University of Amsterdam

UvA-DARE (Digital Academic Repository)

Towards low false discovery rate estimation for protein-protein interactions detected by chemical cross-linking

de Jong, L.; Roseboom, W.; Kramer, G.

DOI
10.1016/j.bbapap.2021.140655

Publication date
2021

Document Version
Final published version

Published in
Biochimica et Biophysica Acta. Proteins and Proteomics

License
CC BY

Link to publication

Citation for published version (APA):
de Jong, L., Roseboom, W., & Kramer, G. (2021). Towards low false discovery rate estimation for protein-protein interactions detected by chemical cross-linking. *Biochimica et Biophysica Acta. Proteins and Proteomics*, *1869*(7), [140655]. https://doi.org/10.1016/j.bbapap.2021.140655

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

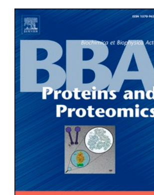
Download date:09 Mar 2023

BBA - Proteins and Proteomics 1869 (2021) 140655

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap



Towards low false discovery rate estimation for protein-protein interactions detected by chemical cross-linking

Luitzen de Jong *, Winfried Roseboom, Gertjan Kramer

Swammerdam Institute for Life Sciences, Mass Spectrometry of Biomolecules, University of Amsterdam, Science Park 904, 1098 HX Amsterdam, the Netherlands

A B S T R A C T

Chemical cross-linking (CX) of proteins in vivo or in cell free extracts followed by mass spectrometric (MS) identification of linked peptide pairs (CXMS) can reveal protein-protein interactions (PPIs) both at a proteome wide scale and the level of cross-linked amino acid residues. However, error estimation at the level of PPI remains challenging in large scale datasets. Here we discuss recent advances in the recognition of spurious inter-protein peptide pairs and in diminishing the FDR for these PPI-signaling cross-links, such as the use of chromatographic retention time prediction, in order to come to a more reliable reporting of PPIs.

1. Introduction

Protein-protein interactions (PPIs) on a proteome wide scale can be detected by chemical cross-linking in living cells or cell free extracts combined with mass spectrometric analysis of peptides obtained after proteolysis (CXMS, CLMS or XLMS). Identification of PPIs is important to understand the dynamic functional organization of living material under different conditions like health and disease and to discover possible new targets for drugs, in particular PPIs in which essential proteins are involved. Especially cross-linking in vivo can trap dynamic assemblies that may dissociate upon cell extraction, implying that transient PPIs may be encountered that have thus far escaped detection by in vitro approaches. In vivo CXMS is an attractive approach since it offers the possibility to get a systems level view on dynamic protein interactions obtained in a single experiment and at the cross-linked amino acid residue pair level. As a bonus, the distance constraint imposed by the length of the cross-link between residue pairs favors structural modelling of the interactions if the 3D structures of the interacting proteins are known.

After the first reviews about CXMS [1,2], an era of development of dedicated cross-linkers, software and analytical approaches followed, culminating in identifications of PPIs in cell free extracts and in living cells [3–5]. Recently, the CXMS community has also come to the point where agreement exists about the aspects for which standards should be developed to be observed in publications in the field [6,7]. One of the aspects concerns a reliable error assessment. Identifying inter-protein cross-links at a low FDR for PPIs in complex samples is a challenging subject, but recent progress in this area enables recognition of spurious candidates in a dataset [8–10]. In addition, filters have been developed to discriminate between true and false positive inter-protein peptide pairs [8–10]. Here we will discuss results presented in these papers in the framework of current problems of reliable PPI identification by CXMS in large scale experiments. It is noteworthy that the CXMS community stated that it is of the utmost importance that the field arrives at a consensus for procedures that return a reliable error assessment and it is also important that the limits of these procedures be mapped out [6]. Therefore we think that the topics discussed here underscore the timeliness of this minireview.

2. Recent publications using CXMS for PPI discovery in complex systems reveal discrepancies in error assessment

Table 1 lists recent examples of large scale CXMS studies. Until now the type of reagent most used with complex systems are cross-linkers containing two N-hydroxysuccinimidyl ester groups connected by a spacer. These homobifunctional reagents attack predominantly the ε-amino group of lysine residues and the *N*-terminal amino group of proteins. To a lesser extent, also Ser, Thr an Tyr can react with N-hydroxysuccinimidyl esters [11]. Several variations of gas-phase cleavable and non-cleavable N-hydroxysuccicimidyl ester-based cross-linkers have been described that enable identification of cross-linked peptide pairs from complex protein samples. Usually trypsin or a combination of trypsin and LysC are employed to generate peptides from cross-linked proteins. Sequential digestion with trypsin and proteases cutting at other amino acids than lysine and arginine is also used [12].

A much chosen subject of study concerns PPIs in isolated mitochondria represented in Table 1 by five recent examples [13–17]. There is partial overlap between the PPIs identified in these studies, but the

* Corresponding author.
E-mail address: l.dejong@uva.nl (L. de Jong).

https://doi.org/10.1016/j.bbapap.2021.140655
Received 15 December 2020; Received in revised form 28 March 2021; Accepted 29 March 2021
Available online 31 March 2021
1570-9639/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

fraction of identified inter-protein cross-linked peptides varied from about 3% to about 60%. This large variation may, at least partly, be explained by differences in FDR estimations, in particular with respect to separate criteria used for assignment of intra-protein and inter-protein peptides, respectively [15,18,19]. Furthermore, FDR can be expressed at different levels, varying from the cross-linked spectrum (CSM) level down to the PPI level (section 4) [10,15,20]. The different levels used to express the FDR in similar types of CXMS studies like the mitochondrial interactome examples [13–17] has been confusing. It underscores the importance of the summons of the CXMS community for the development of standards for reliable FDR estimation for inter-protein peptide pairs [6].

3. Principles of LC-MS/MS identification of cross-linked peptide pairs using non-cleavable and cleavable reagents; implications for FDR estimation

Identification of a cross-linked peptide pair by LC-MS/MS and database searching is more demanding than identification of regular peptides. For identification of a regular peptide in complex mixtures by LC-MS/MS, the mass and charge of the peptide, determined in the MS1 stage, and the masses and charges of fragments obtained by peptide bond cleavages induced by collision with gas molecules (collision-induced dissociation, CID, or collision-activated dissociation, CAD) or other fragmentation methods like electron transfer dissociation (ETD) [21] and higher-energy C-trap dissociation (HCD) [22] determined in the MS2 stage are combined and used for searching a sequence database for a match. For a cross-linked α-β peptide pair identification by this approach is challenging for four reasons. Firstly, MS2 spectra are relatively complex by the presence of signals from fragments of two peptides. This complicated circumstantiality requires unambiguousness of fragment ions to be taken into account in the assignment of the identity of α and β peptides [23], otherwise it will increase the FDR [9]. Secondly, cross-linked peptides are present in sub-stoichiometric amounts compared to regular peptides. Since selection for MS/MS is dependent among others on signal intensity, enrichment by extensive fractionation or isolation is required. Thirdly, because of the relatively large size of cross-linked peptides, ionization results in relatively high charge states. This may hamper correct identification of the monoisotopic mass from weak signals of highly charged cross-linked precursor ions. Sequential digestion to shorten peptide length can increase the fraction of cross-linked peptides susceptible to identification by mass spectrometry [12].

