynamic UbR-Proteasome Subcomplexes by In vivo Cross-linking
919 (X) Assisted Bimolecular Tandem Affinity Purification (XBAP) and Label-free Quantitation. *Mol.*
920 *Cell. Proteomics* (2016). doi:10.1074/mcp.M116.058271

921 50. Walker-Gray, R., Stengel, F. & Gold, M. G. Mechanisms for restraining cAMP-dependent protein
922 kinase revealed by subunit quantitation and novel crosslinking approaches. *Proc. Natl. Acad. Sci.*
923 *U. S. A.* (2017). doi:10.1073/pnas.1701782114

924 51. Cox, J. *et al.* Accurate Proteome-wide Label-free Quantification by Delayed Normalization and
925 Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **13,** 2513–2526 (2014).

926 52. Walzthoeni, T. *et al.* xTract: software for characterizing conformational changes of protein
927 complexes by quantitative cross-linking mass spectrometry. *Nat. Methods* **12,** 1185–90 (2015).

928 53. Fischer, L., Chen, Z. A. & Rappsilber, J. Quantitative cross-linking/mass spectrometry using

929 isotope-labelled cross-linkers. *J. Proteomics* **88,** 120–128 (2013).

930 54. Chen, Z. A., Fischer, L., Cox, J. & Rappsilber, J. Quantitative Cross-linking/Mass Spectrometry
931 Using Isotope-labeled Cross-linkers and MaxQuant. *Mol. Cell. Proteomics* **15,** 2769–2778 (2016).

932 55. Schmidt, C. *et al.* Comparative cross-linking and mass spectrometry of an intact F-type ATPase
933 suggest a role for phosphorylation. *Nat. Commun.* **4,** 1–11 (2013).

934 56. Chavez, J. D. *et al.* Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat.*
935 *Commun.* **6,** 7928 (2015).

936 57. Yu, C. *et al.* Developing a Multiplexed Quantitative Cross-linking Mass Spectrometry Platform for
937 Comparative Structural Analysis of Protein Complexes Developing a Multiplexed Quantitative
938 Cross-linking Mass Spectrometry Platform for. (2016). doi:10.1021/acs.analchem.6b03148

939 58. Rampler, E. *et al.* Comprehensive Cross-Linking Mass Spectrometry Reveals Parallel Orientation
940 and Flexible Conformations of Plant HOP2-MND1. *J. Proteome Res.* **14,** 5048–5062 (2015).

941 59. Kahraman, A. *et al.* Cross-Link Guided Molecular Modeling with ROSETTA. *PLoS One* **8,** e73411
942 (2013).

943 60. Giansanti, P., Tsiatsiani, L., Low, T. Y. & Heck, A. J. R. Six alternative proteases for mass
944 spectrometry–based proteomics beyond trypsin. *Nat. Protoc.* **11,** 993–1006 (2016).

945 61. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-
946 fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2,** 1896–1906
947 (2007).

948 62. Jones, L. J., Haugland, R. P. & Singer, V. L. Development and characterization of the NanoOrange
949 protein quantitation assay: A fluorescence-based assay of proteins in solution. *Biotechniques* **34,**
950 850–861 (2003).

951 63. Hennrich, M. L., Groenewold, V., Kops, G. J. P. L., Heck, A. J. R. & Mohammed, S. Improving depth
952 in phosphoproteomics by using a strong cation exchange-weak anion exchange-reversed phase
953 multidimensional separation approach. *Anal. Chem.* **83,** 7137–7143 (2011).

954 64. Ferber, M. *et al.* Automated structure modeling of large protein assemblies using crosslinks as
955 distance restraints. *Nat. Methods* **13,** 515–523 (2016).

956 65. Meiring, H. D., van der Heeft, E., ten Hove, G. J. & de Jong, A. Nanoscale LC-MS(n): technical
957 design and applications to peptide and protein analysis. *J. Sep. Sci.* **25,** 557–568 (2002).

958 66. Villén, J. & Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by
959 mass spectrometry. *Nat. Protoc.* **3,** 1630–1638 (2008).

