Development of enrichment methods for cross-linked peptides to study the dynamic topology of large protein complexes by mass spectrometry

Buncherd, H.

Citation for published version (APA):

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.
General Discussion

Abstract

Chemical cross-linking combined with mass spectrometry (CXMS) is a powerful approach providing detailed insights into structural flexibility and interaction of proteins/complexes which often escape detection by high-resolution techniques. In this thesis, the development of enrichment approaches aimed to exploit the full potential of CXMS has been addressed. Despite the success of the devised methods to CXMS, further improvements in separative capacity of approaches are expected. Conjoint efforts towards the improvement in software tools and analysis strategies will enhance CXMS. The CXMS-applications proposed in this chapter can be used to obtain novel information aiding understanding processes in the cell at the molecular level.
Introduction

In this thesis, the development and application of approaches aimed to selectively isolate azide-containing cross-linked peptides have been described. The enrichment of azide-containing cross-linked peptides via a cycloaddition reaction using an azide reactive cyclooctyne (ARCO) resin has allowed the conformation and architecture of a macromolecule as complex as a bacterial RNA polymerase to be successfully illustrated (described in Chapter 2). This has demonstrated the value of peptide enrichment for the benefit of CXMS. A completely different approach using diagonal strong cation exchange chromatography for enrichment of azide containing cross-linked peptides has been described in Chapter 4. This novel enrichment approach combined with new bioinformatics strategies to enhance CXMS has enabled the mapping of protein-protein interaction sites from a complex biological system i.e. a HeLa nuclear extract. This is the first time that protein-protein interactions have been identified by CXMS using a database generated from the whole human proteome. Furthermore, in Chapter 5, we have extended the use of our CXMS strategies to shed light on an unresolved biological question, i.e., identification of the interaction sites between bacterial RNA polymerases and helicases. In this chapter, further improvements of the enrichment approaches, strategies and future applications are discussed.

Optimization of conditions used for azide reduction will improve the enrichment of azide-containing cross-linked peptides

In Chapter 2, it has been noticed that after enrichment by the ARCO-resin, azide-containing cross-linked peptides tend to aggregate which was attributed to the hydrophobic nature of the benzyl-group containing cyclooctyne moiety. Moreover, reaction of the ARCO resin with azides is relatively slow. This led to the development of a resin conjugated with a monofluorinated cyclooctyne (MFCO). Compared to ARCO, MFCO is more soluble since it lacks the benzyl group. In addition, an electron-withdrawing fluorine substituent allows the reaction to occur at a faster rate. However, in Chapter 3, it has been shown that the coupling of azide-containing peptides to the MFCO-resin yielded unknown compounds generated from side reactions. Furthermore, gas phase loss of HF has been observed both prior to and during peptide fragmentation by CID, thereby hindering CXMS. For these reasons, MFCO is not suitable for enrichment of azide-containing peptides under the used experimental conditions.

A novel approach for the enrichment of azide containing cross-linked peptides relying on charge-charge interaction has been described in Chapter 4. In this approach, an azido group tagged on the spacer of a cross-linker is converted to a primary amine group by TCEP-induced reduction (TCIR) [1]. Under acidic condition, the reduced-form of azide-containing cross-linked peptides accommodates an additional positive charge derived from a protonated amine yielded from TCIR. By virtue of diagonal strong cation exchange chromatography, the reduced peptides could be selectively separated from unmodified species. An additional advantage of the conversion of the relatively apolar azido group to an amine group is an increased solubility in aqueous solutions, facilitating handling of peptides. Aside from the reduction of the azido group, TCEP also induces a competing cleavage reaction of one of the two amide bonds
connecting formed in the cross-linking reaction between two amino acid residues. A mechanism for the reduction and cleavage reaction has been proposed by Back et al [2]. In our study (Chapter 4) the ratio of reduced/cleavage products based on the UV absorption of early and late shifted material in 2SCX was approximately 1:1. Consistent results have been reported for the BAMG-cross-linked model peptide neurotensin by Kasper et al [1]. For the Reang approach (Chapter 4) and the total cross-linked database approach (chapter 5) a high reduced/cleavage product ratio would be beneficial.