Fourthly, only knowledge of the mass of the intact cross-link is obtained in MS1, the masses of α and β in this stage remaining unknown. To circumvent this problem, in case a non-cleavable cross-linker like DSS [24] or BS3 [25] is used, both α and β are considered regular peptides containing a modification of unknown mass at the cross-linked residue to which the other peptide has been connected. With this hypothesis, the sequence database is first searched with MS2 data in a wide peptide mass window followed by keeping a few high scoring candidates for α or β. The unmodified mass of a high scoring candidate α or β peptide subtracted from the known mass of the intact peptide pair minus the mass of the cross-link remnant reveals the candidate mass of the other composite peptide. With this candidate peptide mass information the MS2 data are used to find possible candidates for the second peptide. The kept α and β candidates are recombined and the highest scoring α-β pair is assigned if criteria for sufficient fragment ions and matched peak intensity, expressed in a score, are fulfilled [12,26].

Determination of the masses of the composite peptides in a linked peptide pair can be determined more directly with use of gas-phase cleavable cross-linkers [19,27–31]. Cleavage can occur at two sites and precursor ions that have underwent a cleavage event at one of the two cleavage sites are represented in the MS2 spectrum by two pairs of peaks, so-called signature ions, each pair having a characteristic mass difference dependent on the type of cross-linker used. From the two pairs of peaks the masses of the composite peptides can be deduced directly, while the presence of only one pair enables calculation of the mass of the other peptide by mass subtraction from the precursor mass. Fig. 1 shows an example of the products formed by CID of a linked peptide pair using the cleavable cross-linker BAMG [19]. In this example the mass difference of a pair of signature ions is 125.084 Da. For DSBU [30] and DSSO [29] this mass difference is 25.979 Da and 31.972 Da respectively. If cleavages of peptide bonds occur along with cleavage of the cross-link, the MS2 spectrum can be directly used for identification of the peptide pair. Alternatively, or additionally, an MS3 stage is included, consisting of selection of the signature ions for fragmentation. Various combinations of CID, HCD, ETD and EThcD (ETD combined with HCD) in MS2 and MS3 stages have been studied to optimize cross-

Table 1
Overview of recent CXMS publication using cleavable and non-cleavable cross-linkers to identify protein-protein interactions in complex systems.

Cross-linker	Cross-linker type	Isolation method for cross-linked peptides	Organism	Material	Reference
BNP-NHP (PIR) [27]	cleavable	Biotin-avidin affinity purification	*Escherichia coli*	intact cells	Mohr e.a. [56]
			Mus musculus	isolated mitochondria	Schweppe e.a. [13]
CBDPS [28]	cleavable	Biotin-avidin affinity purification	*Saccharomyces cerevisiae*	isolated mitochondria	Makepeace e.a. [17]
DSSO [29]	cleavable	–	*Escherichia coli*	cell free extract	Yugundhar e.a. [8,34][a]; Lenz e.a. [10][a]; Liu e. a. [32]
		–	*Homo sapiens*	cell free extract	Yugundhar e.a. [8,34][a]; Liu e.a. [32]
		–	*Mus musculus*	isolated mitochondria	Liu e.a. [14]
DSBU [30]	cleavable	–	*Drosophila melanogaster*	cell free extract	Götze e.a. [36]
BAMG [31]	cleavable	Diagonal SCX chromatography	*Bacillus subtilis*	intact cells	de Jong e.a. [9][a]
BS3 [25]	non-cleavable	–	*Escherichia coli*	cell free extract	Lenz e.a. [10][a]
		–	*Saccharomyces cerevisiae*	isolated mitochondria	Linden e.a. [16]
DSS [24]	non-cleavable	–	*Saccharomyces cerevisiae*	isolated mitochondria	Linden e.a. [16]
			Homo sapiens	isolated mitochondria	Ryl e.a. [15]

[a] These papers both describe how potential false positive inter-protein cross-links can be recognized in a dataset subjected to a target-decoy approach and show development of effective filters to lower the FDR for inter-protein peptide pairs. BNP-NHP, biotin aspartate proline-N-hydroxyphtalamide also known as protein interaction reporter (PIR); CBDPS, cyanurbiotinediproprionylsuccinimide; DSSO, disuccinimidyl sulfoxide; DSBU, discuccimidyl dibutyric urea; BAMG, bis (succinimidyl)-3-azidomethyl-glutarate; BS3, bis(sulfosuccinimidyl) suberate; DSS, disucininimidyl suberate.

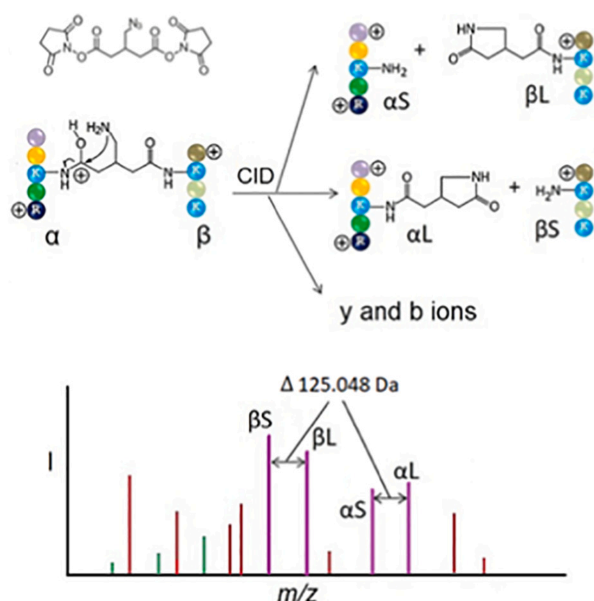


Fig. 1. Gas phase cleavage reactions of BAMG-cross-linked peptides in which the azido group has been reduced to an amine group. Upper left corner, structure of BAMG. Middle part, collision induced dissociation (CID) of a cross-linked peptide pair leads to cleavages of the two cross-link amide bonds along with formation of y and b ions by cleavages of peptide bonds. Cleavage of the cross-link results in formation of an unmodified peptide or short version of the cleavage product (αS or βS), the other peptide being modified by the remnant of the cross-linker in the form of a γ-lactam adding 125.048 Da to the mass of the peptide. This is the longer version of the cleavage product (αL or βL). Amino acids are depicted as colored candies. The indicated gas phase charge states of the cross-linked peptide and the cleavage products are arbitrarily. The lower part is a cartoon of a fragment mass spectrum with two pairs of cleavage products with the characteristic 125.048 Da mass difference (purple sticks) and some peaks of b (green) and y (red) ions. Copied from [9].

link identification and FDR [32–35].

Both cleavable and non-cleavable cross-linkers can be used for analysis of complex systems, although the use of the cleavable cross-linker DSSO resulted in about 20% more unique inter-protein (heteromeric) cross-link spectrum matches than the non-cleavable BS3 in a large scale experiment with a cell free extract from *E. coli* [10].

