Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage

Stefan Jennebach1, Franz Herzog2, Ruedi Aebersold2,3,* and Patrick Cramer1,*

1Gene Center and Department of Biochemistry, Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, 81377 Munich, Germany, 2Department of Biology, Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule (ETH) Zurich, Wolfgang-Pauli-Strasse 16, 8093 Zurich and 3Faculty of Science, University of Zurich, Zurich, Switzerland

Received December 13, 2011; Revised February 1, 2012; Accepted February 21, 2012

ABSTRACT

RNA polymerase (Pol) I contains a 10-subunit catalytic core that is related to the core of Pol II and includes subunit A12.2. In addition, Pol I contains the heterodimeric subcomplexes A14/43 and A49/34.5, which are related to the Pol II subcomplex Rpb4/7 and the Pol II initiation factor TFIIF, respectively. Here we used lysine-lysine crosslinking, mass spectrometry (MS) and modeling based on five crystal structures, to extend the previous homology model of the Pol I core, to confirm the location of A14/43 and to position A12.2 and A49/34.5 on the core. In the resulting model of Pol I, the C-terminal ribbon (C-ribbon) domain of A12.2 reaches the active site via the polymerase pore, like the C-ribbon of the Pol II cleavage factor TFIIS, explaining why the intrinsic RNA cleavage activity of Pol I is strong, in contrast to the weak cleavage activity of Pol II. The A49/34.5 dimerization module resides on the polymerase lobe, like TFIIF, whereas the A49 tWH domain resides above the cleft, resembling parts of TFIIE. This indicates that Pol I and also Pol III are distantly related to a Pol II–TFIIS–TFIIF–TFIIE complex.

INTRODUCTION

RNA polymerase (Pol) I synthesizes the ribosomal RNA precursor in the nucleolus of eukaryotic cells (1,2). Pol I is a 14-subunit, 589 kDa enzyme that consists of a 10-subunit core and two peripheral heterodimeric subcomplexes, A14/43 and A49/34.5. The Pol I structure was investigated by electron microscopy (3,4), and a homology model for the Pol I core was derived from the related Pol II structure. Crystal structures are available for A14/43, which resembles the Pol II subcomplex Rpb4/7, for the dimerization module of A49/34.5, which resembles part of the Pol II transcription factor (TF) IIF, and for the C-terminal tandem winged helix (tWH) domain of A49, which may be related to parts of TFIIE (4,5).

In the absence of a crystal structure for the complete Pol I, three open questions remain on the enzyme’s domain architecture (Figure 1). First, what is the location of the core subunit A12.2? A12.2 contains two zinc ribbon domains that are homologous to those in the Pol II subunit Rpb9. However, the C-terminal zinc ribbon (C-ribbon) of A12.2 is also homologous to the C-ribbon of the Pol II-associated factor TFIIS (6). TFIIS stimulates RNA cleavage by inserting into the Pol II pore and complementing the active site (7–9). Since the A12.2 C-ribbon is required for strong RNA cleavage activity of Pol I (3), does it also bind the pore? Second, what is the location of the A49/34.5 dimerization module? Does it correspond to the location of the TFIIF dimerization module on Pol II (10,11), indicating a functional similarity to TFIIF? Third, is there a defined location of the A49 tWH domain, and does this location support a functional similarity to a region in TFIIE?

Here we addressed these questions by lysine–lysine crosslinking of Pol I, identification of the crosslinked sites by mass spectrometry (MS), and molecular modeling based on X-ray crystallographic information. Such crosslinking-MS analysis has become a powerful tool to study the domain architecture of very large multiprotein complexes (13). Proximal lysine residues in neighboring protein subunits of a multiprotein complex are crosslinked with a bivalent chemical reagent, and the crosslinked sites are identified by mass spectrometry after protein digestion. The crosslinked sites can then be used to position known crystal structures of subcomplexes with respect to each other and derive the three-dimensional architecture of very large assemblies.

