ARTICLE

Received 6 Jun 2014 | Accepted 2 Sep 2014 | Published 14 Oct 2014

DOI: 10.1038/ncomms6132

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

Sung-Hoon Jun¹, Akira Hirata^{1,2}, Tamotsu Kanai³, Thomas J. Santangelo⁴, Tadayuki Imanaka⁵
& Katsuhiko S. Murakami¹

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, TFIIH and Mediator. Such comparisons provide a framework for dissecting interactions between RNAP and these factors during formation of the pre-initiation complex.

¹Department of Biochemistry and Molecular Biology, The Center for RNA Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ²Department of Materials Science and Biotechnology, Graduate School of Science and Engineering, Ehime University, Matsuyama, Ehime 790-8577, Japan. ³Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510 Japan. ⁴Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523, USA. ⁵Department of Biotechnology, Ritsumeikan University, Noji-Higashi, Kusatsu 525-8577, Japan. Correspondence and requests for materials should be addressed to K.S.M. (email: kum14@psu.edu) or to A.H. (email: hirata.akira.mg@ehime-u.ac.jp).

Archaeal proteins associated with genome maintenance and gene expression have extensive functional and structural similarities with their eukaryotic counterparts^{1,2}. This congruence is especially true for the archaeal transcription machinery, and there is a striking structural similarity between archaeal and eukaryotic RNA polymerases (RNAPs)^{3–5}. 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 transcription⁶. All factors are auxiliary for the archaeal and Pol II transcription systems; however, some factors are bona fide RNAP subunits for the Pol I and Pol III transcription systems. Archaeal RNAP is 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*^{1,3,7,8}. PIC formation with Pol II requires a more complex set of GTFs, with minimally six GTFs (TFIIA, TFIIB, TFIID/TBP, TFIIE, TFIIF and TFIH) and the Mediator complex is also required for promoter-specific transcription^{7,9,10}.

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^{11,12}. 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 (refs 13,14). 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 postulates 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 multisubunit 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 analyse the structure and evolution of archaeal–eukaryotic RNAP family¹⁵. *Tko* RNAP purified directly from cells contains substoichiometric amounts of TFE¹⁶, and this heterogeneity likely precluded crystallization attempts. *Tko* RNAP purified from a *Δrpo4* strain yields an enzyme that lacks Rpo4, Rpo7 and TFE¹⁶. 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)¹ 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 that of 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^{17,18}, reveals that the *Tko* RNAP clamp is in

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 (β' 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 represents 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 *Pfu* and *Tko* is highlighted.

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 1,058–1,123) 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*) RNAP¹⁹ (Fig. 1 and Supplementary Table 1), with the exception of the DNA-binding clamp (Fig. 3b). This difference adumbrates that 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 $\sim 21.3^\circ$ 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 $\sim 12^\circ$, 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 opening¹⁷. 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 alone²⁰ and Pol II in complex with GTFs²¹ were best fitted 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^{4,5}.

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 (grey) 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.

Structure of the Rpo3/Rpo11 heterodimer. Rpo3 and Rpo11 of *Tko* RNAP form a two-fold pseudosymmetrical 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)^{4,5,18,22}. However, the architectures of domains 2 and 3 are distinct in each Rpo3, Rpb3 and AC40 subunits. 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 α -helices covered by four β -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 the Mediator²³ (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 four 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 structure²⁴. In all, 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 (refs 4,5) and a homology model of Pol III (ref. 25) 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 (for example, TFIIF, TFIIF and Mediator)^{23,26–28} and do not map to the binding sites of common GTFs shared among the archaeal RNAP and Pol II (for example, 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

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; grey, unique in yeast Pol II). The binding sites of TFIIB, TFIIE, TFIIF, TFIIF and Mediator are indicated. Asterisks indicate positions of four 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 (colours same as in **a**). Circles show approximate locations of disordered insertions at N- or C termini of subunits (for example, Pol II CTD) and their diameters represent their approximate lengths. Seven groups of insertions (I–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 nonspecific to Pol II are indicated in red and pink, respectively. Each GTF is denoted by a unique colour and labelled. (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 grey. Subunits and domains of TFIIF are labelled. Four insertions participated in the TFIIF binding are indicated. This PIC model is adapted from ref. 37.

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)	TFIIH	28
II	i5, i6 and i7	Jaw (Rpb1)		
III	i2 and i25	Clamp head (Rpb1) and jaw (Rpb5)		
IV	i10, i15, i16 and i17	External 1 (Rpb2)	TFIIF	26
V	i11, i12, i13 and i14	Protrusion, lobe and fork (Rpb2)		
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)		
VII	i1, i9, i23 and i24	Clamp core and CTD (Rpb1), tip-associated (Rpb4)	Mediator	27

CTD, C-terminal domain; GTF, general transcription factor.

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–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 *Tko*. 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 the 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^{21,29,30}. The clamp is a conserved mobile domain in all multisubunit cellular RNAPs, and the conformation changes of the clamp are the 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 (refs 31,32). A nascent RNA emerging from the active

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/TFIIEα, shown as a grey ellipse, on the clamp and stalk may stabilize the open conformation of the clamp and stalk in the PIC.

site of archaeal RNAP/Pol II also interacts with the stalk^{33,34}. The coupled conformation change of the stalk with the clamp in archaeal RNAP and Pol II suggests that 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*³⁴. 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 α of the Pol II system are known to interact with the base of the stalk and the tip of the clamp in PIC^{21,26,35}, and it has been observed that the binding of TFIIE stabilizes the open-clamp conformation in human PIC²¹. The binding of TFE or TFIIE α 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 (ref. 6), and several domains of Tfg2 were biochemically mapped to the Pol II lobe and protrusion domains, respectively^{26,36} (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 (Figs 5b and 7a,

Table 1 and Supplementary Movie 3)^{26,37}. 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^{21,28,37}. The Pol II-specific insertions i2, i5, i7 and i23 bridge the jaw, clamp and stalk domains of Pol II, suggesting that 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

Figure 8 | Pol II-specific insertions and TFIIF- and Mediator-binding interfaces on Pol II. (a) Pol II-specific insertions on the Pol II backbone model on the TFIIF-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 the key subunits and domains are indicated. Magnified views of these insertions are shown in boxes.

PICs corresponds to i23 (ref. 21), and i5 and i7 are positioned to potentially interact with the C-terminal domain (CTD) of Ssl2 (refs 28,37; Fig. 7b). Insertion 2 in the clamp head domain may represent an addition contact site as suggested by a cryo-EM/crosslinking study of yeast PIC²⁸.

The cryo-EM structures of Pol II in complex with Mediator suggested extensive interactions between Pol II and Mediator^{38,39} 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 CTD of Rpb1 (ref. 27) and Cys92/Ala159 in domain 2 of Rpb3 (ref. 23). Our structural analysis shows that Cys92 is located within Pol II-specific insertion i18 and participates in coordinating a zinc ion with three 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 homologues in Pol I/Pol III (AC40 subunit) or archaeal RNAP (Rpo3 subunit) share no structural similarity (Supplementary Fig. 3a) 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. *Escherichia coli* BL21-CodonPlus(DE3)-RPL (Stratagene) cells were transformed with pET21a-Rpo3-Rpo11, and transformants were grown in LB media supplemented with 100 µg ml⁻¹ of ampicillin at 37 °C to an OD₆₀₀ of ~0.8 before the addition of isopropyl-β-D-thiogalactoside to 0.5 mM final to induce expression. Cells were harvested 5 h 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 with sonication. The Rpo3/Rpo11 complex was purified from the lysate by heat treatment at 65 °C for 30 min, 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 biosynthesis⁴⁰ during culture growth, followed by an identical purification scheme as for the native complex. Both native and SeMet-labelled Rpo3/Rpo11 were concentrated to 10 mg ml⁻¹ with buffer (10 mM Tris–HCl (pH 8), 50 mM NaCl, 1 mM EDTA and 2 mM dithiothreitol (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 2 weeks. Cryoprotection 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 data sets Native and SeMet were collected at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY, USA) Beamline X25 at 100 K. All data sets were processed by HKL2000 (ref. 41). For SeMet multiwavelength anomalous dispersion phasing, 14 Se atom positions were identified by the programme SnB⁴² and the initial phase was calculated by SOLVE⁴³ followed by automated model building by RESOLVE⁴⁴. The partial model was refined using the native protein data set and the final model was built manually using O⁴⁵ and refined using CNS⁴⁶ 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 favoured 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⁻¹ 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 vapour 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 using liquid nitrogen.

