



Published in final edited form as:

Nat Commun. ; 5: 5132. doi:10.1038/ncomms6132.

## The X-ray crystal structure of the euryarchaeal RNA polymerase in an open clamp configuration

Sung-Hoon Jun<sup>1</sup>, Akira Hirata<sup>1,2</sup>, Tamotsu Kanai<sup>3</sup>, Thomas J. Santangelo<sup>4</sup>, Tadayuki Imanaka<sup>5</sup>, and Katsuhiko S. Murakami<sup>1</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, The Center for RNA Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA

<sup>2</sup>Department of Materials Science and Biotechnology, Graduate School of Science and Engineering, Ehime University, Matsuyama, Ehime, Japan

<sup>3</sup>Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto, 615-8510 Japan

<sup>4</sup>Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523, USA

<sup>5</sup>Department of Biotechnology, Ritsumeikan University, Noji-Higashi, Kusatsu 525-8577, Japan

### Abstract

The archaeal transcription apparatus is closely related to the eukaryotic RNA polymerase II (Pol II) system. Archaeal RNA polymerase (RNAP) and Pol II evolved from a common ancestral structure and the euryarchaeal RNAP is the simplest member of the extant archaeal/eukaryotic RNAP family. Here we report the first crystal structure of euryarchaeal RNAP from *Thermococcus kodakarensis* (*Tko*). This structure reveals that the clamp domain is able to swing away from the main body of RNAP in the presence of the Rpo4/Rpo7 stalk by coordinated movements of these domains. More detailed structure-function analysis of yeast Pol II and *Tko* RNAP identifies structural additions to Pol II that correspond to the binding sites of Pol II-specific general transcription factors including TFIIF, TFIID and Mediator. Such comparisons provide a framework for dissecting interactions between RNAP and these factors during formation of the pre-initiation complex.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

To whom correspondence should be addressed: Katsuhiko S. Murakami (kum14@psu.edu), Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA; Akira Hirata (hirata.akira.mg@ehime-u.ac.jp), Department of Materials Science and Biotechnology, Graduate School of Science and Engineering, Ehime University, Matsuyama, Ehime, Japan.

#### Author contributions

A.H. and K.S.M. initially designed research. T.K., T.J.S. and T.I. contributed to experimental design. S.-H.J. crystallized and solved the *Tko* RNAP structure. A.H. and K.S.M. crystallized and solved the structures of *Tko* Rpo3/Rpo11 and *Tko* Rpo4/Rpo7 subcomplexes, respectively. S.-H.J., A.H., T.J.S. and K.S.M. wrote the manuscript. All authors discussed the results and commented on the manuscript.

#### Additional information

**Accession codes:** Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes: 4QIW, *Tko* RNAP; 4QJV, Rpo3/Rpo11; 4QJF, Rpo4/Rpo7.

**Competing financial interests:** The authors declare no competing financial interests.

## Introduction

Archaeal proteins associated with genome maintenance and gene expression have extensive functional and structural similarities with their eukaryotic counterparts<sup>1,2</sup>. This congruence is especially true for the archaeal transcription machinery and there is striking structural similarity between archaeal and eukaryotic RNAPs<sup>3-5</sup>. Comparing the pre-initiation complex (PIC) formation of the archaeal and three eukaryotic transcription systems (Pol I, II and III) revealed that all RNAPs use a core subset of structurally and functionally related transcription factors to initiate promoter-dependent transcription<sup>6</sup>. All factors are auxiliary for the archaeal and Pol II transcription systems, but some factors are *bona fide* RNAP subunits for the Pol I and Pol III transcription systems. Archaeal RNAP is the most closely related to Pol II in subunit composition and their requirements for general transcription factors (GTFs) exactly match a subset of GTFs required for the activities of Pol II. Archaeal RNAP requires only two monomeric GTFs – TBP and TFB – for PIC formation and transcription *in vitro*, although a third monomeric GTF – TFE – can assist PIC formation *in vitro* and appears essential factor *in vivo*<sup>1,3,7,8</sup>. PIC formation with Pol II requires a more complex set of GTFs, with minimally 6 GTFs (TFIIA, TFIIB, TFIID/TBP, TFIIE, TFIIIF and TFIIF) and Mediator complex is also required for promoter-specific transcription<sup>7,9,10</sup>.

Archaea consists of two major phyla, Euryarchaeota and Crenarchaeota, and phylogenetic analyses of the essential components in DNA replication, transcription and translation suggested that Euryarchaeota have retained a set of features that more likely represent the ancestral form present in the last common ancestor of eukaryotes and archaea<sup>11,12</sup>. Euryarchaeal RNAP is composed of 11 subunits and all subunits are conserved in the archaeal/eukaryotic RNAP family (Supplementary Table 1) whereas crenarchaeal RNAP contains two additional subunits Rpo8 and Rpo13<sup>13,14</sup>. Although structural and functional similarities between archaeal and Pol II transcription machineries have been known for decades, precise comparison of these RNAPs generate new insights about structural motifs of RNAP that participate in the assembly of the PIC and transcription regulation.

We report the crystal structure of *Thermococcus kodakarensis* (*Tko*) RNAP at 3.5 Å resolution, which reveals the molecular details of the open clamp state of the RNAP in the presence of the Rpo4/Rpo7 stalk. Structure-guided sequence alignment between *Tko* RNAP and yeast Pol II postulate how retained insertions and modifications to Pol II during RNAP evolution have been utilized to establish interactions with Pol II-specific GTFs and Mediator. Our structure-function analysis provides insight regarding the evolution of multi-subunit RNAPs with their binding factors and also serves as a guide for studying the physical interactions between Pol II and transcription regulators.

## Results

### *Tko* RNAP purification and crystallization

The phylogenetic analysis of the largest subunit of cellular RNAPs indicates that among Euryarchaeota, Thermococcales including *Tko* is the closest forms of RNAP to the common ancestor of the archaeal/eukaryotic RNAP family (Fig. 1). Therefore, *Tko* RNAP can be

used as an ideal reference to analyze the structure and evolution of archaeal/eukaryotic RNAP family<sup>15</sup>. *Tko* RNAP purified directly from cells contains sub-stoichiometric amounts of TFE<sup>16</sup> and this heterogeneity likely precluded crystallization attempts. *Tko* RNAP purified from a *rpo4* strain yields an enzyme that lacks Rpo4, Rpo7 and TFE<sup>16</sup>. Introduction of recombinant Rpo4 and Rpo7 into this TFE-free RNAP reformed the full 11-subunit enzyme (Supplementary Fig. 1) that could be crystallized successfully. The structure was determined by molecular replacement using the *Sulfolobus solfataricus* RNAP structure (PDB ID 3HKZ)<sup>1</sup> as a search model. We also solved the high-resolution structures of heterodimers formed by *Tko* RNAP subunits including Rpo3/Rpo11 (1.6 Å) and Rpo4/Rpo7 (2.3 Å) (Supplementary Table 2), and replacement with these structures allowed refinement of the final structure of *Tko* RNAP at 3.5 Å resolution with high quality (Supplementary Fig. 2 and Supplementary Table 2).