*To whom correspondence should be addressed. Tel: +41 44 633 31 70; Fax: +41 44 633 10 51; Email: aebersold@imsb.biol.ethz.ch
Correspondence may also be addressed to Patrick Cramer. Tel: +49 89 2180 76965; Fax: +49 89 2180 76999; Email: cramer@genzentrum.lmu.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2012. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
were crosslinked using isotope-labeled disuccimidyl suberate (DSS-d0/d12, Creative Molecules Inc.). To determine the optimal ratio of DSS to Pol I, we mixed 8 μg of Pol I with 25 mM DSS dissolved in dimethylformamide (Pierce Protein Research Products) to a final crosslinker concentration of 0.02, 0.04, 0.08, 0.16, 0.4, 1, 2 or 4 mM (DSS concentration refers to the concentration of one isotope) and analyzed the products using SDS-PAGE (Figure 2A). The optimum concentration of DSS was considered to result in a higher molecular weight band, disappearance of most bands for Pol I subunits and no evidence for Pol I oligomers that would remain in the gel pockets. We decided to use a final concentration of 0.6 mM DSS (1.2 mM DSS for the Pol I–Rrn3 complex) and incubated for 30 min at 30°C. The reaction was stopped by addition of 1 M NH₄HCO₃ to a final concentration of 100 mM and incubation for 30 min at 30°C. To improve the crosslink yield, an additional measurement at 3.5 mM DSS was performed. The sample crosslinked with 3.5 mM, DSS was quenched with Superose 6 buffer containing 100 mM NH₄HCO₃ and subjected to size-exclusion chromatography, to remove Pol I oligomers.

Mass spectrometry analysis

The crosslinked proteins were linearized by addition of two volumes 8 M urea, reduced using 5 mM Tris(2-carboxyethyl)phosphine (TCEP), and alkylated with 10 mM iodoacetamide. The sample was digested using trypsin following standard protocols. Purified samples were reconstituted in 20 μl of size-exclusion chromatography (SEC) mobile phase buffer containing 70% (v/v) water, 30% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid (TFA). Fifteen microliters were applied on a Superdex Peptide PC 3.2/30 column at a flow rate of 50 μl/min. For LC-MS/MS, fractions of interests (retention volumes 0.9–1.4 ml) were pooled and evaporated to dryness. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis was carried out on an Eksigent 1D-NanoLC-Ultra system connected to a LTQ Orbitrap XL mass spectrometer (Thermo Scientific) equipped with a standard nanoelectrospray source. Fractions from SEC were reconstituted in mobile phase buffer containing 97% (v/v) water, 3% (v/v) acetonitrile and 0.1% (v/v) formic acid. The injection volume was chosen according to the 215 nm absorption signals from SEC separation. A fraction corresponding to an estimated amount of 1 μg was loaded onto a 11 cm × 0.075 mm I.D. column pre-packed with Michrom Magic C₁₈ material (3 μm particle size, 200 Å pore size) (Michrom Bioresources, Inc.). Peptides were separated at a flow rate of 300 nL/min using a stepwise gradient from 0.05% (v/v) to 92% (v/v) acetonitrile.

Ion source and transmission setting of the mass spectrometer were as follows: spray voltage 2 kV, capillary temperature 200°C, capillary voltage 60 V and tube lens voltage 135 V. The mass spectrometer was operated in data-dependent mode, selecting up to five precursors from a MS(1) scan (resolution 60 000) in an m/z range from 350 to 1600 for collision-induced dissociation

Figure 1. Structural information on Pol I. The 9-subunit core enzyme is modeled based on the Pol II structure and shown as a gray surface. The structure of the A12.2 homolog Rpb9 is shown in orange on the left near its presumed binding surface. Note that its C-terminal zinc ribbon domain is also homologous to TFIIIS (not shown). The crystal structures of the two domains of the Pol I subcomplex A49/A34.5 (4) are shown on the top (slate blue, A49 dimerization domain; light pink, A34.5 dimerization domain; blue, A49 tWH domain). The crystal structure of subcomplex A14/43 that is related to the Pol II subcomplex Rpb4/7 (3,12) is on the right near its presumed docking site below the clamp.

Recently, we used this approach to locate the Pol I-specific initiation factor Rrn3 on Pol I (14). Our results reveal the domain architecture of Pol I and provide answers to the above questions. They show that the A12.2 C-ribbon domain can reside in the Pol I pore, like the TFIIIS C-ribbon domain in Pol II, that the A49/34.5 dimerization module resides on the Pol I lobe, like the TFIIIF module in Pol II, and that the A49 tWH domain is flexible and can reside above the cleft, to some extent resembling TFIIE in the Pol II system. These results provide structural and functional relationships between Pol I domains and their Pol II counterparts, and explain why Pol I has strong intrinsic RNA cleavage activity, whereas Pol II does not. A comparison with recent topological data on Pol III reveals a conserved functional architecture of all three eukaryotic RNA polymerase machineries.