4. Challenges of FDR estimation

To estimate the FDR for cross-linked peptide pairs a target-decoy database approach is used. In most cases listed in Table 1 the decoy database consists of the reversed or shuffled protein sequences of the target database, but reversed sequences of tryptic peptides with the original C-terminal amino acids [17,36] are also applied. FDR is often expressed at the level of cross-link spectrum matches (CSM) according to FDR = (TD – DD)/TT (Eq. (1)) [20], in which TD is the number of CSMs in which one peptide is from the target (T) database and the other is from the decoy database (D), DD is the number of CSMs in which both peptides are from the decoy database and TT is the number of CSMs with both peptides from the target (forward) database. With a species-specific database of several thousand proteins the search space for inter-protein cross-links is at least three to four orders of magnitude larger than the search space for intra-protein peptide pairs, dependent among others on database and protein size [19], irrespective whether both peptides of an intra-protein peptide pair are from the same protein molecule or from two identical protein molecules [19,37]. This implies that practically all false positives are confined to inter-protein cross-links, while hits containing one forward and one reversed sequence or two reversed sequences from the same protein are completely or nearly completely absent, even if a large number of decoy sequences is generated by imposing a lenient score requirement on a data set from complex protein samples [18,19,38]. So, inter-protein peptide pairs have a much higher FDR than intra-protein cross-links at the FDR at CSM level, if an equal score stringency is used for both type of cross-links. It is recommended by the CXMS community that in results files generated by cross-link search engines both target and decoy sequences of cross-linked peptide pairs should be reported [6], a practice that has only been applied yet in a few instances, e.g., [19,23]. Such a report reveals how the FDR is experimentally distributed between intra-and inter-protein peptide pairs under the chosen stringency for assignment. For instance, from a result file of a CXMS analysis by interrogation of the entire human sequence database at 10 ppm mass accuracy of both MS1 and MS2 signals of a complex sample cross-linked by BAMG [19] we found 459 intra-protein CSMs and no intra-protein decoy CSMs, while 50 inter-protein CSMs were detected (homodimers excluded), along with 36 inter-protein decoy CSMs (30 TD and 6 DD CSMs), corresponding to an overall FDR of 4.7% and a FDR for inter-protein CSMs of 48%. This result was obtained by applying the same criteria for assignment of intra- and inter-protein peptide pairs. Under these conditions lowering the FDR is best obtained by increasing the stringency for assignment of inter-protein peptide pairs.

In cases where the identities of decoy hits are not reported by the chosen search engine one can use the so-called entrapment database approach to validate to which extent the reported inter-protein peptide pairs and intra-protein peptide have been assigned in agreement with the set FDR. In all four approaches discussed in sections 5–8 aimed to recognize spurious inter-protein peptide pairs in complex samples and to lower the FDR [8–10,36], an entrapment database has been used in which the target database is concatenated with a database of protein sequences not existing (N) in the target organism. In this case NN sequences from the same protein reflect false positive intra-protein cross-linked peptide pairs, whereas NN sequences from different proteins and NT and TN sequences reflect false positive inter-protein peptide pairs. The FDR as calculated according to Eq. (1) is proportional to the size of the decoy database. Therefore, in case T and N differ in size, the FDR value based on the number of reported decoy sequences. i.e., TN, NT and NN, should be corrected accordingly. Importantly, the MS/MS features like matched intensities and the number of unambiguously assigned y and b ions in false identifications revealed by the entrapment approach enable formulation of MS/MS criteria to distinguish false from true positives, in cases where decoy hits from the reversed database are not reported. However, if such decoy hits are reported an entrapment database can also be used as an independent way to determine the FDR.

For reliable identification of PPIs it must also be taken into account that the FDR for inter-protein peptide pairs increases from the CSM level to other specific levels [20]. This holds for the level of non-redundant cross-linked peptides, since target hits may result from multiple MS/MS selections, while decoy hits are often assigned from a single MS/MS selection. Filtering the dataset for unique amino acid residue pairs, omitting variations caused by post-translational modifications and missed cleavage sites, further increases the FDR and the same holds for hits revealing specific PPIs, often the ultimate goal in CXMS work with complex samples. Table 2 shows an example of this error propagation from the CSM to the PPI level.

Since the search space in an entire species-specific sequence database for intra-protein cross-links is much smaller than that of inter-protein species, distinction has been made in the assignment of these two categories by several search engines. In work with XlinkX [32] this was done by applying a low 1% FDR at CSM level for inter-protein cross-links. The search engines pLink 2 [24] and MeroX 2.0 [30] separately control the FDR of intra- and inter-protein peptide pairs. Unfortunately these measures alone are insufficient to solve the problem as to how reliably detect PPIs from complex mixtures preferably to a FDR below 5% at PPI level. This has become clear by recent methods to recognize the presence of spurious inter-protein peptide pairs in large data sets

Table 2
Example of error propagation from CSM level to PPI level.

Level of FDR estimation	Peptide pairs	Potentially false positive peptide pairs	(%)
Total CSMs	1856	41	2.2
Intra-protein CSMs	1581	10	0.6
Inter-protein CSMs	275	31	11.3
Inter-protein peptide pairs	93	29	31.2
Inter-protein residue pairs	85	29	34.1
PPIs	76	29	38.1

This dataset was obtained from an in vivo CXMS experiment with *B. subtilis* using pLink 2 as a search engine set at 5% FDR at CSM level [9]. The entire *B. subtilis* sequence database was interrogated with separate control of FDR for intra-protein and inter-protein peptide pairs. Potentially false positive target nominations were recognized based on database-size dependent analysis (about 26% of inter-protein amino acid residue pairs and about 2% of intra-protein residue pairs, section 6) and anomalous retention times in SCX chromatography (about 19% of inter-protein residue pairs and no intra-protein residue pairs, section 8). The different percentages of potential false positives between intra-protein and inter-protein CSMs suggests that the criteria used by pLink 2 for separate control of FDR were not yet effective enough. By application of a composite filter 40 PPIs (48 PPIs combined with results in [39]) were identified and the FDR at PPI level was reduced to about 2% (section 9). All except one potential false positive PPI failed to pass the filter. In contrast, all 10 potentially false intra-protein cross-links passed the filter, in line with their correctly predicted SCX chromatographic retention times.

described in sections 5 to 8.

5. Structural assay-based error recognition in the assignment of inter-protein peptide pairs

For a long time up to now a generally used criterion for the reliability of identified inter-protein cross-links was to which extent the distance constraints imposed on linked residue pairs fit with the know 3D structure of a corresponding protein complex. For instance, for the cross-linker DSSO this is considered to be the case if the distance between the Cα atoms of linked residues is less than 30 Å [8]. In general, the large majority (often around 80%–90%) of the cross-linked peptide pairs from a protein complex in data sets of complex samples fits to known structures, a minor fraction of linked residue pairs that might violate the maximally allowed spatial distance being explained, presumably correct, by structural flexibility. However, recently it has been shown that structural agreement of the subset of cross-links from known protein complexes is not at all a useful criterion for the reliability of the entire data set of identified inter-protein peptide pairs [8]. This was demonstrated with CXMS datasets obtained from protein extracts from cultured human and *Escherichia coli* cells. The data sets were analyzed with the search engine XlinkX at different FDR settings at CSM level. Since the number of false positives increases from the lowest to the highest FDR it would be expected that the fraction of cross-links fitting with known structures would decrease from high to low FDR stringency if structural agreement would be a valid measure of reliability. However the fraction of cross-links (about 80%) that fitted within 30 Å to representative 3D structures was independent on FDR, thereby disqualifying this structural validation as a measure of trustworthiness of the entire collection of identified inter-protein peptide pairs. A better structural assay-based metric for reliability of PPI detection would be the fraction of inter-protein peptide pairs with at least one peptide belonging to an assembly that fits to the known structure of the protein complex in question, since this fraction decreases with an increase in FDR [8]. As an independent way to determine the underlying error of the cross-link search engine an entrapment database approach was used in which the target *E. coli* database was combined with the *Saccharomyces cerevisiae* data base as a source of decoy sequences to estimate the fraction of misidentified inter-protein peptide pairs (FMI), an approach also used by others [9,10,36] (sections 6–8). For instance the dataset filtered at 1% FDR revealed a FMI of no less than 15%. As a solution to the high error rate the authors developed a new search engine named MaxLinker, resulting in fewer errors at higher sensitivity and specificity [34].