960 67. Udeshi, N. D., Mertins, P., Svinkina, T. & Carr, S. A. Large-scale identification of ubiquitination
961 sites by mass spectrometry. *Nat. Protoc.* **8,** 1950–1960 (2013).

962 68. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A ‘Proteomic Ruler’ for Protein Copy Number
963 and Concentration Estimation without Spike-in Standards. *Mol. Cell. Proteomics* **13,** 3497–3506
964 (2014).

965 69. Koukourakis, M. I. *et al.* Lactate dehydrogenase-5 (LDH-5) overexpression in non-small-cell lung
966 cancer tissues is linked to tumour hypoxia, angiogenic factor production and poor prognosis. *Br.*

967 *J. Cancer* **89,** 877–885 (2003).

968 70. Zimmerman, H. J., Pincus, M. R. & Henry, J. B. in *Clinical diagnosis and management by*
969 *laboratory methods* (ed. Henry, J. B.) 281–284 (WB Saunders, Philadelphia, USA, 1996).

970 **TABLES**

971 **Table 1 |** HPLC-SCX fractionation settings used during 65- or 120-min gradient.

Pump 1 (trap column)	
Time Interval (min)	**LC gradient (% B)**
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)	0-0
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

972 For the methods 65 and 120 min the indicated times refer to the total analysis time rather than actual
973 gradient time, which is 39 and 58 min respectively. In parentheses LC and MS parameters are provided
974 for the 120 min method.

975 **Table 2 |** LC and MS parameters used during 85-, 115- or 175-min methods in Orbitrap Fusion and Orbitrap
976 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**
Orbitrap Resolution	60 000
Scan Range, m/z	375-1500
RF lens, %	30
AGC Target	4.0e5
Maximum IIT, ms	50
dd settings MS2	
dd mode	TOP10
MIPS	Peptide
Intensity	Intensity Threshold, 2.0e4
Charge State	3-8
Dynamic Exclusion	
Exclude after n times	
Exclusion Duration, s	12s (16s, 24s)
Mass Tolerance	10 ppm low, 10 ppm high
Sort by Charge	Highest charge state
dd-MS2 OT CID	
Isolation Window, m/z	1.6

CID collision energy, %	30
Orbitrap Resolution	30 000
AGC Target	5.0e4
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
Orbitrap Resolution	30 000
First mass, m/z	120
AGC Target	1e5
Maximum IIT, ms	120
AAPT	ON
dd settings MS3	
dd mode	4 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, 5.0e3
Precursor Ion Exclusion, m/z	Low 18, High 5
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	3.0e4
AAPT	ON
Maximum IIT, ms	90

977 The methods 85, 115 and 175 min, the indicated times refer to the total analysis time rather than actual
978 gradient time, which is 68, 98 and 148 min, respectively. In parentheses LC and MS parameters are
979 provided for the 115 and 175 min methods.

980

981 **Table 3 |** Troubleshooting table.

Steps	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	Too concentrated crosslinker stock solution	Dilute crosslinker solution
6	Low MW region is still present on a gel in case of high crosslinker concentrations	Low activity of the crosslinker	This could be due to hydrolysis of the crosslinker. Therefore, prevent exposing the stock solution to multiple freeze and thaw cycles. Make sure that the crosslinker stock solution is prepared immediately before the crosslinking reaction or use aliquoted stock solutions only once If the crosslinking reagent requires activation (in the case of e.g. acidic crosslinker or zero-length crosslinker), verify whether the correct activation conditions were applied
		Presence of interfering components in the buffer	Define all possible substances which may interfere with the crosslinking reaction. Avoid these in the buffer. In case of NHS-crosslinkers e.g. DSSO avoid buffers containing amines and primary alcohols
		Concentration of the crosslinking reagent is too low	Increase concentration of crosslinking reagent
10	Precipitate is formed	Insufficient enzyme activity to shear all nucleic acid's artefacts	Increase concentration of the enzyme.
		Protein concentration is too high	Dilute reaction mixture to a final protein concentration ~ 1 mg/ml
15	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 used protease, thus checking

