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.
Structural analysis of RNA polymerase-helicase complexes by chemical cross-linking and mass spectrometry

Abstract

Information about dynamics and interactions within multi-protein complexes are essential to understand biological processes. Unfortunately, high-resolution structure determination techniques i.e. X-ray crystallography and NMR are often not applicable to these macromolecular complexes. Chemical cross-linking of protein complexes coupled to mass spectrometry (CXMS) of peptides is a technique that can provide spatial amino acid distance constraints imposed by the length of the spacer of the cross-linking agent, to obtain low resolution (10-30 Å) structural information. However, in particular the low abundance of cross-linked peptides makes this approach analytically challenging. In Chapter 4, we have developed an approach enhancing the efficiency of CXMS. In this chapter, the usefulness of the approach was further demonstrated by mapping several cross-links between bacterial RNA polymerases (RNAPs) and interacting proteins. The interaction between the helicase UvrD and RNAP from E. coli was revealed by 10 cross-linked peptides. Seven cross-links, combined with other evidence about the interaction of the orthologous B. subtilis helicase PcrA to RNAP indicates a binding site for UvrD on the upstream face of RNAP. The same binding site is used by Mfd, a helicase that dislocates stalled RNAP from the template upon transcription-coupled repair. This may imply similar roles of UvrD and Mfd during transcription. UvrD/PcrA may dislodge RNAP from the template to enable DNA replication upon collision with a replication fork. Three other RNAP-UvrD cross-links suggest an alternative binding interaction on the downstream face. One cross-link between B. subtilis RNAP and HelD was also detected on the downstream site of RNAP. Helicases at the downstream face may either remove DNA binding proteins that otherwise might prevent RNAP elongation, or dislocate RNAP upon transcript termination or upon replication-transcription collisions. We also identified a cross-link between RPOE, also called δ subunit of RNAP, and the β' clamp region of RNAP, suggesting how δ can compete with the binding of DNA. This work has demonstrated the use of our CXMS approach to obtain novel information about the bacterial transcription process.

*This work was done in collaborations with Prof. Evgeny Nudler, Medical center, NewYork University, (E. coli-UvrD) and Dr. Peter Lewis, Newcastle University, (B. subtilis-PcrA).
Introduction

Knowledge about the architecture of protein complexes is essential to understand their functions at the molecular level. X-ray crystallography has been accounted for a long time as the main approach for providing detailed structures of biological samples. Often structures can be resolved by x-ray crystallography at high resolution. However, the conditions for protein crystallization, particularly for large complexes, are unpredictable. In general, only stable and rigid complexes can be crystalized. This leaves the information about the interactions of protein subunits in many multi-protein complexes mysterious.

As seen throughout this thesis, chemical cross-linking coupled with mass spectrometry (CXMS) of peptides has been progressively employed in protein structure studies. It was originally introduced by Rappsilber et al. in 2000 [1]. This simple but elegant approach combined the advantages of two techniques i.e. protein conjugation and mass spectrometry (MS) to identify the position of linked amino acids in the protein, while the maximal distance of linked residues is determined by the length of the spacer of the bifunctional reagent. However, harvesting the rich structural information of a cross-linked protein complex is analytically challenging. Over the past few years, numerous efforts have been made to isolate cross-linked peptides prior to MS analysis. We have developed, as presented in Chapter 4, a new analysis strategy aimed to enrich cross-linked peptides and enhance CXMS. We successfully identified linkages in large biological assemblies in a complex protein mixture, extracted from isolated HeLa cell nuclei. In this chapter, we extend the use of our CXMS strategy to identify novel interactions in isolated and reconstituted protein complexes.

Bacterial RNA polymerases (RNAP) are large, dynamic and highly conserved complexes. The core RNAP composed of 5 subunits (α₂ββ’ω) has a molecular weight of ~400 KDa. In transcription, RNAP is responsible for the synthesis of RNA chains, e.g., tRNA, mRNA and rRNA, using a complementary DNA as a template. Transcription is highly regulated and three main phases can be distinguished, i.e., initiation, elongation and termination. Briefly, to initiate the transcriptional process, core RNAP requires an additional subunit, σ factor, to form a holoenzyme for recognizing a specific DNA region located on the upstream of transcription start site, known as a promoter. After binding to the promoter, the RNAP holoenzyme initiates the melting of dsDNA surrounding the transcriptional start site and unwinding of downstream dsDNA. Transcription elongation occurs after the synthesis of a short RNA chain (~ 13 nucleotides) and the abortion of the σ subunit thus results in an elongation complex (EC). During transcription elongation, EC works in concert with transcription factors and additional proteins. For instance, the elongation factor NusA is needed for stable formation of ECs [2], while helicases are required for removal of protein roadblocks. Finally, the termination of RNA transcription can be caused either by intrinsic (RNA “hairpin” formation) or extrinsic termination (Rho dependent), resulted in the dislodging of RNAP from the template.

Although transcription has been studied extensively, a complete set of proteins and their functions involved in formation of a full length transcript remains unclear. Several attempts
have been made to characterize the interactions between RNAP and those transcriptional-related proteins. Recently, several partners of *B. subtilis* RNAP under various growth conditions were identified by tandem affinity purification (TAP) and sequential peptide affinity (SPA) purification [3]. In this study, various types of σ factors were found associated with tagged RNAP at different conditions. For instances, a high abundance of RNAP-σ^A^ and RNAP-σ^B^ was observed during the stationary phase and in the exponential/transition phase, respectively. Among many others proteins, also the helicases PcrA and HelD co-isolated with RNAP when either the β or β’ subunit was provided with the TAP tag or SPA tag. Similarly, in *E. coli*, more than hundred transcription factors and related proteins were found to interact with RNAP during transcription [4]. Despite a large number of proposed interactions, much less is known about the sites of interactions, required for detailed insight into the mechanisms of action at the molecular level of proteins associated with RNAP. Since *E. coli* RNAP has been refractory towards crystallization for a long time, detailed models were mainly based on crystal structures of the homologous enzymes from the *Thermus genus*. [5]. However, recently, a crystal structure of the *E. coli* RNAP holoenzyme (αββ’ωσ) was published (PDB: 4IGC, hereafter called Eco4IGC) (Fig. 1).