6. Differential behavior of inter-protein and intra-protein cross-links as a function of the size of the database to recognize false positive inter-protein peptide pairs

An LC-MS/MS dataset of a digest of a high molecular weight SEC protein fraction from exponentially growing *Bacillus subtilis* cells treated with the cleavable cross-linker BAMG [39] was analyzed with the search engine pLink 2 [26] using the entire *B. subtilis* sequence database [9]. In contrast to the previous pLink version [40], pLink 2 enables analysis of cleavable cross-linkers like DSSO and DSBU and is also adapted to the use of BAMG [9]. Another new feature of pLink 2 is the separate control of FDR for inter-protein and intra-protein cross-links [26].

Despite this separate control, the presence of a relatively large number of potential false positive inter-protein cross-linked peptide pairs, but not of false positive intra-protein peptide pairs, was noticed in a pLink 2 search at 5% FDR at CSM level (Table 2). It appeared that about 26% of the 85 non-redundant inter-protein cross-linked residue pairs contained one or both peptides from proteins not belonging to a set of 673 proteins that were independently identified in the cross-linked sample. On the contrary only 10 of the 449 intra-protein cross-links (2.2%) contained sequences from proteins not belonging to the actually identified 673 proteins. This different behavior of inter- and intra-protein cross-links suggests the presence of false positives among the former cross-link type. Since pLink2 does not report decoy sequences from the reversed database, an independent determination of the error rate with a search at 5% FDR with the target database combined with the *S. cerevisiae* database revealed several decoy hits with one or both composite peptides from the yeast database. None of the decoy hits contained peptides from the same yeast protein, indicating a relatively high FDR for the inter-protein cross-linked peptides and an extremely low FDR for the intra-protein cross-links. Based on the MS/MS spectra of the decoy hits and predictable SCX chromatographic retention times (section 8) a composite filter (section 9) was defined that removed all but one of inter-protein peptide pairs of which at least one composite peptide did not belong to the 673 independently identified proteins. However, this cross-link is probably not a false positive since the MS/MS spectrum shows long y- ion series for both composite peptides along with a high matched intensity [9]. Moreover the cross-link is biologically consistent as it reveals an interaction between the SPbeta prophage-derived uncharacterized protein YopJ and the α subunit of RNA polymerase, thus representing a possible example of xenogeneic regulation of the transcription machinery [41] (Table 3). Only one decoy hit survived the composite filter. Surprisingly, decreasing the FDR to only 0.1% still results in a large number of decoy hits with the entrapment database approach, and additional filtering was required to reliably detect PPIs even under these conditions [9].

A similar database size-dependent approach revealed the presence of potentially false positive inter-protein peptide pairs using the cleavable cross-linker DSBU to detect PPIs in a *Drosophila melanogaster* cell free extract [36] using the search engine MeroX 2.0. Approximately the same numbers of intra-protein cross-links were found in the search at 1% FDR against the entire *D. melanogaster* database (21,973 proteins) and in the search with a database of independently identified protein in the extract (9535 proteins). An equal number of intra-protein peptide pairs detected in these two searches can be explained by assuming a combination of a slightly decreased number obtained with the entire database at the same 1% FDR setting due to the large search space, and an equal slight increase of intra-protein cross-links from proteins that had escaped independent identification. However, by searching the entire database the fraction of unique inter-protein cross-links increased by about 12% and

Table 3
Novel PPIs detected in *B. subtilis* by in vivo cross-linking.

Interacting proteins	Inter-protein peptide pair; numbering of linked Ks between brackets	Physiological context of the interaction
RplQ-PepA	KEWEDVVGLVGK(1)-SVVEKMITLGK(5)	Regulation of post-transcriptional modification of nascent polypeptides [9]
TufA-AbrB	KLLDYAEAGDNIGALLR(1)-KVDELGR(1)	Protein phosphorylation-dependent regulation of transcription and translation [9]
YopJ-RpoA	FAKDIAEEVYYSLK(3)-KSLEEVK (1)	Possible example of xenogeneic regulation of RNA polymerase [9,41]
Fur-YlaN	VVDKINFGDGVSR(4)-VDAEKILK(5)	Regulation of iron metabolism and iron sulfur biogenesis [9,52–54]
NusA-GudB	VTPKGVTELTAEER(4)-IAAQTAKQVVTQR(7)	Regulation of transcription coupled repair and gene decryption [39]
RplL-MtnK	EAKELVDNTPKPLK(3)-LALETGTAFIEKR(12)	Unknown [39]

RplQ, 50S ribosomal protein L17; PepA, probable cytosol aminopeptidase; TufA, elongation factor Tu; AbrB, transition state regulatory protein AbrB; YopJ, SPbeta prophage-derived uncharacterized protein YopJ; RpoA, RNA polymerase subunit alpha; Fur, ferric uptake regulator; YlaN, uncharacterized protein YlaN (essential protein); NusA, transcription termination/antitermination protein NusA; GudB, cryptic catabolic NAD-specific glutamate dehydrogenase GudB; RplL, 50 S ribosomal protein L7/L12; MtnK, methylthioribose kinase.

the number of unique PPIs even by more than 20%. These results suggest a relatively high FDR for inter-peptide proteins and an even higher FDR for PPIs. This was corroborated by FDR estimation using the *E. coli* sequence database in an entrapment database approach. Several decoy hits were put forward without any intra-protein decoy hit. For inter-protein cross-links detected with the small database a FDR of 10.3% was calculated. For a low FDR for PPIs, it might be worthwhile to further filter this otherwise valuable dataset, for instance by defining additional MS/MS criteria based on inspection of the decoy hit MS/MS spectra.

7. Pre-fractionation by SEC of a cell free extract followed by cross-linking in SEC fractions and recombining of fractions to recognize false positive PPIs

In a recent preprint still another approach was used to recognize false positive PPIs in a cell free *E. coli* extract detected by CXMS [10]. To this end the extract was first fractionated by SEC. Proteins in the fractions were identified and quantified followed by cross-linking in the separate fractions. The cross-linked fractions were then recombined, digested and extensively fractionated before LC-MS/MS. PPIs detected between proteins in different SEC fractions cannot have been cross-linked and were used for independent determination of the FDR for PPIs. Also an entrapment database approach using the *S. cerevisiae* sequence database was used for FDR estimations. This workflow was implemented in the open source FDR estimation software tool, xiFDR, v2.0, enabling identification of nearly 600 PPIs at 1% FDR at PPI level [10]

8. Anomalous retention times in SCX chromatography to recognize false positive inter-protein peptide pairs

The large set of intra-protein cross-links that are usually identified at (close to) 0% FDR in complex mixtures by searching an entire species-specific database can be considered as true positives enabling prediction of chromatographic retention times. SCX chromatographic retention time as a possible criterion to identify spurious candidates in a dataset has been used before for regular peptides [42,43]. Recently, this approach has been applied using a dataset obtained from in vivo CXMS of *Bacillus subtilis* [9] based on mass and charge properties of cross-linked peptides, as suggested earlier [23]. See also the legends to Table 2. An even more discriminating effect of SCX chromatography can be expected if the contribution of all 20 amino acid residues on retention time is also taken into account [44]. Likewise RPLC and HILIC may be suitable for retention time prediction provided that cross-linked peptides are amenable for this approach like regular peptides [45,46].