			protease activity. Follow the instructions for optimal enzymatic digestion[60]
		Too little enzyme used	Incubate reaction for a longer time or increase amount of enzyme
		Over-crosslinking	Ensure that lowest possible crosslinking reagent concentration is used
21B (iii)	No or low UV-signal	HPLC-SCX system is not performing correctly	Check BSA quality control runs (**Supplementary Fig. 1**)
		Peptide concentration is too low	Increase injected peptide amount. If necessary, increase amount of initial cell material and repeat the crosslinking and digestion steps.
		Loss of peptides during desalting	Analyze flow-through solutions. If necessary, repeat the loading procedure with collected flow-through solutions and further washing and elution steps. Ensure that procedure is done according to published protocols[60,61]
		Insufficient elution during desalting	Repeat elution of the desalting cartridges. Ensure that procedure is done according to published protocols[60,61]
22	Weak MS signal intensity	Low amount of injected peptide material	Increase the loading amount to a maximum 10%. If 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 further washing and elution steps. Ensure that the procedure is done according to published protocols[60,61]
		Insufficient elution during desalting	Repeat elution step on the desalting cartridges. Ensure that the procedure is done according to published protocols[60,61]
		Problems with LC-MS/MS setup	Please refer to detailed evaluation and troubleshooting of LC-MS/MS system[60]
27	Fewer crosslinks identifications than expected	Instrument is out of calibration	Broaden mass window tolerances to e.g. 50 ppm on all levels. Calibrate the instrument.
		Suboptimal software search settings	Ensure that all the setting are filled according to **Supplementary Table 1**. Compare MS acquisition strategy with the setting in “XlinkX Detection” node. If

			necessary, search data with correct MS strategy (**Supplementary Tutorial**) Check crosslinker modification in "XlinkX Detection" node is set to the reacted version of the linker. It is also recommended to check that crosslinker signature peaks are set correctly under "Extended Properties" tab in chemical modifications list (**Supplementary Fig. 2**)
		Low efficiency of crosslinking reaction	Too few amino acids, which can be captured by used crosslinking reagent. Perform experiment using crosslinking reagent with alternative chemistry and involving other amino acids
		Reactive residues are out of the crosslinkers maximum linking distance	Perform experiment using crosslinking reagent with alternative chemistry and involving other amino acids
		Problems with LC-MS/MS setup	Please refer to detailed evaluation and troubleshooting of LC-MS/MS system[60]
	Most of the crosslinks are detected within the early SCX fractions	Leaking pump B during HPLC-SCX fractionation	Substitute leaking pump
29	Validated less than 90% of crosslinks	Wrong PDB structure	Choose an appropriate PDB structure according to your experimental conditions
		Over-crosslinking	Choose lower concentration of reagent for crosslinking experiment.
30	Most of the crosslinks are not confirmed	PDB structures of the subunits are not chosen correctly	Check whether submitted structures are chosen accordingly to your experimental conditions
		Wrong sequences numbering in PDB file	Adjust numbering of amino acids accordingly to their sequence in the search database