In this chapter, CXMS was applied to investigate the interaction of RNAP and two helicases. To this end, CXMS was conducted in two reconstituted RNAP-helicase complexes, namely the *E. coli* RNAP-UvrD complex and the *B. subtilis* RNAP-PcrA complex. The *B. subtilis* RNAP used for these experiments also contains, as an unintentional contaminant, another helicase, HelD, that has remained tightly bound to the RNAP during purification. Helicases are motor proteins that are capable of unwinding duplex nucleic strands as well as moving along either duplex or single-stranded nucleic strands using energy obtained from ATP hydrolysis. These proteins are involved in processes requiring bare single-stranded nucleic strands, like replication and transcription. Classification of helicase proteins into superfamilies are basically based on conserved motifs of primary structures. UvrD and its orthologue PcrA, as well as HelD, are members of helicase superfamily 1A. UvrD and PcrA are the best-studied helicases in Gram-negative and Gram-positive bacteria respectively [6]. These motor machineries are multi-functional proteins. They perform single stranded DNA translocase and double stranded DNA helicase (unwinding) activities in the 3’→5’ direction. In *B. subtilis*, PcrA is an essential helicase required for rolling circle replication/cell growth whereas lack of UvrD in *E. coli* did not affect cell viability. Interestingly, although lack of rep does not affect cell viability, *E. coli rep* and *uvrD* double mutant is lethal. UvrD and PcrA share a large range of conserved sequences (~ 40 % identity) and may also perform similar functions as the expression of PcrA in an *E. coli rep* and *uvrD* double mutant could restore the viability of the cell. Previous studies have shown that UvrD is directly involved in repair of ultraviolet (UV)-induced DNA damage, methyl-directed mismatch repair, replication and recombination [7] including a displacement of DNA protein roadblocks [8, 9].

Although a comprehensive analysis of the interaction between RNAP and helicase has never been performed, ample evidence for a binary interaction has been put forward over the past few years. The *E. coli* RNAP-UvrD interaction was initially identified via yeast-based tandem affinity purification procedure [10]. A congruent result was obtained by Hu *et al.* using
systematic large-scale tandem-affinity purification [11]. In this approach, UvrD was identified in a protein complex isolated using the RNAP β subunit as the TAP-tag containing protein. Direct binding of the UvrD-orthologue PcrA to RNAP β was discovered by a genome-wide yeast two-hybrid screen using bait proteins involved in replication, among which PcrA [12]. PcrA and HeID were also detected in the protein complex obtained by tandem affinity purifications using TAP-tagged RNAP β and β′ subunits [3]. Recently, the interaction of _B. subtilis_ RNAP and PcrA was studied in more detail using a combination of techniques, including the yeast two-hybrid system, Far-Western blotting and single particle analysis using negative staining transmission electron microscopy. It was found that the N-terminal region of the β subunit interacts with the C-terminal domain (also called 2A domain) of PcrA, β(1-400)–PcrA(577-739) and the C-terminal region of β′ subunit interacts with an internal segment (called 1B domain) of PcrA, β′(1-102)–PcrA(85-189). However, no unambiguous detailed picture of the PcrA–RNAP interaction could be obtained based on these results, due to the low resolution of the single particle analysis [13]. In our study, more detailed insights of RNAP-helicase interactions were obtained using CXMS.

![Fig. 1](image)

**Fig. 1** Three-dimensional crystal structure of the _E. coli_ RNAP (PDB: 4IGC) without the σ70 subunit. Panel A shows the subunits of RNAP in pink (β), yellow (β′) and red (ω) from different views. β′ bridge helix located in between β and β′ subunit. β′ jaw and β lobe 1 are indicated in the dashed oval. Panel B shows green (αI), white (αII). The dashed oval shows αI CTD. This is the first time that αI CTD is included in the intact RNAP.

### Results and Discussion

**MS cross-link analysis of the _E. coli_ core RNAP**

We first mapped cross-links in the core RNAP with the approach developed in _Chapter 4_ and schematically depicted in Fig. 2. Briefly, RNA polymerase was cross-linked with bis(succinimidyl)-3-azidomethyl glutarate (BAMG) at a low protein concentration to prevent aspecific cross-linking between different protein molecules during random collisions [14]. To identify juxtaposed, cross-linked lysines, the BAMG-treated protein complex is subjected to trypsin digestion. Cross-linked lysines are identified by one of the three analytical strategies...
described in the previous chapter, namely the total cross-linked peptide database approach. In this strategy, enrichment of cross-linked peptides is performed through diagonal strong cation exchange chromatography (2SCX) in three steps. First, the peptide mixture is fractionated using SCX (primary SCX, Fig. 2, panel A (a)). Each collected fraction is treated with the reducing agent TCEP, aimed to reduce the azido group (-N₃) on the spacer of BAMG to a primary amine group (-NH₂). This conversion changes the chromatographic behavior of cross-linked peptides enabling enrichment by SCX chromatography (secondary SCX) using identical condition as performed in the primary run (Fig. 2, panel A (b)). Finally, the isolated peptides are subjected to LC-FTMS₁MS₂ analysis. Cross-linked peptides are identified with peptide fragment fingerprinting using a database of all possible cross-linked peptides calculated by the xComb program [15] (lower part of Fig. 2). For determination of the false discovery rate (FDR), the protein database used for the xComb program contained both forward and reversed sequences. Candidate cross-linked species of which one or both of the composing peptides had a reversed sequence were considered as false positives.