Among conventional reducing agents, dithiothreitol (DTT) would be one of the most promising candidates to increase the reduction/cleavage ratio. DTT is commonly used to reduce disulfides bond in proteins and has also been studied for the reduction of an azido group to the corresponding amine group [2]. In addition to the latter reaction, DTT has been shown to induce a reduction of azido groups on the non-natural amino acid azidohomoalanine (azhal) in peptides in competition with the cleavage of peptides at azhal. Interestingly, a ratio of reduced/cleavage product after completion of the reaction (as evidenced by the disappearance of starting compounds) was about 2.5:1 [2]. This relatively high ratio would improve both the enrichment of azide-containing cross-linked peptides and the subsequent mass spectrometric analysis. However, the used reaction conditions [2] required a long incubation time (~16 hr.), while DTT is susceptible to air oxidation, implying the necessity for optimization of the experimental conditions for reduction of BAMG-cross-linked peptides.

An alternative reducing agent that could improve the ratio of reduced/cleavage product is 2- mercaptoethanol (2ME). Similar to DTT, 2ME has also been observed to induce the reduction of azido groups in peptidyl azhal competition with the cleavage of peptides at azhal [2]. In this work, after 48 hr. about 15% of starting material was converted. Intriguingly, only 1% of the converted material was identified as a cleavage product. This high ratio of reduced/cleavage product would significantly enhance the enrichment of the azide-containing cross-linked peptides. However, the reaction rate of the conversion of azides by 2ME is very low. Therefore, an optimization of the condition used for reduction of azides by 2ME is indispensable. One possibility is to increase the concentration of 2ME as this will result in a higher reaction rate [3].

**Combination of peptide fragmentation techniques will improve the identification of azide-containing cross-linked peptides**

Although CXMS has made great progress in the last decade, the identification of cross-linked peptides remains a critical challenge. This may be caused by the large number of theoretical possible cross-link candidates and the lack of full understanding mass spectra of cross-linked peptides.

The identification of linear and cross-linked peptides in MS analysis usually shares a common scheme of work. The first step is the proteolysis of proteins via a protease (e.g. trypsin, chymotrypsin, Glu C, Lys C etc.) to generate smaller peptides followed by a separation through liquid chromatographic (LC) techniques (e.g. reversed phase-LC, ion exchange
chromatography-LC) prior to MS analysis. Ionization using ESI yields primarily 2+ and 3+ ions for linear peptides and up to 6+ for cross-linked peptides) enabling sequence-informative fragmentation of peptides by low-energy collision induced dissociation (CID). The collision of peptide ions with neutral gases results in a vibrational excitation that induces cleavage of an amide bond of the peptide backbone between the carbonyl and the amine group. As a result, b- and y-fragment ions are formed, representing N-terminal and C-terminal fragment ions, respectively. The reliability of peptide identification is tightly connected to the number of the fragment ions from the peptide backbone. However, peptides that are highly suitable for conventional CID are relatively small (~6-20 amino residues) and have a charge ≤ 3+ [4, 5] corresponding to the general characteristic of linear peptides. In contrast to the regular species, cross-linked peptides are larger (primarily, 10-35 amino residues) and more highly charged in ESI (predominantly 3+ ≤ z ≤ 5+) [6]. It is known that CID produces a lower sequence coverage when the peptide becomes larger and more highly charged [7]. We have observed that sometimes no fragments at all from one of the two peptides are detected, preventing unambiguous identification. In the data set in Chapter 4 and 5, we found that nearly 20% of the selected precursor ions were both larger than 3,000 Da and highly charged peptides (≥ 4+). Of this 20%, only a quarter could be identified as cross-linked peptides. Although not all these precursor ions may correspond to cross-linked peptides, an improvement of peptide identification is expected to be obtained by the use of complementary approaches effectively generating informative fragment ions from large and highly charged peptides.