MATERIALS AND METHODS

Pol I preparation and crosslinking

Endogenous 14-subunit Pol I was purified as described (3) except that the size-exclusion chromatography step was performed in 20 mM HEPES, pH 7.8, 1 mM MgCl₂, 300 mM potassium acetate, 10% glycerol, and 5 mM DTT (Superose 6 buffer). Pol I-containing fractions were pooled and concentrated to 0.8–1.0 mg/ml. A Pol I–Rrn3 complex was purified as described (14). Pol I complexes
Single- and double-charged precursor ions and precursors of unknown charge states were rejected. CID was performed for 30 ms using 35% normalized collision energy and an activation of $q = 0.25$. Dynamic exclusion was activated with a repeat count of 1, an exclusion duration of 30 s, list size of 300 and a mass window of ± 50 ppm. Ion target values were 1 000 000 (or maximum 500 ms fill time) for full scans and 10 000 (or maximum 200 ms fill time) for MS/MS scans, respectively.

Database searching

For data analysis, Thermo Xcalibur .raw files were converted to the open mzXML format with ReAdW version 4.3.1 using default settings. The mzXML files were used as input for xQuest searches (15). For the following MzXML2Search [part of Trans-Proteomics Pipeline, (16)], the files were converted to the .mgf (Mascot generic file) format. MzXML2Search was executed with the option "-T10000" to export precursors with a mass above the default value of 4200 Da. Unmodified peptides from the protein mix were identified by searching an in-house Mascot server (ver. 2.3.0) against the Uniprot/SwissProt database using the following parameters: maximum number of missed cleavages = 2, taxonomy = chordata, fixed modifications = carbamidomethyl-Cys, variable modification = Met oxidation, MS1 tolerance = 15 ppm, MS2 tolerance = 0.6 Da, instrument type = ESI-TRAP and decoy mode set to on. For validation, the peptide probability was set to $P < 0.05$ and additional filters were used (require bold red = yes, peptide score $> 20$).

Crosslinked peptides and peptide monolinks were identified using an in-house version of the dedicated search engine xQuest using an advanced scoring model (Walzthoeni et al., in preparation). Tandem mass spectra of precursors differing in their mass by 12.075321 Da (mass difference of DSS-d0 and DSS-d12) were paired if they had a charge state of 3+ to 8+ and were triggered above the default value of 4200 Da. Unmodified peptides from the protein mix were identified by searching an in-house Mascot server (ver. 2.3.0) against the Uniprot/SwissProt database using the following parameters: maximum number of missed cleavages = 2, taxonomy = chordata, fixed modifications = carbamidomethyl-Cys, variable modification = Met oxidation, MS1 tolerance = 15 ppm, MS2 tolerance = 0.6 Da, instrument type = ESI-TRAP and decoy mode set to on. For validation, the peptide probability was set to $P < 0.05$ and additional filters were used (require bold red = yes, peptide score $> 20$).

Crosslinked peptides and peptide monolinks were identified using an in-house version of the dedicated search engine xQuest using an advanced scoring model (Walzthoeni et al., in preparation). Tandem mass spectra of precursors differing in their mass by 12.075321 Da (mass difference of DSS-d0 and DSS-d12) were paired if they had a charge state of 3+ to 8+ and were triggered above the default value of 4200 Da. Unmodified peptides from the protein mix were identified by searching an in-house Mascot server (ver. 2.3.0) against the Uniprot/SwissProt database using the following parameters: maximum number of missed cleavages = 2, taxonomy = chordata, fixed modifications = carbamidomethyl-Cys, variable modification = Met oxidation, MS1 tolerance = 15 ppm, MS2 tolerance = 0.6 Da, instrument type = ESI-TRAP and decoy mode set to on. For validation, the peptide probability was set to $P < 0.05$ and additional filters were used (require bold red = yes, peptide score $> 20$).