In the *E. coli* apoenzyme a total of 291 cross-linked spectral matches were identified, resulting in 92 intraprotein and 32 interprotein cross-linked peptides at a FDR of 0.7% (Supplementary Table 1). We next determined the distances between de Cα atoms of linked amino acid residues based on the recent crystal structure of the *E. coli* holoenzyme (PDB file 4IGC, further called *Eco4IGC*) [5]. It is important to note that no large differences were observed between the apoenzyme and the holoenzyme crystal structures from the *Thermus* RNAP. In Fig. 3 the distribution of cross-linked peptides from the apoenzyme is depicted as a function of the measured distances between Cα atoms of linked residues, as far as present in Eco4IGC. A significant difference between the distance distribution of our cross-linked peptides and a random selection of theoretical possible cross-linked peptides was demonstrated by a *P*-value of 2.46 x 10⁻⁷ Most of the cross-links spanned distances between Cα of the linked residues less than 28.7 Å. This is considered the maximal distance for structural agreement with Eco4IGC, assuming a spacer length of 7.7 Å, a lysine side chain length of 6.5 Å, and a coordinate error of 4 Å. However, no less than 21 cross-links spanned distances exceeding 28.7 Å (Table 1). These cross-linked peptides were basically grouped into four regions, i.e., β₁91-331, β₈₄₄-909, β’₉-96 and β’₁₁₇₀-1₃₄₀. The detection of cross-links spanning large distances in certain discrete domains suggest local conformational flexibility, a well-known property of bacterial RNA polymerases [16]. However, apparent movements over more than 50 Å (Table 1) are unexpected and require further investigation.
Chapter 5

Fig. 2 Overview of the procedure and chemicals used for the enrichment of cross-linked peptides. To map interaction sites between different proteins in a complex, by identifying juxtaposed, reactive lysines, the protein complex is cross-linked with bis(succinimidyl)-3-azidomethyl glutarate (BAMG) and then subjected to trypsin digestion. Cross-linked peptides were identified by one of the three analytical strategies described in the previous chapter, namely the total cross-linked peptide database approach. Enrichment of cross-linked peptides is performed through diagonal strong cation exchange chromatography approach (2SCX) in three steps. First, the peptide mixtures are fractionated using SCX (primary SCX). Each collected fraction is treated with the reducing agent TCEP, aimed to reduce the azido group (-N₃) on the spacer of BAMG to a primary amine group (NH₂). This conversion changes the chromatographic behavior of cross-linked peptides enabling enrichment by SCX chromatography (secondary SCX) using identical condition as performed in the primary run. This is schematically shown for one primary fraction, depicted as a grey bar in the primary chromatogram a). In the secondary run (chromatogram b) the reduced cross-linked peptides elute in a broad peak marked # at a later retention time. Finally, the isolated peptides are subjected to LC-FTMS/MS analysis. Cross-linked peptides are identified with peptide fragment fingerprinting using a database of all possible cross-linked peptides calculated by the xComb program. A, Workflow of the cross-linked peptide isolation approach. a) primary SCX chromatogram, b)

![Graph showing distribution of cross-linked peptides identified in E. coli RNAP and E. coli RNAP-UvrD complexes as a function of the distance between the Cα atoms of linked residues (PDB: 4IGC, [5]). White bars, all theoretical possible linkages of Lys-Lys expressed as percentage of the total number. Black bars, number of identified cross-linked peptides in E. coli RNAP. Grey bars, number of identified cross-linked peptides in E. coli RNAP in complex with UvrD. 35 and 32 cross-links from RNAP and RNAP-UvrD, respectively, are not represented in this histogram, since the linked residues are in disordered areas not included in 4IGC.]

### Table 1

List of cross-linked peptides which span distances larger than 28.5 Å between Cα of linked residues both in E. coli RNAP and E. coli RNAP-UvrD complex. Boldface, cross-linked residues.

<table>
<thead>
<tr>
<th>Peptide A</th>
<th>Peptide B</th>
<th>Linked residue peptide A</th>
<th>Linked residue peptide B</th>
<th>RNAP distance (Å)</th>
<th>RNAP-UvrD (PDB: 4IGC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSWLDFEFDPKGNLFVR</td>
<td>ETGSK</td>
<td>βK191</td>
<td>βK1170</td>
<td>28.64</td>
<td></td>
</tr>
<tr>
<td>DTKGPEEITADIPVGAEALSK</td>
<td>AITGSRKPLK</td>
<td>βK844</td>
<td>βK321</td>
<td>32.32</td>
<td></td>
</tr>
<tr>
<td>TFKPER</td>
<td>NEKR</td>
<td>βK50</td>
<td>βK296</td>
<td>32.43</td>
<td></td>
</tr>
<tr>
<td>VYVEKGR</td>
<td>ETGSK</td>
<td>βK265</td>
<td>βK1170</td>
<td>34.00</td>
<td></td>
</tr>
<tr>
<td>AVKER</td>
<td>KPETIN YR</td>
<td>βK476</td>
<td>βK40</td>
<td>42.28</td>
<td></td>
</tr>
<tr>
<td>GETQLTPEEKLRL</td>
<td>KPETIN YR</td>
<td>βK900</td>
<td>βK40</td>
<td>42.30</td>
<td></td>
</tr>
<tr>
<td>FLKAOPTK</td>
<td>TFKPER</td>
<td>βK9</td>
<td>βK50</td>
<td>44.89</td>
<td></td>
</tr>
<tr>
<td>SGESSLFSR</td>
<td>AVAVDSGVTAVAKR</td>
<td>βK639</td>
<td>βK719</td>
<td>45.38</td>
<td></td>
</tr>
</tbody>
</table>
Preparation and characterization of RNAP-helicase complexes

To map the cross-links and to use the information to determine how the helicases are spatially arranged with respect to the other subunits in the core RNAP, *B. subtilis* RNAP-PcrA and *E. coli* RNAP-UvrD complexes were prepared. Both helicases were added to each RNAP in 1:1 molar ratio. The mixtures were incubated at RT for 30 minutes. The complexes were then subjected to the cross-linking reaction (see Materials and Methods). The formation of cross-links between subunits in the *B. subtilis* RNAP-PcrA and *E. coli* RNAP-UvrD complex were investigated using SDS-PAGE (data not shown). The cross-linked complexes were then subjected to tryptic digestions. Cross-linked peptides were isolated using 2SCX, and MS analysis was conducted as described in the previous section.
### Table 2
List of cross-linked peptides involved in RNAP-UvrD interaction shown with number of ions selection for MS/MS. Boldface, cross-linked residues.