982

983

984 **Supplementary information submitted:**

985 Supplementary Figure 1

986 Supplementary Figure 2

987 Supplementary Figure Legends

988 Supplementary Table 1

989 Supplementary Tutorial

990 Supplementary Results 1

991 Supplementary Results 2

992 Supplementary Results 3

993 Supplementary Results 4

994 Supplementary Results 5

995 Supplementary Results 6

996 Supplementary Results 7

997 Supplementary Results 8

998 Supplementary Results 9

999 Supplementary Methods 1

1000 Supplementary Methods 2

1001 Supplementary Methods 3

1002 **Information available for the reviewers:**

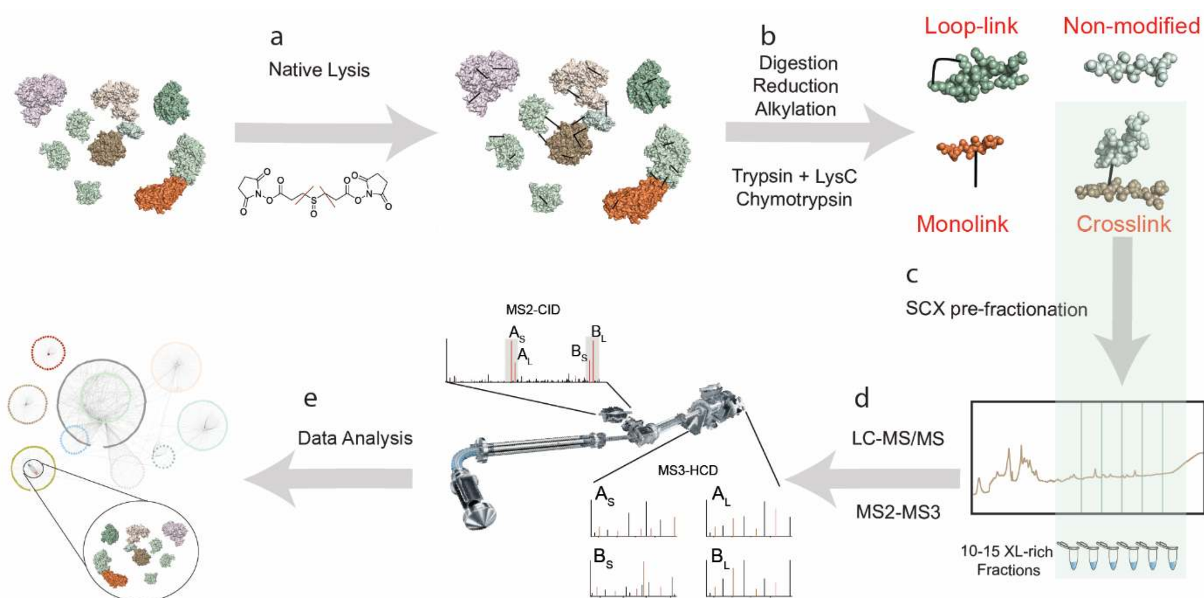
1003 Supplementary Methods 3 is a zip-archive containing the latest version of the XlinkX node compiled for
1004 the currently widely available Proteome Discoverer 2.2.0.388 and which has been used to process the
1005 data. Please find installation instructions in the README.txt file contained in the zip-archive. We expect
1006 the release of Proteome Discoverer 2.3 with the XlinkX nodes early 2018.

1007 PRIDE repository with a project accession PXD008418 and can be assessed through following credentials:

1008 username: reviewer36114@ebi.ac.uk, password: yLqwhaHr

1009

1010 **FIGURES:**

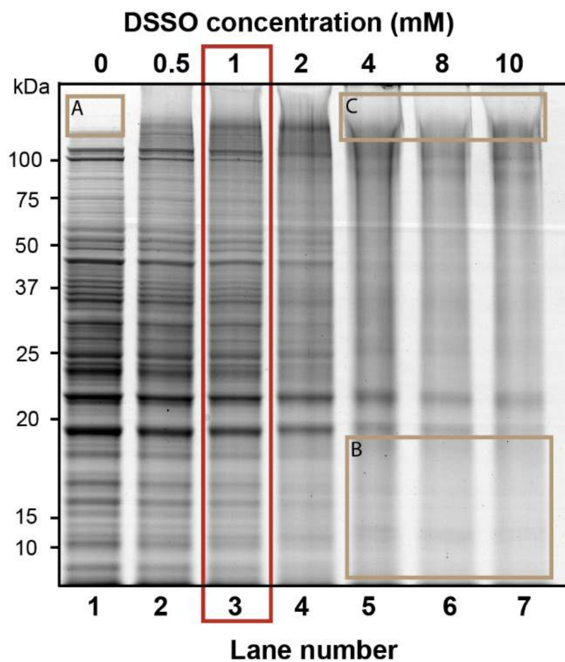


1011

1012 **Figure 1 | Generic workflow for XL-MS experiments. (a)** Cells or tissue are lysed gently, leaving protein
1013 complexes intact. **(b)** After optimized incubation with the cross-linking reagent, and proteolytic digestion
1014 4 peptide products can be categorized. **(c)** Enrichment and pre-fractionation of XL-peptides using
1015 techniques like SCX. **(d)** Advanced data acquisition techniques utilizing multiple steps of fragmentation
1016 techniques (CID, HCD) are used to identify the peptides. **(e)** XlinkX for Proteome Discoverer v. 2.3 is used
1017 to identify the crosslinked peptides. The resulting data can consequently be integrated into structural
1018 modeling software (e.g. HADDOCK, I-TASSER, DisVis).