Crosslinked peptides and peptide monolinks were identified using an in-house version of the dedicated search engine xQuest using an advanced scoring model (Walzthoeni et al., in preparation). Tandem mass spectra of precursors differing in their mass by 12.075321 Da (mass difference of DSS-d0 and DSS-d12) were paired if they had a charge state of 3+ to 8+ and were triggered above the default value of 4200 Da. Unmodified peptides from the protein mix were identified by searching an in-house Mascot server (ver. 2.3.0) against the Uniprot/SwissProt database using the following parameters: maximum number of missed cleavages = 2, taxonomy = chordata, fixed modifications = carbamidomethyl-Cys, variable modification = Met oxidation, MS1 tolerance = 15 ppm, MS2 tolerance = 0.6 Da, instrument type = ESI-TRAP and decoy mode set to on. For validation, the peptide probability was set to $P < 0.05$ and additional filters were used (require bold red = yes, peptide score $> 20$).
within 2.5 min of each other. These spectra were then searched against a preprocessed .fasta database. For the protein mixture, the database contained the UniProt/SwissProt entries of the target proteins. xQust search parameters were set as follows: maximum number of missed cleavages (excluding the crosslinking site) = 2, peptide length = 4-40 amino acids, fixed modifications = carbamidomethyl-Cys (mass-shift = 57.02146 Da), mass shift of the light crosslinker = 138.06808 Da, mass-shift of monolinks = 156.078644 and 155.096428 Da, MS\textsuperscript{1} tolerance = 15 ppm, MS\textsuperscript{2} tolerance = 0.2 Da for common ions and 0.3 for crosslink ions, search in enumeration mode (exhaustive search). The search results were filtered (MS\textsuperscript{1} tolerance window = -4 to +7 Da), and all spectra were manually validated. Identifications were only considered for the result list when both peptides had at least four bond cleavages in total or three adjacent ones, respectively, and a minimum length of six amino acids.

RESULTS

Crosslinking-MS analysis of 14-subunit Pol I

To address outstanding questions on the functional architecture of Pol I, we carried out crosslinking-MS analysis for Pol I from the yeast Saccharomyces cerevisiae. Endogenous Pol I was purified as described (3,17) except that the final size-exclusion chromatography step was carried out in the presence of potassium acetate (Materials and Methods). For crosslinking, 110 µg of Pol I were incubated with isotope-labeled disuccinimidyl suberate (DSS, Creative Molecules, Inc.). DSS reacts with primary amines in lysine side chains and protein N-termini. Crosslinking efficiency was monitored by SDS-PAGE (Figure 2A). After digestion with trypsin, crosslinked peptides were enriched by size-exclusion chromatography, and peptides and their fragments were detected with high-resolution MS (Materials and Methods). After initial measurements at 0.6 mM DSS, we performed an additional measurement using 3.5 mM DSS, in order to obtain additional crosslinks (Materials and Methods), and combined the data in our final analysis. Crosslink data within Pol I that were obtained from a Pol I–Rrn3 complex (14) were also included in our analysis, because the addition of Rrn3 did not alter the crosslinking pattern on Pol I. Measurement of the four samples, of which two were Pol I–Rrn3 complexes, resulted in 1047 mass spectra that matched crosslinked peptides, an example of which is shown in Figure 2B.

Confirmation of the Pol I core model

We identified 239 unique linkage pairs within nine subunits of the Pol I core (subunits A190, A135, AC40, AC19, Rpb5, Rpb6, Rpb8, Rpb10, Rpb12; excluding A12.2, A49/34.5, and A14/43) (Figure 3A). We analyzed the crosslinked residues with the atomic Pol II structure (18) and the Pol I core homology model (3). We assumed that the distance between C\textsubscript{z} atoms of crosslinked lysines must be ≤30 Å, corresponding to the length of the DSS spacer (11.4 Å) plus two times the length of a lysine side chain (6.5 Å) plus an estimated coordinate error of 3 Å for flexible lysine side chain ends. Crosslinking sites that fell in regions that adopt the Pol II fold (3) were assigned to Category A (Table 1). Sites outside these regions were assigned to Category B. Of the 239 crosslink pairs, 73 (30.5%) comprised only Category A sites (A–A pairs), 80 (33.5%) contained one Category B site (A–B pairs) and 86 (36%) were B–B pairs. Of the 73 A–A pairs, 70 could be analyzed (7) as both crosslinked residues were present in the structure. Among the 70 A–A crosslink pairs, 68 (97.1%) fell within the acceptable C\textsubscript{z} distance of ≤30 Å (Figure 2C). The two remaining pairs exceeded the maximum distance by only 1.9 Å and 4.3 Å, respectively, and this can be explained by structural flexibility. One crosslink involves the bridge helix, which undergoes conformational changes (19–21), whereas the other involves the clamp, which is mobile in Pol I (3) and Pol II (19,20,22). We additionally obtained crosslinks within the A14/43 subcomplex, and between A14/43 and the Pol I core, which were consistent with the previously obtained structure and location of A14/43 (3,12). These results demonstrate the validity of our method and confirm the previous Pol I model.