<table>
<thead>
<tr>
<th>Peptide A</th>
<th>Peptide B</th>
<th>Linked residue peptide A</th>
<th>Linked residue peptide B</th>
<th>Number of ions selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTKLPEEITADIPVGEAALS</td>
<td>GEKQOQR</td>
<td>βK844</td>
<td>UvrDK501</td>
<td>9</td>
</tr>
<tr>
<td>GETQLTPEEKLLR</td>
<td>AMNLDEKQWPPR</td>
<td>βK900</td>
<td>UvrDK135</td>
<td>2</td>
</tr>
<tr>
<td>GETQLTPEEKLLR</td>
<td>GEKQOQR</td>
<td>βK900</td>
<td>UvrDK501</td>
<td>2</td>
</tr>
<tr>
<td>AIFGEKASDVK</td>
<td>LLKR</td>
<td>βK909</td>
<td>UvrDK124</td>
<td>1</td>
</tr>
<tr>
<td>AIFGEKASDVK</td>
<td>AMNLDEKQWPPR</td>
<td>βK909</td>
<td>UvrDK135</td>
<td>5</td>
</tr>
<tr>
<td>AIFGEKASDVK</td>
<td>GEKQOQR</td>
<td>βK909</td>
<td>UvrDK501</td>
<td>4</td>
</tr>
<tr>
<td>KPETINMR</td>
<td>ELLQEKALAGR</td>
<td>βK40</td>
<td>UvrDK448</td>
<td>1</td>
</tr>
<tr>
<td>EELENTSETKR</td>
<td>AMNLDEKQWPPR</td>
<td>βK213</td>
<td>UvrDK135</td>
<td>2</td>
</tr>
<tr>
<td>KGLADTALK</td>
<td>LLAGDTGVMVGDDDQSIYWR</td>
<td>βK781</td>
<td>UvrDK242</td>
<td>1</td>
</tr>
<tr>
<td>GLADTALKTANGSYLTR</td>
<td>LLAGDTGVMVGDDDQSIYWR</td>
<td>βK789</td>
<td>UvrDK242</td>
<td>21</td>
</tr>
</tbody>
</table>

**Fig. 4** Four residues (shown in blue) involved in 7 linkages of *E. coli* RNAP-UvrD complex on the upstream face of *E. coli* RNAP.

Besides the core subunits α, β, β’ and ω, the *B. subtilis* RNAP preparation used for the reconstitution with PcrA contained also the proteins RPOD (σ), RPOE (δ) and HelD, all three
known to interact with the apoenzyme. We identified 130 mass spectral matches resulting in 34 intraprotein and 19 interprotein cross-linked peptides (Supplementary Table 2). The CXMS was performed at a FDR of 0 %. The distance distribution of cross-linked peptides identified in this study was significantly different from a random selection of all theoretical possible cross-links at P-value of 7.9x10^-7. Unfortunately, no linkages between RNAP and PcrA were identified. Surprisingly, a cross-link between a lysine residue in the region close to DNA entry channel, β'K1011 and K97 in HelD was detected. Additionally, we also identified a cross-link between δK48 and β'K208 and between σK107 and -β'K153.

For the E. coli RNAP-UvrD complex, we identified 486 mass spectral matches resulting in 90 intraprotein and 56 interprotein cross-linked peptides (Supplementary Table 3). CXMS was performed at a FDR of 0.4 %. The distance distribution of cross-linked peptides identified in this study (Fig. 3) was significantly different from the distribution of all theoretically possible cross-links at P-value of 2.17x10^-7. Out of 146 linkages, 14 cross-linked peptides span distances larger than 28.7 Å (Table 1). Interestingly 10 linkages of RNAP-UvrD complexes were identified (Table 2). Out of those, 7 linkages were located on the upstream face of RNAP. The residues on RNAP involved in these linkages are shown in Fig. 4. Another linkage showed the interaction of the β' jaw to UvrD, namely the cross-link between β'K213 and UvrD_K135 (Table 2). Intriguingly, two linkages showed an interaction of the β' bridge helix and UvrD, β'K781-UvrD_K242 and β'K789-UvrD_K242.

**Location of HelD on B. subtilis RNAP and its possible function in transcription**

HelD, also known as helicase IV, is a member of the helicase superfamily 1A, sharing seven conserved motifs with PcrA and UvrD. HelD catalyzes the unwinding of dsDNA in 3'→5' direction using energy obtained from ATP hydrolysis. In a recent work, HelD was co-purified with RNAP by the TAP technique using tagged β and β' subunits [3]. In E. coli, the functions of UvrD and helicase IV, an orthologue of B. subtilus HelD are overlapping [17]. Recently, HelD was shown to enhance the removal of RNAP from the template through its ATP-dependent activity, and a synergistic effect was observed with the δ subunit [18] (discussed below in more detail). In our study, the interaction between HelD and RNAP was confirmed by CXMS. Unfortunately, the orientation of HelD in complex with RNAP was ambiguous since only one linkage was identified, allowing rotation of HelD with respect to RNAP around the cross-link. Nevertheless the cross-link implies binding of HelD to the downstream face of RNAP. The strong binding of HelD to RNAP suggests that is remains permanently bound to RNAP in vivo. Besides removal of RNAP upon transcription termination, the position of HelD at the downstream face is ideal for removal of DNA binding proteins that might otherwise obstruct transcription by the elongating RNAP.