1019

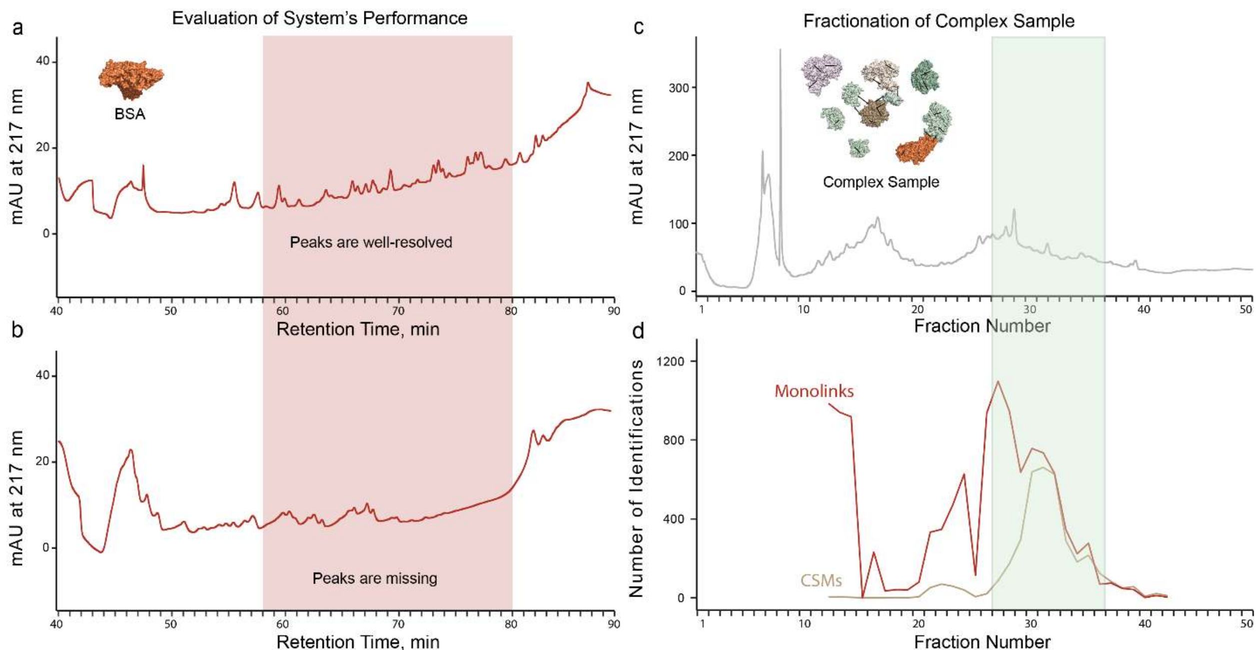
1020



1021 **Figure 2 | Optimization of crosslinking reaction conditions by titration.** Lane 1 corresponds to the
1022 experiment when no crosslinking agent was added; the high MW region is empty (box A). For lanes 5-7
1023 the crosslinking agent was added in high concentrations, leading to a less intense low MW region (box B),
1024 while the upper part of the gel is over-saturated with high MW species (box C). Lane 3 with intermediate
1025 1mM concentration of crosslinker shows the presence of proteins in both regions and is selected as the
1026 optimal crosslinker concentration.