9. A composite filter for a low FDR for PPIs

Recent methods to recognize spurious inter-protein cross-linked peptides in large datasets (sections 5–8) combined with the use of an entrapment database to generate decoy sequences enable the formulation of MS/MS criteria to discriminate between true and false positives. Likewise chromatographic behavior of the true positive intra-protein cross-links can be used to predict retention times to be applied as a filter for false positives. For an in vivo CXMS experiment with *B. subtilis* a composite filter was developed of which SCX chromatographic retention time as one of the assignment criteria is an important component. Mass spectrometric components of the filter include matched signal intensity and the number of unambiguously assigned y and b ions required for the least scoring composite peptide. Since on average much more y ions than b ions are present in MS/MS spectra [19,23,39], distinction is made between these fragment ions in the relative contribution to the composite filter. These chromatographic and MS/MS criteria can be incorporated in existing and novel software tools and can be applied to previously analyzed datasets. Another part of the filtering concerns a limited number of candidates with one common peptide and with the same precursor mass or a different precursor mass caused by a post-translational modification or point mutation. An example is shown in Fig. 2. Also these filtering principles can be incorporated in existing and new software tools. Application of the composite filter to the combined, largely overlapping, data [9,39] resulted in 48 PPIs (homodimers not taken into account) and one decoy hit, i.e., 2% FDR at PPI level, at 77% sensitivity determined by applying the composite filter to the large set of true positive unique intra-protein peptide pairs.

10. Importance of gas-phase cleavage and isolation of target peptides

A number of cleavable [19,27,28,47,48] and non-cleavable cross-linkers [49,50] has been designed to enable purification of target peptides. Purification may increase the fraction of target peptides selected for fragmentation in a data-dependent approach (section 3), thereby expecting to lead to more assignments, at least at sufficient high yield obtained with the isolation procedure. It may lower the FDR by decrease of the fraction of regular peptides of which some fragment spectra by accident may resemble that of a target peptide [35]. In a critical report [51] it was noticed that in many instances both the effect of cleavability and the effect of addition of chemical groups aimed at isolation of the cross-linked peptides on the number of assigned unique residue pairs has not yet been underpinned by data. However, with respect to the cleavable cross-linker CBDPS it was recorded that the number of assignments after the purification step increased 2.7 fold in a study of mitochondrial PPIs [17]. For the cleavable cross-linker BAMG, requiring TCEP-induced reduction of the azido group after cross-linking to enable isolation of cross-linked peptides by 2D SCX chromatography and to render the cross-link cleavable, it can be noticed that all known 13 cross-linked peptides from a small protein were recovered in the expected shifted SCX fractions [23], while in complex samples the shifted reduced cross-links belong to the peptides with the highest signal intensities [23]. It was also shown that 88% out of 400 reduced crosslinks in a complex sample, independently identified as being non-cleavable in the gas phase, contained at least two mass signals of expected cleavage products, enough to deduce the masses of the two composite peptides [19]. In conclusion both isolation and, possibly to a lesser extent, gas-phase cleavage of cross-linked peptides [10] in general favor both the

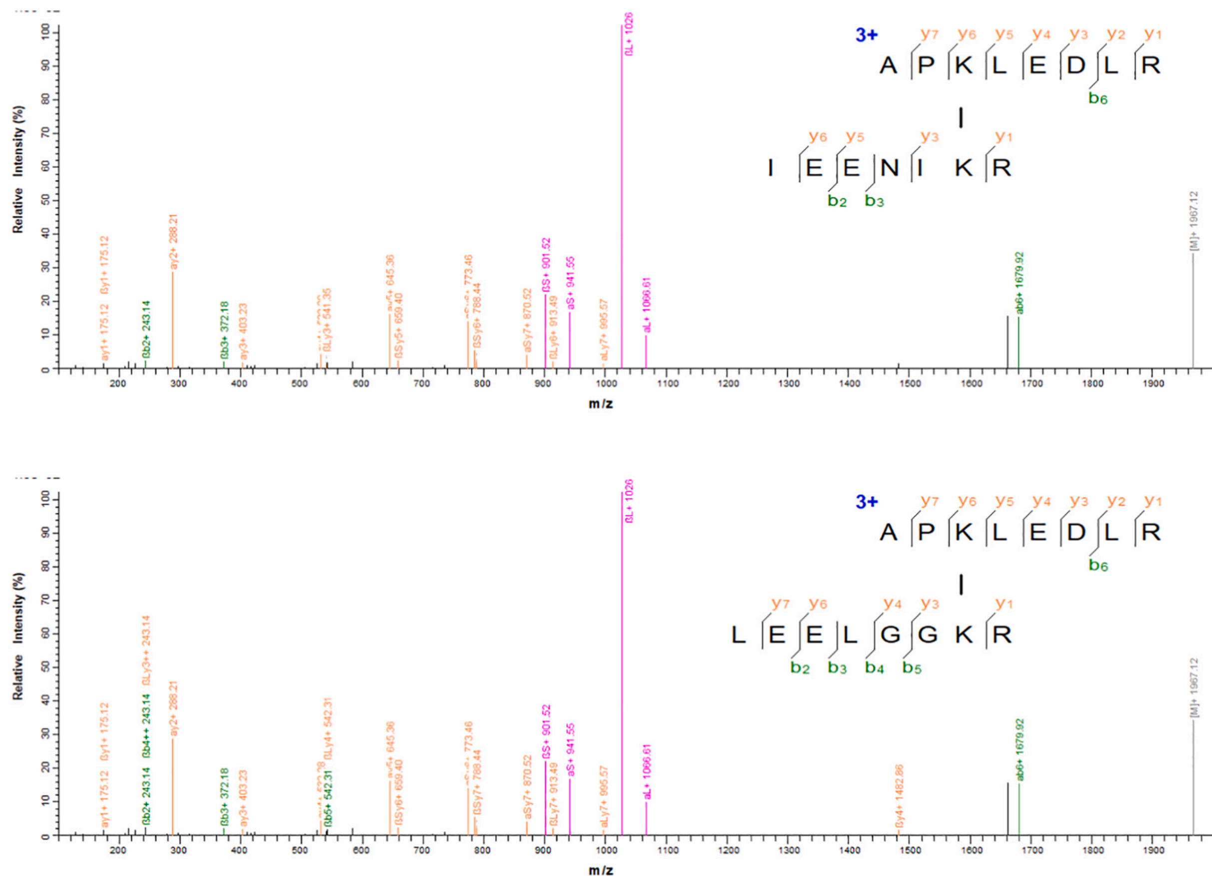


Fig. 2. One nomination but two possible identities for the same precursor ion. Mass spectra are depicted with the pLink 2-associated software tool pLabel. A precursor ion with MH^+ 1967.12 was selected 3 times for MS/MS within 17 s during LC-MS/MS. Two times only the BAMG-cross-linked intra-protein peptide pair APKLEDR-LEELGGKR from the protein CYSJ_BACSU was identified by pLink 2, and one time only the inter-protein peptide pair APKLEDLR-IEENIKR (upper panel) from CYSJ_BACSU and YPBG_BACSU. With three unambiguously assigned y ions for the least scoring peptide and a matched intensity of 87.20%, the MS/MS spectrum satisfies the criteria formulated in the composite filter for assignment [9]. However the lower panel shows that the same MS/MS spectrum also fits with the intra-protein peptide pair, likewise with three unambiguously assigned y ions for the least scoring peptide and in this case a matched intensity of 87.79%. Because of the uncertainty of the true identity of this precursor none were assigned. It is worthwhile to carefully screen MS/MS spectra with one peptide in common, to see whether alternatives exist, like in this case, or for instance by assuming a post-translational modification not taken into account during the search with pLink 2. This check including a decision tree for assignment can be incorporated in a software tool for filtering datasets for spurious inter-protein peptide pairs.

number of identified PPIs and a low FDR.