Model extension reveals a unique jaw region

To extend the Pol I model, we analyzed Category A–B and B–B crosslink pairs (Table 1, Figure 3A). A–B pairs connect residues in regions of the homology model that share the Pol II fold (Category A) with residues in sequence regions with no or very weak conservation (Category B). Within the nine core subunits, we observed 80 A–B crosslinks, of which 15 could not be analyzed as they contain residues within specific insertions or residues that are not present in the Pol II structure. A total of 39 A–B pairs (60%) showed C\textsubscript{z} distances below 30 Å. The involved Category B residues were reclassified as B*, and their surrounding region (overall 125 residues within A190, A135 and AC40) was included in the Pol I homology model, as it was likely that the region containing the B* site adopts a Pol II-like fold. Of the resulting 23 B*–B crosslinks, 17 pairs contained lysine residues present in the Pol II structure, and two showed a permissible C\textsubscript{z} distance. Based on these findings, we extended the Pol I core model to parts of the clamp core, pore, funnel, foot, dock and cleft domains of the largest subunit A190, to small parts of the lobe, fork and wall domains of the second largest subunit A135, and to parts of domain 2, the loop domain, and the dimerization domain in subunit AC40, the counterpart of the Pol II subunit Rpb3 (Figure 3A, Supplementary Figure S1). The extended homology model relates 69.6% of the nine core Pol I subunit sequences to their Pol II counterparts, although the large Pol I subunits A190 and A135 show only 25.5% and 25.6% sequence identity to their Pol II counterparts, respectively.

Of the 166 A–B and B–B crosslinks within the Pol I core, 95 could be analyzed with the Pol II structure, of which 43 (26 A–B and 17 B–B crosslinks) did not fall within the C\textsubscript{z} distance restraint of 30 Å (Table 1). Of these crosslinks, 35 involve residues that fall in a region that may correspond to the Pol II jaw domain (A190...
residues 1252–1487). This region is conserved among yeast species but is poorly conserved in higher eukaryotes. Secondary structure prediction (23) indicates that this region is only partially related to the Pol II jaw. Residues 1251–1337 are predicted to form three helices and two β-strands, residues 1338–1438 are apparently unstructured, and residues 1439–1495 may form two helices and two β-strands. Crosslinks to residues K1260,
A12.2 C-ribbon binds Pol I like the Rpb9 C-ribbon binds Pol II, only one out of 12 crosslinks falls within the allowed Cz distance (Figure 4C). These results show that the A12.2 N-ribbon binds the Pol I surface similar to the N-ribbon of Rpb9, whereas the A12.2 C-ribbon binds the Pol I pore, like the TFIIS C-ribbon binds Pol II.

**49/34.5 binds the lobe like TFIIF**

Our data also position the dimerization module of the A49/34.5 subcomplex (4) on Pol I. We obtained a total of 92 crosslinks involving the Pol I subcomplex A49/34.5 (Figure 5A). Within the dimerization module, 13 out of the 14 crosslinks agreed with the A49/34.5 dimerization module structure (PDB 3NFF) (4) (Figure 5B). The distance for one crosslink pair was slightly above the limit, but this was likely due to a difference in structures, since the structure was obtained from a different species, *Candida glabrata*. Crosslinks between the dimerization module and the Pol I core indicate that the module is positioned on one side of the Pol I cleft on the lobe domain of A135. A crosslink between A49 residue K116, which is seven residues beyond the A49 C-terminal residue in the dimerization module structure (PDB 3NFF), connects to the protrusion of A135. Lysines located 6 and 12 residues beyond the A34.5 C-terminal residue in the structure crosslink to the external domains of A135, which are adjacent to the lobe. All these crosslinks can be explained when we assume that the A49/34.5 dimerization module occupies the location of the TFIIF dimerization module on the lobe of Pol II (10), but a 30° rotation of the module structure led to an even better fit (Figure 6A).

