Development of an enrichment method for azide-containing peptides to study proteome dynamics by mass spectrometry

Nessen, M.A.

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.
Chapter 5

Selective enrichment and identification of cross-linked peptides to study protein structures by mass spectrometry*

Abstract

Chemical cross-linking and mass spectrometry can be used to obtain structural information of proteins and protein complexes. After chemically linking residues within or between the protein(s), followed by proteolytic digestion, mass spectrometric analysis can reveal the residues that are in close proximity. However, data analysis is hampered by the complexity of the samples, containing both unmodified peptides and different types of peptides modified by the cross-linking reagent: peptides containing a single modified residue (type 0), intrapeptide cross-links (type 1) and interpeptide cross-links (type 2). In this chapter, we show that the combination of enrichment of cross-link-modified peptides, labeled by an azido group, via covalent attachment to an azide-reactive cyclooctyne-conjugated resin (ARCO-resin) and strong cation exchange (SCX) can be used to enrich and identify type 2 cross-link peptides. The observation of a reporter-ion for type 0 modified peptides further enhances the efficiency of data analysis. For two different proteins: cytochrome c and NK1, a natural occurring splice variant of the polypeptide growth factor HGF/SF, nearly all residues were found to be modified by the cross-linker. Furthermore, for NK1 a type 2 cross-link was identified that suggests the protein in solution can have a different structure than observed in crystals.

* This chapter is in preparation for publication.
Introduction

Knowledge of the tertiary and quaternary structure of a protein (complex) is important to understand its (biological) function and action on the molecular level. Common techniques to obtain high resolution structures of proteins are X-ray diffraction and nuclear magnetic resonance (NMR). However, not all proteins are suited to be studied by these methods. Other techniques frequently employed in order to obtain structural information are cryo electron microscopy followed by 3D reconstruction and small angle X-ray scattering (SAXS). These methods give rise to low resolution structures, and by modeling, different modules of the protein or protein complex can be located in the structure. To optimize the models, distance constraints within a protein or between proteins can be obtained by chemical cross-linking and mass spectrometry\(^1\text{-}^3\). The most commonly used cross-linkers are bifunctional N-hydroxy-succinimidyld (NHS) esters that are reactive towards the primary amines in a protein, as present in lysine side chains and the N-terminus of a protein\(^2\text{-}^3\). From the length of the spacer of the cross-linker, the maximum distance between the amines can be determined and used as constraints for modeling.

After digestion of a cross-linked protein (complex), four major types of peptides are expected in the digest: i) unmodified linear peptides, and cross-link modified peptides: ii) mono-links or type 0 cross-links, modification of peptides by a partially hydrolyzed cross-linker, iii) loop-links or type 1 cross-links, a cross-link within one peptide and iv) cross-links or type 2 cross-links, a cross-link between two different peptides\(^4\text{-}^5\). In addition, peptides with one or more additional modifications by the cross-linker may exist. Mainly type 2 cross-links provide powerful 3-D structural information and can be used to refine models. However, the complexity of the peptide mixture makes it hard to detect and identify the low-abundant type 2 cross-links, hampering structural analysis of larger proteins and complexes.

We have previously described a method to simplify the complexity of cross-linked peptide mixtures by combining an azide-containing cross-linker with a tris(2-carboxyethyl)phosphine (TCEP)-dependent reaction and diagonal chromatography\(^6\). However, the method is laborious and some TCEP-induced cleavage products may be too small for detection and/or identification by mass spectrometry.

Alternative methods that have been described use a biotin moiety attached to the spacer of a bifunctional cross-linking reagent, enabling affinity purification of peptides modified by the cross-linker on (strept)avidin resin\(^7\text{-}^8\). However, the
Enrichment of cross-linked peptides

relatively large size of the biotin group may interfere with cross-linking. To avoid this potential problem, while still enabling introduction of an affinity group for isolation, a cross-linker has been described containing a terminal alkyne group in the linker. This alkyne was modified after cross-linking and digestion by a biotin moiety via the (3+2) copper catalyzed cycloaddition of terminal alkynes to azides\textsuperscript{9}. The copper-catalyzed coupling reaction, though, may induce oxidation of the peptides, as has been described before, hampering identification\textsuperscript{10,11}.

As an alternative to biotin affinity purification, in Chapter 2 and 4 a solid phase enrichment method is presented that uses the azide-reactive cyclooctyne (ARCO) resin\textsuperscript{11}. Peptides modified by azide-containing cross-linkers, such as bis(succinimidyl)-3-azidomethyl glutarate (BAMG), were specifically isolated from a protein digest. However, type 2 cross-links remain a minor component of the peptide mixture dominated by the structurally less informative type 0 and type 1 cross-links, making cross-link mapping still laborious.