11. Many transient PPIs have yet to be discovered; the role of in vivo cross-linking

There is little doubt that many PPIs, in particular those of a transient nature, have yet to be discovered. An advantage of in vivo cross-linking is the possibility to trap dynamic interactions that can be easily lost upon cell extraction. Indeed, a relatively large fraction of the PPIs detected in an in vivo study [9,39] of *B. subtilis* using BAMG as a cross-linker relate to dynamic interactions. By adding BAMG directly to an exponentially growing culture, harvesting and washing the cells before cross-linking are avoided, so that dynamic interactions occurring only under defined growth conditions can be trapped [39]. In this study 61 unique inter-protein cross-links were detected that revealed 48 PPIs (section 9). Thirty seven PPIs have been reported before, 31 of which are stable and 6 are dynamic interactions. Of the remaining 11 PPIs, 5 concerned proteins known to dynamically interact with ribosomes or RNA polymerase, but the actual interacting subunit in these complexes had not been determined before. Of the 6 novel interactions, 5 can be placed in a regulatory context, the biological relevance of the remaining interaction being yet unclear (Table 3). Of great interest are interactions of essential proteins as possible targets for new drugs. An interaction was found between the essential protein YlaN of unknown function (YLAN_BACSU) [52] and the ferric uptake repressor Fur (FUR_BACSU) [53]. Recent data point to a role of YlaN in FeS cluster biogenesis [54]. How the YlaN-Fur interaction may regulate the formation of FeS clusters remains to be elucidated. The biological significance and the relatively large fraction of transient interactions detected by this approach carried out at very low FDR at PPI level with a subset of proteins underscores the power of in vivo cross-linking. The depth of analysis of this approach can be further improved by covering a wider molecular weight window of SEC fractionated protein complexes and by application of the usual 250 or 500 mm column length for on line reversed phase LC instead of a 70 mm column used in these studies [9,39].

It should be noted that the Gram-positive *B. subtilis* is a suitable bacterium for in vivo cross-link studies of intracellular PPIs, thanks to its high membrane permeability for the cross-linkers DSG and BAMG [39], and probably others like DSSO and DSBU as well. On the contrary, the double membrane of Gram-negative organisms may form a barrier towards efficient cross-linking of cytoplasmic proteins. This can be explained both by a relatively low permeability in Gram-negative bacteria of the outer membrane for lipophilic compounds [55] like most cross-linkers and by scavenging of the cross-linker molecules that succeeded to pass this barrier by the high concentration of proteins in the periplasmic space. These circumstances may be responsible for the overrepresentation of outer membrane and periplasmic proteins in PPIs detected by in vivo cross-linking of Gram-negative organisms, for

instance *E. coli* [56] and the multi drug-resistant nosocomial pathogen *Acinetobacter baumannii* [57].

12. Perspectives and summary

High standards with respect to FDR estimation are necessary for CXMS to play an important role in the discovery of PPIs in living material in the near future. First of all optimal conditions of cross-linking, digestion and state of the art LC-MS/MS are required. Furthermore a search engine should be used that efficiently identifies cross-links with the chosen cross-linker. To this end a peptide library for benchmarking can be applied [58]. Criteria for a search engine choice include the number of identified cross-links, the experimental FDR, the speed of analysis and the format of the output file, enabling easy application of additional filters. After in vivo cross-linking, targeted isolation of cross-linked peptides, or extensive fractionation using either HILIC, SCXC or SEC or a combination of these approaches are needed preceding LC-MS/MS analysis to increase the number of identifications and to lower the FDR. A possible workflow to efficiently and reliably identify PPIs could start with data analysis to detect PPI-signaling inter-protein peptide pairs at a lenient 5% or 10% FDR at CSM level. Output files should list both target (T) and decoy (D) hits. This enables separate assessment of the experimental FDR, the presence of false intra-protein cross-linked peptide pairs being signaled by the presence of DD, TD or DT peptide pairs from the same protein and false inter-protein peptide pairs by the presence DD, TD or DT peptide pairs from different proteins. In this stage retention time prediction of SCX chromatography can be used as a filter for part of spurious inter-protein peptide pairs. The large fraction of intra-protein cross-links can be used as true positives, since even at moderate FDR at CSM level usually no false positives can be detected in this cross-link type [9,19,32]. Increasing the stringency of MS/MS criteria to filter remaining false positive inter-protein peptide pairs can be based on characteristic features of inter-protein decoy hits with respect to matched intensity and unambiguously assigned y ions, and to a lesser extent b ions. If decoy hits are not reported by the search engine in results files, or if an additional independent method is required, the entrapment database approach can be applied. Importantly, these aspects could in principle be applied to previously analyzed datasets and be incorporated into novel and existing software tools easily. This likewise holds for determination of the correct identity in similar cases as shown in Fig. 2.

Also of importance are the recently developed methods to recognize false positives among inter-protein crosslinks. Of these methods both structural-assay based error recognition [8], a differential behavior of intra-protein and inter-protein crosslinks as function of the sequence database search space [9,36] and anomalous chromatographic retention times [9] are applicable to in vitro as well as in vivo approaches. For an approach with cell free extracts, pre-fractionation of proteins by SEC followed by cross-linking and recombining of SEC fractions before LC-MS/MS is applicable [10]. Final presentation of data should be accompanied by FDR estimation at the PPI level.