One of the promising candidates is electron transfer dissociation (ETD) [8]. It is an adapted version of electron-capture dissociation (ECD) [9]. In the ETD process, positively charged peptide ions are converted into radical cations via electron transfer reaction, mostly through anthracene radical anions. Cleavage of the bond between the amine and the α-carbon in the peptide backbone occurs as a consequence of the instability of the radical cation. The fragment ions generated from the C-terminal and N-terminal part of peptides are termed c- and z-fragment ions respectively. ETD is rather suitable for larger peptides (>16 amino residues) and higher charged peptide ions (z ≥ 3+) compared to CID [4, 5]. In addition, it has been shown that the fragmentation coverage of cross-linked peptides produced by ECD was significantly improved for higher charge states [10]. As the characteristics for peptide identification by CID and ETD techniques are complementary, the combination of these techniques is a promising approach for comprehensive analysis of CXMS [11].

Another promising candidate used to improve fragmentation of cross-linked peptides is infrared multiple photon dissociation (IRMPD) [10]. In this technique, peptide precursors are bombarded with infrared photons. Peptide ions subsequently absorb and accumulate internal energy until the energy of activation is achieved, leading to peptide bond dissociation. This primarily resulted in b- and y-type product ions. Similar to ETD/ECD, IRMPD leads to improved fragmentation coverage for large and highly charged peptide ions [10]. Therefore, the combination of CID and IRMPD would enhance the efficacy of cross-linked peptide identification.
Software toolkits for CXMS

One of the most important requirements enabling a successful CXMS, particularly of large complex systems, is a software suite aiding data analysis. In Chapter 4, software tools have been developed to support our CXMS strategies and validation of cross-linked peptides, i.e., Biner, Reang and Yeun Yan. Here, some features that may improve CXMS strategies are discussed.

Further filtering of fragment ions may enhance the nomination of peptides by the Reang approach

As outlined in Chapter 4, briefly, Reang identifies precursor ions, potentially corresponding to a cross-linked peptide pair A and B, showing evidence for cleavage of the amide bonds in the presumed cross-link. Such cleavage events result in product ions of the unmodified peptides A and B and in modified peptides $A_m$ and $B_m$ fulfilling the mass relationships (equation 1, 2 and 3) shown in Chapter 4. Each entry in the MS$_1$MS$_2$ data file identified in this search is then replaced by new entries in pkl file format. The original precursor ion m/z value is replaced by the m/z values of the candidate composing peptide ions, and the list of fragment ions is sorted to contain only species not exceeding the mass of the new precursor ion. The new pkl files are input for Mascot to nominate candidate peptides for $A$, $A_m$, $B$ and $B_m$ by interrogating a protein database. This approach implies that the pkl file entries for peptide $A$ and $A_m$ still contain fragment ions belonging to peptide B and $B_m$ and vice versa. The questions arises how efficient both peptides $A$ and $B$ or their lactam modifications can be identified and to which extent the “contaminating” fragments from the other composing peptide prevent identification by Mascot. An answer to the first question is provided by the result mentioned in Chapter 4 that 17 cross-linked peptides from the T complex were identified with the Reang approach, while the total cross-linked peptide database approach yielded 25 cross-links. Inspection of the Yean Yan output of the 8 cross-links not identified in the Reang approach suggests that in four cases both composing peptides built up from at least 6 amino acids yielded enough fragments for assignment by Mascot. Possibly one or both peptides escape nomination by Mascot due to the presence of fragment ions from the other peptide. In cases in which only one composing peptide is nominated by Mascot it would be worthwhile to remove the fragment ions belonging to the nominated composing peptide and its lactam-modified from the pkl file entry of the other candidate composing peptide, followed by a Mascot search with the thus filtered pkl file entry (Fig. 1).

Reorganization of codes may improve the speed of data analysis

Software tools used for our CXMS strategies have been written as prototypes and operated on a Microsoft Visual Basic platform. For the validation of cross-linked candidates nominated by the Reang approach from HeLa nuclear extract sample (Chapter 4), the rate of validation by Yeun Yan was about 10 candidates/ min, performed on Windows 7, Version 2009, service pack 2, Intel(R) Core(TM)2 Duo CPU P8600 @ 2.4GHz and 4 GB DDR2 memory. In this study, > 100,000 candidates needed to be validated by Yeun Yan, thus requiring > 166 hr. runtime to complete the calculation. In addition, several codes should be reorganized or rewritten for more
efficient calculation. Although the speed of validation is not definitely a prerequisite for CXMS, faster software would be more preferable. Other computer languages e.g. C, C++ or Perl would be optional as they consume less memory thus performing at higher speed.