### The *Tko* RNAP structure

The overall shape of *Tko* RNAP resembles the crenarchaeal RNAP and eukaryotic Pol I and Pol II (Fig. 2). All subunits of *Tko* RNAP are conserved in archaeal/eukaryotic RNAPs supporting that the *Tko* RNAP structure represents the closest form to their common ancestor (Fig. 2c). Superposition of the *Tko* RNAP structure with the *Sso* RNAP and yeast Pol II structures, both captured in the closed clamp conformation<sup>17,18</sup>, reveals that the *Tko* RNAP clamp is in an open state (Fig. 3a). In the *Tko* RNAP structure, the position of DNA binding clamp (Rpo1' residues 1–322, Rpo1'' residues 332–391 and Rpo2 residues 1058–1123) is widely opened and hinged away from the main channel. The *Tko* RNAP structure fits nicely into the cryo-EM map of the closely related *Pyrococcus furiosus* (*Pfu*) RNAP<sup>19</sup> (Fig. 1 and Supplementary Table 1) with the exception of the DNA binding clamp (Fig. 3b). This difference adumbrates the archaeal/eukaryotic RNAPs can readily adopt different clamp conformations in solution. The clamp of *Tko* RNAP swings away from the main channel and undergoes a clockwise rotation of ~21.3° compared with the clamp position in *Sso* RNAP (Fig. 3c). The repositioning of the clamp – termed opening – is coupled with the movement and counterclockwise rotation of Rpo4/Rpo7 stalk of ~12°, which allows the clamp to open without a steric hindrance with the stalk (Supplementary Movie 1). This concerted movement resolves, in molecular detail, two concerns raised from the interpretations of the crystallographic studies of yeast Pol II: 1) the suggestion that the clamp may only be opened in the absence of the stalk, and 2) the suggestion that a tip loop of the stalk binding underneath the clamp may serve as a wedge to restrict clamp opening<sup>17</sup>. The *Tko* RNAP structure indicates that the clamp opening is possible in the presence of the stalk in archaeal RNAP and likely Pol II (Fig. 3d), and there is substantial evidence in support of such from cryo-EM structural studies of Pol II. Pol II alone<sup>20</sup> and Pol II in complex with GTFs<sup>21</sup> were best fit by a model of Pol II in the open-state clamp configuration, and in each of these studies, the stalk was present. Using the *Tko* RNAP structure, a model of Pol II showing probable concerted transitions of the clamp and stalk domains was developed indicating that these domains are able to open without a steric clash (Supplementary Movie 2).

In the Pol I crystal structure, the main cleft adopts a wide conformation despite the clamp domain adopting a closed configuration. The Pol I stalk – containing the A43/A14 subunits

– is tightly associated with the main body of Pol I suggesting that Pol I and Pol II rely on partially overlapping but likely distinct conformational rearrangements to alter the conformation of the DNA binding cleft<sup>4,5</sup>.

### Structure of the Rpo3/Rpo11 heterodimer

Rpo3 and Rpo11 of *Tko* RNAP form a two-fold pseudo-symmetrical heterodimer comprising full-length Rpo11 and domain 1 of Rpo3 that is flanked by the two additional domains of Rpo3 (Fig. 2b, left), and this domain organization is well-conserved in the archaeal-eukaryotic RNAP family (AC40/AC19 in Pol I/Pol III, Rpb3/Rpb11 in Pol II) (Supplementary Fig. 3a)<sup>4,5,18,22</sup>. However, the architectures of domains 2 and 3 are distinct in each Rpo3, Rpb3 and AC40 subunit. Comparisons of crenarchaeal and euryarchaeal Rpo3 reveal no structural homology within domain 3. A ferredoxin-like 4Fe-4S cluster binding domain containing the 3Fe-4S cluster plus a disulfide bond dominates the *Sso* Rpo3 folding, whereas domain 3 of *Tko* Rpo3 adopts a structure containing a pair of  $\alpha$ -helices covered by four  $\beta$ -strands. Substantial differences within domain 2 are also evident, with domain 2 of *Sso* Rpb3 containing two pairs of disulfide bonds whereas no Cys residues are present in *Tko* Rpo3. Similarly disparate folds of domains 2 and 3 are present in the AC40 and Rpb3. The Rpo3/Rpo11 heterodimer is on the opposite surface of the DNA binding channel, and as such these different domains of Rpo3 are surface exposed and in position to provide unique and specific interfaces for transcription factors binding near upstream DNA in the PIC. Consistently, in the Pol II-dependent transcription, domain 2 of Rpb3 is one of the binding targets of Mediator<sup>23</sup> (further description in Discussion).

### Structural differences between *Tko* RNAP and yeast Pol II

The overall sequence identity of *Tko* RNAP and yeast *Saccharomyces cerevisiae* Pol II is only 39 % (Supplementary Table 1), complicating simple amino acid sequence alignments. The strong conservation of the overall fold of these RNAP structures permitted structure-guided sequence alignments (Supplementary Figs. 4 and 5) that provided precise comparisons of amino acid sequences of these RNAPs. These alignments identified regions where insertions – defined as 4 residues – were retained in each RNAP. *Tko* RNAP has 4 unique insertions compared with Pol II whereas yeast Pol II contains 30 distinct insertions compared with *Tko* RNAP (Fig. 4 and Supplementary Fig. 4). Most Pol II insertions are on the surface of Pol II and 18 insertions are fully or partially disordered in the yeast Pol II crystal structure<sup>24</sup>. 27 of the 30 Pol II insertions are conserved in human Pol II suggesting that these insertions play fundamental roles in the Pol II-dependent transcription (Fig. 4 and Supplementary Table 3). We also compared these 30 insertions with the crystal structure of Pol I<sup>4,5</sup> and a homology model of Pol III<sup>25</sup> and identified that 13 insertions are unique in the Pol II structure (Pol II-specific insertions) (Supplementary Fig. 6 and Supplementary Movie 3).

Notably, Pol II-specific insertions map precisely to the previously established binding sites of transcription factors unique to the Pol II transcription system (*e.g.*, TFIIF, TFIIFH and Mediator)<sup>23,26–28</sup> and do not map to the binding sites of common GTFs shared among the archaeal RNAP and Pol II (*e.g.*, TFB/TFIIB and TFE/TFIIE) (Figs. 4a and 5 and Supplementary Movie 3). This disparity indicates that the insertions retained in extant Pol II

might have been adopted as unique binding surfaces for specific transcription factors in the Pol II transcription system. For clarity, the 30 insertions (i1-i30) were separated into seven groups (Groups I to VII) (Fig. 4b and Table 1) for further structure-function analysis based on their clustered locations and interactions with TFIIF, TFIIH and Mediator (in Discussion).

## Discussion

Here, we report the first crystal structure of euryarchaeal RNAP, from *Thermococcus kodakarensis*. The open clamp conformation adopted by *Tko* RNAP is coupled with a rotational and swinging movement of the stalk. These coordinated movements resolve a long-standing question of how potential steric clashes between the clamp and stalk of archaeal RNAP and Pol II can be reconciled. The *Tko* RNAP structure gives a structural basis for the understanding of the clamp conformation changes of archaeal RNAP and Pol II during the transcription cycle<sup>21,29,30</sup>. The clamp is a conserved mobile domain in all multi-subunit cellular RNAPs, and the conformation changes of the clamp are key structural features throughout the transcription cycle including transcription initiation, transition to a stable elongation complex, and transcription pausing and termination. The stalk is a unique structure of archaeal-eukaryotic RNAPs and it is located near the clamp and the RNA exit channel.

