

# Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure

Yakov Korkhin<sup>1,2,3,4</sup>, Ulug M. Unligil<sup>3,4</sup>, Otis Littlefield<sup>1,2</sup>, Pamlea J. Nelson<sup>1,2,3,4</sup>, David I. Stuart<sup>5</sup>, Paul B. Sigler<sup>1,2†</sup>, Stephen D. Bell<sup>6</sup>, Nicola G. A. Abrescia<sup>5,7\*</sup>

**1** Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, United States of America, **2** Howard Hughes Medical Institute, Yale University, New Haven, Connecticut, United States of America, **3** Harvard Medical School, Boston, Massachusetts, United States of America, **4** Howard Hughes Medical Institute, Harvard University, Boston, Massachusetts, United States of America, **5** Division of Structural Biology and the Oxford Protein Production Facility, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **6** Sir William Dunn School of Pathology, University of Oxford, Oxford, United Kingdom, **7** Structural Biology Unit, CIC bioGUNE, Derio, Spain

**The archaeal RNA polymerase (RNAP) shares structural similarities with eukaryotic RNAP II but requires a reduced subset of general transcription factors for promoter-dependent initiation. To deepen our knowledge of cellular transcription, we have determined the structure of the 13-subunit DNA-directed RNAP from *Sulfolobus shibatae* at 3.35 Å resolution. The structure contains the full complement of subunits, including RpoG/Rpb8 and the equivalent of the clamp-head and jaw domains of the eukaryotic Rpb1. Furthermore, we have identified subunit Rpo13, an RNAP component in the order Sulfolobales, which contains a helix-turn-helix motif that interacts with the RpoH/Rpb5 and RpoA'/Rpb1 subunits. Its location and topology suggest a role in the formation of the transcription bubble.**

Citation: Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, et al. (2009) Evolution of complex RNA polymerases: The complete archaeal RNA polymerase structure. *PLoS Biol* 7(5): e1000102. doi:10.1371/journal.pbio.1000102

## Introduction

Gene expression in cellular organisms across the three kingdoms of life is carried out by multisubunit RNA polymerase (RNAP) enzymes. Eukaryotes have three different multisubunit nuclear RNAPs (Pol I, II, and III), whereas Archaea and Bacteria have single enzymes [1]. A wealth of structural information has been gathered in the past decade allowing the visualization of RNAP II in isolation, in the act of transcription, and in complex with transcription factors transcription factor II S (TFIIS) or transcription factor II B (TFIIB) [2–5]. The archaeal transcription machinery is orthologous to that of eukaryotes, but initiation only requires two accessory factors: transcription factor B (TFB) (an ortholog of TFIIB) and TATA-box binding protein (TBP) [6–8], and thus provides a simplified model system for studying transcription initiation. In eukaryotes, the additional basal factors are needed, in part to facilitate DNA melting at the initiation site, a functional complexity that Archaea must overcome by other means [6,7]. Recent structural studies on archaeal polymerases [9,10] have shed light on the basic architecture, but the information gathered thus far remains incomplete.

We have therefore determined the crystal structure of the complete DNA-directed RNA polymerase from the archaeon *Sulfolobus shibatae* (SshRNAP) at 3.35 Å resolution (see Materials and Methods and Table 1) revealing the complete 13 subunit set of the functional enzyme including two subunits so far undetected: RpoG/Rpb8 and Rpo13 (Figure 1). The highly positively charged C-terminus of Rpo13 extends into the DNA entry channel, suggesting its involvement in binding to nucleic acids. Our intact RNAP structure allows us to propose a model for the archaeal preinitiation complex formation.

## Results

### Overall RNA Polymerase Structure

The structure determination in two distinct crystal forms, to 3.35 Å resolution is described in Materials and Methods; 3,334 residues in 13 subunits are seen, but the resolution achieved limits the accuracy of the position of the side chains. For the most flexible subunits, Rpo4 and Rpo7, accounting for some 265 residues, the register of the sequence is, in places, uncertain ( $\pm 1$ ). The model quality is good, it lies in the top 15th percentile of structures solved at between 3.1–3.6 Å as judged by MolProbity [11], 78% of the residues are in the most favoured region of the Ramachandran plot (see Table 1). This compares favourably with the equivalent value of 61% for the structure of RNAP from *S. solfataricus* [9]. The overall architecture, subunit arrangement, composition, and topology closely follow those of the eukaryotic counterpart RNAP II [2]; for this reason, we propose a new subunit nomenclature applicable to all archaeal RNAPs and based on the eukaryotic terminology (Figures 1 and 2). The basic assembly resembles (root mean square deviation [rmsd] 1.1 Å,

**Academic Editor:** Martin Egli, Vanderbilt University, United States of America

**Received** October 16, 2008; **Accepted** March 19, 2009; **Published** May 5, 2009

**Copyright:** © 2009 Korkhin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abbreviations:** HTH, helix-turn-helix; rmsd, root mean square deviation; RNAP, RNA polymerase; TBP, TATA-box binding protein; TFB, transcription factor B; TFE, transcription factor E; TFIIS, transcription factor II S; TFIIB, transcription factor II B

\* To whom correspondence should be addressed. E-mail: nabrescia@cicbiogune.es

† Deceased.

## Author Summary

Transcription, the process of converting DNA into RNA (which in turn is translated into proteins by ribosomes) is carried out by the multisubunit RNA polymerase (RNAP) enzyme. Transcription is fundamental to all organisms across the three kingdoms of life—Eukarya, Bacteria, and Archaea—and can be divided into three major steps: initiation, transcription/elongation, and termination. Eukaryotes have three different nuclear RNAPs, whereas Archaea and Bacteria have one. Archaeal transcription is similar to that of eukaryotes, but initiation requires only two accessory proteins bound to DNA: transcription factor B (TFB) and TATA-box binding protein (TBP). It is believed that studies of the archaeal enzyme may shed light on the more complex eukaryotic RNAP. Our complete structure of the archaeal RNAP from *Sulfolobus shibatae* has fully elucidated its architecture, confirming its close evolutionary relationship with the eukaryotic RNAP II and at the same time revealed a new subunit, Rpo13, with no ortholog in the eukaryotic enzyme. The location and topology of Rpo13 allow us to suggest a mechanism by which Archaea bypass the additional eukaryotic cofactors required for transcription initiation.