1027

1028

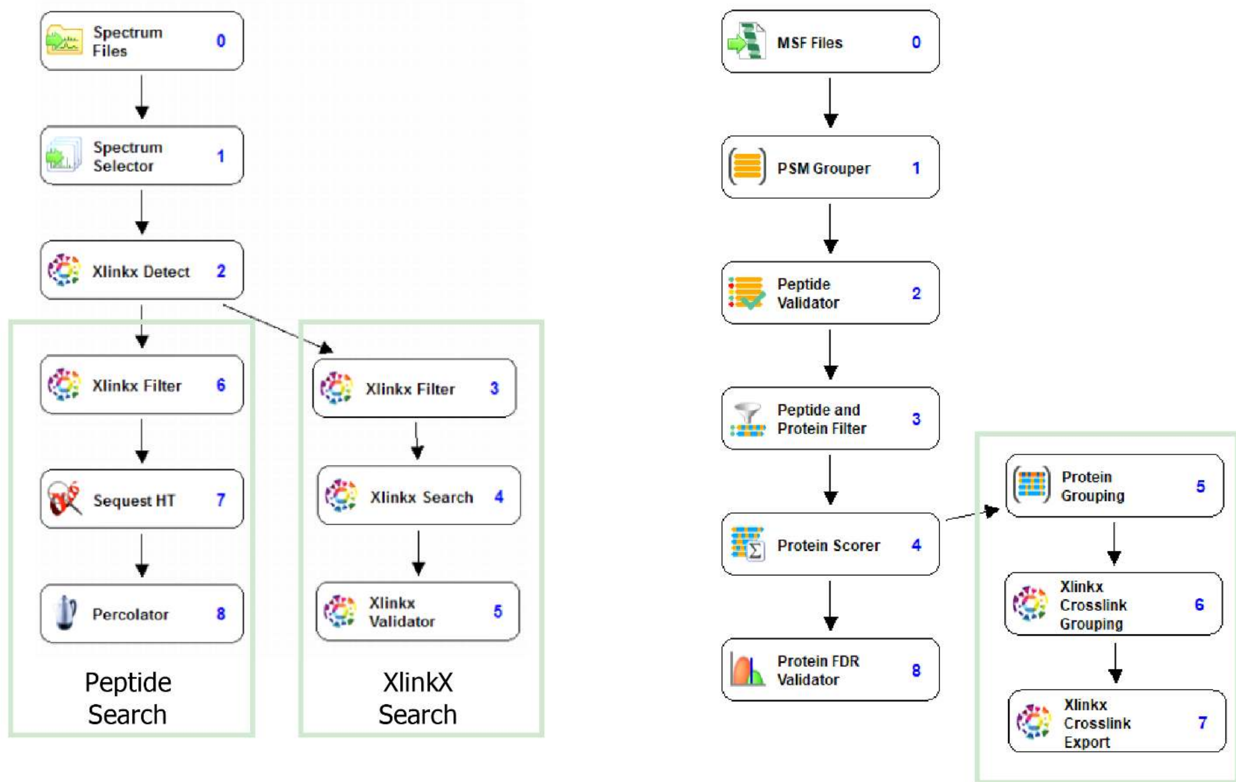


1029

1030 **Figure 3 | SCX fractionation profile for whole PC9 cell lysate samples. (a)** SCX chromatogram on a well
1031 performing system recorded at wavelength 217 nm. **(b)** SCX chromatogram for separation on un-
1032 equilibrated SCX column recorded at wavelength 217 nm. **(c)** Snapshot of SCX chromatogram with 50
1033 collected fractions recorded at wavelength 217 nm. **(d)** Distribution of the number of identified crosslinks
1034 (number of entries from the Crosslink Spectral Match, or CSM, table) and monolinks (number of entries
1035 annotated with water-quenched linker as a modification in the Peptide Spectral Match, or PSM, table)
1036 over the analyzed fractions; the most CSM-rich fractions are highlighted.