The stalk also serves as a binding platform for GTFs of archaeal RNAP and Pol II<sup>31,32</sup>. A nascent RNA emerging from the active site of archaeal RNAP/Pol II also interacts with the stalk<sup>33,34</sup>. The coupled conformation change of the stalk with the clamp in archaeal RNAP and Pol II suggests the stalk may be used to control the clamp conformation as a leverage-like structure. The binding of a nascent RNA to the stalk of archaeal RNAP has been shown to increase the processivity of transcription *in vitro*<sup>34</sup>. The interaction between a nascent RNA and the stalk may stabilize the closed clamp state and this provides a plausible explanation for the enhancement of transcription processivity (Fig. 6a). On the other hand, TFE of archaeal system and TFIIE- $\alpha$  of Pol II system are known to interact with the base of the stalk and the tip of the clamp in PIC<sup>21,26,35</sup>, and it has been observed that the binding of TFIIE stabilizes the open clamp conformation in human PIC<sup>21</sup>. The binding of TFE or TFIIE- $\alpha$  on the tip of the clamp and base of the stalk may stabilize the open clamp conformation with the stalk as a leverage-like role (Fig. 6b).

Structure-guided sequence alignments between *Tko* RNAP and yeast Pol II revealed ~30 Pol II insertions (Fig. 4b and Table 1) and there is a correlation between their locations and the proposed binding surfaces for the Pol II-specific GTFs (TFIIF and TFIIH) and Mediator. The molecular details of such interactions are largely unknown, and it is illustrative to highlight some of the known and predicted interactions that occur between Pol II-specific GTFs and these grouped insertions. Such structure-guided analyses may establish targets for more specific biochemical assays probing the molecular mechanisms of Pol II-specific GTFs during transcription initiation and early elongation.

TFIIF consists of two conserved subunits, Tfg1 and Tfg2<sup>6</sup>, and several domains of Tfg2 were biochemically mapped to the Pol II lobe and protrusion domains, respectively<sup>26,36</sup>

(Fig. 5b). Group V insertions are located on the lobe, protrusion, and fork domains, and may participate in the binding of the Tfg2 subunit of TFIIF to Pol II (Fig. 5b, Fig. 7a, Table 1, and Supplementary Movie 3)<sup>26,37</sup>. It should be noted that the amino acid residues on Pol II that interact with TFIIF are not conserved in Pol I and Pol III (Supplementary Fig. 7a).

TFIIH was proposed to bind to a surface connecting the jaw, clamp and stalk domains of Pol II, however the molecular details of the interface remain unknown<sup>21,28,37</sup>. The Pol II-specific insertions i2, i5, i7, and i23 bridge the jaw, clamp and stalk domains of Pol II suggesting these insertions might be involved in direct interactions with TFIIH (Fig. 8a). The density connecting the stalk and TFIIH in the cryo-EM study of human PICs corresponds to i23<sup>21</sup>, and i5 and i7 are positioned to potentially interact with the C-terminal domain of Ssl2<sup>28,37</sup> (Fig. 7b). Insertion 2 in the clamp head domain may represent an additional contact site as suggested by a cryo-EM/cross-linking study of yeast PIC<sup>28</sup>.

The cryo-EM structures of Pol II in complex with Mediator suggested extensive interactions between Pol II and Mediator<sup>38,39</sup> and no fewer than seven Pol II subunits (Rpb1, Rpb2, Rpb3, Rpb4, Rpb6, Rpb7 and Rpb11) were proposed to be involved in this interaction. Although a massive Pol II-Mediator interaction surface is likely, the molecular details facilitating such interactions are currently limited to a few surfaces including the structurally unresolved C-terminal domain (CTD) of Rpb1<sup>27</sup> and Cys92/Ala159 in domain 2 of Rpb3<sup>23</sup>. Our structural analysis shows Cys92 is located within Pol II-specific insertion i18 and participates in coordinating a zinc ion with 3 other cysteines to form the Zn loop of Rpb3 (Fig. 7c). This structure is conserved in Pol II from yeast to humans but is not conserved in the AC40 counterpart of Pol I/Pol III (Supplementary Fig. 7b). As noted earlier, the domain 2 structures of the Rpb3 homologs in Pol I/Pol III (AC40 subunit) or archaeal RNAP (Rpo3 subunit) share no structural similarity (Supplementary Fig. 4a) indicating that unique surface exposed insertion regions may represent the binding sites for Pol II specific regulatory complexes. Pol II-specific insertions in Group IV, VI and VII insertions locate to the opposite side of the enzyme from the DNA binding cleft (Fig. 8b and Supplementary Movie 3) containing the well-characterized Mediator-binding sites including the CTD (i9) and the Zn loop of Rpb3 (i18). The insertions elucidated from these structural analyses provide logical positions to further investigate the Pol II – Mediator interaction.

## Methods

### Purification and crystallization of the *Tko* Rpo3/Rpo11

A polycistronic plasmid (pET21a-Rpo3-Rpo11) was generated to simultaneously overexpress the genes encoding *Tko* Rpo3 and Rpo11. *E. coli* BL21-CodonPlus(DE3)-RIPL (Stratagene) cells were transformed with pET21a-Rpo3-Rpo11, and transformants were grown in LB media supplemented with 100 µg/mL of ampicillin at 37 °C to an OD600 of ~0.8 before addition of IPTG to 0.5mM final to induce expression. Cells were harvested 5 hours post-induction, suspended in lysis buffer [20 mM Tris-HCl (pH 8.0), 50 mM KCl, 10 mM β-mercaptoethanol, 5 % glycerol and protease inhibitor cocktail (Roche)] and lysed by sonication. The Rpo3/Rpo11 complex was purified from the lysate by heat treatment at 65 °C for 30 minutes, followed by passage and fractionation of the cleared supernatant through two separate chromatographic columns (Q-sepharose and Superdex-75 gel filtration column

chromatography, GE Healthcare). A selenomethionine (SeMet) substituted Rpo3/Rpo11 complex was prepared by suppression of methionine biosynthesis<sup>40</sup> during culture growth, followed by an identical purification scheme as for the native complex. Both native and SeMet-labeled Rpo3/Rpo11 were concentrated to 10 mg/mL with buffer (10 mM Tris-HCl (pH 8), 50 mM NaCl, 1 mM EDTA and 2 mM DTT) for crystallization. Microbatch crystallization, mixing protein and crystallization solutions was performed under a thin layer of paraffin oil at 4 °C against a reservoir containing 0.1 M CAPS (pH 10), 0.1 M ammonium dihydrogen phosphate and 34 % (w/v) PEG4000. Crystals reached their full size (0.15 × 0.10 × 0.10 mm, diamond shape) within two weeks. Cryo-protection of the crystal was achieved by stepwise transfer to a crystallization solution containing 45 % (w/v) PEG4000, and the crystals were flash-frozen using liquid nitrogen.

### Structure determination of *Tko* Rpo3/Rpo11

The datasets Native and SeMet were collected at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY) Beamline X25 at 100 Kelvin. All data sets were processed by HKL2000<sup>41</sup>. For SeMet Multi-wavelength anomalous dispersion (MAD) phasing, 14 Se atom positions were identified by the program SnB<sup>42</sup> and the initial phase was calculated by SOLVE<sup>43</sup> followed by automated model building by RESOLVE<sup>44</sup>. The partial model was refined using the native protein data set and the final model was built manually using O<sup>45</sup> and refined by CNS<sup>46</sup> at 1.6 Å resolution (Supplementary Table 2). The crystal belongs to the primitive orthorhombic space group and contains two structurally identical Rpo3/Rpo11 complexes in each asymmetric unit. Ninety-eight per cent of the residues fall in favored regions of the Ramachandran plot and none of them is in disallowed regions.