![Diagram](image)

**Fig. 1** An additional strategy to enhance the Reang approach. Briefly, the Reang approach identifies pairs of gas-phase cleavage products of each parent ion (P) using equations shown in the black line box. Potential cleavage products are subsequently used to generate a new pkl file for database searching. For instance, a pair of potential cleavage products identified by Reang is composed of $M_A$ and $M_B$. The new pkl file contains a new set of MS$_1$MS$_2$ spectra which each MS$_1$MS$_2$ spectrum comprises 1) a potential cleavage product as a new precursor ion ($M_A$) and 2) fragment ions derived from a parent ion (P, i.e., a type 2 cross-linked peptide candidate) of the new precursor ion ($M_A$). The fragment ions exceeding the mass of the new precursor ($M_A$) ion are removed. The new precursor ion ($M_A$) with the sorted fragment ions is then submitted to a search engine for screening the sequence database. Fragment ions assigned for ($M_A$) by the search engine are removed from of the MS$_1$MS$_2$ pkl file entry of $M_B$. 

120
Future plans

A success in mapping protein-protein interaction sites of both complex mixtures (Chapter 4) and isolated complexes (Chapter 5) has opened opportunities for obtaining further information to understand the molecular basis of biological processes. Here we suggest possible applications using our CXMS strategy as a core approach to extricate cellular protein interaction networks.

Mapping of protein-protein interaction sites in complexes/large scale

Formation of large complexes is a basis of several biological processes. Of each interaction involved in such complicated formation, the characterization is indispensable for understanding their molecular functions in the cell. Despite several methodological approaches [12] for mapping protein-protein interaction (PPI) developed during the past years, two techniques are the most pervasive for large scale analysis [13]. The yeast two-hybrid (Y2H) system is a genomic-based approach primarily used to map binary interactions. Another approach used to identify proteins involved in the complex is affinity purification coupled with mass spectrometry (AP/MS). Although these two or even perhaps all large-scale mapping approaches, thus far, can identify proteins involved in the interactions/complexes, none of these approaches provide mapping of protein-protein interaction sites at the resolution that can be obtained by CXMS. Moreover, due to the high false positive/negative rate and often irreproducibility of the results, the efficiency and reliability of PPI identified by Y2H are currently critically questioned [12]. In addition, AP/MS cannot distinguish between direct and indirect interaction of proteins in the complex [14]. Therefore, considerable care has to be taken to interpret AP/MS derived data.

As demonstrated in Chapter 4 and 5, CXMS could provide mapping of protein-protein interaction sites either in a small system as isolated complexes or large-scale on a (sub)/proteomic level. Furthermore, CXMS is capable of distinguishing type of interactions (direct/indirect interaction) as well as can also illustrate the dynamic flexibility of complexes under physiological condition. Moreover, since our cross-linker, BAMG is a cell membrane permeable reagent, capturing of the PPI network in living cell (in vivo CXMS) is possible. The use of CXMS developed in this thesis will, therefore, certainly provide additional/novel information for PPI networks (Fig. 2).
Fig. 2 Characterization of protein-protein interaction (PPI). CXMS can be applied to study PPI in various sources of samples e.g. isolated complexes, whole cell extracted. In addition, the use of cross-linkers that permeate through a cell membrane, *in vivo* mapping of PPI networks can be achieved.