**Location of the δ subunit on B. subtilis RNAP and its possible function in transcription**

The Gram-positive specific δ subunit is a small protein (21.4 kDa) encoded by the rpoE gene. It possesses a highly negatively charged unstructured C-terminal domain. RPOE binds to RNAP competitively with nucleic acids [19]. In transcription, the δ subunit enhances the specific binding of RNAP to promoters as well as promotes the RNAP recycling process through an
RNA displacement activity [20]. Although it is a non-essential protein under normal conditions, the δ subunit is a crucial factor responsible for rapid gene expression regulated by [iNTP]-sensitive promoters [21]. Despite extensive functional studies, the interaction between the δ subunit and RNAP remains mysterious. In our study, a cross-link between δ_K48 and β'_K208 was identified. As shown in Fig. 5, the β'_K208 residue lies inside the DNA entry channel relatively close to the dsDNA. This suggests that the δ subunit can bind in the DNA entry channel, thereby preventing association of RNAP with downstream DNA. In a recent work, the δ subunit together with HelD was observed to synergistically promote RNAP removal from the template and RNAP recycling in transcription, thereby increasing RNA synthesis [18]. Additionally, the polyanionic C-terminal region of δ was observed to displace RNA bound to RNAP by mimicking ssRNA [19]. This activity could randomly occur both to the DNA and RNA strand. However, efficient displacement of the bound RNA requires both N- and C-terminal regions. Therefore it was proposed that the N-terminal region may bind to the surface of RNAP to orientate the C-terminal domain [19]. The interaction of the δ subunit and core RNAP identified in this study, combined with functional evidence suggest that the N-terminal region of the δ subunit interacts with the β’ jaw lobe to direct the C-terminal domain to interact with RNAP (Fig. 5).

Fig. 5 Proposed orientation of the δ subunit in the RNAP-δ subunit complex. Panel A shows the interaction of δ and RNAP, observed on the upstream face of RNAP. Panel B shows that the N-terminus of δ lies in the active region of RNAP. Red and blue spheres are δ_K48 and β'_K208 respectively. As a point of reference, an arrow locates β’ bridge helix.

Locations of UvrD on E. coli RNAP and possible roles in transcription

The linkage sites between RNAP and UvrD were revealed by 10 cross-linked peptides, of which 7 linkage sites were positioned to the upstream side of RNAP. Out of these 7 linkages, three cross-linked peptides involved β_K900 and UvrD i.e. UvrD_K124, UvrD_K135 and UvrD_K501. In a coherent manner, β_K900 was detected to interact with UvrD_K135 and UvrD_K501. In addition, β'_K40 and β_K844 were shown to interact with UvrDK_448 and UvrDK_501, respectively. The program
PyMOL [22] was applied to obtain a picture of the UvrD-RNAP interaction based on these cross-links. As models for the structures of RNAP and UvrD we used pdb files 4IGC and 2IS1 respectively. Docking of the two proteins was achieved manually. Since *E. coli* UvrD and *B. subtilis* PcrA are orthologues, we also took into account the $\beta(1-400)$-$PcrA(577-739)$, $\beta'(1-102)$-$PcrA(85-310)$ and $\beta'(228-310)$-$PcrA(85-310)$ interactions identified by Far-western blotting [13]. In line with these interactions, single particle analysis showed PcrA on the upstream side of RNAP [13]. Fig. 6 shows a model in which the 7 *E. coli* RNAP-UvrD cross-links and the *B. subtilis* RNAP-PcrA interaction data are integrated. In the model, RNAP and UvrD were considered as rigid bodies. The best result was obtained by placing the C-terminal domain of UvrD, as far as present in 2IS1 (the last 67 amino acids in the C-terminus are absent in the structure) in direct contact with the region $\beta(50-127)$, to account for the $\beta(1-400)$-$PcrA(577-739)$ interaction. The distances between C$\alpha$ of linked residues measured from the manual-docking complex are shown in Table 3. Notably, only the distance between $\beta'_{K40}$ and UvrD$_{K448}$ exceeded 28.7 Å. In addition, $\beta'_{K40}$ was also involved in several linkages spanning large distances. This suggests that $\beta'_{K40}$ is located on a flexible (mobile) domain.

The $\beta(50-127)$ region is also the binding site of Mfd [23], a member of helicase superfamily 2. Mfd plays an important role in transcription-coupled repair where it removes a stalled RNAP caused by a DNA lesion in the ssDNA template. By binding to the $\beta$ subunit on the upstream face of a stalled RNAP, Mfd translocates a ssDNA template in the direction 3'→5' to push a stalled RNAP forward. This results in a dislodging of a stalled RNAP from the damaged DNA.

**Fig. 6** Location of UvrD on the upstream face of RNAP. The interaction of UvrD to RNAP is shown based on the crystal structure of UvrD (PDB: 2IS1) and *E. coli* RNAP (PDB: 4IGC). Panel A shows an overview of UvrD-RNAP complex. UvrD binds to RNAP on the upstream face. RNAP $\beta(1-450)$ is shown in pale green, surface representation. RNAP $\beta'(1-102)$ and RNAP $\beta'(228-310)$ are shown in pink, surface representation. UvrD$_{(571-720)}$ is shown in orange, ribbon representation. UvrD$_{(85-189)}$ is shown in yellow, ribbon representation. Panel B shows the enlarged view of the rectangular area in Panel A.
Table 3 List of cross-linked peptides involved in RNAP-UvrD interaction shown with distances between C\textalpha\ of linked residues measured from the manual-docking complex. Boldface, cross-linked residues.

<table>
<thead>
<tr>
<th>Peptide A</th>
<th>Peptide B</th>
<th>Linked residue peptide A</th>
<th>Linked residue peptide B</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTGLPGEITADINVGEAALS K</td>
<td>GEKQTR</td>
<td>βK844</td>
<td>UvrDK501</td>
<td>25.3</td>
</tr>
<tr>
<td>GETQLTPEEKLLR</td>
<td>AMNLDEKKWPPR</td>
<td>βK900</td>
<td>UvrDK135</td>
<td>28.3</td>
</tr>
<tr>
<td>GETQLTPEEKLLR</td>
<td>GEKQTR</td>
<td>βK900</td>
<td>UvrDK501</td>
<td>28.5</td>
</tr>
<tr>
<td>AIFGEKASDVK</td>
<td>LLKR</td>
<td>βK909</td>
<td>UvrDK124</td>
<td>27.9</td>
</tr>
<tr>
<td>AIFGEKASDVK</td>
<td>AMNLDEKKWPPR</td>
<td>βK909</td>
<td>UvrDK135</td>
<td>28.5</td>
</tr>
<tr>
<td>AIFGEKASDVK</td>
<td>GEKQTR</td>
<td>βK909</td>
<td>UvrDK501</td>
<td>18.5</td>
</tr>
<tr>
<td>KPETINYR</td>
<td>ELLQERALAGR</td>
<td>βK40</td>
<td>UvrDK448</td>
<td>49.5</td>
</tr>
</tbody>
</table>