### Purification and crystallization of the *Tko* Rpo4/Rpo7

A polycistronic plasmid (pET21a-Rpo4-Rpo7) was generated to simultaneously overexpress the genes encoding *Tko* Rpo7 and Rpo4. Rpo4/Rpo7 was expressed and purified as described for the preparation of *Tko* Rpo3/Rpo11. The purified Rpo4/Rpo7 complex was concentrated to 30 mg/mL with buffer (10 mM Tris-HCl (pH 8), 50 mM NaCl, 0.1 mM EDTA and 1 mM DTT) for crystallization. Crystals were obtained using hanging drop vapor diffusion by mixing equal volumes of Rpo4/Rpo7 and crystallization solution [0.1 M NaAcetate (pH 5.0) and 30 % glycerol] and incubating at 22 °C over the same crystallization solution. Crystals were directly frozen by liquid nitrogen.

### Structure determination of *Tko* Rpo4/Rpo7

X-ray diffraction data was collected at the X-ray core facility at Pennsylvania State University at 100 Kelvin and data set was processed by HKL2000<sup>41</sup>. The structure of the *Sso* Rpo4/Rpo7 complex from the complete *Sso* RNAP<sup>18</sup> was used as a search model for molecular replacement. Positional refinement was performed using Refmac<sup>547</sup> and Phenix<sup>48</sup> and the resulting map was used for building the final model manually by Coot<sup>49</sup>. The final structure was refined at 2.3 Å resolution (Supplementary Table 2). The crystal belongs to the primitive orthorhombic and contains one Rpo4/Rpo7 in an asymmetric unit. Ninety-four per

cent of the residues fall in favored regions of the Ramachandran plot and two per cent of them are in disallowed regions.

### Purification and preparation of *Tko* RNAP

*Tko Rpo4* strain, KUWLFB<sup>16</sup>, was grown under anaerobic conditions at 75 °C in nutrient-rich media (ASW-YT) containing 0.5% yeast extract (Y) and 0.5% trypton (T) in artificial seawater<sup>16</sup>. 2 L seed culture was inoculated into 200 L batch cultures, and cells were grown for ~20 hrs until reaching mid-log phase. For RNAP purification, 50 g of cells were suspended in 200 ml lysis buffer (10 mM Tris-HCl (pH 8.0), 500 mM KCl, 10% glycerol, 10 mM imidazole, 10 μM ZnCl<sub>2</sub>, 5 mM 2-mercaptoethanol, 0.3 μM leupeptin, 1 μM pepstatin, 1.5 mM benzamidine hydrochloride and 0.5 mM PMSF) and lysed by an Emulsiflex C3 homogenizer (Avestin, Inc.) at 20,000 psi. After centrifugation (27,000 x g for 1hr), the supernatant was loaded to 2 x 5 mL tandemly linked Ni-NTA affinity columns (Qiagen) equilibrated with the lysis buffer and washed with the same buffer containing 20 mM imidazole. Proteins were eluted with the lysis buffer containing 200 mM imidazole and precipitated by ammonium sulfate (final 80% saturation). The pellet was suspended in TGED buffer (20 mM Tris-HCl (pH 8.0), 10% glycerol, 0.5 mM EDTA, 5 mM DTT) until its conductivity was below 10 S/m and RNAP was further purified by binding and elution from a 5 ml HiTrap Q HP (GE Healthcare) column following a linear KCl gradient from 0.1 to 0.4 M. SDS-PAGE analysis of fractions resultant from HiTrap Q chromatography revealed a mixture of RNAP complexes lacking Rpo4/Rpo7 and RNAP complexes lacking Rpo4. To reconstitute RNAP containing all subunits, both RNAP pools from HiTrap Q were mixed with recombinant Rpo4/Rpo7 and Rpo4 at a ratio of 1 : 4 : 1 (RNAP : Rpo4/Rpo7 : Rpo4) for 1 hr at 20 °C and were further purified by successive passage and elution from 5 ml HiTrap Heparin, 8 ml MonoQ and Superdex200 columns (GE healthcare). Approximately 5 mg of 11-subunit *Tko* RNAP was obtained from each 50 g preparation.

### Crystallization and structure determination of *Tko* RNAP

*Tko* RNAP was concentrated to 10 mg/mL in buffer (10 mM Tris-HCl (pH 8.0), 200 mM KCl, 5 % glycerol, 10 μM ZnCl<sub>2</sub>, 5 mM DTT, 0.1 mM EDTA) and crystals were grown by hanging-drop vapor diffusion by mixing 1.2 μL of RNAP and 1 μL of reservoir solution (0.1 M imidazole (pH 8.0), 0.2 M CaCl<sub>2</sub>, 0.2M NaNO<sub>3</sub> and 12% PEG8000) at 22 °C. The crystals appeared in 3 days and grew to full size (0.1 x 0.05 x 0.3 mm) in 2 weeks. For cryo-crystallography, the crystals were transferred stepwise over a period of 5 min to 20 % ethylene glycol in 5 % increments and flash-frozen in liquid N<sub>2</sub>. The crystals belong to the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> and contain two *Tko* RNAPs per asymmetric unit. Diffraction data were collected at the Macromolecular Diffraction line at the Cornell University High Energy Synchrotron Source (MacCHESS) F1 beamline (Cornell University, Ithaca, NY) at 100 Kelvin and the data were processed by HKL2000<sup>41</sup>.

The structure of *Tko* RNAP was determined by molecular replacement using AutoMR in Phenix<sup>48</sup>. A search model for the molecular replacement was prepared from the *Sso* RNAP structure (PDB: 3HKZ)<sup>1</sup> with the following modifications: i) *Tko* Rpo3/Rpo11 and Rpo4/Rpo7 subcomplexes replaced their counterparts in *Sso* RNAP, ii) RpoG and Rpo13 were removed from the *Sso* coordinates. In the course of the structure determination of *Tko* RNAP



a substantial difference was noted in the position of the clamp domain compared with the clamp position determined for *Sso* RNAP. Therefore, we removed the clamp domain of *Sso* RNAP from the search model. After rigid body refinement and deformable elastic network (DEN) refinement using CNS<sup>50</sup>, the electron density map was interpreted and traced with Coot<sup>49</sup>. The crystal structure of clamp domain of *P. furiosus* RNAP (PDB ID 3QQC, chain A) had been determined<sup>51</sup> and we therefore fitted this structure manually into the electron density map corresponding to the region specified for the *Tko* RNAP clamp domain. Further refinement was performed using Phenix<sup>48</sup> with non-crystallographic symmetry and secondary structure restraints and the resulting model was manually rebuilt with Coot<sup>49</sup>. The final position and orientation of the clamp domain of *Tko* RNAP was confirmed by the locations of Zn ions. Ninety-five per cent of the residues fall in favored regions of the Ramachandran plot and five per cent of them are in disallowed regions.

### Phylogenetic analysis

Amino acid sequences of largest subunit of RNAPs from bacteria, archaeal and eukaryote were aligned by Muscle with default parameters, and phylogenetic tree was constructed using the Molecular Evolutionary Genetics Analysis (MEGA6)<sup>52</sup> with maximum likelihood method using Jones-Taylor-Thornton (JTT) model, uniform rate and bootstrap replication of 500 times.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank the staff at the Penn State Fermentation Facility for supporting *Tko* cell culture and the NSLS and MacCHESS for supporting crystallographic data collections. This work was supported by NIH grants GM087350-A1 (to K.S.M.) and GM100329 (to T.J.S.), Grants-in-Aid for Scientific Research on Innovative Areas “Transcription Cycle”, JSPS KAKENHI Grant Number 25118516 (to A.H), and JSPS KAKENHI Grant Number 26292038 (to T.K.). TJS contributions were initiated at Ohio State University. Figures were prepared using PyMOL. We thank Steve Hahn for critical reading of the manuscript, and Yuichiro Takagi, Kenji Murakami, Francisco Asturias and Finn Werner for helpful discussions.