It this chapter, we report that the inclusion of strong cation exchange (SCX) in our enrichment method can be used to improve the mass spectrometric analysis of cross-linked peptides. In addition, the observation of a reporter-ion for type 0 cross-links allows exclusion of this type of cross-links from the analyses, reducing the complexity of the analysis of cross-linked samples even further. We demonstrate that type 2 cross-links are selectively isolated and identified for two different proteins: cytochrome c and NK1, a natural occurring splice variant, composed of two domains, of the polypeptide growth factor HGF/SF.
Results and Discussion

Recently we have described a method for the specific isolation of azide-containing peptides, using the azide-reactive-cyclooctyne (ARCO) resin (Figure 1c). Via a strain-promoted (3+2)-cycloaddition, peptides are captured on the resin via their azide-moiety to the cylooctyne, forming a triazole (Figure 1d). Reductive cleavage of the disulfide bond liberates a thiol that can be alkylated by iodoacetamide. The mass of the released peptide will increase with the residue mass of the cyclooctyne, linker and alkylating reagent, resulting in an addition of 358.2 Da (Figure 1d). An overview of the experimental setup is given in Figure 1a.

![Figure 1](image.png)

**Figure 1, Overview of procedure and chemicals used for the enrichment of cross-linked peptides.** To obtain spatial distance information, proteins are cross-linked with bis(succinimidyl)-3-azidomethyl glutarate (BAMG). After cross-linking, proteins are digested and the obtained peptide mixture is incubated with the ARCO-resin. After cleavage from the resin, enriched peptides are separated by SCX into two fractions and analyzed by mass spectrometry. **a**, Workflow of the enrichment method. **b**, Structure of the the azide-containing cross-linker bis(succinimidyl)-3-azidomethyl glutarate (BAMG). BAMG adds 151.0 Da to type 1 and type 2 cross-links and 169.1 Da to type 0 coss-links. **c**, ARCO-resin, consists of a poly-dimethylamide solid support, a disulphide as cleavable linker and a cyclooctyne as reactive group towards azides. Via the strain-promoted (3+2) cycloaddition azide-containing peptides are captured on the resin. **d**, Enriched cross-linked peptides. The modification adds 509.2 Da to type 1 and type 2 cross-links and 527.2 Da to type 0 coss-links.
Enrichment of cross-linked peptides

Strong cation exchange enriches type 2 cross-links from cytochrome c

To optimize our enrichment method for cross-linked peptides, strong cation exchange (SCX) was included after the covalent capturing of cross-link modified peptides. SCX can be used to increase the content of type 2 cross-linked peptides in a peptide mixture digested by trypsin\textsuperscript{13-15}. At low pH, these cross-links can accommodate at least four positive charges and thus will bind more strongly to the SCX resin than linear peptides bearing only two positive charges (Figure 2b). Type 2 cross-links, therefore, elute at higher salt concentrations from the column than most linear peptides (Figure 2). SCX alone, however, does not enrich type 2 cross-linked peptides specifically, since (unmodified) peptides with a charge state of more than 3+ will also elute at a higher salt concentration. Therefore, removal of unmodified peptides with use of the ARCO-resin, before SCX chromatography, will result in a much more pure preparation of cross-linked peptides.

To appreciate the value of SCX in the method, cytochrome c was cross-linked by BAMG as a model system. After digestion, BAMG-modified peptides were enriched by the ARCO-resin. The obtained peptides were then separated on an SCX HPLC column, and eluted at two salt concentrations. Mass spectrometric analysis of the two fractions showed that as expected the low salt fraction (50 mM KCl) contained 2+ peptides, mainly type 0 and type 1 cross-links, while the high salt fraction (250 mM KCl) contained the 3+ and 4+ peptides, mainly type 2 cross-links (Figure 2 and Table 1). These results show that the inclusion of SCX facilitates the analysis of cross-linked peptides considerably; a significant enrichment of type 2 cross-links has been reached by separating them from most type 0 and type 1 cross-links and unmodified peptides.

### Table 1, Enrichment by ARCO-resin and SCX of BAMG-modified peptides shows enrichment in type 2 cross-links in the high salt fraction.

Overview of number of modified lysine residues identified after enrichment by the ARCO-resin and SCX of BAMG-cross-linked cytochrome c and NK1. After SCX two fractions are obtained, the low salt fraction of (modified) peptides eluting at 50 mM KCl and the high salt fraction of (modified) peptides eluting at 250 mM KCl. FTMS data is matched with at least 3 ppm accuracy. Number of linked residues confirmed by fragmentation is given between brackets (x).