**The A49 tWH domain lies above the cleft**

We obtained seven crosslinks within the tWH domain of A49 that were consistent with the structure of the isolated domain (4) (Figure 5C). We further observed five crosslinks between the tWH domain and the Pol I core, namely to the lobe and protrusion of A135 on one side of the cleft, and to the Pol I-specific insertion in the clamp head domain of A190 on the other side (Figure 6B). Extended alignments (Supplementary Figure S1) based on HHpred predictions (24) position the crosslinked residues K259 and K267 of A190 6 and 14 residues beyond strand β4 in a Pol I-specific 52 amino acid insertion in the β4–β5 loop of the clamp head (20). Additionally, the A49 residue K170, which is 15 residues N-terminal of the first-ordered tWH residue, crosslinks to the dimerization domain of A34.5, indicating proximity between the dimerization module and the tWH domain. The location of the tWH domain over the cleft is consistent with a role of this domain in DNA binding (4), although a repositioning of this domain is required during promoter DNA loading into the cleft. It also corresponds to the location of the Pol III subunit C34 (25), which may be evolutionary related to the A49 tWH domain (4), and is similar to the position of TFIIE on the Pol II clamp (26).

**The A12.2 C-ribbon binds the pore like TFIIIS**

Our crosslinking data provide the desired topological insights into the enigmatic A12.2 subunit, which is essential for the strong intrinsic cleavage activity of Pol I (3). We identified a total of 27 crosslinks that contained sites in A12.2 (Figure 4A). Of those, 8 were intra-subunit crosslinks and 19 were crosslinks between A12.2 and other Pol I subunits, which were analyzed with the coordinates of Rpb9 and TFIIIS from the Pol II–TFIIS complex structure (PDB 1Y1V) (7). The intra-subunit crosslinks can be explained with the Rpb9 structure and those in the C-ribbon also with the TFIIIS structure. Crosslinks between residue K46 of A12.2, which is located five residues beyond the N-ribbon, and two A190 residues (K1459 and K1473) in the Pol I jaw (Figure 4A) and residue K298 in the lobe of A135 indicate a position of the A12.2 N-ribbon similar to that of the Rpb9 N-ribbon on Pol II. Four crosslinks to K64, K68 and K73 fall within the linker between the N- and C-ribbon, which likely follows the path taken by the linker of TFIIIS. The remaining 12 inter-subunit crosslinks unambiguously position the A12.2 C-ribbon in the Pol I pore. If one assumes that the A12.2 C-ribbon binds in the pore like the TFIIIS C-ribbon, 11 of the 12 crosslinks fall within the maximum distance restraint, and one crosslink exceeds the limit by only 2 Å (Figure 4B). In contrast, if one assumes that the
head domain (residues K111/K309 and K250/K267 in the clamp head and the Pol I-specific insertion, respectively) and with residue K1012 in the bridge helix near the polymerase active center (Figure 6C). This location of the A49 linker is consistent with the location of the corresponding region in the largest TFIIF subunit on Pol II (10,26). One crosslink connects the A49 linker to Rpb5 and another one to residue K5 in the mobile N-terminal tail of subunit A43, consistent with a genetic interaction between A49 and the N-terminus of A43 (27). If extended, K5 in the mobile tail could be up to 60 Å away from the first-ordered residue in the A43 structure (PDB 2RF4) (3,12). Only a small number of crosslinks are observed to the C-terminal extension of A34.5, including crosslinks to subunits AC40 and AC19 (Figure 6C). The broad distribution of crosslinks involving the A49 linker around the central cleft indicate that this region is mobile, consistent with NMR and circular dichroism measurements, which show that the linker is unfolded in isolation (S. Geiger and P. Cramer, unpublished data.)