### References

1. Hirata A, Murakami KS. Archaeal RNA polymerase. *Curr Opin Struct Biol.* 2009; 19:724–31. [PubMed: 19880312]
2. Beattie TR, Bell SD. Molecular machines in archaeal DNA replication. *Curr Opin Chem Biol.* 2011; 15:614–9. [PubMed: 21852183]
3. Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol.* 2011; 9:85–98. [PubMed: 21233849]
4. Fernandez-Tornero C, et al. Crystal structure of the 14-subunit RNA polymerase I. *Nature.* 2013; 502:644–9. [PubMed: 24153184]
5. Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P. RNA polymerase I structure and transcription regulation. *Nature.* 2013; 502:650–5. [PubMed: 24153182]
6. Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell.* 2012; 45:439–46. [PubMed: 22365827]
7. Jun SH, Reichlen MJ, Tajiri M, Murakami KS. Archaeal RNA polymerase and transcription regulation. *Crit Rev Biochem Mol Biol.* 2011; 46:27–40. [PubMed: 21250781]

8. Sarmiento F, Mrazek J, Whitman WB. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A*. 2013; 110:4726–31. [PubMed: 23487778]
9. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol*. 2004; 11:394–403. [PubMed: 15114340]
10. Takagi Y, Kornberg RD. Mediator as a general transcription factor. *J Biol Chem*. 2006; 281:80–9. [PubMed: 16263706]
11. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. The archaeobacterial origin of eukaryotes. *Proceedings of the National Academy of Sciences*. 2008; 105:20356–20361.
12. Elkins JG, et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences*. 2008; 105:8102–8107.
13. Korkhin Y, et al. Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol*. 2009; 7:e1000102. [PubMed: 19419240]
14. Wojtas MN, Mogni M, Millet O, Bell SD, Abrescia NG. Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic Acids Res*. 2012; 40:9941–52. [PubMed: 22848102]
15. Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*. 2013; 504:231–6. [PubMed: 24336283]
16. Hirata A, et al. Archaeal RNA polymerase subunits E and F are not required for transcription *in vitro*, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol Microbiol*. 2008; 70:623–33. [PubMed: 18786148]
17. Armache KJ, Kettenberger H, Cramer P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A*. 2003; 100:6964–8. [PubMed: 12746495]
18. Hirata A, Klein BJ, Murakami KS. The X-ray crystal structure of RNA polymerase from Archaea. *Nature*. 2008; 451:851–4. [PubMed: 18235446]
19. Kusser AG, et al. Structure of an archaeal RNA polymerase. *J Mol Biol*. 2008; 376:303–7. [PubMed: 18164030]
20. Kostek SA, et al. Molecular architecture and conformational flexibility of human RNA polymerase II. *Structure*. 2006; 14:1691–700. [PubMed: 17098194]
21. He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. *Nature*. 2013; 495:481–6. [PubMed: 23446344]
22. Cramer P, Bushnell DA, Kornberg RD. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*. 2001; 292:1863–76. [PubMed: 11313498]
23. Soutourina J, Wydau S, Ambroise Y, Boschiero C, Werner M. Direct interaction of RNA polymerase II and mediator required for transcription *in vivo*. *Science*. 2011; 331:1451–4. [PubMed: 21415355]
24. Armache KJ, Mitterweger S, Meinhart A, Cramer P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem*. 2005; 280:7131–4. [PubMed: 15591044]
25. Jasiak AJ, Armache KJ, Martens B, Jansen RP, Cramer P. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell*. 2006; 23:71–81. [PubMed: 16818233]
26. Chen HT, Warfield L, Hahn S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat Struct Mol Biol*. 2007; 14:696–703. [PubMed: 17632521]
27. Robinson PJ, Bushnell DA, Trnka MJ, Burlingame AL, Kornberg RD. Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci U S A*. 2012; 109:17931–5. [PubMed: 23071300]
28. Murakami K, et al. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science*. 2013; 342:1238724. [PubMed: 24072820]
29. Chakraborty A, et al. Opening and closing of the bacterial RNA polymerase clamp. *Science*. 2012; 337:591–5. [PubMed: 22859489]
30. Weixlbaumer A, Leon K, Landick R, Darst SA. Structural basis of transcriptional pausing in bacteria. *Cell*. 2013; 152:431–41. [PubMed: 23374340]

31. Edwards AM, Kane CM, Young RA, Kornberg RD. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem.* 1991; 266:71–5. [PubMed: 1985924]
32. Naji S, Grunberg S, Thomm M. The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J Biol Chem.* 2007; 282:11047–57. [PubMed: 17311916]
33. Ujvari A, Luse DS. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol.* 2006; 13:49–54. [PubMed: 16327806]
34. Hirtreiter A, Grohmann D, Werner F. Molecular mechanisms of RNA polymerase--the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic Acids Res.* 2010; 38:585–96. [PubMed: 19906731]
35. Grohmann D, et al. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell.* 2011; 43:263–74. [PubMed: 21777815]
36. Eichner J, Chen HT, Warfield L, Hahn S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* 2010; 29 :706–16. [PubMed: 20033062]
37. Grunberg S, Warfield L, Hahn S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol.* 2012; 19:788–96. [PubMed: 22751016]
38. Davis JA, Takagi Y, Kornberg RD, Asturias FA. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell.* 2002; 10 :409–15. [PubMed: 12191485]
39. Cai G, et al. Mediator head module structure and functional interactions. *Nat Struct Mol Biol.* 2010; 17:273–9. [PubMed: 20154708]
40. Doublet S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* 1997; 276:523–30. [PubMed: 9048379]
41. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A.* 1997; 276:307–326.
42. Smith GD, Nagar B, Rini JM, Hauptman HA, Blessing RH. The use of SnB to determine an anomalous scattering substructure. *Acta Crystallogr D Biol Crystallogr.* 1998; 54:799–804. [PubMed: 9757093]
43. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr.* 2000; 56:965–72. [PubMed: 10944333]
44. Terwilliger TC. Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement. *Acta Crystallogr D Biol Crystallogr.* 2003; 59:1174–82. [PubMed: 12832760]
45. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A.* 1991; 47 ( Pt 2):110–9. [PubMed: 2025413]
46. Brunger A. Version 1.2 of the Crystallography and NMR system. *Nat Protoc.* 2007; 2:2728–33. [PubMed: 18007608]
47. Winn MD, Isupov MN, Murshudov GN. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta crystallographica Section D, Biological crystallography.* 2001; 57:122–33.
48. Adams PD, et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr.* 2002; 58:1948–54. [PubMed: 12393927]
49. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–32. [PubMed: 15572765]
50. Schroder GF, Levitt M, Brunger AT. Super-resolution biomolecular crystallography with low-resolution data. *Nature.* 2010; 464:1218–22. [PubMed: 20376006]
51. Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* 2011; 30:1302–10. [PubMed: 21386817]