<table>
<thead>
<tr>
<th></th>
<th>type 0</th>
<th>type 1</th>
<th>type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cytochrome c</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low salt fraction</td>
<td>13 (12)</td>
<td>10 (10)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>high salt fraction</td>
<td>4 (3)</td>
<td>2 (2)</td>
<td>11 (11)</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td>17 (15)</td>
<td>12 (12)</td>
<td>12 (12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>type 0</th>
<th>type 1</th>
<th>type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NK1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low salt fraction</td>
<td>12 (8)</td>
<td>6 (3)</td>
<td>0 (n.a.)</td>
</tr>
<tr>
<td>high salt fraction</td>
<td>8 (6)</td>
<td>1 (0)</td>
<td>17 (8)</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td>20 (14)</td>
<td>7 (3)</td>
<td>17 (8)</td>
</tr>
</tbody>
</table>

Strong cation exchange enriches type 2 cross-links from cytochrome c

To optimize our enrichment method for cross-linked peptides, strong cation exchange (SCX) was included after the covalent capturing of cross-link modified peptides. SCX can be used to increase the content of type 2 cross-linked peptides in a peptide mixture digested by trypsin\textsuperscript{13-15}. At low pH, these cross-links can accommodate at least four positive charges and thus will bind more strongly to the SCX resin than linear peptides bearing only two positive charges (Figure 2b). Type 2 cross-links, therefore, elute at higher salt concentrations from the column than most linear peptides (Figure 2). SCX alone, however, does not enrich type 2 cross-linked peptides specifically, since (unmodified) peptides with a charge state of more than 3+ will also elute at a higher salt concentration. Therefore, removal of unmodified peptides with use of the ARCO-resin, before SCX chromatography, will result in a much more pure preparation of cross-linked peptides.

To appreciate the value of SCX in the method, cytochrome c was cross-linked by BAMG as a model system. After digestion, BAMG-modified peptides were enriched by the ARCO-resin. The obtained peptides were then separated on an SCX HPLC column, and eluted at two salt concentrations. Mass spectrometric analysis of the two fractions showed that as expected the low salt fraction (50 mM KCl) contained 2+ peptides, mainly type 0 and type 1 cross-links, while the high salt fraction (250 mM KCl) contained the 3+ and 4+ peptides, mainly type 2 cross-links (Figure 2 and Table 1). These results show that the inclusion of SCX facilitates the analysis of cross-linked peptides considerably; a significant enrichment of type 2 cross-links has been reached by separating them from most type 0 and type 1 cross-links and unmodified peptides.
Figure 2, Combination of enrichment by the ARCO-resin and SCX allows identification of cross-linked peptides by FT-ICR-MS and Q-TOF-MSMS. After enrichment by the ARCO-resin, cross-link modified peptides are separated by SCX, generating two fractions. The low salt fraction (50 mM KCl) contains 2+ and 3+ charged peptides, comprising type 0 and type 1 cross-links. The high salt fraction (250 mM KCl) contains mainly 4+ charged peptides, consisting of type 0 and type 1 cross-links containing additional basic residues and type 2 cross-links (major component). From the SCX fractions LC-FTMS data is obtained, which is analysed using CoolToolBox, resulting in a list of candidate cross-link modified peptides identified by accurate mass (< 3 ppm). LC-Q-TOF-MSMS analysis of the SCX fractions allows identification of the candidate cross-link modified peptides by fragmentation.
Enrichment of cross-linked peptides

To test the robustness of the enrichment method, the cross-linked peptide mixtures derived from cytochrome c were incubated under different conditions with the ARCO-resin. Enrichment was carried out in i) acetonitrile and phosphate buffer pH 7.5, ii) under acidic conditions, in acetonitrile and 0.1% TFA, and iii) denaturing conditions, in acetonitrile and 3 M guanidine. Analysis of the enriched cross-link containing peptides showed that similar results were obtained for all conditions. All identified 12 type 2 cross-links were detected under the three different conditions. Of the remaining 29 modified peptides, only five species of low abundance (one type 0 cross-link and four type 1 cross-links) were not detected under all conditions. After enrichment also four unmodified peptides were detected in the sample when coupling in phosphate buffer at pH 7.5. However, under acidic and denaturing conditions only two peptides were detected at relatively low intensity (<1.5%). This shows that a high enrichment of BAMG-modified peptides has been obtained under different coupling conditions, underscoring the robustness of the method.

Analysis of mass spectrometric data of enriched cross-linked peptides from cytochrome c

Cross-linked peptides from relatively small proteins can be reliably identified by accurate mass measurement. Comparison of the identified peptides based on accurate mass and our in-house designed software program VIRTUALMSLAB with the fragmentation data, shows that nearly all candidate cross-link modified peptides were confirmed by MS/MS. Out of the 17 type 0 cross-links assigned by VIRTUALMSLAB in the low and high salt fractions, 15 were selected for fragmentation and could be confirmed. For type 1 cross-links, all 12 masses were selected and could be assigned correctly. For type 2 cross-links, similar results were obtained. Eleven masses could be assigned to a single type 2 cross-link,

Figure 2, continued. a, UV absorbance trace of an SCX run of enriched cross-linked cytochrome c. After loading the sample, a two-step gradient (50 mM KCl and 250 mM KCl) is used to separate type 2 cross-links from the other modified peptides. b, Overview of expected products in SCX fractions after enrichment. 0+: unreacted cyclooctyne is uncharged at pH 2.9 and will not bind to the SCX column, eluting in the void volume. 2+, 3+: type 0 and type 1 cross-links. 3+, 4+: type 0, type 1 and type 2 cross-links (majority). c, Total Ion Count of LC-FTMS run of the high salt fraction. d, FTMS spectrum at t = 23.0 min. e, Q-TOF-MSMS spectrum of a type 2 cross-link at m/z 1013.8543 (= 3039.54 Da), with the sequences (a) MIFAGIKK and (b) EETLMYENPKK. f, Overview of observed and assigned fragments in peptide sequence. y*: loss of the cyclooctyne with the linker. ▲: cyclooctyne with linker at m/z 402.2.