for 2,938 residues aligned, 97% of the residues in common between the two structures) the recently published archaeal RNAP structure [9], but our structure adds considerable new information. More specifically, we have located the clamp-head domain in Rpo1N (Figure 2A), the jaw domain in subunit Rpo1C (Figure 2B), and the entire Rpo8 subunit (Figure 2C). Furthermore, we observe density corresponding to a helix-turn-helix (HTH) motif in a groove created by Rpo5 and the clamp-head domain of Rpo1N (Figures 2D and 3). Mass

spectrometry analysis (see Materials and Methods) confirmed the presence of a previously reported RNAP subunit, named “component F” [12], comprising 104 residues of which the 45-residue HTH motif constitutes an ordered fragment. Fitting of the electron density yielded to a satisfactory alignment (Figures 2D and S1). We rename this subunit Rpo13, as RpoF has been used to refer to the distinct archaeal Rpb4 homolog. Uniquely in the archaeal RNAP, the Rpo13 subunit does not have an ortholog in the eukaryotic RNAP II.

Of the approximately 370-kDa archaeal RNAP, subunits Rpo1 (split into two subunits Rpo1N and Rpo1C) and Rpo2 represent more than two-thirds of the mass and are equivalent to bacterial  $\beta'$  and  $\beta$  and to the eukaryotic Rpb1 and Rpb2 [9,13]. These subunits are composed of different domains that perform specific roles during RNA polymerisation in the active site of the Rpo1N subunit (catalytic residues D456, D458, and D460 are conserved across cellular RNAPs [2,8]). The second largest subunit, Rpo2, contains three  $Zn^{2+}$  atoms (two coordinated with His570 and with His696/His997, respectively; the third located in the clamp). Its polypeptide chain is largely ordered, and it provides, with Rpo1, the catalytic activity of the RNAP. Both subunits in the cellular RNAPs contain a double- $\Psi$   $\beta$ -barrel domain involved in the polymerization process, whose heterodimeric structure has been suggested as the ancestral core enzyme [13].

Subunits Rpo3 (containing a 4Fe-FS cluster [9]), Rpo6 and Rpo11 constitute, along with Rpo1 and Rpo2, the core RNAP, conserved across the three domains of life [8,9,13]. Structure-based phylogenetic analysis between the homologous components illustrates this evolutionary relationship (Figure S2) and strengthens the idea of a transcription apparatus that has

**Table 1.** Data Collection and Refinement Statistics

Data Collection	Crystal_1	Crystal_2
X-ray source	Brookhaven	Advanced Photon Source
Beamline	X25	BioCARS 14C
Wavelength (Å)	1.1000	0.9000
Space group	$P2_12_12_1$	$P2_12_12$
Unit cell parameters (Å)	$a = 196.1, b = 211.9, c = 238.9$	$a = 193.5, b = 212.6, c = 129.0$
Molecule per AU	2	1
Resolution range (Å)	30.2–3.52	42.8–3.35
N° of images	363	*
Unique reflections	114,084	59,401
Redundancy <sup>a</sup>	6.2(3.2)	*
Completeness <sup>a</sup> (%)	93.0(62.1)	77.0(44.8)
$I/\sigma(I)$ <sup>a</sup>	10.1(2.0)	8.5(1.7)
$R_{\text{merge}}$ (%) <sup>a</sup>	15.7(47.7)	*
Refinement	Crystal_1	Crystal_2
Resolution range (Å)	30.2–3.52	42.8–3.35
N° of reflections (working/test)	108172/5721	56389/3012
$R_{\text{work}}/R_{\text{free}}$ (%)	33.3/38.4	27.1, 34.1
N° of atoms: Protein/ $Zn^{2+}$ / $Mg^{2+}$ /Fe-S cluster	52,721, 10, 2, 2	26,458, 9, 1, 1
Rms $\Delta$ bond length (Å)	0.009	0.006
Rms $\Delta$ bond angles (°)	1.280	1.027
$\langle B\text{-factor} \rangle$ (Å <sup>2</sup> ): Protein/ $Zn^{2+}$ / $Mg^{2+}$ /Fe-S cluster	100.8, 82.2, 80.2, 87.4	88.8, 61.7, 58.7, 68.0
Ramachandran statistics <sup>b</sup> : Most favoured, additional allowed, generously allowed, disallowed (%)	75.8, 19.2, 2.8, 2.1	77.6, 17.5, 2.7, 2.2

\*The statistics for Crystal\_2 are derived from merged observed intensities and corresponding errors [I(σ)] present in the only available mtz file; raw images and processing log-files have been lost.

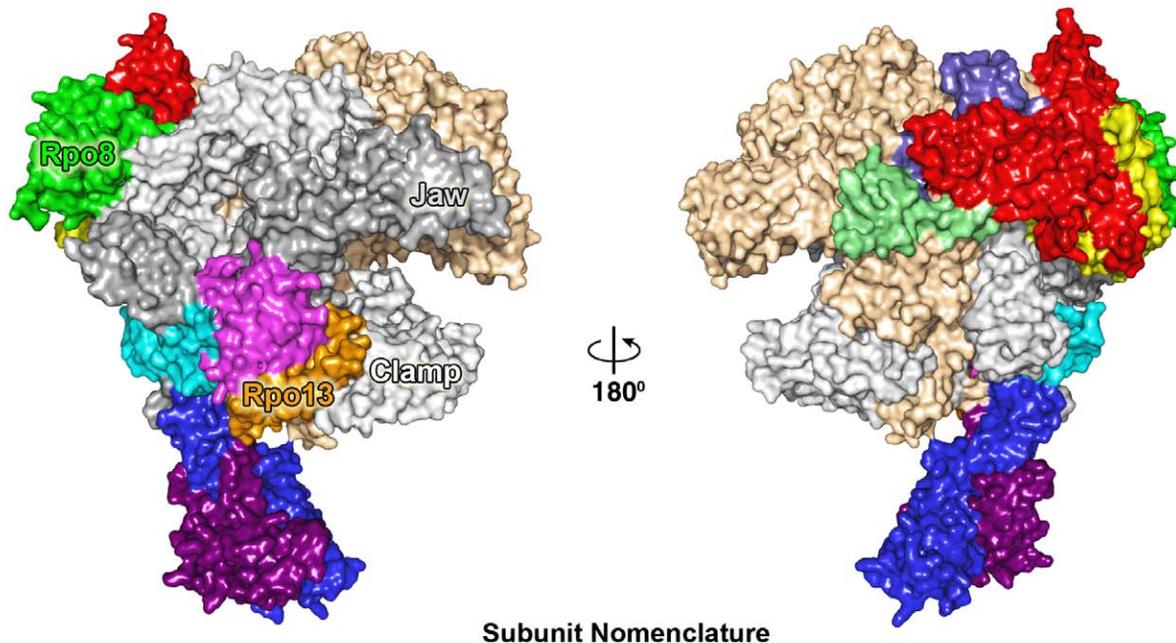
<sup>a</sup>In parentheses, the values for the last resolution shell.