DISCUSSION

Here we used lysine–lysine crosslinking, mass spectrometry, and modeling based on crystallographic structures, to unravel the domain architecture of Pol I. The data allowed for an extension of the previous model of the Pol I core, confirmed the location of the A14/43 subcomplex on the core and positioned the remaining subunit A12.2 and the two domains of the peripheral subcomplex A49/34.5 on the core (Figure 7). From these data, a view emerges that Pol I is evolutionarily related to a partial Pol II–TFIIS–TFIIF–TFIIE complex. The relationship extends to Pol III, which contains the A12.2-related subunit C11 that is required for RNA cleavage (28), a heterodimeric subcomplex, C37/53, which is related to TFIIF and A49/34.5 (29–31), and an additional subcomplex, C82/34/31, which is related to TFIIE (32,33). Below we discuss the three key findings from this work.

First, the C-ribbon of A12.2 can reside in the Pol I pore and corresponds to the C-ribbon of TFIIS, although it is also homologous to the C-ribbon of Rpb9. The A12.2
C-ribbon contains the charged residues R102, D105 and E106 at its tip, which correspond to TFIIS residues R287, D290 and E291, which complement the Pol II active site and induce strong RNA cleavage (9). These results are consistent with recent mutagenesis data indicating that the corresponding Pol III subunit C11 also enters the pore with its C-ribbon to induce strong RNA cleavage (6). The results thus explain the role of A12.2 and C11 in transcript cleavage (28,34), suggest a close evolutionary relationship between Pol I and Pol III, and confirm that Pol I and Pol III differ from Pol II in their mode of RNA cleavage (6).

Second, the A49/34.5 dimerization module is located on the lobe of Pol I, and the linker of subunit A49 reaches into the central cleft of the enzyme. These findings are consistent with the localization of the corresponding TFIIF dimerization module on the lobe of Pol II and the TFIIF linker in the cleft (10,11). The conserved location of the dimerization modules in all three polymerases is consistent with a similar function of the TFIIF-like subcomplex in transcription (10,29). The observed stimulatory effect of the dimerization module on RNA cleavage (3,4) can now be explained as an indirect effect resulting from its proximity to subunit A12.2, which is essential for cleavage (3) and likely tends to dissociate when A49/34.5 is lacking. This model is supported by the observation that deletion of C37 from Pol III leads to a loss of C11 (30).

Finally, the data indicate that the mobile A49 C-terminal tWH domain can reside above the cleft. This position is similar to that of C34 in the Pol III system, as revealed by cryo-EM (25,35). Bioinformatic analysis (36) and homology modeling (4) suggested an evolutionary

---

**Figure 5.** Crosslinks confirm A49/A34.5 domain structures. (A) Schematic overview of crosslinks involving the A49/A34.5 subcomplex. A49 and A34.5 domains of known structure are labeled. The color code is as in Figures 1 and 3A. (B) Crosslinks (green and red lines) are consistent with the known A49/A34.5 dimerization module structure (PDB 3NFG). Cα atoms of crosslinked residues are depicted as slate and pink spheres for A49 and A34.5, respectively. (C) Crosslinks are consistent with the structure of the A49 tWH domain (PDB 3NFI). Cα atoms of crosslinked residues are represented as spheres.
relationship of C34 and the A49 tWH domain to the β subunit of TFIIE. Since TFIIIE crosslinks to the clamp of Pol II (26), the A49 tWH domain, C34 and TFIIIE can all adopt similar locations on their respective polymerase cores. Consistent with this, the A49 tWH domain binds single-stranded DNA and may have a role in promoter binding and/or opening (4). Since the Pol III subunit C82 also binds single-stranded DNA (33) and the
archaeal TFIIE homolog TFE stabilizes an open promotor complex (37,38), we suggest that the distantly related A49 tWH domain, the Pol III subcomplex C82/34/31 and TFIIE share an old function in binding the melted DNA region above the active center cleft in an open promotor complex during initiation. Prior loading of the DNA into the cleft may be enabled by the observed mobility of these proteins.

Comparison of the crosslinking data presented here with previous EM data on Pol I strongly suggests that the A49/34.5 subcomplex, like its counterpart TFIIF, maintains a considerable degree of mobility on the polymerase surface. In a cryo-EM reconstruction at 12 Å resolution, densities were observed spanning from the funnel of Pol I to the AC19/40 heterodimer, consistent with some promotor complex interactions with NTP and TFIIS. Mol. Cell., 131, 583–594.