**Quantitative CXMS analysis**

*The region of interaction*

Protein-protein interaction surfaces can indirectly be located by the absence/presence of modified/linked residues using quantitative CXMS analysis. To this end, $^{14}$N-labeled subunit ($^{14}$P) and $^{15}$N-labeled intact complex ($^{15}$C) are separately cross-linked with BAMG. Similar to conventional quantitative proteomic approaches, the resulting cross-linked $^{14}$P and $^{15}$C are mixed in 1:1 molar ratio. The mixture is digested and further analyzed using CXMS approach. Comparison of mass spectral peak intensities from $^{14}$P and $^{15}$C may reveal mono-linked and cross-linked peptides located on the surfaces of interaction. In this sense, modified residues in $^{14}$P must be accessible by cross-linkers (Fig. 3). The absence of these residues in $^{15}$C implies that they are no longer accessible, thus being buried on the interaction surface (Fig. 3).

*Protein-protein interactions under two different experimental conditions and changes in protein-protein interactions in time*

$^{15}$N/$^{14}$N CXMS analysis can also be applied to study protein-protein interactions occurring in the cell under two different conditions e.g. changes in protein-protein interaction profile under stress condition. For instance, $^{14}$N treated with stress of interest and $^{15}$N reference preparation are separately cross-linked with BAMG. After cell extractions, $^{15}$N and $^{14}$N cross-linked
material is mixed in equimolar amount followed by XL identification and relative quantification of all identified cross-linked peptides based on their $^{15}$N/$^{14}$N ratios. This provides information to reveal the dynamics of protein-protein interactions. In addition, the changes in protein-protein interaction profile (from isolated proteins/complex system) can be followed in time by using the $^{15}$N-labelled system as a reference. Alternatively, one can use a $^{13}$C ($^{15}$N)-labeled cross-linker for the reference and $^{12}$C ($^{14}$N) for the experimental system.

**Fig. 3** Identification of interaction surface using quantitative CXMS analysis. Comparison of cross-links identified in free subunit and its complex form may reveal the interaction surface. For instance, Panel A, five cross-linked peptides are visible in free form of subunit. Panel B, the absence of cross-links ID 2, 3, 4 (shown in red) after
forming complex with its counterpart (shown in blue) can be quantitated by 15N/14N-labeling technique. Inset shows the example of expected mass spectral peak intensities 15N/14N ratio from 15C and 14P.

**Conclusion**

In this thesis we have explored two strategies to isolate cross-linked peptides from complex peptide mixtures obtained by digestion of (mixtures of) BAMG-treated protein complexes: capturing to ARCO beads (*Chapter 2 and 3*) and diagonal chromatography (*Chapter 4 and 5*). The latter method turned out to be superior over the first for two mean reasons. The first reason is the tendency of the cross-linked peptides released from the ARCO beads to aggregate in aqueous media, in contrast to cross-linked peptides isolated by diagonal chromatography. This problem could not be met by using a less hydrophobic and more reactive monofluorinated cycloctyne, since side reactions and HF loss in the gas phase severely hampered mass spectrometric analysis of cross-linked peptides (*Chapter 3*). The second reason is a bonus that the diagonal chromatography approach kept in store, namely the gas phase cleavability of the cross-link amide bonds upon reduction of the azido group to an amine group. This property enabled determination of the masses of the two composing peptides of BAMG-cross-linked species. Knowledge of the masses of composing peptides tremendously facilitates CXMS from large sequence databases, as shown in *Chapter 4*. Further improvement of the method is expected by the use of a reducing agent predominantly converting azide-containing cross-linked peptides to the corresponding amine group.

The use of CID-complementary approaches effectively generating informative fragment ions from large and highly charged peptides e.g. ETD, IRMPD is expected to improve the peptide fragmentation of cross-linked peptides that lack the sequence information of one of the two composing peptides. Furthermore, an additional strategy to further filter fragment ions is expected to aid the nomination by the Reang approach of peptides containing only a few fragment ions of low signal intensity. Finally, the identification and validation of cross-linked peptide candidates can be less time consuming by reorganization of the codes or operating on more suitable platforms.

Methodology coupled with strategies presented in this thesis makes possible to obtain insight into conformational changes of proteins and also enable mapping of protein-protein interaction sites either in particular complexes or on a large scale. The acquired data will aid understanding molecular functions as well as the network of communications of proteins in the cell.

**References**