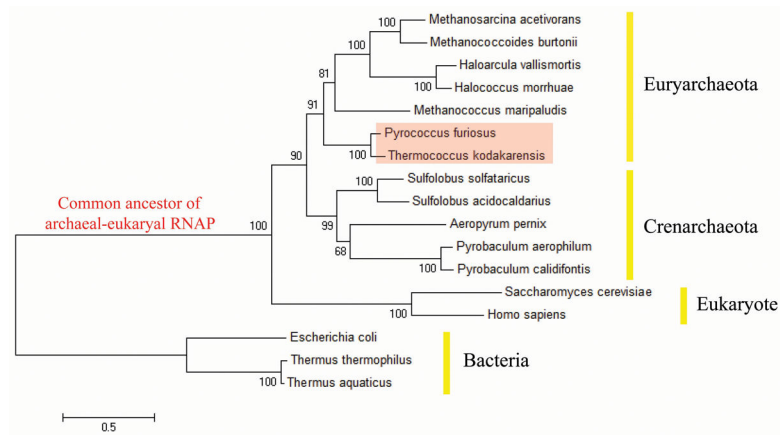
52. Hall BG. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol.* 2013; 30 : 1229–35. [PubMed: 23486614]

Author Manuscript

Author Manuscript

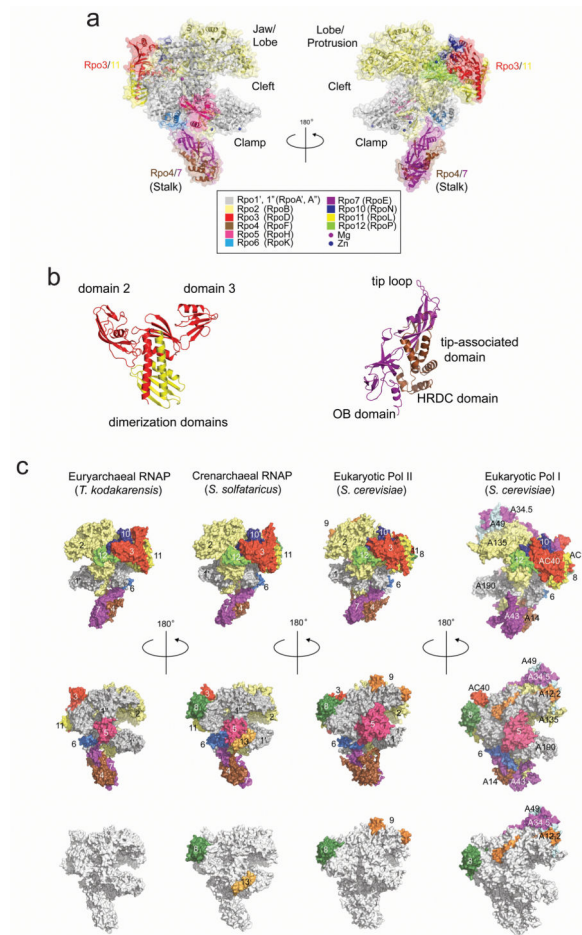
Author Manuscript

Author Manuscript



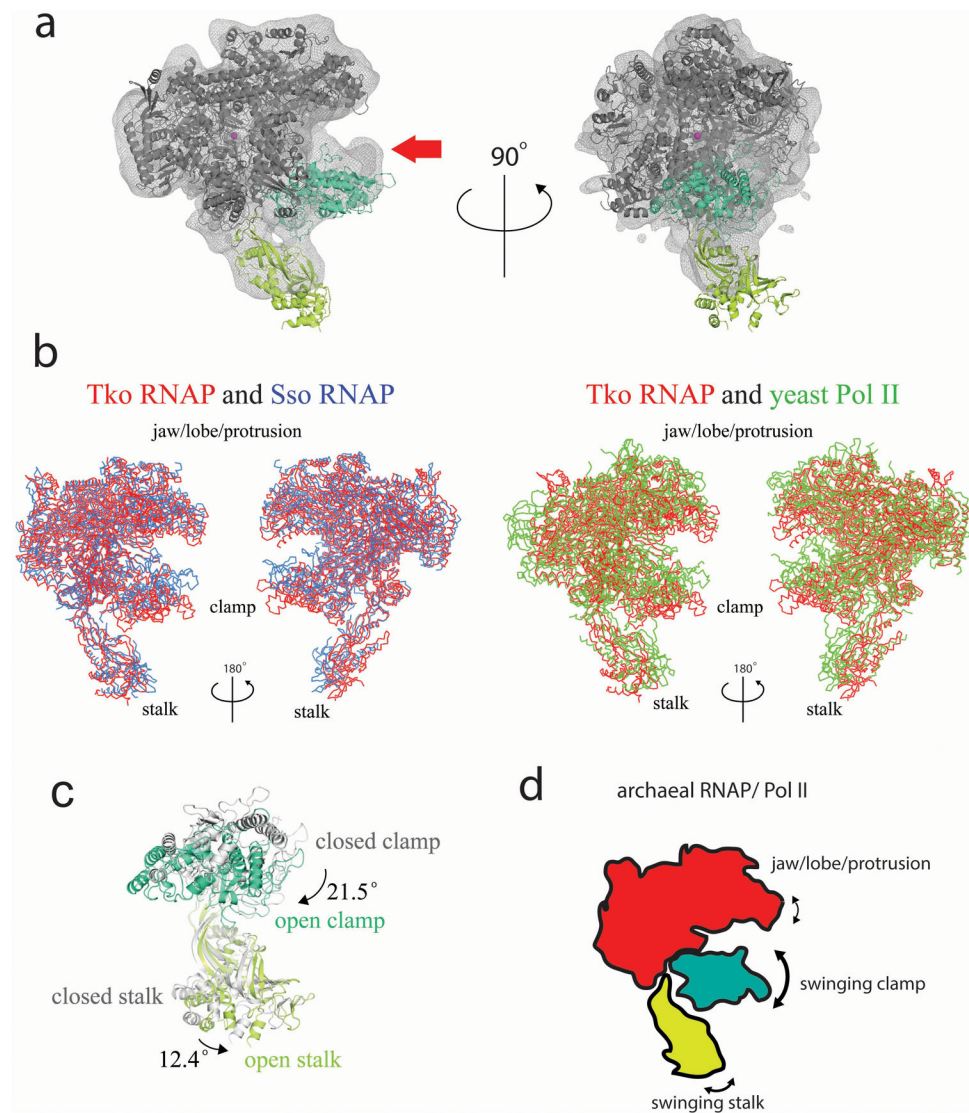
**Figure 1. Phylogenetic analysis of the largest subunit of RNAP in Bacteria, Euryarchaeota, Crenarchaeota and Eukaryotes**

Maximum-likelihood phylogenetic tree made with the largest subunit of RNAP ( $\beta'$  of bacterial RNAP, Rpo1'+Rpo1'' of archaeal RNAP, and Rpb1 of eukaryotic RNAP II) rooted with bacterial sequences. Bootstrap support based on 500 replicates is shown at each node. Scale bar represent the average number of substitutions per residues. The position of common ancestor of archaeal-eukaryal RNAP is indicated in red. An order of *Thermococcales* including *Pyrococcus furiosus* (*Pfu*) and *Thermococcus kodakarensis* (*Tko*) is highlighted.