<sup>b</sup>Values calculated using Procheck [37].

AU, asymmetric unit.

doi:10.1371/journal.pbio.1000102.t001

## Complete Archaeal RNAP



Subunit Nomenclature

Archaeal	Rpo1N	Rpo1C	Rpo2	Rpo3	Rpo4	Rpo5	Rpo6	Rpo7	Rpo8	Rpo10	Rpo11	Rpo12	Rpo13
	(RpoA')	(RpoA'')	(RpoB)	(RpoD)	(RpoF)	(RpoH)	(RpoK)	(RpoE')	(RpoG)	(RpoN)	(RpoL)	(RpoP)	
Eukaryotic	Rpb1	Rpb2	Rpb3	Rpb4	Rpb5	Rpb6	Rpb7	Rpb8	Rpb10	Rpb11	Rpb12	--	

**Figure 1.** Surface Representation of the Complete Archaeal RNA Polymerase from *S. shibatae*

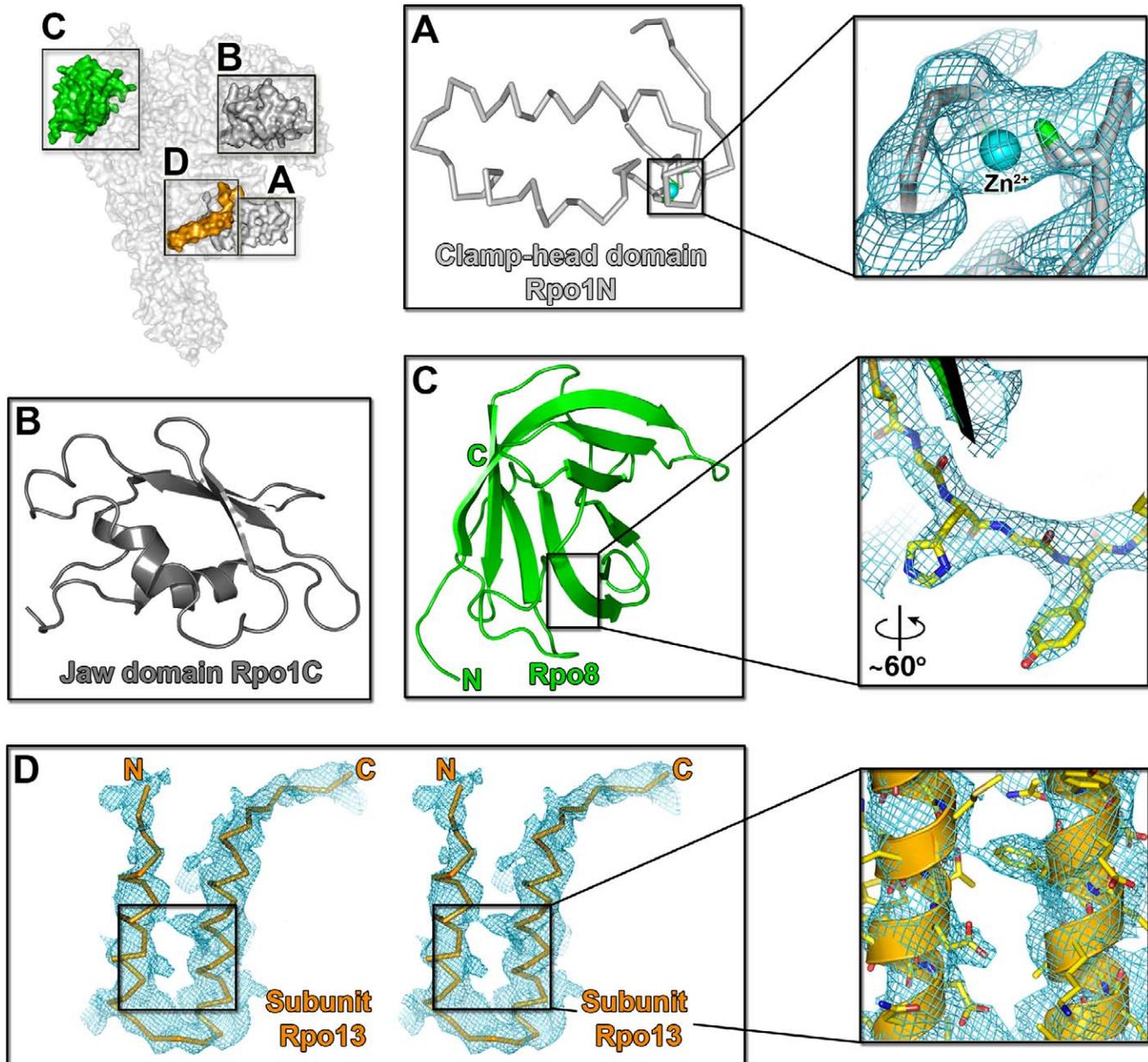
The overall architecture of the RNAP is shown with the different subunits colour-coded as illustrated in the Subunit Nomenclature (below). The newly located structural elements are labelled.  
doi:10.1371/journal.pbio.1000102.g001

increased cellular specificity associated with the addition of new functional modules. The remaining known archaeal subunits Rpo4/7, Rpo5, Rpo10, and Rpo12, with homologs only in Eukarya (class II subunits), decorate the core enzyme as shown in Figure 1 (Rpo10 and Rpo12 each bind a  $Zn^{2+}$ ). Of all the subunits common to Archaea/Eukarya, Rpo5/Rpb5 differs the most in size. Rpb5 is composed of a jaw and assembly domain [2], of which only the latter is present in the archaeal Rpo5 (which lacks the first ~130 residues of Rpb5). The absence of the jaw domain allows access to a positively charged groove between the assembly domain and helices 3 and 4 of the clamp-head domain (see below and Figure 3A and 3B).