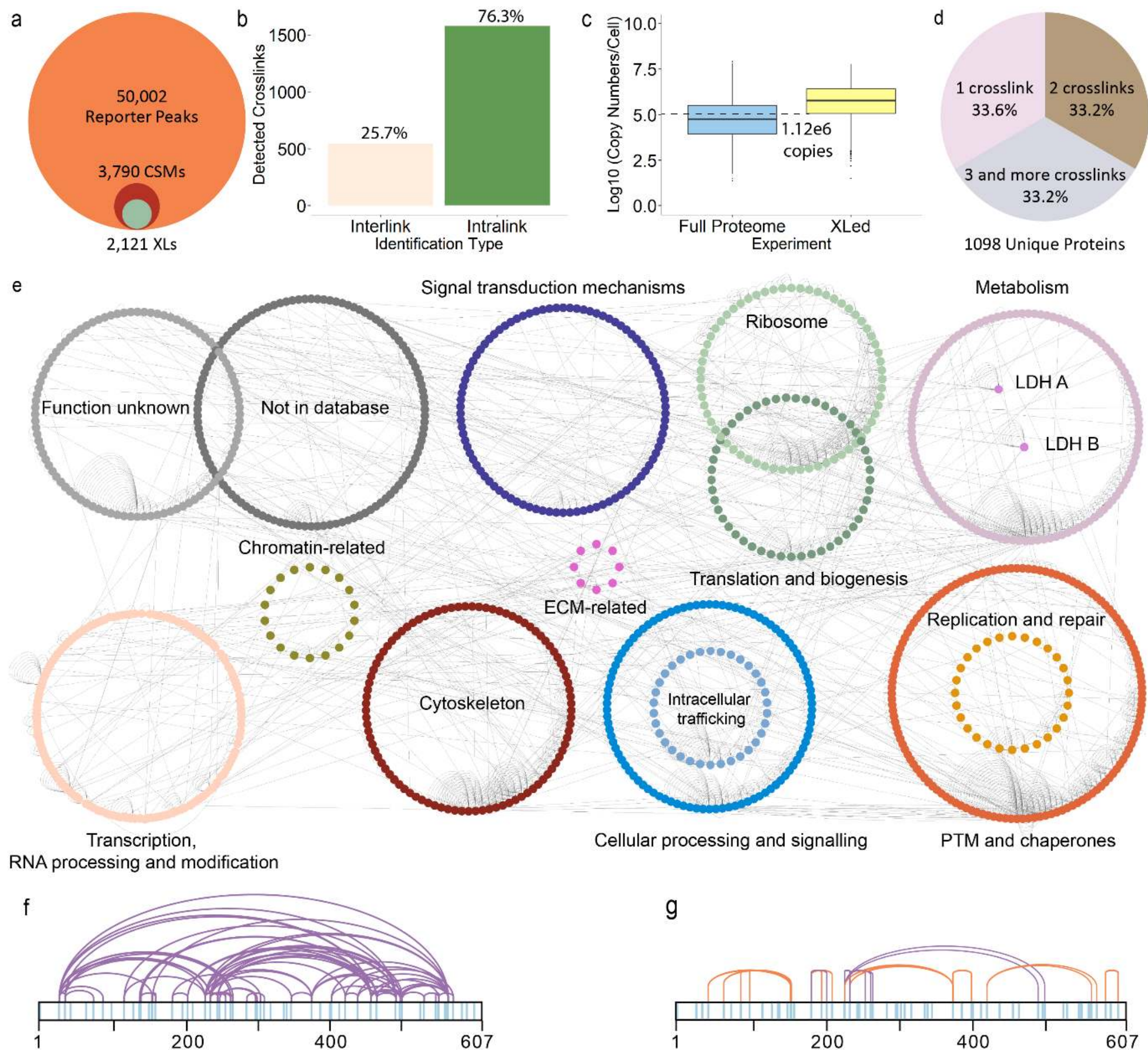
1037

1038



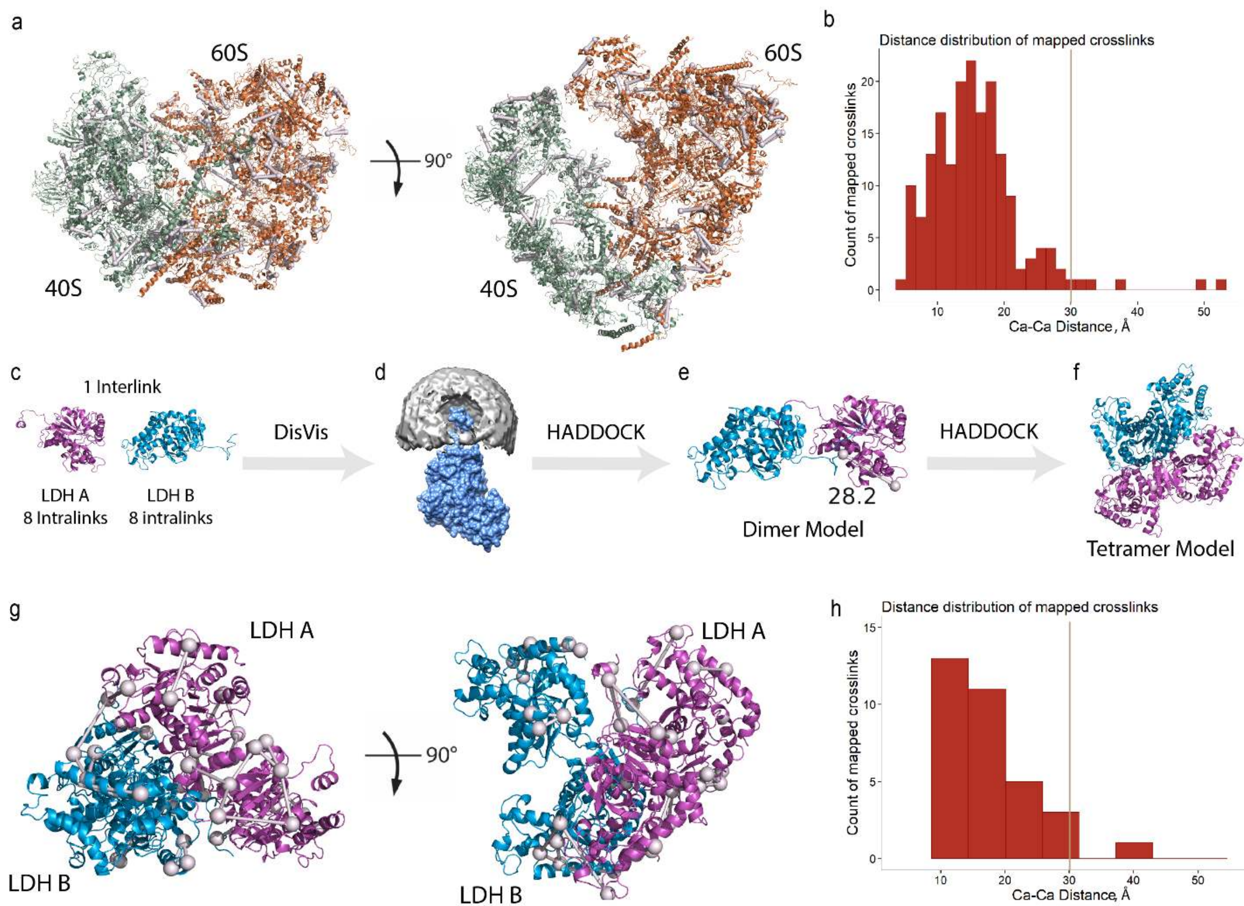
1039

1040 **Figure 4 | Schematic representation of Proteome Discoverer workflows for identification of crosslinked**
1041 **peptides**. **(a)** Processing workflow **(b)** Consensus workflow**.**



1042

1043 **Figure 5 | Interaction network and (a)** Overview of detected Reporter Peaks, CSMs and Crosslinks. **(b)**
1044 Types of detected crosslinks. **(c)** Estimation of sensitivity by correlating detected crosslinked proteins to
1045 protein copy numbers per HeLa cell. **(d)** Distribution of detected proteins by number of detected
1046 crosslinks. **(e)** Interaction network generated with Cytoscape for the whole cell lysate, where proteins
1047 grouped by their function according to EggNOG database. **(f)** Visualization of detected crosslinks for the
1048 low complexity BSA sample by xiNET digested by trypsin and **(g)** chymotrypsin. Crosslinks exclusively
1049 detected with chymotrypsin are represented in orange.



1050

1051 **Figure 6 | Validation of identified crosslinks on the known Ribosome Structure. (a)** Ribosome with
1052 mapped crosslinks. **(b)** Distance distribution of crosslinks mapped on a ribosome. **(c)** DisVis input. **(d)**
1053 Predicted interaction interface generated by DisVis. **(e)** Dimer model obtained using HADDOCK with
1054 mapped confirmed interlink. **(f)** HADDOCK output with a final tetramer model. **(g)** LDH-3 complex
1055 consisting from 2 LDH A and 2 LDH B subunits with mapped crosslinks. **(h)** Distance distribution of
1056 crosslinks mapped on a modelled LDH-3 complex.

1057