It is conceivable that UvrD has the same role as Mdf in transcription-coupled repair under other circumstances when an elongating RNAP complex encounters situations causing transcriptional arrest, e.g., replication-transcription collisions. Both head on and colinear collision with a replication fork occur, since replication is about one order of magnitude faster than transcription [24]. Blockages formed by transcription-replication collisions markedly affect the viability of the cell since DNA replication and RNA synthesis fail to complete, thus ultimately cause cell death. To overcome this, additional transcription factors (TFs) are required. It has been shown that UvrD place a role in the continuation of replication upon replication-transcription collisions [28]. Fig. 7 shows the structure of UvrD and depicts the different functional domains that be designated. Although under debate, the 2B subdomain is thought to be involved in the helicase activity. However, in the model depicted in Fig. 6 the close contact the 2B subdomain to RNAP may obstruct the helicase activity. This would imply that UvrD in this configuration cannot be active as a helicase and cannot interact with duplex DNA in the RNAP-UvrD complex. A reorientation is required to be able to function as a helicase so that the UvrD active cleft is exposed to the DNA exit channel close to ssDNA template. Under these conditions UvrD may bind to the ssDNA template and perform its ssDNA translocase activity (3'[→]5') direction using energy derived from ATP hydrolysis to mechanically remove a stalled RNAP from DNA strand (Fig. 8). A permanent association of UvrD to RNAP has the advantage that transcriptional road blocks for DNA replication can be rapidly dismantled. However, the mechanism by which the required translocase activity of UvrD is activated under these conditions to remove the stalled RNAP is not understood.
Fig. 7 Structure of UvrD in complex with DNA (PDB: 2IS1). Subdomain 1A, 1B, 2A and 2B are shown in orange, blue, light blue and yellow respectively. Asterisk indicates the truncated C-terminal domain. dsDNA is shown in white. In short, a cycle of helicase/translocase activity of UvrD starts with its interaction with single strand DNA. This induces conformational changes in domains 1B and 2B to form interfaces for duplex DNA binding. In ATP binding state, the cleft formed in between domain 1A and 2B is in the close formation, close to nucleotide. Also as a consequence of the cleft closure, domains 1B and 2B move closure to duplex DNA thus causing DNA duplex onto negatively charged surface and destabilize the duplex formation of DNA. Furthermore, the cleft closure also causes the flipping of bases in the cleft region (in between domain 1A and 2A). This results in a movement of helicase along single strand DNA. This figure is adapted from [25].

Of the 10 linkages identified, three cross-linked peptides were incompatible with the model depicted in Fig. 6. These three cross-links suggest an alternative binding to RNAP. Two cross-links indicate interactions between the β’ bridge helix and UvrD, namely the β’K781-UvrD_{K242} and β’K789-UvrD_{K242} linkages. Although MS analysis in this study is not a quantitative analysis, the large number of ions selected compared with other linkages, suggests a high rate of the cross-linking reaction between β’K789 and UvrD_{K242}. The MS/MS spectrum leaves no doubt about its identification (Fig. 9). The third linkage (β’K213-UvrD_{K135}) indicates binding of UvrD to the tip of the β’ jaw region (Fig. 10). Interestingly, this linkage site is at least 82 Å away from the interactions of UvrD_{K135} at the upstream face of RNAP. So, it is unlikely that UvrD can be cross-linked to β’900 and β’909 and, in the same position, also to β’213. This suggests multiple interactions of UvrD molecules with RNAP, either simultaneously or separately. The β’ bridge helix composed of amino acids 770-805 is an integral component of the catalytic center. It spans the active site and is involved in the nucleotide addition cycle [26, 27]. During the cycle its conformation switches from kinked to straight. The linkages identified in this study suggest that UvrD may interact with the β’ bridge helix and the β’ jaw while binding to dsDNA using its 2B domain, on the downstream face of RNAP. However, it is questionable whether UvrD can be in close contact with the bridge helix under turnover conditions, given the crucial role of this structure in catalysis, with K781 absolutely conserved. Therefore, cross-links with residues K781 and K789 may only be formed under non-turnover conditions. In a recent study, UvrD either in concert with DinG or Rep can remove stalled
RNAP occurred in head-on replication-transcription collisions [28]. It was proposed that UvrD migrates in the 3’-5’ direction on the leading strand template, heading toward the downstream face of transcription unit [28]. Although considerably less is known about how UvrD removes the stalled RNAP, this model supports the interaction and orientation of UvrD on the downstream face of RNAP revealed in our study. A role for UvrD as discussed above for HelD in the removal of DNA binding protein in front of an elongating RNAP is also possible.

Interestingly, the UvrD<sub>K448</sub>-UvrD<sub>K448</sub> cross-link in our dataset (Supplementary Table 3) indicates that two molecules of UvrD can be in close vicinity. Dimerization of UvrD provides an explanation for the two binding sites on RNAP, with one monomer binding to the upstream face and the other interacting at the downstream face of RNAP.