Structure determination of *Tko* Rpo4/Rpo7. X-ray diffraction data were collected at the X-ray core facility at Pennsylvania State University at 100 K and the data set was processed by HKL2000 (ref. 41). The structure of the *Sso* Rpo4/Rpo7 complex from the complete *Sso* RNAP¹⁸ was used as a search model for molecular replacement. Positional refinement was performed using Refmac5 (ref. 47) and Phenix⁴⁸ and the resulting map was used for building the final model manually by Coot⁴⁹. 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 favoured regions of the Ramachandran plot and 2% of them are in disallowed regions.

Purification and preparation of *Tko* RNAP. *Tko* Δ*Rpo4* strain, KUWLEB¹⁶, 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 seawater¹⁶. Two litres seed culture were inoculated into 200 l batch cultures, and cells were grown for ~20 h 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 ZnCl₂, 5 mM 2-mercaptoethanol, 0.3 µM leupeptin, 1 µM pepstatin, 1.5 mM benzamidine hydrochloride and 0.5 mM phenylmethyl sulphonyl fluoride) and lysed by an Emulsiflex C3 homogenizer (Avestin Inc.) at 20,000 p.s.i. After centrifugation (27,000 g for 1 h), the supernatant was loaded to 2 × 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 and 5 mM DTT) until its conductivity was below 10 S m⁻¹ 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–polyacrylamide gel electrophoresis 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 h 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⁻¹ in buffer (10 mM Tris–HCl (pH 8.0), 200 mM KCl, 5 % glycerol, 10 µM ZnCl₂, 5 mM DTT and 0.1 mM EDTA), and the crystals were grown by hanging-drop vapour diffusion by mixing 1.2 µl of RNAP and 1 µl of reservoir solution (0.1 M imidazole (pH 8.0), 0.2 M CaCl₂, 0.2 M NaNO₃ and 12% PEG8000) at 22 °C. The crystals appeared in 3 days and grew to full size (0.1 × 0.05 × 0.3 mm) in 2 weeks. For cryocrystallography, the crystals were transferred stepwise over a period of 5 min to 20% ethylene glycol in 5% increments and flash-frozen in liquid N₂. The crystals belong to the space group P2₁2₁2₁ 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, USA) at 100 K and data were processed by HKL2000 (ref. 41).

The structure of *Tko* RNAP was determined by molecular replacement using AutoMR in Phenix⁴⁸. A search model for the molecular replacement was prepared from the *Sso* RNAP structure (PDB: 3HKZ)¹ with the following modifications: (i) *Tko* Rpo3/Rpo11 and Rpo4/Rpo7 subcomplexes replaced their counterparts in *Sso* RNAP and (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 Crystallography & NMR System (CNS)⁵⁰, the electron-density map was interpreted and traced with Coot⁴⁹. The crystal structure of clamp domain of *Pfu* RNAP (PDB ID 3QQC, chain A) had been determined⁵¹, 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 Phenix⁴⁸ with noncrystallographic symmetry and secondary structure restraints, and the resulting model was manually rebuilt with Coot⁴⁹. The final position and orientation of the clamp domain of *Tko* RNAP were

confirmed by the locations of Zn ions. Ninety-five per cent of the residues fall in favoured regions of the Ramachandran plot and five per cent of them are in the disallowed regions.

Phylogenetic analysis. Amino-acid sequences of the largest subunits of RNAPs from bacteria, archaea and eukaryote were aligned by Muscle with default parameters, and a phylogenetic tree was constructed using the Molecular Evolutionary Genetics Analysis (MEGA6)⁵² with maximum-likelihood method using Jones–Taylor–Thornthorn model, uniform rate and bootstrap replication of 500 times.

References

- Hirata, A. & Murakami, K. S. Archaeal RNA polymerase. *Curr. Opin. Struct. Biol.* **19**, 724–731 (2009).
- Beattie, T. R. & Bell, S. D. Molecular machines in archaeal DNA replication. *Curr. Opin. Chem. Biol.* **15**, 614–619 (2011).
- Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9**, 85–98 (2011).
- Fernandez-Tornero, C. *et al.* Crystal structure of the 14-subunit RNA polymerase I. *Nature* **502**, 644–649 (2013).
- Engel, C., Sainsbury, S., Cheung, A. C., Kostrewa, D. & Cramer, P. RNA polymerase I structure and transcription regulation. *Nature* **502**, 650–655 (2013).
- Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).
- Jun, S. H., Reichlen, M. J., Tajiri, M. & Murakami, K. S. Archaeal RNA polymerase and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* **46**, 27–40 (2011).
- Sarmiento, F., Mrazek, J. & Whitman, W. B. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc. Natl Acad. Sci. USA* **110**, 4726–4731 (2013).
- Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**, 394–403 (2004).
- Takagi, Y. & Kornberg, R. D. Mediator as a general transcription factor. *J. Biol. Chem.* **281**, 80–89 (2006).
- Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. & Embley, T. M. The archaeobacterial origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **105**, 20356–20361 (2008).
- Elkins, J. G. *et al.* A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107 (2008).
- Korkhin, Y. *et al.* Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* **7**, e1000102 (2009).
- Wojtas, M. N., Mogni, M., Millet, O., Bell, S. D. & Abrescia, N. G. Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic Acids Res.* **40**, 9941–9952 (2012).
- Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013).
- 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.* **70**, 623–633 (2008).
- Armache, K. J., Kettenberger, H. & Cramer, P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl Acad. Sci. USA* **100**, 6964–6968 (2003).
- Hirata, A., Klein, B. J. & Murakami, K. S. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**, 851–854 (2008).
- Kusser, A. G. *et al.* Structure of an archaeal RNA polymerase. *J. Mol. Biol.* **376**, 303–307 (2008).
- Kostek, S. A. *et al.* Molecular architecture and conformational flexibility of human RNA polymerase II. *Structure* **14**, 1691–1700 (2006).
- He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
- Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001).
- Soutourina, J., Wyda, S., Ambrose, Y., Boschiero, C. & Werner, M. Direct interaction of RNA polymerase II and mediator required for transcription *in vivo*. *Science* **331**, 1451–1454 (2011).
- Armache, K. J., Mitterweber, S., Meinhart, A. & Cramer, P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**, 7131–7134 (2005).
- Jasiak, A. J., Armache, K. J., Martens, B., Jansen, R. P. & Cramer, P. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol. Cell* **23**, 71–81 (2006).
- Chen, H. T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat. Struct. Mol. Biol.* **14**, 696–703 (2007).
- Robinson, P. J., Bushnell, D. A., Trnka, M. J., Burlingame, A. L. & Kornberg, R. D. Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc. Natl Acad. Sci. USA* **109**, 17931–17935 (2012).
- Murakami, K. *et al.* Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**, 1238724 (2013).
- Chakraborty, A. *et al.* Opening and closing of the bacterial RNA polymerase clamp. *Science* **337**, 591–595 (2012).
- Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431–441 (2013).
- Edwards, A. M., Kane, C. M., Young, R. A. & Kornberg, R. D. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. *J. Biol. Chem.* **266**, 71–75 (1991).
- 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.* **282**, 11047–11057 (2007).
- Ujvari, A. & Luse, D. S. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat. Struct. Mol. Biol.* **13**, 49–54 (2006).
- 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.* **38**, 585–596 (2010).
- 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* **43**, 263–274 (2011).
- Eichner, J., Chen, H. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* **29**, 706–716 (2010).
- 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.* **19**, 788–796 (2012).
- Davis, J. A., Takagi, Y., Kornberg, R. D. & Asturias, F. A. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol. Cell* **10**, 409–415 (2002).
- Cai, G. *et al.* Mediator head module structure and functional interactions. *Nat. Struct. Mol. Biol.* **17**, 273–279 (2010).
- Doublet, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. A* **276**, 307–326 (1997).
- Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A. & Blessing, R. H. The use of SnB to determine an anomalous scattering substructure. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 799–804 (1998).
- Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
- Terwilliger, T. C. Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1174–1182 (2003).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**(Pt 2): 110–119 (1991).
- Brunger, A. Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733 (2007).
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 122–133 (2001).
- Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Schroder, G. F., Levitt, M. & Brunger, A. T. Super-resolution bio-molecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
- Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 (2011).
- Hall, B. G. Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **30**, 1229–1235 (2013).

Acknowledgements

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.). The contributions of T.J.S. were initiated at the 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 their helpful discussions.

Author contributions

A.H. and K.S.M. initially designed the 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.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

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

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Jun, S.-H. *et al.* The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat. Commun.* 5:5132 doi: 10.1038/ncomms6132 (2014).