Enrichment by ARCO-resin under different conditions is robust
which all were confirmed. For one mass, two different type 2 cross-links could be matched within 3 ppm, after which fragmentation showed one of them to be correct. These results show that the identification of cross-linked peptides by high accuracy mass data has great potential and can be used for a first analysis of cross-linked protein samples.

However, fragmentation will be necessary for the identification of some type 2 cross-links: residues involved in the cross-link or the difference between type 0 and type 2 cross-links can sometimes not be made by accurate mass. Furthermore, for larger systems the number of possibilities for type 2 cross-links will increase exponentially with the number of lysine residues and thereby the number of false positives. This makes the analysis of the data solely based on accurate mass practically impossible. Therefore, the combination of high mass accuracy and fragmentation will be very useful: the first limits the possibilities for a given mass, when analyzing the acquired data from the second.

Figure 3, Fragmentation of enriched type 0 cross-links shows a distinctive reporter ion at m/z 528.2. a, Overview of observed and assigned fragments of type 0 cross-link observed in NK1 with the sequence EFGHEFDLYENK(mod)DYIR. b, Fragmentation spectrum of the type 0 cross-link at m/z 2602.20. y*: loss of the cyclooctyne with the linker. ▲: cyclooctyne with linker at m/z 402.2. ▼: further fragmentation of the cyclooctyne gives rise to a benzoylium ion with 268.1 Da. ■: reporter ion at m/z 528.2, consisting of the cyclooctyne, linker and cross-linker. ♦: loss of reporter ion.
Enrichment of cross-linked peptides

Fragmentation behavior of BAMG-modified peptides

As discussed in previous section, cross-linked peptides from relatively small proteins can be reliably identified at the high mass accuracy (<3 ppm) achieved under our experimental conditions. However, for the applicability of our technique to more complex systems, yielding more than one candidate cross-link for a measured mass, high quality MS/MS data are required for unambiguous identification of the peptides involved in the cross-link. For the type 2 cross-links isolated via our enrichment procedure, information-rich CID fragmentation spectra were obtained as exemplified in Figure 2e and f. The observation of dominant y- and b-ion series in most spectra suggests that our approach enables high throughput cross-link analysis with software programs calculating a score to indicate the chance that hits are the result of random events.15, 17-21

In addition to the backbone fragmentation of both peptides, two characteristic reporter ions can be observed. These ions of the linker at m/z 402.2 and m/z 268.2 and the loss of 401.2 are specific to ARCO-enriched peptides and often observed in fragmentation spectra, enabling discrimination between enriched peptides and those that are nonspecifically isolated.11

For type 0 cross-linked peptides another reporter-ion is observed, which enables distinction between type 0 and type 2 cross-links. Inspection of 55 MS/MS spectra of type 0 cross-links, from both cytochrome c and NK1, revealed the presence of an ion at m/z 528.2 in ~80% of the cases (Figure 3). This ion originates from dissociation of the amide bond of the lysine epsilon-amine with the BAMG-cross-linker, and corresponds to the mass of the cyclooctyne-linker, attached to the BAMG-cross-linker, which is hydrolyzed on one side yielding a free carboxylic acid moiety. For the other two types of cross-links an ion at m/z 528.2 was observed in less than 3% of all analyzed spectra. This renders observation of this ion useful; spectra containing the reporter ion signal can be excluded from further study if only type 1 and type 2 cross-links are of interest. So, in addition to SCX, reporter ion analysis will improve efficiency of mass spectrometric analysis of type 2 cross-links even further.

Cross-links in cytochrome c

The cross-linking with BAMG is carried out under very mild conditions; on average one to two cross-links are introduced per cytochrome c molecule.6 This assures both efficient digestion of the cross-linked protein and limits the possibility of changing the protein structure, potentially introduced by cross-linking reagents. Yet, the large majority of the 19 lysine residues in cytochrome c is found to be...
modified by BAMG (Figure 4a). Apart from modification by the BAMG cross-linker of the lysine residues, it was found that the methionine residues in the cytochrome \( c \) preparation were partially oxidized. In one example, a cross-link between two peptides, each containing a methionine, was identified with none of the methionine residues oxidized, one of the residues oxidized and both residues modified by oxidation. These observations show that the combination of cross-linking with BAMG and our enrichment method is sensitive and comprehensive.