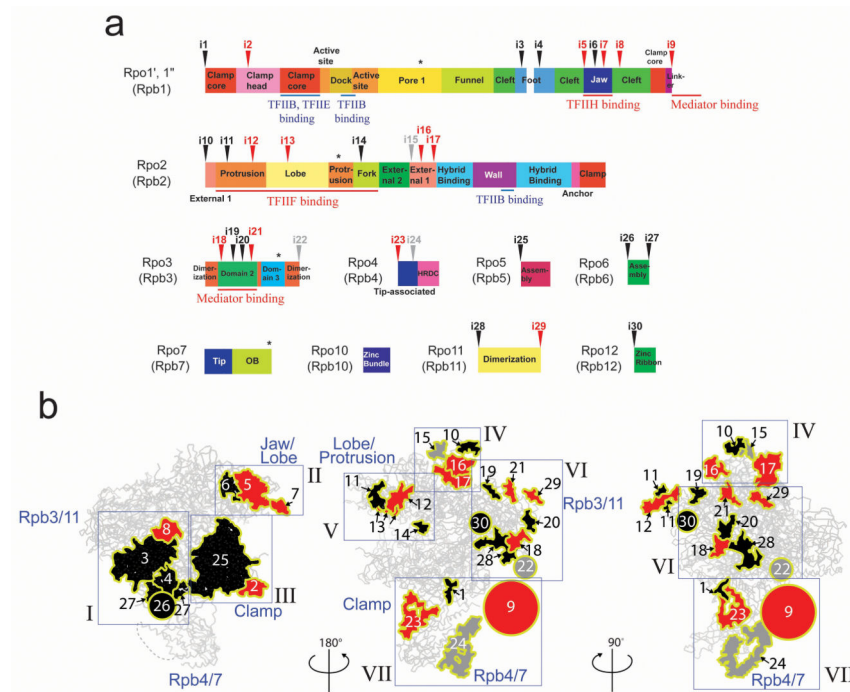
**Figure 2. Crystal structures of the *Tko* RNAP and other members of archaeal/eukaryotic RNAP family**

(a) *Tko* RNAP structure in a ribbon model along with a transparent molecular surface. Each subunit is denoted by a unique color and labeled. Two sets of nomenclatures, one for traditional (in parenthesis) and the other based on the eukaryotic terminology, are shown. Some structural features are labeled. (b) Crystal structures of Rpo3/Rpo11 (left) and Rpo4/Rpo7 (right) complexes. Domains are labeled. Each subunit is colored as in a. (c) Surface representations of the archaeal and eukaryotic RNAPs. Each subunit is denoted by a unique color and labeled. Orthologous subunits are depicted by the same color. In a bottom panel, common subunits are shown in gray.



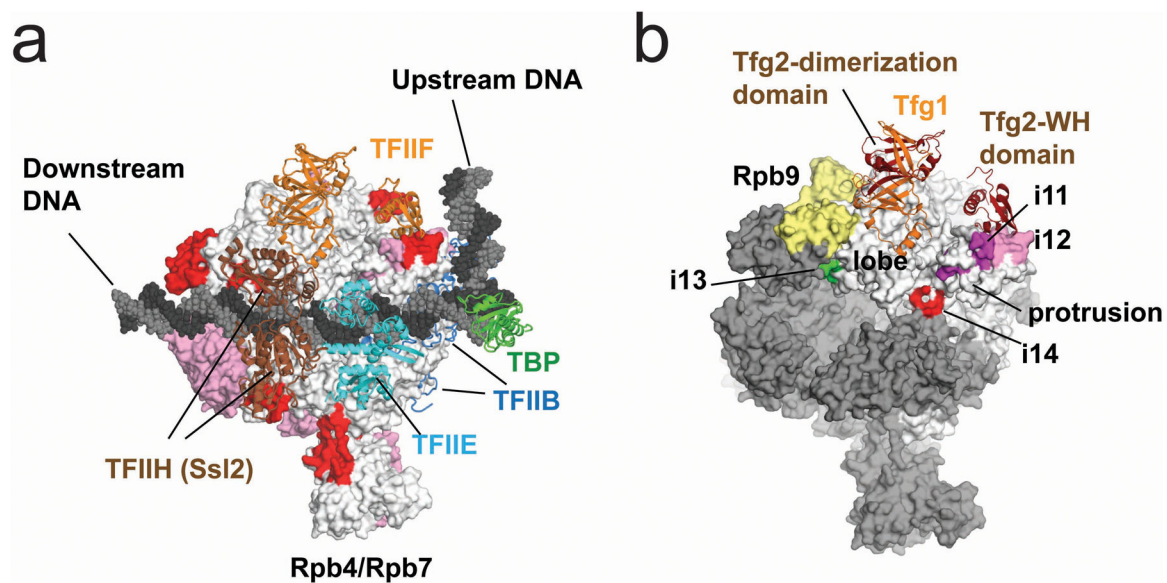
### Figure 3. Opening and closing of the DNA binding clamp of RNAP

(a) The structure of *Tko* RNAP is superimposed on the structures of *Sso* RNAP (left, blue, PDB: 3HKZ) and yeast Pol II (right, green, PDB: 1WCM) using the enzyme active sites as references. (b) Cryo-EM density map of *Pfu* RNAP (light grey mesh, EMD-1711) is overlaid with the *Tko* RNAP structure (main body of RNAP, grey; clamp, dark cyan; Rpo4/Rpo7 stalk, yellow green; active site Mg, magenta sphere). The clamp of *Pfu* RNAP is indicated by a red arrow and in a closed-state. (c) Proposed motions of the clamp and stalk of archaeal RNAP. The open clamp (dark cyan) and stalk (yellow green) observed in the *Tko* RNAP structure are superimposed on the closed clamp and stalk (gray) in the *Sso* RNAP structure. Motions and angles of the clamp and stalk from the closed to the open conformations are indicated. (d) Schematic representation of archaeal RNAP/Pol II showing the mobility of the DNA-binding cleft and the stalk.



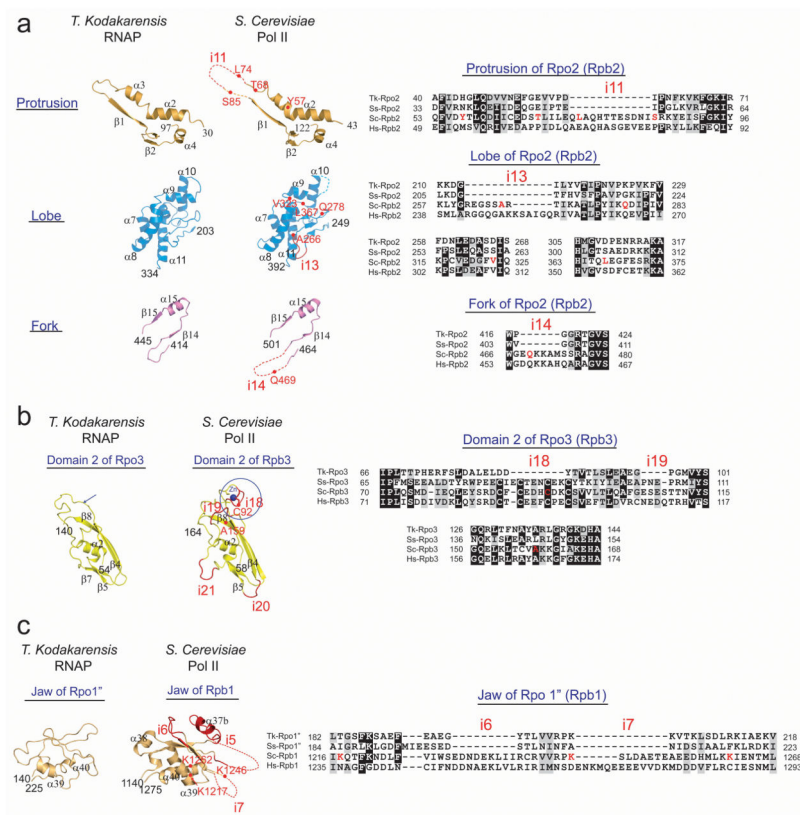
**Figure 4. Structural differences between the *Tko* RNAP and yeast Pol II**  
**(a)** Schematic diagrams of domains and domain-like regions of the *Tko* RNAP based on the Pol II nomenclature. Inverted triangles indicate positions of 30 insertions found in yeast Pol II compared with *Tko* RNAP (black, insertions in eukaryotic RNAPs; red, Pol II-specific insertions; gray, unique in yeast Pol II). The binding sites of TFIIB, TFIIE, TFIIF, TFIIH and Mediator are indicated. Asterisks indicate positions of 4 insertions found in the *Tko* RNAP subunits. **(b)** The structure of yeast Pol II and the three-dimensional representation of the 30 insertions in **a**. Insertions are depicted by plain-surfaces with yellow boundaries (colors same as in **a**). Circles show approximate locations of disordered insertions at N- or C-termini of subunits (*e.g.*, Pol II CTD) and their diameters represent their approximate lengths. 7 groups of insertions (I to VII) are indicated in blue rectangles.