In conclusion, recent approaches have been described to recognize the presence of false-positive inter-protein cross-linked peptides in datasets and to obtain an FDR at the PPI level in the 1–2% range using stringent criteria for assignment. Under these conditions the sensitivity can be assessed with the true positive intra-protein cross-links reliably determined with a lower stringency [9]. When in the near future more PPIs can be identified at this high degree of reliability, CXMS can play a major role in understanding of the structure, function and dynamics of living material.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.W. Back, L. de Jong, A.O. Muijsers, C.G. de Koster, Chemical cross-linking and mass spectrometry for protein structural modeling, J. Mol. Biol. 331 (2003) 303–313.
[2] A. Sinz, Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes, J. Mass Spectrom. 38 (2003) 1225–1237.
[3] J.D. Chavez, J.E. Bruce, Chemical cross-linking with mass spectrometry: a tool for systems structural biology, Curr. Opin. Chem. Biol. 48 (2019) 8–18.
[4] C. Iacobucci, M. Götze, A. Sinz, Cross-linking/mass spectrometry to get a closer view on protein interaction networks, Curr. Opin. Biotechnol. 63 (2020) 48–53.
[5] M. Matzinger, K. Mechtler, Cleavable cross-linkers and mass spectrometry for the ultimate task of profiling protein–protein interaction networks in vivo, J. Proteome Res. 20 (2021) 78–93.
[6] A. Leitner, A.M.J.J. Bonvin, C.H. Borchers, R.J. Chalkley, J. Chamot-Rooke, C. W. Combe, J. Cox, M.-Q. Dong, L. Fischer, M. Götze, F.C. Gozzo, A.J.R. Heck, M. R. Hoopmann, L. Huang, Y. Ishihama, A.R. Jones, N. Kalisman, O. Kohlbacher, K. Mechtler, R.L. Moritz, et al., Toward increased reliability, transparency, and accessibility in cross-linking mass spectrometry, Struct. Lond. Engl. 2020 (28) (1993) 1259–1268.
[7] C. Iacobucci, C. Piotrowski, R. Aebersold, B.C. Amaral, P. Andrews, K. Bernfur, C. Borchers, N.I. Brodie, J.E. Bruce, Y. Cao, S. Chaignepain, J.D. Chavez, S. Claverol, J. Cox, T. Davis, G. Degliesposti, M.-Q. Dong, N. Edinger, C. Emanuelsson, M. Gay, et al., First community-wide, comparative cross-linking mass spectrometry study, Anal. Chem. 91 (2019) 6953–6961.
[8] K. Yugandhar, T.-Y. Wang, S.D. Wierbowski, E.E. Shayhidin, H. Yu, Structure-based validation can drastically underestimate error rate in proteome-wide cross-linking mass spectrometry studies, Nat. Methods 17 (2020) 985–988.
[9] L. de Jong, W. Roseboom, G. Kramer, A composite filter for low FDR of protein-protein interactions detected by in vivo cross-linking, J. Proteomics 230 (2021) 103987.
[10] S. Lenz, L.R. Sinn, F.J. O'Reilly, L. Fischer, F. Wegner, J. Rappsilber, Reliable identification of protein-protein interactions by crosslinking mass spectrometry, BioRxiv (2020), https://doi.org/10.1101/2020.05.25.114256.
[11] S. Kalkhof, A. Sinz, Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters, Anal. Bioanal. Chem. 392 (2008) 305–312.
[12] M.L. Mendes, L. Fischer, Z.A. Chen, M. Barbon, F.J. O'Reilly, S.H. Giese, M. Bohlke-Schneider, A. Belsom, T. Dau, C.W. Combe, M. Graham, M.R. Eisele, W. Baumeister, C. Speck, J. Rappsilber, An integrated workflow for crosslinking mass spectrometry, Mol. Syst. Biol. 15 (2019), e8994.
[13] D.K. Schweppe, J.D. Chavez, C.F. Lee, A. Caudal, S.E. Kruse, R. Stuppard, D. J. Marcinek, G.S. Shadel, R. Tian, J.E. Bruce, Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 1732–1737.
[14] F. Liu, P. Lössl, B.M. Rabbitts, R.S. Balaban, A.J.R. Heck, The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes, Mol. Cell. Proteomics MCP 17 (2018) 216–232.
[15] P.S.J. Ryl, M. Bohlke-Schneider, S. Lenz, L. Fischer, L. Budzinski, M. Stuiver, M.M. L. Mendes, L. Sinn, F.J. O'Reilly, J. Rappsilber, In situ structural restraints from cross-linking mass spectrometry in human mitochondria, J. Proteome Res. 19 (2020) 327–336.
[16] A. Linden, M. Deckers, I. Parfentev, R. Pflanz, B. Homberg, P. Neumann, R. Ficner, P. Rehling, H. Urlaub, A cross-linking mass spectrometry approach defines protein interactions in yeast mitochondria, Mol. Cell. Proteomics MCP 19 (2020) 1161–1178.
[17] K.A.T. Makepeace, Y. Mohammed, E.L. Rudashevskaya, E.V. Petrotchenko, F.-N. Vögtle, C. Meisinger, A. Sickmann, C.H. Borchers, Improving identification of in-organello protein-protein interactions using an affinity-enrichable, isotopically coded, and mass spectrometry-cleavable chemical crosslinker, Mol. Cell. Proteomics MCP 19 (2020) 624–639.
[18] M.J. Trnka, P.R. Baker, P.J.J. Robinson, A.L. Burlingame, R.J. Chalkley, Matching cross-linked peptide spectra: only as good as the worse identification, Mol. Cell. Proteomics MCP 13 (2014) 420–434.
[19] H. Buncherd, W. Roseboom, L.J. de Koning, C.G. de Koster, L. de Jong, A gas phase cleavage reaction of cross-linked peptides for protein complex topology studies by peptide fragment fingerprinting from large sequence database, J. Proteomics 108 (2014) 65–77.
[20] L. Fischer, J. Rappsilber, Quirks of error estimation in cross-linking/mass spectrometry, Anal. Chem. 89 (2017) 3829–3833.
[21] J.E.P. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 9528–9533.
[22] J.V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann, Higher-energy C-trap dissociation for peptide modification analysis, Nat. Methods 4 (2007) 709–712.
[23] H. Buncherd, W. Roseboom, B. Ghavim, W. Du, L.J. de Koning, C.G. de Koster, L. de Jong, Isolation of cross-linked peptides by diagonal strong cation exchange chromatography for protein complex topology studies by peptide fragment fingerprinting from large sequence databases, J. Chromatogr. A 1348 (2014) 34–46.
[24] P.F. Pilch, M.P. Czech, Interaction of cross-linking agents with the insulin effector system of isolated fat cells covalent linkage of 125I-insulin to a plasma membrane receptor protein of 140,000 daltons, J. Biol. Chem. 254 (1979) 3375–3381.

[25] J.V. Staros, N-hydroxysulfosuccinimide active esters: bis(N-hydroxysulfosuccinimide) esters of two dicarboxylic acids are hydrophilic, membrane-impermeant, protein cross-linkers, Biochemistry 21 (1982) 3950–3955.

[26] Z.-L. Chen, J.-M. Meng, Y. Cao, J.-L. Yin, R.-Q. Fang, S.-B. Fan, C. Liu, W.-F. Zeng, Y.-H. Ding, D. Tan, L. Wu, W.-J. Zhou, H. Chi, R.-X. Sun, M.-Q. Dong, S.-M. He, A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides, Nat. Commun. 10 (2019) 3404.

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

[28] E.V. Petrotchenko, J.J. Serpa, C.H. Borchers, An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics, Mol. Cell. Proteomics MCP 10 (2011) (M110.001420).

[29] A. Kao, C. Chiu, D. Vellucci, Y. Yang, V.R. Patel, S. Guan, A. Randall, P. Baldi, S. D. Rychnovsky, L. Huang, Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes, Mol. Cell. Proteomics MCP 10 (2011) (M110.002212).

[30] M.Q. Müller, F. Dreiocker, C.H. Ihling, M. Schäfer, A. Sinz, Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS, Anal. Chem. 82 (2010) 6958–6968.

[31] P.T. Kasper, J.W. Back, M. Vitale, A.F. Hartog, W. Roseboom, L.J. de Koning, J. H. van Maarseveen, A.O. Muijsers, C.G. de Koster, L. de Jong, An aptly positioned azido group in the spacer of a protein cross-linker for facile mapping of lysines in close proximity, Chembiochem. Eur. J. Chem. Biol. 8 (2007) 1281–1292.

[32] F. Liu, P. Lössl, R. Scheltema, R. Viner, A.J.R. Heck, Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification, Nat. Commun. 8 (2017) 15473.