The structural information obtained from cross-linking is derived from the distance spanned by the cross-linker used. BAMG spans maximally 20 Ångstrom, measured between the alpha carbon atoms of the lysine residues. This distance can only be reached when the lysine residue side chains are maximally stretched on the protein surface. As most lysine side chains will be prevented from maximal stretching by sterical constraints, it is expected to find most of the cross-links to span a shorter distance. Analysis of the distances between the alpha carbon atoms of type 1 and 2 cross-links in the structure of horse cytochrome \( c \), showed that the expected distance of 20 Ångstrom is not exceeded and that no cross-links further apart than 17.5 Ångstrom are observed (Figure 4c).

![Figure 4](image_url)

**Figure 4**, Residues involved in cross-linking show the method to be sensitive. Overview of residues involved in cross-linking with BAMG. a, Cytochrome \( c \) contains 19 lysine residues. The N-terminus is acetylated and is not involved in cross-linking. Numbering as in sequence (pdb file: 1AKK). b, NK1 contains 21 lysine residues. The N-terminus is involved in cross-linking. Numbering as in sequence (pdb file: 1GP9). \( \text{K}: \text{Lysine. Y: Tyrosine. Black: Confirmed by Q-TOF-MSMS. Dashed: Identified by FTMS. a and d, Distribution of theoretical (grey) and experimental (black) found type 1 and 2 cross-links, as the distance between lysine residues in crystal structure of c, horse cytochrome \( c \) (pdb file: 1AKK) and d, NK1 (pdb file: 1GP9). For NK1 19 out of the 21 lysine residues are resolved in the crystal structure 1GP9. Distance is measured between the alpha carbons of lysine residues and is through space. The maximal distance spanned by BAMG is 20 Ångstrom.)
Out of the 70 theoretical type 1 and 2 cross-links that are less than 17.5 Ångstrom apart from each other, 31% is found back in our dataset. This is comparable to the results of Lee et al., who found 33% of the lysine pairs within 20 Ångstrom as cross-link\textsuperscript{37}, using a cross-linker with a 3.5 Ångstrom longer spacer, a different fractionation scheme and different mass analysis equipment.

Cross-links in NK1

To test the applicability of the method to other proteins, the splice variant of hepatocyte growth factor/scatter factor (HGF/SF) – NK1 – was chosen. This protein is a potential antagonist for the receptor Met, an oncogenic protein involved in different types of cancer\textsuperscript{22-24}. NK1 is 21 kDa in size, contains 20 lysine residues and consists of two domains. It behaves as a monomer in solution, but forms a dimer in the crystal structures\textsuperscript{25-27} as well as in solution in the presence of heparin\textsuperscript{28, 29}. Still, the orientation of the two domains in solution is not known and the mechanism by which NK1 binds the Met receptor is not yet completely understood\textsuperscript{30}. It has been proposed that NK1 changes from a dimeric structure to a monomeric structure via “domain swapping”\textsuperscript{27}, such that the interactions between the protomers in the dimer are replaced by equivalent interdomain interactions within the monomer. Firstly, hinge movement between the N and K1 domains has been observed. Secondly, strong interactions between the protomers in the dimer are lacking between the domains within the protein. Chemical cross-linking might aid in the study of the solution structure of the NK1 protein and thereby gain more insight in its function. NK1 has been studied before by chemical cross-linking; however, at that time only four type 2 cross-links could be identified\textsuperscript{16}.

As for cytochrome c, nearly all lysine residues of NK1 are involved in cross-links (Figure 4b). From the FTMS analysis, a list of 17 unique candidates for type 2 cross-links was obtained, that were detected both fully alkylated and without alkylation of the linker (see Material and methods). Four of these cross-links comprise the N-terminus, which is not resolved in the crystal structures. In fragmentation experiments eight of the 17 type 2 cross-links were selected for fragmentation and the proposed structure of all of them could be confirmed (Table 1).

When locating the linked lysine residues involved in type 1 and 2 cross-links in the crystal structures, it appears that most are found within one of the two domains of NK1. This is reflected in the distances spanned by the linked residues: for 19 out of the 21 cross-links that can be mapped in the crystal structures, distances
Figure 5, Fragmentation spectrum of enriched type 2 cross-link derived from NK1 that suggests the protein can have a different structure in solution compared to the available crystal structures. The type 2 cross-link depicted in a - b points to a different structure of the protein in solution compared to the available crystal structures. The alpha carbon atoms of the linked lysine residues within one protomer in the dimeric crystal structures (PDB-files: 1GP9, 1BHT, 1NK1, 2QJ2) are separated by 25.4 to 26.3 Ångstrom, largely exceeding the maximal distance of 20 Å that can be spanned by the cross-linker BAMG.
Enrichment of cross-linked peptides