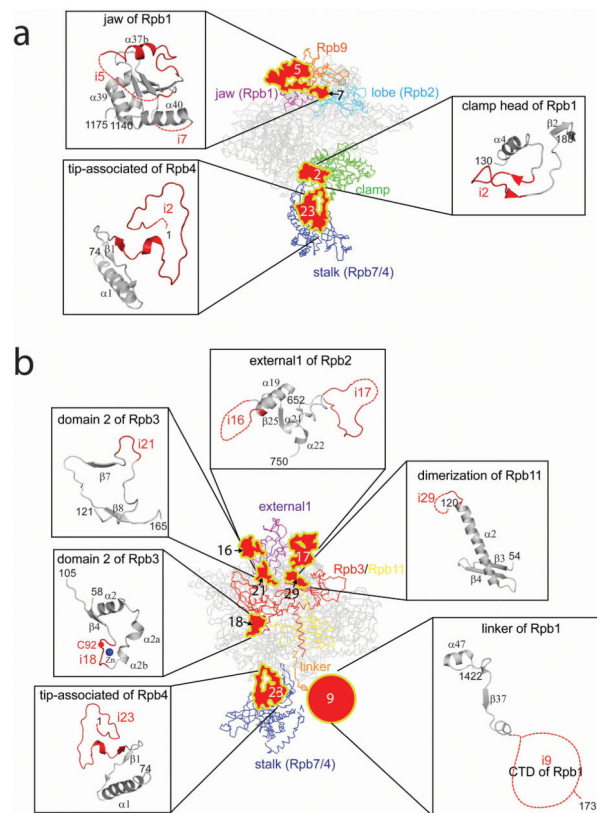


**Figure 5. The PIC model of Pol II**

(a) Pol II, GTFs and DNA are depicted by surface, cartoon and cpk models, respectively. Insertions specific and non-specific to Pol II are indicated in red and pink, respectively. Each GTF is denoted by a unique color and labeled. (b) The Pol II and TFIIF interactions in the PIC. Pol II and TFIIF are depicted by surface and cartoon models, respectively. Rpb2 and Rpb9 are in white and yellow, respectively and all other subunits are in gray. Subunits and domains of TFIIF are labeled. Four insertions participated in the TFIIF binding are indicated. This PIC model is adapted from ref. 37.

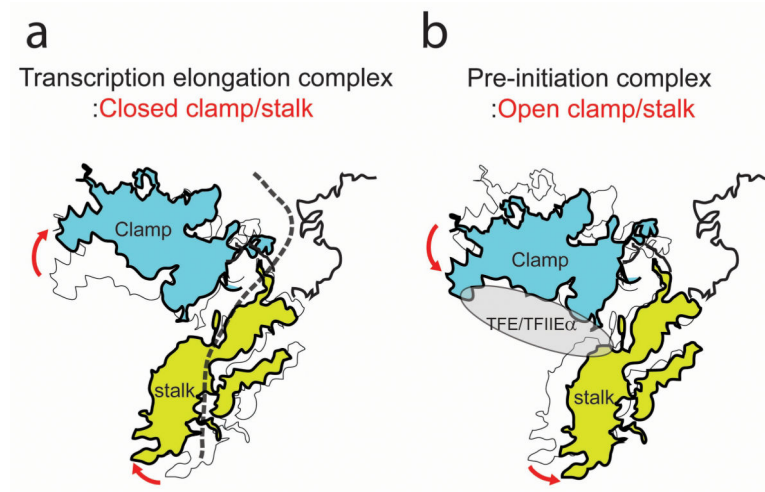


**Figure 6. Model of clamp conformation control through the stalk**  
**(a)** A nascent RNA, depicted as a dashed line, may stabilize the closed conformation of the clamp in the transcription elongation complex by the interaction with the stalk. **(b)** Binding of TFE/TFIIE $\alpha$ , shown as a gray ellipse, on the clamp and stalk may stabilize the open conformation of the clamp and stalk in the PIC.



**Figure 7. Structure and sequence comparisons of *Tko* RNAP and yeast Pol II around binding sites of the Pol II-specific GTFs/Mediator**

Structures of Pol II around the binding sites of TFIIF (a), TFIIH (b) and Mediator (c) and their counterparts in the *Tko* RNAP structure are shown. Insertions are colored and indicated in red and disordered regions are depicted as dashed lines. Amino acid sequence alignments around these regions are shown on their right (Tk, *Tko*; Ss, *Sso*; Sc, *Sce*; Hs, *Homo sapiens*). Amino acid residues cross-linked to TFIIF (a) and TFIIH (b) are indicated in red. Cys92 and Ala159 of Rpb3, which interact directly with Mediator, are indicated in red and the positions of i18 are indicated by a blue arrow on *Tko* RNAP and a blue circle on Pol II (c).



**Figure 8. Pol II-specific insertions and TFIH and Mediator binding interfaces on Pol II**  
**(a)** Pol II-specific insertions on the Pol II backbone model on the TFIH binding interface.  
**(b)** Pol II-specific insertions on the Mediator binding interface. In both panels, insertions are shown as red plain-surfaces with yellow-green boundaries on the Pol II backbone model and key subunits and domains are indicated. Magnified views of these insertions are shown in boxes.

**Table 1**

Pol II insertions (I-VII) and their interactions with GTFs and Mediator

| Group | Insertions                                 | Domains having these insertions  | Binding factors | References |
|-------|--|--|-----------------|------------|
| I     | i3, i4, i8, i26, and i27                   | Foot and cleft (Rpb1), N- and C- terminal tails (Rpb6)                           |                 |            |
| II    | i5, i6, and i7                             | Jaw (Rpb1)   | TFIIH           | 28         |
| III   | i2 and i25                                 | Clamp head (Rpb1) and jaw (Rpb5)   |                 |            |
| IV    | i10, i15, i16, and i17                     | External 1 (Rpb2)  |                 |            |
| V     | i11, i12, i13, and i14                     | protrusion, lobe and fork (Rpb2)   | TFIIF           | 26         |
| VI    | i18, i19, i20, i21, i22, i28, i29, and i30 | Domain 2 and tail (Rpb3), N-and C-terminal tails (Rpb11) and zinc ribbon (Rpb12) | Mediator        | 23         |
| VII   | i1, i9, i23, and i24                       | Clamp core and CTD (Rpb1), tip- associated (Rpb4)                                | Mediator        | 27         |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript