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):
Buncherd, H. (2013). Development of enrichment methods for cross-linked peptides to study the dynamic topology of large protein complexes by mass spectrometry

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: http://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 (http://dare.uva.nl)
Selective enrichment and identification of cross-linked peptides to study 3-D structures of protein complexes by mass spectrometry

Abstract

Chemical cross-linking of protein complexes combined with mass spectrometry is a powerful approach to obtain 3-D structural information by revealing amino residues that are in close spatial proximity. To increase the efficiency of mass spectrometric analysis, we have demonstrated the selective enrichment of cross-linked peptides from the 350 kDa protein complex RNA polymerase (RNAP) from Bacillus subtilis. Bis(succinimidyl)-3-azidomethyl glutarate was used as a cross-linker along with an azide-reactive cyclooctyne-conjugated resin to capture target peptides. Subsequently released peptides were fractionated by strong cation exchange chromatography and subjected to LC-MS1MS2. We mapped 10 different intersubunit and 24 intrasubunit cross-links by xComb database searching supplied with stringent criteria for confirmation of the proposed structure of candidate cross-linked peptides. The cross-links fit into a homology model of RNAP. Cross-links between β lobe 1 and the β′ downstream jaw, and cross-links involving the N-terminal and C-terminal parts of the α subunits suggest conformational flexibility. The analytical strategy presented here can be applied to map protein–protein interactions at the amino acid level in biological assemblies of similar complexity. Our approach enables the exploration of alternative peptide fragmentation techniques that may further facilitate cross-link analysis.

**Introduction**

Knowledge about the tertiary and quaternary structure of a protein complex is important to understand its action on the molecular level. Common techniques to obtain high resolution structures of proteins include X-ray diffraction and nuclear magnetic resonance (NMR). However, not all proteins are suitable for study 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 placed within the structure. To validate and optimize the models, distance constraints within a protein or between proteins can be obtained by chemical cross-linking and mass spectrometry [1-8]. Also results from information-driven docking approaches can be experimentally validated with cross-linking [9]. The most commonly used cross-linkers are bifunctional N-hydroxy-succinimidyl (NHS) esters that are reactive towards the primary amines in a protein, as present in lysine side chains and N-termini. 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), three types of cross-linker-modified peptides are expected in the digest, along with unmodified linear peptides, i) mono-links or type 0 cross-links, modification of peptides by a partially hydrolyzed cross-linker, ii) loop-links or type 1 cross-links, a cross-link within one peptide and iii) cross-links or type 2 cross-links, a cross-link between two different peptides [10, 11]. Mainly type 2 cross-linked peptides provide powerful 3-D structural information, especially those containing an intersubunit cross-link. However, type 2 cross-linked peptides are usually both rare, and of low abundance, hampering efficient mass spectrometric identification.

Recently, several approaches have been described for enrichment of type 2 cross-linked peptides to facilitate mass spectrometric analysis. Enrichment of target peptides after digestion of a cross-linked protein by trypsin digestion can be achieved by strong cation exchange chromatography at low pH [11, 12]. However, separation from the other peptides is only partial. A new amine specific cross-linker, diethyl suberthioimidate, enables a more effective enrichment of type 2 cross-links based on strong cation exchange chromatography [13, 14]. However, the cross-link efficiency of diethyl suberthioimidate is low, implying the necessity of reaction conditions that may be unfavourable for labile protein complexes. Enrichment approaches using covalent capturing to a thiol reactive resin [15], affinity chromatography [16-20] combined with the use of a cleavable cross-linker [21] or diagonal chromatography combined with a cleavable cross-linker [22] have up to now only been applied to relatively small proteins. Another cleavable cross-linker provided with a biotin moiety for affinity purification has a very long spacer [23, 24] and, therefore, is less suitable than short cross-linkers for obtaining distance constraints to validate detailed 3-D models.

The limitations of existing methodology to enrich the low abundant information rich type 2 cross-links suggests that their mass spectrometric detection and structural elucidation would be tremendously facilitated by a more effective isolation procedure. We recently
presented a solid phase enrichment method that uses the azide-reactive cyclooctyne (ARCO) resin. Peptides containing an azido group, including peptides modified by an azide-containing N-hydroxysuccinimide cross-linker, could be specifically isolated from a protein digest via a selective covalent linkage to the ARCO resin by the strain-promoted azide–alkyne cycloaddition reaction [25]. Based on this approach, we describe here the development of a workflow for the isolation and identification of type 2 cross-links from a multisubunit enzyme complex, RNA polymerase (RNAP) from Bacillus subtilis.

**Results**

**Choice of the ARCO resin**

The aim of this work is to assess the suitability of a solid phase copper-free click chemistry reaction for selective enrichment and efficient identification of cross-linked peptides from large protein complexes, according to the scheme depicted in Fig. 1A. Besides the ARCO resin, we tested the usefulness of two other cyclooctynes, an azacyclooctyne and a monofluorocyclooctyne coupled to polydimethylacrylamide beads (supplemental results, supplemental Fig. S1).

![Fig. 1](image_url) Overview of the 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 fractionated by SCX chromatography and analyzed by mass spectrometry. A, workflow of the enrichment method. B, structure of 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 cross-links. C, ARCO-resin, consisting of a polydimethylacrylamide solid support, a disulphide as cleavable linker and a cyclooctyne as reactive group towards azides. Via the strain-promoted azide–
alkyne 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 cross-links.

The xComb/Mascot approach to identify cross-linked peptides

Mapping of cross-links in large protein complexes, like RNAP, requires calculation of a database of all possible cross-linked peptides and a search engine to identify candidate peptides based on tandem mass spectrometric data. For database generation we use the software tool xComb [26]. In the xComb-created library, each pair of peptides in a cross-link is represented by a linear structure. The database is interrogated with MS$_1$MS$_2$ data from a cross-linked protein digest by a search engine for peptide identification. Coupling of the cross-linker, hydrolyzed at one side, to a lysine residue is included as a variable modification, resulting in equal masses of target peptides and database entries. A candidate cross-linked peptide may be nominated based on the partial resemblance of experimental mass data and the calculated list of fragment ions of its linearized counterpart [26, 27]. xComb has been made freely accessible via the Internet [26]. It can be combined with several search engines for peptide identification. Here we use Mascot as a search engine.

Effectiveness of the xComb/Mascot approach to identify cross-linked peptides isolated by capturing to the ARCO-resin

To test the effectiveness of the xComb/Mascot strategy for identification of BAMG-cross-linked peptides modified with the cyclooctyne–triazole-spacer moiety, a peptide sample containing cross-links of known identity is required. For this purpose we used a tryptic digest of BAMG-treated cytochrome c subjected to the ARCO resin-mediated enrichment procedure depicted in Fig. 1. The combination of capturing by the ARCO-resin and SCX chromatography of released peptides leads to a high enrichment of type 2 cross-links (supplemental results, supplemental Table S1, supplemental Figs. S2 and S3). The small size of cytochrome c enables identification of most type 2 cross-links based on accurate mass measurement of peptides. Matching the experimental accurate mass data with a calculated database of type 2 cross-linked peptides resulted in the nomination of 11 unique candidates and 2 candidates for a peptide at 2221.23 m/z, with cross-links between K73 and K86 and between K79 and K87 respectively (supplemental Table S1). LC-MS$_1$MS$_2$ experiments amply confirmed the identity of the 11 unique candidates, including a number of variants due to partial oxidation of methionine residues, as well as that of the K73–K86 cross-link and provided no evidence for the existence of a cross-linked peptide connecting K79 and K87 (supplemental Table S2 and Supplementary materials). These results show that BAMG-cross-linked peptides can be highly enriched and that the enriched peptides modified with a triazole–cyclooctynes-spacer moiety can yield informative fragmentation spectra, enabling unambiguous identification.

To test the xComb/Mascot strategy, a database was calculated from a list of proteins composed of cytochrome c and the four subunits of B. subtilis RNA polymerase. Upon searching the database with MS$_1$MS$_2$ data, all cross-linked peptides, identified as described in the previous paragraph, were repeatedly nominated by Mascot as candidates in three different LC- MS$_1$MS$_2$ runs. (supplemental Table S2). Also a number of false positives were put
forward, but these could be discarded by applying additional criteria with respect to the number of fragment ions from both peptide moieties required for confirmation, as formulated in Materials and Methods (supplemental results, supplemental Table S2). These results provided a firm basis to explore the ARCO-beads-based enrichment combined with the xComb/Mascot approach for mapping cross-links in more complex systems.

**Identification of cross-linked peptides from RNA polymerase**

As a model system for cross-link analysis of large biological assemblies we study RNA polymerase (RNAP) from Bacillus subtilis. This is a useful and representative model system for the development of cross-link analysis strategies for several reasons. It is a relatively large assembly of 345 kDa, built up from 4 different subunits, α (34.8 kDa), β (133.6 kDa), β’ (134.2 kDa) and ω (7.7 kDa) assembled in a 2:1:1:1 stoichiometry. The complex contains 195 lysine residues, corresponding to 6.3% of the total number of residues, i.e., a percentage close to the proteome wide occurrence of this amino acid. Furthermore, *B. subtilis* RNAP can be conveniently expressed in *Escherichia coli* (30). A homology model [28] based on the crystal structure of *Thermus aquaticus* RNAP [29] is available.

Recombinant *B. subtilis* RNAP, C-terminally tagged with nine consecutive histidine residues, was obtained by overexpression in *E. coli* and purified as described before [30]. The enzyme was subjected to cross-linking with BAMG at the relatively low cross-linker concentration of 0.2 mM. The effect of cross-linking on the formation of intersubunit cross-linking was studied by SDS PAGE (Fig. 2). For these studies DSG was chosen as a cross-linker since it has the same cross-link efficiency as BAMG [22], while DSG-induced cross-links, in contrast to BAMG-induced cross-links [22, 31], are stable under conditions of SDS PAGE in the presence of β-mercaptoethanol. Under the chosen cross-link conditions the amount of protein in the β/β’ band decreased by approx. 50%, while material of lower mobility appeared on the gel, indicating that intersubunit cross-links had formed (Fig. 2). Also the α subunits were converted to material of lower mobility. After digestion, the cross-linked peptides were enriched via the ARCO-resin and subjected to fractionation by SCX chromatography according to the procedure depicted in Fig. 1. The chromatogram obtained after SCX chromatography of released material from the ARCO beads is depicted in supplemental Fig. S4. Mass spectrometric analysis of SCX fractions with LC-MS<sub>1</sub>MS<sub>2</sub> revealed the presence of 46 unmodified peptides, 112 type 0 cross-links and 20 type 1 cross-links (supplemental Table S3).
Fig. 2 SDS-PAGE of RNA polymerase before (lane A, 1.25 μg RNAP) and after cross-linking with 0.2 mM DSG (lane B, 1.25 μg; lane C, 2.5 μg, and lane D, 5 μg). The β and β' subunit migrate in the same band. Material of low mobility represents cross-linked subunits. At the low cross-linker concentration used, and the mild reaction conditions (2 hr. on ice) roughly 50% of β/β' had not yet formed intersubunit cross-links. These conditions were chosen to limit the risks of cross-linker-induced protease resistance and conformational changes. The ω subunit (8 kDa) is outside the separation area of the gel.

Sixty eight number one candidate type 2 cross-linked peptides for MS$_1$MS$_2$ spectra from 64 different precursor m/z values were nominated by Mascot using the xComb database created as described in Materials and Methods. More than 90% of type 2 cross-link candidates were found in SCX fractions 6 to 10, while approx. 70% of the unmodified peptides and type 0 and type 1 cross-links eluted in the first 5 SCX fractions (supplemental Table S3). This underscores the usefulness of SCX chromatography to further enrich type 2 cross-links after capturing to and release from the ARCO resin.

MS$_1$MS$_2$ spectra of part of the candidates contained only fragments from one peptide. The lack of sequence information from one of the two peptides involved in a cross-link prevents distinction between a true positive and a false positive, the latter being possibly the result, for instance, of unknown experimentally introduced chemical changes, naturally occurring post-translational modifications or cross-linking at other residues than lysine or the N-terminal amino acid [32]. Thirty six candidate cross-linked peptides, comprising 34 different cross-links, remained upon subjection of MS$_1$MS$_2$ spectra to additional criteria with respect to the presence of unambiguous fragment ions from both peptides (supplemental Table S4). Fig. 3 shows the effect of this filtering on the distribution of candidate cross-links in relation to the distances between C$_\alpha$ atoms of connected lysine pairs in the model of RNAP, as far as the cross-links are not located in unresolved areas in the crystal structure of the template used for modeling. Before filtering a number of candidates had distances that exceed the span of the cross-linker by tens of Ångstroms. Such candidates may either be in conflict with the model or be false positives. After applying the criteria for confirmation, only candidates with cross-link distances less than 30 Å remained (Fig. 3, supplemental Table S4), although the identity of still 12 candidates in this group nominated by Mascot could not be confirmed by the applied criteria. The complete disappearance of candidate cross-links spanning distances exceeding 30
Å upon an increase in the stringency of criteria for confirmation suggests that true positives are confined to the group of cross-links spanning distances less than 30 Å.

Fig. 3 Distribution of type 2 cross-links in RNA polymerase as a function of the distance between the C\(_\alpha\) atoms of linked residues. White bars, all possible K–K linkages expressed as percentage of the total number. Grey bars, number of cross-linked peptides nominated by Mascot upon screening an xComb database with tandem mass spectrometry data. Black bars, identified cross-linked peptides after application of criteria for the presence of MS\(_1\)MS\(_2\) signals from both peptides in the cross-link pairs.

**Reporter ions detected in MS\(_1\)MS\(_2\) spectra of cross-linked peptides**

It is noteworthy that most MS\(_1\)MS\(_2\) spectra of type 0, type 1 and type 2 cross-links revealed the presence of an \(m/z\) 402.2 fragment ion resulting from cleavage of the C–N bond between the spacer of the cross-linker and the triazole–cyclooctyne moiety. Also neutral loss of 401.2 Da from the precursor ion due to the same cleavage reaction is frequently observed. Further fragmentation of the \(m/z\) 402.2 ion by cleavage of its amide bond results in the formation of an ion at \(m/z\) 268.1. This fragmentation behavior of the triazole–cyclooctyne moiety was noticed before for azidohomoalanine-containing peptides enriched via the ARCO resin [25]. In addition, cleavage of the amide bond formed by reaction of the \(\varepsilon\)-amine group of lysine residues with the active ester of the cross-linker is also frequently observed. For type 2 cross-links this type of cleavage yields fragment ions that add confidence to peptide identification. An example of the different types of ions formed upon fragmentation of a type 2 cross-link is shown in Fig. 4A. For type 0 cross-links the cleavage of the amide bond between a lysine and the cross-link remnant yields the unmodified parent peptide and a reporter ion at \(m/z\) 528.2 (Fig. 4B).
Fig. 4. Fragment ion spectra of cross-linked peptides showing reporter ions and cleavages of the cross-link along with peptide bond cleavages. A, type 2 cross-linked peptide (m/z 529.287, 5+) connecting α(I)M1 and α(II)K88; filled diamonds, fragments from MIEIEKPK; open diamonds, fragments from KLALKYSDEEK; asterisk, reporter ion at m/z 402.2; open square, reporter ion at m/z 268.1; hexagon, fragment ion resulting from a neutral loss of 401.2 Da; triangles, fragment ions resulting from cleavage of the cross-link, leading to the unmodified peptides. B, type 0 cross-link (m/z 752.409, 3+) modified at β′K1026; open square, reporter ion at m/z 528.2; asterisk, reporter ion at 402.2 m/z.

Cross-links are in agreement with a model of the 3-D structure of RNA polymerase and reveal conformational flexibility

Most cross-links span distances less than 17.5 Å (supplemental Table S4), and, therefore, are in perfect agreement with the model of RNA polymerase. See supplemental results for the maximal distance between Cα atoms of BAMG-induced cross-linked lysine residues. Cross-links spanning distances between 17.5 Å and 30 Å either may be explained by conformational flexibility or may require refinement of the model. One cross-link connects the aminoterminal methionine of α with K88 of α. This cross-link is present in two cross-linked peptides, one containing a missed cleavage site for trypsin. Since there are two α subunits in the complex, four different combinations are possible for the M1-K88 connection. The intersubunit combination between M1 of αI and K88 of αII (Fig. 4A) spans the shortest possible distance, i.e., 28.9 Å, still exceeding the maximal distance that can be spanned by BAMG by more than 10 Å. This can be explained by assuming that the N terminus is mobile. Indeed, in the Thermus aquaticus RNAP crystals 2–5 residues of the amino termini of the α subunits are not resolved, suggesting that the N-termini have some freedom to move.

Ten other cross-links span distances exceeding 17.5 Å, varying from 0.7 to 11.9 Å (supplemental Table S4). The Cα atoms of the linked residues with mutual distances exceeding 20 Å are often characterized by relatively high crystallographic B factors in corresponding residues in the template structure from T. aquaticus used for modeling (Table 1). This suggests...
that these residues are in areas of relatively high mobility. The cross-link between K156 (β) and K987 (β′) is located in the area where β lobe 1 and the lower downstream jaw formed by part of β′, also called downstream mobile clamp [33], approach each other (Fig. 5). The discrepancy of nearly 12 Å between the distance of cross-linked lysines between β-lobe 1 and the β′ downstream claw region in the model and the maximal cross-linker span distance suggests that these domains can move with respect to each other over considerable distances. This is consistent with a large conformational flexibility in bacterial RNA polymerase revealed by electron microscopy [34, 35] and with regions of disorder detected in solution [36]. A recent study of yeast RNA polymerase II likewise resulted in a number of cross-links spanning distances exceeding those determined by X-ray crystallography [37].

Considering the conformational flexibility in certain parts of the enzyme, and taking into account that the structure of the template for modeling was obtained at a resolution of 3.3 Å, we conclude that the cross-links identified here are in agreement with the model of the structure of *B. subtilis* RNA polymerase.

Table 1. Crystallographic B factors in the template structure used for modeling the RNA polymerase of *C. α* atoms corresponding to cross-linked residues farther than 20 Å apart. Boldface, cross-linked residues.

<table>
<thead>
<tr>
<th>Peptide A</th>
<th>Peptide B</th>
<th>Distance (Å)</th>
<th>Position of cross-link</th>
<th>Crystallographic B-factors (Å²) of corresponding residues in RNAP from <em>T. aquaticus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>DTKLGFEEITR</td>
<td>LLRHFGEKAR</td>
<td>22.11</td>
<td>βK803–βK868</td>
<td>82.12 (βK716), 111.57 (βK781)</td>
</tr>
<tr>
<td>KGPTATVIPNR</td>
<td>DKEQIYVGAVETR</td>
<td>29.40</td>
<td>βK156–βK987</td>
<td>127.43 (βY158), 181.56 (β'R1289)</td>
</tr>
<tr>
<td>KPETINARY</td>
<td>AKVR</td>
<td>23.40</td>
<td>βK30–β86</td>
<td>143.48 (β'K28), 214.08 (β'B84)</td>
</tr>
<tr>
<td>SKLTTVGK</td>
<td>FFLEKGADV</td>
<td>20.06</td>
<td>βK556–βK598</td>
<td>81.86 (β'A849), 49.45 (β'A889)</td>
</tr>
<tr>
<td>VISIWSA</td>
<td>KDGLTALK</td>
<td>22.18</td>
<td>βK709–βK785</td>
<td>89.68 (β'T1001), 58.29 (β'R1080)</td>
</tr>
<tr>
<td>LKEAESGK</td>
<td>VSDLTTVQYLLNEVQKVR</td>
<td>20.09</td>
<td>βK1011–βK1053</td>
<td>186.02 (β'L1312), 55.75 (β'R1354)</td>
</tr>
<tr>
<td>VLTDAIKGK</td>
<td>DELLGKEV</td>
<td>21.68</td>
<td>β'K1152–β'K1162</td>
<td>43.95 (β'A1453), 70.38 (β'A1463)</td>
</tr>
</tbody>
</table>
Fig. 5. Position of a cross-link between K156 in the β-lobe 1 and β'K987 at the tip of the downstream jaw exceeding the span of the cross-linker, suggesting conformational flexibility in this part of the RNA core polymerase (panel A). Panel B shows the MS,MS² spectrum of the m/z 853.710536 (4+) precursor ion and identified fragment ions. Filled diamonds, fragment from KGFTAVIPNR; open diamonds, fragments from DKQOEIVQGAVETR.

The distances of 5 cross-links could not be determined, as all five contain K294 located in the unresolved C terminal domain of the α chains. A remarkable cross-link is found between K294 of αI or αII and K161 of β', located at the tip of the N-terminal part of lower claw region [34] of β', ~120 Å away from the well-structured domain, ending at T228, of the most nearby α subunit on the opposite part of the complex (Supplemental Fig. S5). This suggests that the mobile α C-terminus can extend over large distances. Interactions between different enzyme molecules in a 'head to tail' orientation might provide an alternative explanation for the αK294–β'K161 cross-link. The cross-link between K294 in αI or αII and K704 of β is also far away (> 40 Å) from the well-structured part of the most nearby α subunit.

For understanding of the spatial arrangement of subunits in protein complexes, intersubunit cross-links are highly informative. Assuming that the connection between M1 and K88 is formed between αI and αII, we find cross-links between all subunits that are known to interact with each other, comprising α–α, α–β, α–β', β–β' and β'–ω interactions (supplemental Table S4). The high yield of intersubunit cross-links indicates that our approach is very promising in structural studies of assemblies of similar complexity as RNAP.

Large cross-linked peptides may escape detection

Out of the 36 identified type 2 cross-linked peptides only one species exceeds 4000 Da. On the contrary, no less than 37% of all possible cross-linked tryptic peptides with no missed cleavages spanning < 23.5 Å in the model of RNAP are more than 4000 Da (Supplemental Fig. S6). The underrepresentation of identified type 2 cross-linked peptides of a relatively large size can also be inferred from other recent data obtained with large complexes [37–40]. This implies that decreasing the peptide size by double digestion would bring a substantial fraction of cross-linked peptides within a mass range suitable for structural elucidation. Such a double digest approach may benefit from the enrichment obtained with the ARCO resin.

Discussion

The most important result reported here is the selective enrichment and identification of type 2 cross-linked peptides from a digest of a large biological assembly. Selective enrichment is obtained by using a cross-linker bearing an azido group in the spacer and by capturing the target peptides to an azido-reactive cyclooctyne resin, followed by fractionation of released material by SCX chromatography and mass spectrometric analysis. The mapping of the cross-links in the isolated peptides at the amino acid level was achieved using the software tool xComb that can be combined with several search engines for peptide identification [26]. By our enrichment approach an important limitation has been circumvented in cross-link analysis, namely the rare occurrence and low abundance of type 2 cross-links in proteolytic digests. The enrichment is so effective that the low concentration of the cross-linker used in our experiments
enabled mapping of several intersubunit cross-links. The low average number of cross-links per enzyme molecule under the experimental conditions both prevents the target protein becoming protease-resistant [22] and diminishes the possibility of conformational changes induced by the imposed reactions.

The enrichment procedure via capturing to the ARCO resin is robust, since the strain-promoted azide–alkyne cycloaddition reaction can take place in both aqueous media at different pH values and in organic solvents like DMF. The polymeric material used to conjugate the azide-reactive cyclooctyne moiety, dimethyl polyacrylamide beads [25, 41], is easy to handle in amounts as low as 0.5 to 2 mg in 1:1 water/acetonitrile mixtures of different pH values and ionic strengths. The beads do not stick to polypropylene surfaces under these conditions and can be efficiently washed batchwise in 2 ml vials by resuspension in volumes as large as 1–2 ml of different media, followed by brief centrifugation and convenient complete removal of the supernatant by pipetting. The beads can also be subjected to solvent-dependent swelling and shrinking [41]. An interesting application one can think of is trypsin-mediated $^{18}$O labeling of C-termini of captured peptides on the beads for recognition of mass signals corresponding to type 2 cross-linked peptide ions and for more easy interpretation of fragment spectra [42] or for quantitative purposes [43-45]. To this end, the beads with the covalently attached peptides can be shrunk and fully dehydrated in dry acetonitrile and be swollen in a small volume of a trypsin solution in $^{18}$O water, followed by quenching of the reaction and dilution and washing to remove the trypsin.

The good agreement of the set of 24 identified intrasubunit type 2 cross-links and 10 identified intersubunit cross-links with knowledge about the 3-D structure of prokaryotic RNA polymerases suggest that our analytical approach does not suffer from false positives. Therefore, the methodology is promising for cross-link analysis of biological assemblies as complex as the B. subtilis core RNA polymerase. This enzyme can interact with several different initiation and elongation factors that regulate transcription [46]. Since the precise mechanism of action of these factors is often not known and no high resolution structures are available for many of the enzyme-transcription factor complexes it will be attractive to map the interaction sites with our approach.

While results suggest the absence of false positives, our approach may not fully exclude the presence of false negatives. The presence of false negatives is suggested by the nonrandom distribution of cross-link distances in candidate cross-linked peptides nominated by Mascot of which the proposed structure could not yet be confirmed by employing the criteria for assignment as defined in this work. The reason for excluding a candidate cross-link from the list of identified species is lack of sufficient sequence evidence for one of the two peptides, most often a small peptide. This problem is generally recognized and alternative fragmentation techniques, i.e., electron transfer dissociation [40, 47], electron capture dissociation and infrared multiphoton dissociation have been explored [48-50] to improve the sequence coverage of cross-linked peptides. Further development in these areas may benefit from our isolation procedure for cross-linked peptides.
Conclusion

We have presented a procedure for the enrichment of type 2 cross-linked peptides from a digest of RNA polymerase treated with a low concentration of the amine specific cross-linker bis(sucinimidyl)-2-azidomethyl glutarate (BAMG). BAMG has a short spacer that can reveal highly informative distance constraints, and is provided with an azido group for capturing target peptides to an azide-reactive cyclooctyne resin. The enrichment procedure is robust and fast and the obtained type 2 cross-links can be identified using LC-ESI-MS/MS (CID) analysis and existing bioinformatics tools. Cross-links were detected for all known intersubunit interactions and, together with numerous intrasubunit cross-links, fit into a model of the 3-D structure of RNA polymerases. Our method for efficient mapping of cross-links is applicable to biological assemblies as complex as RNA polymerase. Use of alternative peptide fragmentation techniques like ETD, ECD and IRMPD may also benefit from the enrichment method presented here.

Acknowledgement

We thank Dr. Yusuke Nakasone and Orawan Borirak for their initial contributions in the protein overexpression of B. subtilis RNAP.

Materials and methods

Synthesis of reagents

Monofluorocyclooctyne (MFCO) was synthesized as described [51]. Azacyclooctyne [52] was kindly provided by Drs. Bertozzi and Sletten, University of California, Berkeley. Coupling of cyclooctynes to dimethyl polyacrylamide beads was carried out as described previously [25]. Bis(succinimidyl)-3-azidomethylglutarate (BAMG) [22] was prepared with omission of the purification of the intermediate 3-(azidomethyl)-glutaric acid.

Protein determination; polyacrylamide gel electrophoresis

Protein was measured using bicinchoninic acid [53] using a protein assay kit (Pierce). For polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate [54], 3–8% gradient NuPAGE Tris–acetate gels (Invitrogen) were used and stained with Coomassie Brilliant Blue.

Purification of RNA polymerase

Bacillus subtilis RNAP core enzyme (α2ββ')ω) was overproduced in Escherichia coli using the dual plasmid system [30]. After Histrap column (GE healthcare) purification the protein was loaded onto a 1 ml ResourceQ column (GE healthcare) pre-equilibrated with 50 mM HEPES, 150 mM NaCl, 10% (v/v) glycerol, pH 7.4 (HNG buffer). A linear gradient from 150 mM to 1 M NaCl was applied over 10 min at a flow rate of 1 ml/min [30].

Cross-linking and digestion

For cross-linking, 1 ml of the RNAP peak fraction eluting from the ResourceQ column at 350 mM NaCl, containing 0.7 mg protein, was mixed with 1.5 ml HNG medium. Cross-linking on ice was started by the addition of 0.2 mM BAMG and was continued for 2 hr. Cross-linking with disuccinimidyl glutarate (DSG, Thermo Fisher) for SDS PAGE analysis was carried out on small scale under the same conditions. The reaction was quenched by adding 1 M Tris–HCl, pH 8.0, to a final concentration of 20 mM. Subsequently, the protein was concentrated on a Amicon Ultra 10 kDa filter (Millipore). Sulphydryl groups were alkylated at room temperature in the dark for 30 min in a medium composed of 6 M urea, 50 mM iodoacetamide and 50 mM Tris–HCl pH 8. The preparation was diluted with 50 mM Tris–HCl pH 8 to lower the urea concentration to 2 M and then digested with trypsin (Trypsin
Gold, Promega, Madison, WI, USA)) (1:20 w/w) for 3 h at 37 °C. The digest was desalted on C18 reversed phase pipette tips (TT3 Toptips, Glygen), eluted with 0.1% TFA in 50% acetonitrile and dried in a vacuum centrifuge.

Enrichment of cross-linker modified peptides by the ARCO resin

A BAMG-cross-linked protein digest in 50 μl 0.1% TFA in 50% acetonitrile was added to ~ 2 mg of dry ARCO resin and incubated at 40 °C for 24 hr. After capturing, the resin was washed for 10 min with 1 ml 0.1 M acetic acid in 50% acetonitrile, 1 ml 0.6 M NaCl in 50% acetonitrile and finally 1 ml 0.1 M ammonium bicarbonate in 50% acetonitrile. Captured peptides were released by incubating the resin for 30 min with 50 μl of a solution containing 5 mM tris(carboxyethyl)phosphine (TCEP) and 0.1 M ammonium bicarbonate in 50% acetonitrile, followed by alkylation of free thiols with 20 mM iodoacetamide. After 30 min, peptides were desalted with TT3 Toptips (Glygen), eluted in 50 μl 0.1% TFA in 75% acetonitrile and dried in a vacuum centrifuge.

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

Material recovered from the enrichment by the ARCO resin was solubilized in 50 μl 0.1% TFA and immediately diluted with 1 ml of a solution containing 10 mM potassium phosphate buffer, pH 2.9 in 25% acetonitrile for loading on a polysulfethyl aspartamide column (2.1 mm 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⁻¹, was performed using a linear gradient from 0 to 250 mM KCl over 20 min followed by a gradient over 10 min to 500 mM KCl. Light absorption of the effluent was continuously measured at 214 nm and 280 nm. Material with absorbance at 214 nm eluting upon the start of the KCl gradient was collected in 0.2 ml fractions, desalted on C18 reversed phase pipette tips, eluted with 0.1% TFA in 50% acetonitrile, divided over 3–5 200 μl glass insert vials (Grace), and stored dry after evaporation of the solvents in a vacuum centrifuge.

Mass spectrometry

Sample loading

Dry samples in glass insert vials were reconstituted with a solution containing 0.1% TFA in 5% acetonitrile just prior to LC-MS,MS₂ analysis to prevent possible premature aggregation of peptides [55].