Figure 5, continued. This indicates that the monomeric structure in solution differs from the structure of the protomers in the crystal dimers. In c – f, the fragmentation spectra of two type 2 cross-links comprising both one of the peptides of the type 2 cross-link in a – b support the modification of these lysine residues and the fragmentational behavior these peptides. a, Fragmentation spectrum of the type 2 cross-link at m/z 2484.41 b, Overview of observed and assigned fragments of type 2 cross-link observed in NK1 with the sequences (a) TCK(mod)AFVFDK and (b) NCIIGK(mod)GR. c, Fragmentation spectrum of the type 2 cross-link at m/z 2707.36. d, Overview of observed and assigned fragments of type 2 cross-link observed in NK1 with the sequences (a) TCK(mod)AFVFDK and (b) SYK(mod)GTVSITK. e, Fragmentation spectrum of the type 2 cross-link at m/z 2177.22. f, Overview of observed and assigned fragments of type 2 cross-link observed in NK1 with the sequences (a) HN(mod)YVEGQR and (b) NCIIGK(mod)GR. In these cross-links, the cysteine residues were found to be alkylated. y*: loss of the cyclooctyne with the linker. ▲: cyclooctyne with linker at m/z 402.2. ◊: Deconvolution artefacts of multiple charged ions observed after processing of the data with Masslynx Proteinlynx software.

of less than 17.5 Ångstrom are spanned, 26% of the total number of theoretical type 1 and 2 cross-links possible (Figure 4d). One of the cross-links exceeding the maximal distance originates from a mass that can be explained by two type 2 cross-links, spanning a distance of either 4.9 or 19.6 Ångstrom. By accurate mass no distinction can be made and fragmentation is needed to confirm one of both.

The other type 2 cross-link that does not obey the rule of 17.5 Ångstrom was assigned by accurate mass (1 ppm accuracy) both fully alkylated and with unalkylated linker. Furthermore, the structure of this type 2 cross-link has been confirmed by fragmentation (Figure 5). The cross-link spans a distance of almost 26 Ångstrom, is found between two peptides with lysine residues K85 and K132, located in the two different domains of the protein, and cannot be fitted in the crystal structures. The cross-linking has been carried out in solution in absence of heparin and no dimer has been formed, as judged by SDS-PAGE and in agreement with previous cross-linking experiments23, 26. This implies that NK1 indeed can adopt a structure in solution different from the one observed in the crystal structures27. The cross-link found between K85 and K132 can best be explained by assuming that the protein has adopted a “closed” conformation at the moment of connection of these two lysine residues.

Cross-linking and enrichment of cysteine containing proteins

NK1 is a two domain protein of 21 kDa, containing 10 cysteine residues, which are all involved in disulfide bridges. To facilitate digestion, the BAMG-cross-linked protein was subjected briefly to treatment with a relatively low concentration of TCEP to (partially) reduce the five disulfide bonds, while at the same time preventing reduction of the azide or even cleavage of the cross-linker6, 31, followed by alkylation of free SH-groups. To determine the extent of reduction
of the disulfide bonds in the protein, analysis was carried out with and without alkylation after TCEP-induced release from the resin. Under these conditions, reduction of possible remaining disulfide linked peptides – enriched via the resin through cross-link modified peptides – will occur as well. In the absence of the second alkylation step both the linker and peptides, which survived the first reduction, bear a free thiol group.

Results show that indeed reduction of disulfide bonds in the protein was only partial. However, this did not prevent efficient digestion since we found practically all lysines to participate in one or more of the three types of cross-links. Moreover, similar results were obtained with and without alkylation after release from the ARCO beads. For data analysis, though, inclusion of an alkylation step after enrichment is preferable because all cysteine residues become alkylated after the complete reduction, yielding single products, which reduces the size of the database.

Further improvement

Further improvement of the method for more complex samples could be achieved by collecting multiple fractions during the SCX separation, instead of separating the sample in two fractions as described here. Furthermore, the use of a curved gradient as described recently instead of our two step gradient should gain further in sensitivity.

A second improvement of the method relates to the solubility of the enriched peptides. Recent experiments have shown that avoiding sample drying during workup, and keeping the enriched peptides in a solution containing at least 10% acetonitrile, increases the yield considerably, only an equivalent of 4 μg cross-linked NK1 being sufficient for analysis. It is of interest though that no additional cross-links were detected with the modified procedure suggesting that loss of material by drying is not selective (results not shown).

Conclusion

We have described a robust procedure for the enrichment of azide-containing peptides for cross-linked samples. The enrichment is reproducible under different coupling conditions, making it a suitable method for different systems, and the inclusion of SCX makes this to the best of our knowledge, the first method able to enrich type 2 cross-links selectively. Fragmentation of enriched cross-links give rise to good quality spectra and the observation of an m/z 528.2 reporter ion for type 0 cross-links makes it possible to discriminate between the different cross-
link modified species, aiding in the data analysis of cross-linked protein samples. The number of cross-links identified in this study and by others suggests that in general only about one third of all theoretical cross-links in a protein (or complex) are observed. This should be taken into account when considering a cross-link approach for 3D structure validation of a protein (complex). Structural information for two small proteins (cytochrome c and NK1) has been obtained by our method and conditions have been optimized that will enable the leap to cross-link analysis of more complicated systems.