![Diagram showing the interaction of UvrD with RNA polymerase (RNAP) in transcription.](image)

**Fig. 8** Putative roles of UvrD in transcription based on the previously published information integrated with CXMS data presented in this study. In transcription, elongating RNAP complex may encounter events causing it to be stalled on DNA, thus also blocking other processes e.g. DNA replication. In this case, UvrD may be required for removing the stalled RNAP complex. It is proposed that the DNA-binding domain (2B subdomain) of UvrD tightly interacts with the upstream face of RNAP and leaves the catalytic center exposed to the DNA exit channel near the ssDNA. UvrD may bind to the ssDNA template and subsequently perform its 3’-5’ ssDNA translocase activity thus generating force to push the stalled RNAP complex from the DNA. The blue arrow indicates the direction of ssDNA template moved by UvrD. Black dashed arrows indicate the force generated by UvrD 3’-5’ translocase activity towards the stalled RNAP complex.
Fig. 9 The identification of the linkage between RNAP β' K789 (GLADTALKTANSGYLTR) and UvrD K242 (LLAGDTGKVMIVGDDQSIYGR). The theoretical mass of this cross-linked peptide matches with the relative mass of m/z 1097.0761 (4+) precursor ion, at the mass different of 16 ppm. Panel A shows the MS/MS spectrum of the m/z 1097.0761 (4+) precursor ion and identified fragment ions. Different font colors are used for MS/MS product ions resulting from primary fragmentations (black) and secondary fragmentations of a cleaved off unmodified peptides (blue) and peptide with a remnant of BAMG (red). See Chapter 4 for the role of the amino group in the spacer for the gas phase cleavage behavior in collision induced dissociation of cross-linked peptides. Panel B shows the mapped fragments and structure of the cross-linked peptide. This cross-linked peptide was identified by 29 different precursor ions.
Interaction of RNA polymerase-helicases

Fig. 10 Putative roles of UvrD in transcription based on the previously published information integrated with CXMS data presented in this study. In transcription, elongating RNAP complex may encounter head-on replication-transcription collision resulted in stalled RNAP complex. To allow a completion of replication, the stalled RNAP complex needs to be removed. In this study, the interaction of RNAP-UvrD on the downstream side of transcription was illustrated by linkages of UvrD with β’ bridges helix and β’ jaw region. This orientation supports the model that UvrD translocates on the leading strand of the replication fork (in 3’→5’ direction) towards the transcription unit and works in concert with DinG or Rep to remove stalled RNAP [28]. Oval yellow shows the β’ bridge helix.

Conclusion

In this study, our CXMS approach presented in Chapter 4 was applied to characterize the interaction of RNAPs and helicases. The cross-linking experiments were conducted in two separated biological systems i.e. *E. coli* RNAP-UvrD and *B. subtilis* RNA-PcrA. Although the mapping of the interaction site(s) of PcrA with RNAP was not successful in this study, a cross-link between HelD and RNAP and a cross-link between the δ subunit and RNAP were identified. Unfortunately, with only one cross-link and no additional structural data about the interaction, the orientations of HelD and the δ subunit on RNAP are uncertain. However, this study provides the interaction information which is compatible with evidence from functional studies. Intriguingly, UvrD was shown to directly interact with RNAP at two regions i.e. on the upstream and the downstream face of RNAP. These interactions imply novel biological functions of UvrD in the transcription process. The UvrD interaction on the upstream face of RNAP is compatible with that of PcrA. In both orientations UvrD may be involved in removing stalled RNAP from DNA. This study has emphasized the usefulness of CXMS for studying unresolved biological questions. Although the CXMS is a powerful approach for structural analysis, additional information from other techniques, e.g., single particle analysis, x-ray crystallography, yeast two hybrid system, and functional studies are needed for understanding the functional significance of the interactions between helicases and RNAP.

Acknowledgement

We thank Katelyn McGarry and Behrad Ghavim for their contributions in the protein cross-linking analysis of *E.coli* RNAP-UvrD.
Chapter 5

Materials and Methods

Software

Codes for the software tool Yeun Yan were written in Visual Basic language for Applications. The software tool was operated on a Microsoft Visual Basic platform.

Proteins and preparation of RNAP-helicase complex

_E. coli_ RNA and UvrD _E. coli_ were kindly made available by Prof. Evgeny Nudler, Medical center, NewYork University. _B. subtilis_ RNA and _PcrA_ _B. subtilis_ were kindly made available by Dr. Peter Lewis, Newcastle University. Helicases were added to RNAP in 1:1 molar ratio in 50 mM HEPES 150 mM Nacl 10% Glycerol buffer, pH 7.4. The mixtures were incubated at RT for 30 minutes before the start of the cross-linking reaction.

Cross-linking and digestion

The protein concentration of all samples was 0.2 mg/ml, in a buffer of pH 7.4 containing 50 mM HEPES, 150 mM Nacl and 10% lycerol. The cross-link reaction was started by the addition of a solution containing 80 mM BAMG in acetonitrile to obtain a final concentration of 0.4 mM BAMG and 0.5% acetonitrile. The reaction was quenched by adding 1 M Tris-HCl pH 8.0 to a final concentration of 50 mM. Subsequently, the protein was concentrated with 0.5 ml Amicon Ultra 10 kDa cut off centrifugal filters (Millipore). SH-groups were alkylated by addition of a solution of 0.8 M iodoacetamide, followed by the addition of a solution of 9.6 M urea to obtain final concentrations of 40 mM iodoacetamide and 6 M urea, resp. Incubation was for 30 min at room temperature in the dark. The solution was diluted 6 times by the addition of 40 mM Tris-HCl pH 8.0 and digested overnight at 37°C. Peptides were desalted on C18 reversed phase TT3 top tips (Glygen), eluted with 0.1 % TFA in 50% acetonitrile and dried in a vacuum centrifuge.

Diagonal strong cation exchange (SCX) chromatography

Peptides in 50 μl 0.1% TFA in 20% acetonitrile were diluted with 1 ml of a solution containing 10 mM potassium phosphate buffer, pH 2.9 in 20% acetonitrile (buffer A) for loading on a Polysulfethyl aspartamide 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, was performed using a linear gradient from 0 to 250 mM KCl in buffer A over 20 min followed by a gradient over 10 min to 500 mM KCl in buffer A, _primary run_. Absorbance at 214 nm of the effluent was continuously recorded. Ten minutes after the start of the KCl gradient, the effluent was manually collected in 0.2 ml fractions. Collected peptides were desalted on C18 reversed phase TT2 top tips (Glygen), eluted with 0.1 % TFA in 50% acetonitrile and dried in a vacuum centrifuge. For the secondary SCX runs, peptides were dissolved in 20 μl 40 mM TCEP in 20% acetonitrile and incubated for 2 h at 60°C. The reaction medium was diluted with 1 ml buffer A just before loading. Elution occurred under the same conditions as in the _primary run_. After the start of the KCl gradient, material with absorbance at 214 nm was collected after the end of the elution time of the _primary fraction_. Collected effluent was lyophilized, solubilized in 0.1% TFA, and desalted using C18 Ziptips (Millipore). Peptides were eluted from the Ziptips with 0.1% TFA in 50% acetonitrile and dried in a vacuum centrifuge.