LC-FTICR MS,MS₂

MS,MS₂ data were acquired using an ApexUltra Fourier transform ion cyclotron resonance mass spectrometer (BrukerDaltonic, Bremen, Germany) equipped with a 7 T magnet and an Apollo II dual source coupled to an Ultimate 3000 HPLC system with a 300 μm ID, 250 mm C18 reverse phase column (Dionex, Sunnyvale, CA, USA) operated at a flow rate of 3 μl/min. After injecting samples, the LC gradient profile started at 100% buffer A (0.1% formic acid (v/v) in water). The acetonitrile concentration was raised linearly by mixing with buffer B (0.1% formic acid in 50% acetonitrile) to 12.5% acetonitrile in 5 min, then to 22.5% in 20 min and finally to 50% in 5 min. Ions were selected by data-dependent acquisition in the Q sector and subjected to collision-induced dissociation in the hexapole at an argon pressure of about 6.4–7.1.10⁻⁶ mbar measured at the pressure gauge. Data were processed with Mascot distiller version 2.4.2. (Matrix Science).

LC-Q-TOF-MS,MS₂

For electrospray MS and low energy collision-induced dissociation (MS,MS₂) analyses on a Q-TOF mass spectrometer with a Z-Spray orthogonal ESI source (Micromass, Whytenshawe, United Kingdom), peptide samples 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). After loading the sample 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 of increasing acetonitrile concentrations in 0.1% formic acid [25]. Direct infusion of the flow was supported by a Nanobore Emitter (Proxeon, Odense, Denmark). Survey scans were acquired from m/z 350–1500. Ions were selected for MS,MS₂ in a data-dependent mode, recorded from m/z 50–2500, 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 analysis

Identification of unmodified peptides and type 0 and type 1 cross-links

The SwissProt Bacillus subtilis database was used for identification by Mascot of unmodified peptides, type 0 cross-linked peptides and type 1 cross-linked peptides. For unmodified peptides, up to one missed cleavage was included in the search parameters, while for type 0 and type 1 cross-links two, resp. three missed cleavages were allowed.

Identification of type 2 cross-links

A database of all possible cross-linked peptides for interrogation using standard search engines was generated by xComb version 1.2 [26, 27]. The amino acid sequences of proteins of interest were uploaded in UniProt FASTA format. Trypsin was the chosen enzyme for digestion with two missed cleavages allowed. Both intra- and inter-protein cross-links were taken into account. The minimum peptide length for each peptide of the pair was four amino acids with at least one trypsin missed cleavage for amine cross-linking. The cross-link database was uploaded in Mascot version 2.2. The following parameters were used to identify candidate type 2 cross-links based on the processed data files from LC-MS$_1$MS$_2$ experiments: (i) a “do not cleave” enzyme with the nonexisting amino acid “J” as the cleavage site; (ii) no missed cleavages; (iii) a fixed modification in the form of carbamidomethyl at C; (iv) a variable modification at K with a group of composition C$_{n}$H$_{2n+1}$N$_{2}$O$_{5}$S. For MS$_1$MS$_2$ data obtained with FTMS, the precursor tolerance was set at 20 ppm and the product ion tolerance at 0.05 Da. Identified unmodified peptides and type 0 cross-linked peptides were used for internal calibration to obtain a mass accuracy for precursor ions better than 8 ppm. Criteria used for confirmation of type 2 cross-links were a mass tolerance window of 8 ppm for the intact peptide ion and detection of at least 5 unambiguous fragment ions for cross-linked peptides built up from up to 25 amino acids, and 6 unambiguous fragment ions for peptides composed of more than 25 amino acids. Both peptides in the cross-link should be represented by at least one fragment. All required fragments should belong to the 30 signals of highest intensity. For MS$_1$MS$_2$ data obtained with the Q-TOF mass spectrometer both the precursor tolerance and the product ion tolerance were set at 0.4 Da. For confirmation of the identity of type 2 cross-linked candidates, detection of the following numbers of unambiguous fragment ions dependent on the size of the cross-linked peptide was required: at least 5 fragments of which at least 2 fragments from each of the two peptides for cross-linked peptides up to a total of 15 residues; for cross-linked peptide pairs of more than 15 residues: at least 2 fragments for a peptide if it contains 4 or 5 amino acids, at least 3 fragments if it contains 6 or 7 residues and at least 4 fragments for peptides built up from 8 or more amino acids. The required unambiguous fragment signals should belong to the 55 signals of highest intensity in MS$_1$MS$_2$ spectra.

References

36. Kontur, W.S., et al., Solute probes of conformational changes in open complex (RPo) formation by Escherichia coli RNA polymerase at the lambdaPR promoter: evidence for unmasking of the active site in
the isomerization step and for large-scale coupled folding in the subsequent conversion to RPo. Biochemistry, 2006. 45(7): p. 2161-77.