Acknowledgement

We thank the master students Niels Nouse and Pascal van Alphen for their contributions to the cross-linking studies of respectively cytochrome c and NK1. We thank Sophie Timsit for initial experiments and Hansuk Buncherd and Winfried Roseboom for experimental support. We are grateful to Professor Ermanno Gherardi (Medical Research Council Laboratory of Molecular Biology, Cambridge, UK) for his contributions to the NK1 work.
Material and Methods

Synthesis of the azide reactive cyclooctyne (ARCO)-resin

The ARCO-resin was synthesized as described by Nessen et al.\textsuperscript{11}.

Preparation of BAMG-cross-linked cytochrome c digest

A tryptic digest of BAMG-cross-linked cytochrome c was prepared as described by Kasper et al.\textsuperscript{6}. The obtained digest in 100 mM ammonium bicarbonate was desalted on C18 reverse phase pipette tips, eluting in 75% acetonitrile.

Preparation of BAMG-cross-linked NK1 digest

A tryptic digest of BAMG-cross-linked NK1 was prepared similar to cytochrome c. NK1 (7 µM in 50 mM sodium phosphate buffer pH 7.5 with 100 mM NaCl) was cross-linked with 0.1 mM BAMG. After cross-linking and removal of the reagents with a Biomax filter (10 kDa) as described before, the protein was incubated with TCEP (2 mM) at room temperature for 15 minutes. Then NaN\textsubscript{3} (5 mM) was added, followed by iodocetamide (30 mM) and after 30 minutes of incubation at room temperature in the dark, reagents were removed by washing the cross-linked protein twice with 50 mM potassium phosphate buffer pH 7.5. After concentration of the protein, the mixture was transferred to a Lo-Bind Eppendorf tube and trypsin (Trypsin Gold (Promega, Madison, WI, USA)) was added (1:20 w/w) for overnight digestion. The digest was desalted on C18 reversed phase pipette tips, eluting in 50% acetonitrile, 0.1% TFA.

Enrichment of cross-linker modified peptides by the ARCO-resin

Cross-link-modified peptides were enriched by the ARCO-resin as described by Nessen et al.\textsuperscript{11}. For cytochrome c, 125 µg BAMG cross-linked protein digest in 75 µl was added to ~2 mg of dry ARCO-resin in a final solution of either i) 25 mM potassium phosphate buffer pH 7.5 in 50% acetonitrile, ii) 0.1% TFA, pH 2 in 50% acetonitrile or iii) 3 M guanidine hydrochloride in 25% acetonitrile. After incubation at 40 °C for 24 h, the samples were treated as described before\textsuperscript{11}.

For NK1, 150 µg of BAMG cross-linked protein digest in 75 µl was added to ~2 mg of dry ARCO-resin in 25 mM potassium phosphate buffer pH 7.5 in 50% acetonitrile. After washing of the resin and release of captured peptides, the free thiol was either alkylated by addition of 55 mM iodoacetamide (final concentration) or no alkylation was carried out and samples were treated as described before.

Enrichment of cross-linked peptides by strong cation exchange (SCX)

All material recovered from the enrichment by the ARCO-resin was loaded in 1 mL of a solution containing 10 mM potassium phosphate buffer, pH 2.9 in 20% acetonitrile on a PolySULPHOETHYL Aspartamide\textsuperscript{TM} Column (2.1mm ID, 10 cm length) (PolyLC Inc., Columbia, USA) operated on an Ultimate HPLC system (LC Packings, Amsterdam, The Netherlands). Elution, at a flow rate of 0.1 ml min\textsuperscript{-1}, was performed using a two-step isocratic program which consisted of a first elution step of 50 mM KCl in a buffer containing 10 mM potassium phosphate buffer, pH 2.9 in 20% acetonitrile, for 20 min (low salt) followed by a second elution step with 250 mM KCl for 20 min (high salt) and finally, a linear gradient to 500 mM KCl. Forty fractions of 100 µl were collected. Fractions showing absorbance at 214 nm were pooled within in their respective elution step and selected for further analysis. This resulted in two fractions which were named low salt and high salt. For cytochrome c, 0.1% TFA was added to obtain a final acetonitrile concentration of less than 10%. For NK1, the acetonitrile was evaporated in a vacuum centrifuge. Then, fractions were subsequently desalted on C18 reversed phase pipette tips.
Enrichment of cross-linked peptides

Mass spectrometry

LC-FT-MS

Accurate mass data were acquired using an ApexQ Fourier transform ion cyclotron resonance mass spectrometer (Bruker Daltonic, Bremen, Germany) equipped with a 7T magnet and a CombiSource™ coupled to an Ultimate 3000 (Dionex, Sunnyvale, CA, USA) HPLC system with a 100 µm ID, 150 mm monolithic reverse phase column (Onyx C18, Phenomenex, Torrance, USA). Samples were injected as a 3 µl 0.1% TFA aqueous solution and directly loaded onto the analytical column. About 15% (cytochrome c) or 30% (NK1) of the material was injected. Following injection, a linear gradient was applied in 30 min from 0.1% FA to 0.1% FA and 40% acetonitrile at a flow rate of 2 µl min⁻¹. During elution a chromatogram of 550 high resolution ESI FTMS spectra was recorded using a MS duty cycle of about 3 seconds.

LC-Q-TOF-MS/MS

Electrospray MS and low energy collision-induced dissociation (MS/MS) analyses were performed on a Q-TOF (Micromass, Whytenshawe, United Kingdom) mass spectrometer with a Z-Spray orthogonal ESI source. For LC prior to MS, peptide mixtures were loaded onto an Ultimate nano HPLC-system (LC Packings) and separated on a PepMap100 C18 reversed phase column (75 µm I.D., 25 cm length; Dionex, Sunnyvale, CA, USA). About 30% of the material was transferred to an analysis vial and the solvent was evaporated with a miVac DNA concentrator high performance centrifugal evaporator system (GeneVac, Ipswich, United Kingdom) after which the peptide mixture was redissolved in a 10 µl solution of 5% ACN and 0.1% TFA. After loading on a trap column (Acclaim, PepMap 100, 300 µm i.d., 5 mm length, Dionex, Sunnyvale, CA, United States of America), elution of the peptides, with a flow rate of 0.3 µl min⁻¹, was performed using a step-gradient as described by Nessen et al.¹¹. Direct infusion of the flow was supported by a Nanobore Emitter (Proxeon, Odense, Denmark). Survey scans were acquired from m/z 350-1,500. Most intense ions were selected for MS/MS in a data-dependent mode, recorded from m/z 50-2,500, scan time 1.00 s, interscan delay 0.10 s with argon as collision gas at a pressure of 0.04 mbar measured on the quadrupole pressure gauge. Fragmentation spectra were processed with Masslynx Proteinlynx software, generating peak list files (pkl).

Data processing and analysis

FT-ICR data were processed using the Data Analysis 3.4 software program (Bruker Daltonic, Bremen, Germany). In a batch procedure mass spectra were extracted from the chromatogram and the monoisotopic masses of the peptides were determined using Brukers peak recognition technology SNAP II™. Mass calibration was achieved by selectively extracting and then summing about 6 mass spectra from the chromatogram showing calibrant ions. The summed spectrum was mass calibrated and the resulting calibration parameters were applied to all spectra in the chromatogram. This resulted in a mass calibration of better than 1.5 ppm over the entire chromatogram for all analyses. For each FTMS analyses the resulting array of monoisotopic mass lists was exported as a MASCOT generic file (mgf). Ion abundances in the exported array of monoisotopic mass lists were the sum of abundances of all isotopes over all charge states for each peptide. The exported mgf file was imported in the CoolToolBox software program which is a major update of the in-house developed VIRTUALMSLAB program¹⁶,³².

From the imported array monoisotopic mass spectra the CoolToolBox program constructed peptide ion chromatograms, for which the mass and retention time was taken at the apex of the chromatogram profile and the abundance was summed over the chromatogram profile. The
ultimate LC-MS data processing resulted in a peptide monoisotopic mass list with abundance and LC retention. The processed LC-FTMS data were matched with the CoolToolBox generated virtual experiment peptide database\textsuperscript{16}.

The virtual experiment database generated in CoolToolBox consisted of i) unmodified tryptic peptides, ii) cross-linked peptides, for which each peptide contained one internal lysine residue, iii) loop-linked peptides, containing up to three internal lysine residues, and iv) mono-linked peptides, containing up to two internal lysine residues.

For cytochrome \textit{c}, peptides in the reference database ii), iii) and iv) were modified by BAMG, the cyclooctyne, linker and iodoacetamide. All FTMS data with an intensity of at least 0.75% was matched with an accuracy of 3 ppm against each database separately. To determine false positive identifications, the experimental MS data was matched to a scrambled database, derived from a scrambled sequence of cytochrome \textit{c}, which was repeated ten times. The average number of false cross-links was then divided by the number of real cross-links, obtaining the false positive ratio. For the different types of cross-links false positive ratios of less than 6% (0-6%) were obtained, with exception for the type 2 cross-links in the low salt fraction, with a false positive ratio of 30% (0.3 false cross-links on average divided by 1 identified cross-link). Acquired fragmentation spectra were inspected and assigned manually.

For NK1, cysteine residues were partially alkylated and the N-terminus was added as a group to be modified by the cross-linker. Data was matched against databases containing peptides modified by BAMG, the cyclooctyne, linker with and without alkylation. FTMS data with an intensity of at least 0.25% was matched with an accuracy of 3 ppm against each database separately. Only cross-links found in both experiments (with and without alkylation) were considered for further analysis. For type 2 cross-links a false positive ratio of 2% was obtained. Acquired fragmentation spectra were inspected and assigned manually.

References

Enrichment of cross-linked peptides