Mass spectrometry and Data processing

LC-FTMS/MS data were acquired with an ApexUltra Fourier transform ion cyclotron resonance mass spectrometer (Bruker Daltonic, Bremen, Germany) equipped with a 7T magnet and a nano-electrospray Apollo II DualSource™ coupled to an Ultimate 3000 (Dionex, Sunnyvale, CA, USA) HPLC system. Peptides were injected as a 10 μl 0.1% trifluoroacetic acid, 3% acetonitrile aqueous solution and loaded onto a PepMap100 C18 (5-μm particle size, 100-Å pore size, 300-μm inner diameter x 5 mm length) precolumn. Following injection, the peptides
were eluted via an Acclaim PepMap 100 C18 (3-µm particle size, 100-Å pore size, 75-µm inner diameter x 250 mm length) analytical column (Thermo Scientific, Etten-Leur, The Netherlands) to the nano-electrospray source. A 60 min (early shifted material) and a 120 min gradient profile (late shifted material) was used from 0.1% formic acid / 3% CH₃CN / 97% H₂O to 0.1% formic acid / 50% CH₃CN / 50% H₂O at a flow rate of 300 nL/min. In each MS duty cycle up to 3 data dependent Q-selected peptide ions were fragmented in the hexapole collision cell at an Argon pressure of 6x10⁻⁶ mbar (measured at the ion gauge) using a CID activation kinetic energy profile optimized for different peptide charge states. Target peptides already selected were dynamically excluded for 25 seconds. The resulting fragment ions were detected in the ICR cell at a resolution of up to 60000. The instrument was mass calibrated at 1.5 ppm over an m/z range up to 1400 using the CID fragment ions from off-line electrosprayed human [Glu1]-fibrinopeptide (Sigma-Aldrich). All Raw LC-FTMS/MS data were processed with the MASCOT DISTILLER program, version 2.4.3.3 (64bits), MDRO 2.4.3.1 (MATRIX science, London, UK), including the Search toolbox and the Quantification toolbox. Peak-picking for both MS₁ and MS₂ spectra were optimized for the mass resolution of up to 60000. Peaks were fitted to a simulated isotope distribution with a correlation threshold of 0.75, and an S/N threshold of 2. Processed data were exported in Mascot generic file format (mgf).

Peptide identification

Databases of all possible cross-linked peptides for interrogation using standard search engines were calculated by xComb version 1.2 [15]. For E. coli RNAP and E. coli RNAP-UvrD complex, the forward and reversed sequences of target proteins were uploaded in UniProt FASTA format. For B. subtilis RNAP-PerA, only the forward sequences of target proteins were used. 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 databases were uploaded in Mascot version 2.2. The following parameters were used to identify candidate cross-links based on the processed data files from LC-FTMS/MS experiments: (i) a “do_not_cleave” enzyme with the nonexistent 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 C₆H₅NOS. For nomination by Mascot of cross-linked peptide candidates from E. coli RNAP (i), E. coli RNAP-UvrD (ii) and B. subtilis-PerA complex (iii), only b and y ions without loss of water or ammonia are taken into account. Mass accuracy settings for precursor and fragments in searching by Mascot were 10 ppm and 0.01 Da, resp. for candidates (i), 30 ppm and 0.01 Da resp. for candidates (ii), and 10 ppm and 0.01 Da resp. for candidates (iii). No threshold for the Mascot ions score was taken into account for the nomination of candidates i-iii.

Validation

The Yeun Yan software tool is used for validation of cross-links proposed by xComb approach. Also an ions score is calculated by Yeun Yan to provide a measure for the degree of matching of the experimental MS₁/MS₂ spectrum with the theoretical spectrum. For proposed candidate cross-linked peptides, Yeun Yan calculates the masses of possible b and y fragments, b and y fragments resulting from water loss (b⁰, y⁰) and ammonia loss (b*, y*), fragment ions resulting from cleavage of the amide bonds of the cross-link, and b, b⁰, b*, y, y⁰ and y* fragments resulting from secondary fragmentations of cleavage products. A prerequisite for nomination by Yeun Yan as a candidate and calculation of the corresponding score is the presence of at least one unambiguous y ion per composing peptide. The YY score is calculated according to the equation

\[ YY_{score} = y_A \times y_B \times (f_{tot}/f_{ass})^2 \]

In which; yA and yB are the numbers of unambiguously matching y ions of peptide A and peptide B, respectively. fass is the total number of matching fragment ions, including b and y fragments, b and y fragments resulting from water loss (b⁰, y⁰) and ammonia loss (b*, y*), fragment ions resulting from cleavage of the amide bonds of the cross-link, and b, b⁰, b*, y, y⁰ and y* fragments resulting from secondary fragmentations of cleavage products. ftot is the total number of fragments taken into account for the total number of matching fragments, starting from the fragment ion of highest intensity. An intraprotein candidate for a precursor ion nominated by Yeun Yan is assigned
if the number of unambiguously matching y ions to each of its composing peptides is the same as or higher than the number of unambiguously matching y ions to each of the composing peptides in possible other candidates for that precursor. For a precursor ion of which no intraprotein cross-link was assigned according to the previous criterion, an interprotein cross-link candidate is assigned if it fulfills all of the following criteria (i) the YY score is higher than the YY score of possible other candidates for that precursor; in case two or more candidates have equal highest scores no assignments will be put forward, (ii) the MS2 spectrum contains at least 1 unambiguously matching y ions for each composing peptide and (iii), for E. coli RNAP, the YY score > 0, E. coli RNAP-UvrD complex, the YY score > 0, B. subtilis RNAP-PerA, the YY score is at least 4.

References

13. Harriott, K., The characterisation of the interaction between PcrA and RNA polymerase, in Science and Information Technology, School of Environmental and Life Sciences. 2012, University of Newcastle, Australia.