[33] L. Kolbowski, M.L. Mendes, J. Rappsilber, Optimizing the parameters governing the fragmentation of cross-linked peptides in a tribrid mass spectrometer, Anal. Chem. 89 (2017) 5311–5318.

[34] K. Yugandhar, T.-Y. Wang, A.K.-Y. Leung, M.C. Lanz, I. Motorykin, J. Liang, E. E. Shayhidin, M.B. Smolka, S. Zhang, H. Yu, MaXLinker: proteome-wide cross-link identifications with high specificity and sensitivity, Mol. Cell. Proteomics MCP 19 (2020) 554–568.

[35] Z. Ser, P. Cifani, A. Kentsis, Optimized cross-linking mass spectrometry for in situ interaction proteomics, J. Proteome Res. 18 (2019) 2545–2558.

[36] M. Götze, C. Iacobucci, C.H. Ihling, A. Sinz, A simple cross-linking/mass spectrometry workflow for studying system-wide protein interactions, Anal. Chem. 91 (2019) 10236–10244.

[37] O. Rinner, J. Seebacher, T. Walzthoeni, L.N. Mueller, M. Beck, A. Schmidt, M. Mueller, R. Aebersold, Identification of cross-linked peptides from large sequence databases, Nat. Methods 5 (2008) 315–318.

[38] F. Liu, D.T.S. Rijkers, H. Post, A.J.R. Heck, Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry, Nat. Methods 12 (2015) 1179–1184.

[39] L. de Jong, E.A. de Koning, W. Roseboom, H. Buncherd, M.J. Wanner, I. Dapic, P. J. Jansen, J.H. van Maarseveen, G.L. Corthals, P.J. Lewis, L.W. Hamoen, C.G. de Koster, In-culture cross-linking of bacterial cells reveals large-scale dynamic protein-protein interactions at the peptide level, J. Proteome Res. 16 (2017) 2457–2471.

[40] B. Yang, Y.-J. Wu, M. Zhu, S.-B. Fan, J. Lin, K. Zhang, S. Li, H. Chi, Y.-X. Li, H.-F. Chen, S.-K. Luo, Y.-H. Ding, L.-H. Wang, Z. Hao, L.-Y. Xiu, S. Chen, K. Ye, S.-M. He, M.-Q. Dong, Identification of cross-linked peptides from complex samples, Nat. Methods 9 (2012) 904–906.

[41] A. Tabib-Salazar, N. Mulvenna, K. Severinov, S.J. Matthews, S. Wigneshweraraj, Xenogeneic regulation of the bacterial transcription machinery, J. Mol. Biol. 431 (2019) 4078–4092.

[42] C.-Y. Yen, S. Russell, A.M. Mendoza, K. Meyer-Arendt, S. Sun, K.J. Cios, N.G. Ahn, K.A. Resing, Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: protease cleavage and SCX elution rules from data mining of MS/MS spectra, Anal. Chem. 78 (2006) 1071–1084.

[43] R.C. Dwivedi, V. Spicer, M. Harder, M. Antonovici, W. Ens, K.G. Standing, J. A. Wilkins, O.V. Krokhin, Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics, Anal. Chem. 80 (2008) 7036–7042.

[44] D. Gussakovsky, H. Neustaeter, V. Spicer, O.V. Krokhin, Sequence-specific model for peptide retention time prediction in strong cation exchange chromatography, Anal. Chem. 89 (2017) 11795–11802.

[45] V. Spicer, Y.W. Lao, D. Shamshurin, P. Ezzati, J.A. Wilkins, O.V. Krokhin, N-capping motifs promote interaction of amphipathic helical peptides with hydrophobic surfaces and drastically alter hydrophobicity values of individual amino acids, Anal. Chem. 86 (2014) 11498–11502.

[46] O.V. Krokhin, P. Ezzati, V. Spicer, Peptide retention time prediction in hydrophilic interaction liquid chromatography: data collection methods and features of additive and sequence-specific models, Anal. Chem. 89 (2017) 5526–5533.

[47] R.M. Kaake, X. Wang, A. Burke, C. Yu, W. Kandur, Y. Yang, E.J. Novtisky, T. Second, J. Duan, A. Kao, S. Guan, D. Vellucci, S.D. Rychnovsky, L. Huang, A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells, Mol. Cell. Proteomics MCP 13 (2014) 3533–3543.

[48] M. Matzinger, W. Kandioller, P. Doppler, E.H. Heiss, K. Mechtler, Fast and highly efficient affinity enrichment of Azide-A-DSBSO cross-linked peptides, J. Proteome Res. 19 (2020) 2071–2079.

[49] D. Tan, Q. Li, M.-J. Zhang, C. Liu, C. Ma, P. Zhang, Y.-H. Ding, S.-B. Fan, L. Tao, B. Yang, X. Li, S. Ma, J. Liu, B. Feng, X. Liu, H.-W. Wang, S.-M. He, N. Gao, K. Ye, M.-Q. Dong, et al., Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states, ELife 5 (2016), e12509.

[50] B. Steigenberger, R.J. Pieters, A.J.R. Heck, R.A. Scheltema, PhoX: an IMAC-enrichable cross-linking reagent, ACS Cent. Sci. 5 (2019) 1514–1522.

[51] A. Belsom, J. Rappsilber, Anatomy of a crosslinker, Curr. Opin. Chem. Biol. 60 (2020) 39–46.

[52] L. Xu, S.E. Sedelnikova, P.J. Baker, A. Hunt, J. Errington, D.W. Rice, Crystal structure of S aureus YlaN, an essential leucine rich protein involved in the control of cell shape, Proteins 68 (2007) 438–445.

[53] H. Pi, J.D. Helmann, Sequential induction of Fur-regulated genes in response to iron limitation in Bacillus subtilis, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 12785–12790.

[54] J.M. Peters, A. Colavin, H. Shi, T.L. Czarny, M.H. Larson, S. Wong, J.S. Hawkins, C. H.S. Lu, B.-M. Koo, E. Marta, A.L. Shiver, E.H. Whitehead, J.S. Weissman, E. D. Brown, L.S. Qi, K.C. Huang, C.A. Gross, A comprehensive, CRISPR-based functional analysis of essential genes in bacteria, Cell 165 (2016) 1493–1506.

[55] H. Nikaido, Molecular basis of bacterial outer membrane permeability revisited, Microbiol. Mol. Biol. Rev. MMBR 67 (2003) 593–656.

[56] J.P. Mohr, P. Perumalla, J.D. Chavez, J.K. Eng, J.E. Bruce, Mango: a general tool for collision induced dissociation-cleavable cross-linked peptide identification, Anal. Chem. 90 (2018) 6028–6034.

[57] X. Wu, J.D. Chavez, D.K. Schweppe, C. Zheng, C.R. Weisbrod, J.K. Eng, A. Murali, S.A. Lee, E. Ramage, L.A. Gallagher, H.D. Kulasekara, M.E. Edrozo, C. N. Kamischke, M.J. Brittnacher, S.I. Miller, P.K. Singh, C. Manoil, J.E. Bruce, In vivo protein interaction network analysis reveals porin-localized antibiotic inactivation in Acinetobacter baumannii strain AB5075, Nat. Commun. 7 (2016) 13414.

[58] R. Beveridge, J. Stadlmann, J.M. Penninger, K. Mechtler, A synthetic peptide library for benchmarking crosslinking-mass spectrometry search engines for proteins and protein complexes, Nat. Commun. 11 (2020) 742.