The Rpo4/7 heterodimer, which is conditionally required for initiation [14], protrudes from the main structure interacting mostly through Rpo6 and the C-terminus of Rpo1C (Figure 1). The Rpo4/7 stalk is highly mobile as judged by the electron density and normal mode analysis (Figure 4). Some weak density attributable to the C-terminus of subunit Rpo4 is visible and reminiscent of the position of the C-terminal helix 6 (H6) observed in the isolated Rpo4/7 crystal structure from *Methanococcus jannaschii* [15]. This suggests that H6 can move from its location, possibly contributing to the interaction with accessory cofactors such as transcription factor E (TFE) [7,8].

### Structural Conservation of the Flexible Clamp-Head and Jaw Domains of Rpo1 Subunit

The electron density maps reveal the clamp head domain (residues 97–173) in the Rpo1N subunit to be mostly ordered (residues 97–156). This domain, absent from the recent archaeal X-ray structure [9] is located at the DNA-binding cleft (Figure 2A), as is its eukaryotic counterpart. It adopts a fold similar to its eukaryotic ortholog and contains a  $Zn^{2+}$  ion chelated by C98, C101, and C146 (Figure 2A). It also shows some differences including a longer HTH motif (residues 107–140) that flexes by approximately  $34^\circ$ . This motif, together with the Rpo5 subunit, forms a cavity into which Rpo13 docks (Figure 3B and 3C). The remaining portion of the Rpo1N structure is well ordered, showing a  $Zn^{2+}$  ion in the clamp core domain and a single  $Mg^{2+}$  ion in the active site [2]. Electron density corresponding to the jaw domain of the Rpo1C subunit is also visible. This domain is implicated in binding DNA downstream of the transcription start site [4], and residues 147–236 have been modelled initially by reference to the eukaryotic polymerase (see Materials and Methods and Figure 2B). The bridge helix in Rpo1N and the trigger loop in Rpo1C, which have been proposed to play critical roles in DNA translocation [3], are respectively in a straight conformation (albeit with some differences from that observed by Hirata and co-workers [9] [rmsd 0.8 Å, for 40 aligned C $\alpha$ ]) and disordered. The location and the topology of



**Figure 2.** Newly Identified Subunits and Located Subdomains

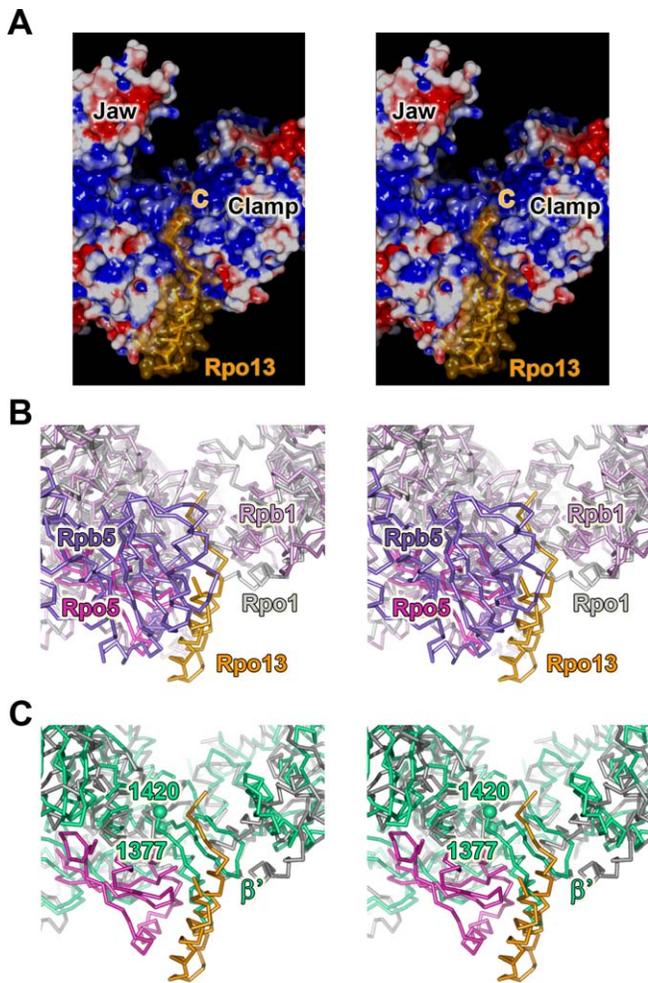
(A) Ribbon representation of the clamp-head domain in Rpo1N (light grey); C98, C101, and C146 as green stick; and Zn ion as a cyan sphere. On the right, sigmaA-weighted 2Fo-Fc electron density countered at  $1\sigma$  (as a blue mesh) corresponding to the  $Zn^{2+}$  and surrounding cysteines (arbitrary view). (B) Cartoon representation of the jaw domain in the Rpo1C subunit. (C) Cartoon representation of the entire Rpo8 subunit. On the right, sigmaA-weighted 2Fo-Fc electron density countered at  $1.2\sigma$  of a structural detail. (D) Stereo view of the sigmaA-weighted 2Fo-Fc map electron density (contoured at  $1.1\sigma$ ) correspondent to the HTH motif of the Rpo13 subunit (as orange ribbon with N and C labelling the N- and C-termini). The inset shows some side chains (as sticks) fitting density.  
doi:10.1371/journal.pbio.1000102.g002

the clamp-head and jaw domain in the Rpo1 subunit reinforce the structural similarity with Rpb1 of naked RNAP II.

### Structure of the Rpo8 Subunit

Until recently, it was believed that archaeal RNAPs from both the Crenarchaea and Euryarchaea kingdoms, the two main groups of Archaea, did not possess orthologs of Rpb8 [8,16], but recent work has identified divergent homologs of Rpb8 in the Crenarchaea [1,17]. This subunit, Rpo8 (15.1 kDa;

132 residues), is seen for the first time in our current study and is a constitutive structural element of the RNAP complex (Figures 1 and 2C). Rpo8 adopts an eight-stranded antiparallel  $\beta$  oligo-binding (OB) fold (Figure 2C) with the first three and last 15 residues disordered. It is located at a peripheral position (Figure 1), similar to eukaryotic Rpb8, with which it shares 112  $C\alpha$  equivalences (out of 114) with 2.9 Å rmsd (Figure S3) and 14% sequence identity (the lowest sequence identity of all archaeal and eukaryotic subunits). Rpo8 interacts with subunit Rpo1N, sitting in the external crevice



**Figure 3.** Binding locus of Subunit Rpo13

(A) Electrostatic surface representation (blue positive, red negative, and white neutral) with Rpo13 as orange ribbon (C labels the C-terminus) and semitransparent surface.

(B) Stereo view of the docking region of Rpo13 (orange), in light-magenta Rpo5 and in grey Rpo1 with the eukaryotic Rpb5 (violet) and Rpb1 (plum; PDB entry 1WCM) superimposed; rmsd 2.4 Å for 1,129 C $\alpha$  equivalences.

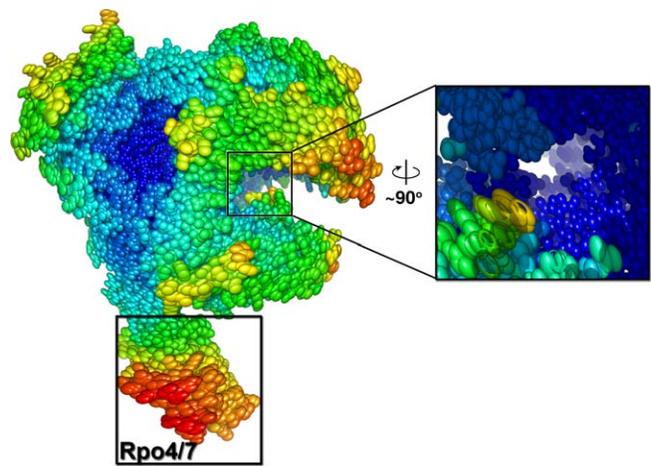
(C) Archaeal Rpo13, Rpo5, and Rpo1 as (B) with the bacterial  $\beta'$  from the *T. aquaticus* RNAP core structure (turquoise; PDB entry 1I6V) superimposed; rmsd 3.4 Å for 748 C $\alpha$  equivalences. The two spheres mark the insertion region 1,377–1,420 in the bacterial  $\beta'$  that spatially correlates with the archaeal subunit Rpo13. Superimpositions were carried out using the Structure Homology Program [44].

doi:10.1371/journal.pbio.1000102.g003

formed by residues 507–596 and burying a surface area of 1,470 Å<sup>2</sup>, equivalent to the interaction of Rpb8 with Rpb1 (Figure 2). The shorter  $\Omega$ -loop motif (residues 63–69) is ordered, and the isosurface of electrostatic potential shows marked segregation of charges with a line of basic residues R20, K24, K52, K54, and K56 decorating the top of the molecule (Figure S4).

### Rpo13, a Novel Constituent RNAP Subunit

Predictions from the amino acid sequence [18,19] indicate an ordered HTH core motif ( $\alpha$ 1 and  $\alpha$ 2) with flexible N- and C-termini, respectively, predicted as mainly coil and  $\alpha$ -helix ( $\alpha$ 3), suggesting a simple three-helix bundle protein proto-



**Figure 4.** Ellipsoid Representation of the Thermal Motions of the Entire RNAP

B ranges from Bmin = 40 Å<sup>2</sup> (dark-blue) to Bmax = 233 Å<sup>2</sup> (red) values obtained by normal mode analysis [27]. Subunits Rpo4/7 are highly mobile, whereas the inset highlights the catalytic site as a rigid frame. All figures in this article were generated in Pymol [45].  
doi:10.1371/journal.pbio.1000102.g004

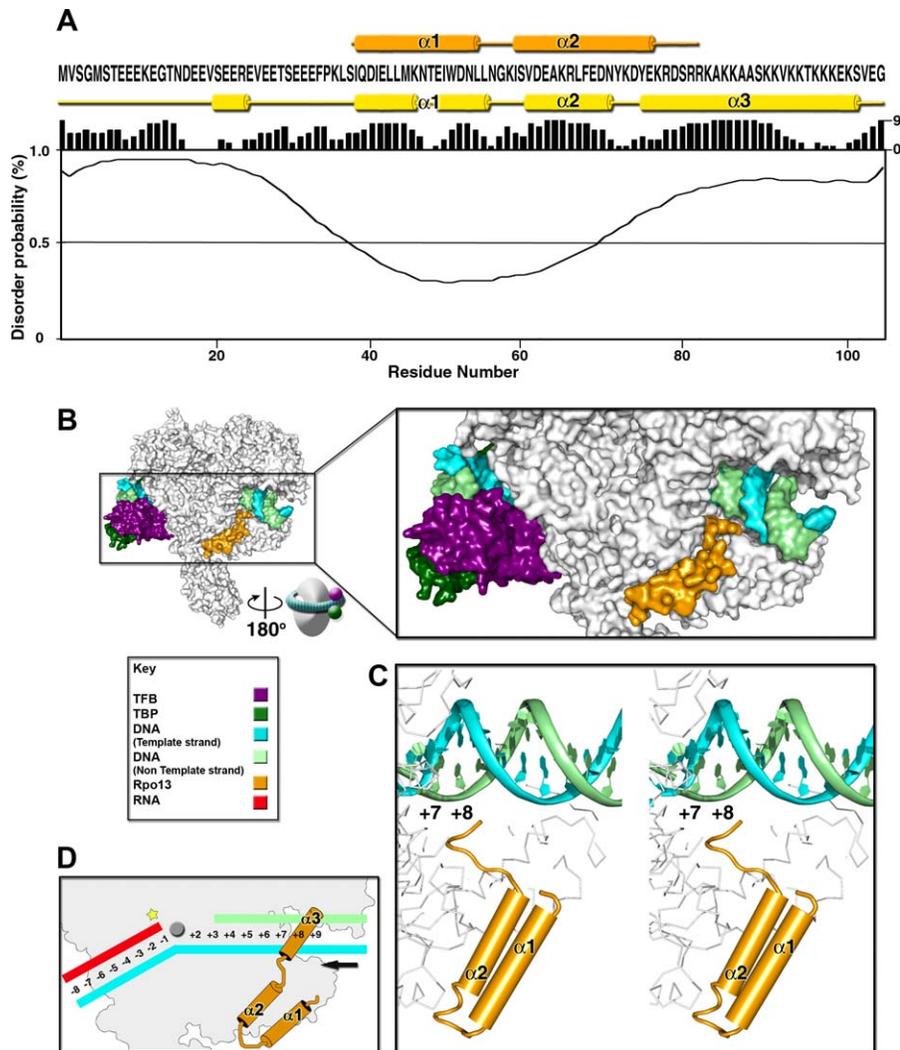
typical of DNA binding proteins [20] with an N-terminal extension (Figure 5A).

The HTH motif of Rpo13 (absent from the recent model of the RNAP from *S. solfataricus* [9]), fits between Rpo5 and the clamp-head domain of Rpo1N (Figures 1 and 3) in the position equivalent to that occupied by residues 1377–1420, in the  $\beta'$  large subunit of the crystal structures of the RNAP from the bacteria *Thermus aquaticus* and *T. thermophilus* RNAP [21,22] (Figure 3C). The corresponding sequence of this bacterial insertion is highly conserved within the Thermus-Deinococcus phylum, but there is also some detectable sequence similarity in other bacterial  $\beta'$  sequences mainly of the Proteobacteria phylum. In *T. aquaticus* and *T. thermophilus*, this structure (whose function has not yet been characterized) folds into two antiparallel  $\beta$  strands followed by an  $\alpha$ -helix; an organisation quite different from our Rpo13 (Figure 3C). In eukaryotes, the equivalent locus is partially occupied by the eukaryote-specific Rpb5 jaw domain (Figure 3B). Significantly, this domain is involved in downstream DNA binding [23], and the entire subunit has also been implicated in contacting transcription factor II B [24].

### RNAP Plasticity

We have used a normal-mode-based protocol [25] for structure refinement (see Material and Methods).

Although it is difficult to recognize biologically relevant modes per se, the principal modes present a simple dynamic picture of the RNAP, which might have relevance in vivo. The low-order modes include a pincer movement of the jaw-lobe module and clamp, in agreement with the structural variability found in static eukaryotic RNAP structures [2,26,27], and a contraction of the same structural elements generating a “ratchet” of the HTH motif and rudder (residues 278–297) in the clamp-core of Rpo1N; movements that can be detected by observing, for example, modes 2, 28, and 29 given in Protocols S1 and S2. This conformational plasticity would facilitate the transition of the RNAP from the apo- to the DNA-bound form for transcription initiation [2]. The Rpo4/7 subunits also swing



**Figure 5.** Archaeal Preinitiation Complex Model

(A) Secondary structure elements of Rpo13 model (orange) aligned with the amino acid sequence (below), yellow secondary structure prediction (Protein Prediction [19]) with reliability histogram (0 = low to 9 = high), and graph of disorder prediction (RONN [18]).

(B) Left, overall architecture with RNAP, TBP-TFB-DNA modelled as described in the main text, represented as a surface and coloured as in the key. Right, enlarged view of the TBP-TFB-DNA complex with the DNA wrapping the RNAP from the back (as viewed from left to right) with the downstream DNA in the entry channel close to Rpo13.

(C) Stereo view of the two ordered helices ( $\alpha 1$  and  $\alpha 2$ ) of Rpo13 and DNA with the C-terminal end proximal to the DNA major groove at nucleotide +8.

(D) Cartoon representation of the model of action of Rpo13 during polymerase activity (elements described in the key); star (yellow) marks the catalytic site, circle (grey) the bridge helix, and arrow (black) the direction of DNA translocation. The nucleic acid arrangement at the fork reflects the RNAP II elongation complex structure [46]. The location and orientation of Rpo13  $\alpha 3$  onto the DNA are speculative.

doi:10.1371/journal.pbio.1000102.g005

along the side of the polymerase, a flexibility underlining the multiple roles played by these subunits [8]. This analysis defines the catalytic site as a rigid ensemble relative to the rest of the structure (Figure 4), providing enzymatic precision at the heart of a flexible machine [28].

### A Model for the Archaeal Preinitiation Complex

To gain insights into the assembly of the archaeal preinitiation complex, we have docked our intact RNAP onto the DNA-RNA hybrid visualized in the RNAP II transcribing complex [29], the eukaryotic RNAP-TFIIB complex [5], and the TFBc-TBP-DNA complex from *Pyrococcus woesei* [30] (TFB and TBP, respectively, share 51% and 46% sequence identity with their *S. shibatae* orthologs) (see Figures

5B and 5C and Text S2). The overall architecture of the archaeal preinitiation complex resembles the minimal initiating eukaryotic complex RNAP-TFIIB-TBP-DNA [5,31], but differs by not requiring homolog basal factors TFIIF and TFIIF and by the influence of TFE [6,8]. Our structure and the preinitiation complex model provide a rationale for this minimal set of cofactors (see Discussion below).

## Discussion

### Evolutionary Implications of the Newly Located RNAP Subunits

Our results clarify the evolutionary relationships of Archaea with Eukarya and Bacteria. The finding that Rpo8

is an integral component of the *Sulfolobus* enzyme, together with the ordered clamp-head and jaw domains in Rpo1, underscores the fact that the crenarchaeal and eukaryotic RNAPs have conserved the same basic enzymatic platform even when the sequence identity is lower than 15%. This implies a closer structural ancestry between RNAPs from Crenarchaea and Eukaryotes. A major structural difference however is the presence of the carboxyl-terminal-domain (CTD) in eukaryotic Rpb1. This feature presumably represents a later evolutionary acquisition, acting as a bolt-on module that facilitates coordination of eukaryote-specific cotranscriptional processing events such as capping, splicing, and polyadenylation.

On the other hand, whereas archaeal RNAP structurally anticipates the enzymatic machinery of the eukaryotic systems, RNAPs from the order Sulfolobales (and others from the Crenarchaea kingdom) have acquired Rpo13, a subunit that is not present in Eukarya and that corresponds architecturally to an insertion into the bacterial  $\beta'$  subunit of *T. aquaticus* and *T. thermophilus*. Sequence analysis [32] shows that the *Rpo13* gene has orthologs in the orders Sulfolobales and Desulfurococcales of the Crenarchaeota phylum, but not in Euryarchaea (Figure S5). We have detected Rpo13 as part of the RNAP in three *Sulfolobus* species: first in our structure from *S. shibatae* (in both crystal forms, see Materials and Methods and Figure S6), then in *S. acidocaldarius* [12] and finally in *S. solfataricus* where we have reanalyzed the electron density (Figure S7). These results indicate that Rpo13 constitutes a stable structural component of the enzyme. Furthermore, the topology and location of the ordered fragment of Rpo13 suggests a mechanism of action in the context of the preinitiation complex assembly model (see below).

### Putative Role of Rpo13 at Initiation and Elongation

Although the function of Rpo13 is still unknown, its location leads us to hypothesize roles at initiation and elongation. Rpo13 could facilitate transcription bubble formation once the archaeal preinitiation complex is formed (modelled as in Figure 5B and 5C) and the N-terminal domain of TFB has driven the DNA towards the RNAP active centre [5,33]. The visible C-terminus of Rpo13 is at about 7 Å distance from the phosphate group of the nucleotide at position +8 from the start site of the modelled non-template DNA strand (Figure 5C). This close juxtaposition combined with the prediction of Rpo13 being a basic trihelical bundle protein, reminiscent of DNA recognition proteins, suggests that the third predicted helix of Rpo13 may interact with DNA. Additionally, the fact that Rpo13 has ten lysines in the last 20 residues supports an interaction of the predicted C-terminal  $\alpha 3$  helix with negatively charged DNA rather than with the positively charged DNA binding cleft of Rpo1 (Figures 3A, 5A, 5C, and 5D). At initiation, the  $\alpha 3$  helix of Rpo13 could provide a lock point, against which the main body of the polymerase cleft can push or twist the DNA; during elongation these locking interactions would be overcome by the translocation forces but may still interfere with DNA duplex stability. In this manner, Rpo13 may perform some of the roles attributed to eukaryote-specific general transcription factors. In this view, the  $\alpha 1$  and  $\alpha 2$  helices of Rpo13 act as constitutive anchors onto the RNAP, whereas  $\alpha 3$  confers additional functionality, prefiguring some of the

capabilities of removable cofactors needed for eukaryotic initiation.

Moreover, the presence of Rpo13 illustrates how the ancestral core enzyme was modulated by incorporation of novel subunits, a process that in eukaryotes has led to the emergence of three distinct classes of nuclear RNAPs.

## Materials and Methods

**Production and purification of RNA polymerase from *S. shibatae*.** Production and purification of archaeal RNAP are described elsewhere [34]. Briefly, *S. shibatae* cells were grown in three steps of 4 d each to a final optical density at 600 nm ( $OD_{600}$ )  $\approx$  3.0. This growth served as inoculum for the final large-scale cell growth, carried on for additional 4 d and to an  $OD_{600} \approx$  4.0. The RNA polymerase was purified from the cell pellet by dialysis, Q Sepharose chromatography, and Hi-Trap heparin column.

**Crystallization and data collection.** Archaeal RNAP crystals were initially obtained with a microbatch under oil technique (Hampton) using purified RNAP at approximately 7.0 mg/ml in 150 mM KCl, 100 mM SrCl<sub>2</sub>, 100 mM Na-Cacodylate (pH 6.5), and 12% PEG MME 5K. This initial condition was expanded using a hanging-drop vapour diffusion technique and several attempts were made to optimize the original fragile crystals by adding 1 mM Zn<sup>2+</sup> to exploring different gradient concentration of PEG 20K. Different datasets were collected either from native or heavy-atom-soaked crystals. Useful data that contributed to the initial phasing came from two crystals, a native close to 3.5 Å and a W11 tungsten cluster soak close to 4.0 Å (unpublished data), belonging to  $P2_12_1$  space group (Crystal<sub>1</sub>) with two RNAP complexes per asymmetric unit (AU). A second native dataset was later collected from a crystal in  $P2_12_2$  space group (Crystal<sub>2</sub>) with one RNAP molecule in the AU. This was obtained by adding 5% glycerol to the above crystallization conditions and diffracted to a resolution of about 3.35 Å. All crystals were flash-frozen and the collected datasets indexed, integrated, and scaled using HKL [35]. A summary of the native data collection statistics is shown in Table 1.

**Structure determination and refinement.** Initially, the Crystal<sub>1</sub> structure was solved to a resolution of 6 Å by first finding a molecular replacement locked-rotation function solutions for two RNAP molecules in the asymmetric unit (GLRF program [36]), and subsequently by finding a phase translation function solution (MOLREP [37]) utilizing low-resolution phases obtained experimentally from crystals soaked in W11 tungsten clusters (SIRAS technique, SOLVE program [38]). Yeast RNAP polyaniline coordinates (Protein Data Bank [PDB] entry 1EN0) were used as a search model. The phases to approximately 4 Å were then 2-fold averaged and solvent flattened using DM [37] allowing the initial manual rebuilding of the different subunits as alanine models. When the data from the crystal in  $P2_12_2$  space group (Crystal<sub>2</sub>) were obtained, cross-crystal averaging between the two crystal forms was performed in GAP (unpublished program, D. I. Stuart and J. Grimes). Prior to this, rigid-body refinement of the initial polyaniline model was carried out in both crystal forms using REFMAC [37] and the correspondent 2Fo-Fc and Fo-Fc maps inspected for integrity of RNAP. The cross-averaged electron density map at 3.52 Å confirmed the presence of all subunits, including the Rpo8 and the additional density resembling a HTH motif close to Rpo1N (residues 97–151).

The sequence for each archaeal subunit was determined by PCR-mediated cloning and sequencing (see below) and docked onto the RNAP polyaniline model. The model was then improved by iterative manual building and refinement as described in detail in Text S1.

When the coordinates of the RNAP structure from *S. solfataricus* (PDB entry 2PMZ) [9] became available, the refinement process accelerated, and we turned our attention to the higher resolution data (3.35 Å; Crystal<sub>2</sub>) encouraged by its higher signal-to-noise ratio at 3.52 Å (2.5 vs. 2.0 of Crystal<sub>1</sub>) and confident of the correct but incomplete new starting model (PDB entry 2PMZ).

We finalized the refinement of our RNAP structure against the data collected from the  $P2_12_2$  crystal by iterative manual adjustment of the model, normal-mode [25] and positional with B overall refinement in REFMAC [37]. A summary of the final statistics for Crystal<sub>2</sub> is shown in Table 1. As a validation tool, we extended the refinement of the Crystal<sub>1</sub> RNAP model whose current statistics are also included in Table 1.

All major findings for the RNAP structure in the  $P2_12_2$  space group are observed in the  $P2_12_1$  structure including the presence of

the Rpo13 subunit bound to the two RNAP molecules in the asymmetric unit (Figure S6). Coordinates and structure factors for both crystal forms have been deposited in the Protein Data Bank under ID codes 2WB1 (Crystal\_1) and 2WAQ (Crystal\_2).

**Sequencing of RNA polymerase subunits.** Open reading frames for the *S. shibatae* RNAP subunits were amplified by PCR from genomic DNA using Pfu DNA polymerase prior to cloning in PCR script (Stratagene) and sequencing (Geneservice). Primers were designed based on the conservation of flanking genes in *S. solfataricus* and *S. tokodaii* and are listed in Table S1. Table S2 shows the sequence identity with *S. solfataricus*.

**Identification of subunit Rpo13.** Inspection of the electron density of neighbouring subunits and superimposition with known RNAP components excluded the 2Fo-Fc electron density map correspondent to the unknown HTH motif of being an ordered or cleaved fragment of an already identified RNAP subunit. This structural element is reminiscent of the HTH domain of basal and specific transcription factors including the archaeal TFE [39] and DNA-binding proteins [20,40]. However, biochemical data rule out the possibility that this represents a fragment of TFE (unpublished data) and its binding locus is inconsistent with it belonging to transcription factor S (TFS) [4] or TFB [5]. To identify the novel archaeal subunit, we carried out liquid chromatography-mass spectrometry with peptide fingerprinting [LC MS/MS; (HCTplus; Bruker Daltonics) coupled to a HPLC system (Ultimate; Dionex/LC-Packings)] of the sample used for crystallization at the Central Proteomics Facility Headington (CPF, University of Oxford). Using the sequenced genome of *S. solfataricus* [41] through an in-house Mascot server [42] (Matrixscience), we were able to fingerprint a protein component rich in lysine residues (Uniprot: SSO0396) corresponding to a previously detected RNAP component of the 13-subunit RNAP from *S. acidocaldarius* [12]. This gene is cotranscribed with *pcna3* (a DNA replication accessory factor), a situation reminiscent of the clustering of genes of other Rpos and information processing components [43].

The peptide mapping was carried out on the trypsin-digested fractions corresponding to molecular weight bands between 5 and 25 kDa extracted from a SDS-polyacrylamide gradient gel electrophoresis of the RNAP used for crystallization. The 2Fo-Fc electron density supported a protein binding to the RNAP in a 1:1 ratio and the packing restraints suggested a molecule no larger than approximately 250 residues. The gene for the homologous protein in *S. shibatae* was then sequenced and used for secondary and disorder structure predictions and model building (Figure 5A). The data search was rerun against our *S. shibatae* sequence database, and Figure S8 shows the Rpo13 peptides identified and the sequence coverage.

## Supporting Information

**Figure S1.** Stereo View of 2Fo-Fc and Fo-Fc Electron Density Maps of Rpo13

Stereo view of 2Fo-Fc (blue, contoured at  $0.9\sigma$ ) and Fo-Fc (red, contoured at  $2.5\sigma$ ) sigmaA-weighted maps prior the assignment of the Rpo13 sequence onto the poly-alanine helix-turn-helix model (as stick orange) in the RNAP structure of the crystal in  $P2_12_12$  space group. The black circle highlights the density corresponding to the residue used as a marker for sequence assignment.

Found at doi:10.1371/journal.pbio.1000102.sg001 (1.26 MB DOC).

**Figure S2.** Structure-Based Phylogenetic Trees of the Core Conserved Subunits in Archaea, Eukarya, and Bacteria

Structures were generated using a modified version of the Structure Homology Program [47] and PHYLIP package [48]. At the centre the surface representation of the archaeal core enzyme. Nomenclature as Figure 1.

Found at doi:10.1371/journal.pbio.1000102.sg002 (1.33 MB DOC).

**Figure S3.** Stereo View of Superimposed Ca Traces of Archaeal Rpo8 and Eukaryotic Rpb8

Stereo view of superimposed C $\alpha$  traces of archaeal Rpo8 (red ribbon) and eukaryotic Rpb8 (yellow ribbon; PDB entry 1L50); N and C label respectively the N terminus and C terminus of Rpo8.

Found at doi:10.1371/journal.pbio.1000102.sg003 (799 KB DOC).

**Figure S4.** Comparison of Surface Charge Distribution of Archaeal Rpo8 versus Eukaryotic Rpb8

Comparison of the surface charge distribution (blue positive, red

negative and white neutral) of archaeal Rpo8 versus the eukaryotic Rpb8 viewed accordingly to ssDNA binding surfaces of Rpb8 [49]. The left panel is related to the Rpo8 viewed in Figure 2A by  $90^\circ$  rotation, and the right panel by a further  $90^\circ$  as indicated by the central arrow. (A) Electrostatic potential isosurface of Rpo8 with the positions of basic residues lining on the crest of the molecule (K56 not visible). (B) Rpb8 potential surface.

The charges assignments were generated with PDB2PQR program using Amber charges [50].

Found at doi:10.1371/journal.pbio.1000102.sg004 (2.77 MB DOC).

**Figure S5.** Multiple Sequence Alignment of Rpo13 Orthologs in Sulfolobales and Desulfurococcales

Multiple sequence alignment (ClustalW [51] and Esript [52]) between the protein sequences of Rpo13 in *Sulfolobus* and *Desulfurococcales*. *Sulfolobus solfataricus* (gene: SSO0396), *Sulfolobus tokodaii* (gene: ST0398), *Sulfolobus acidocaldarius* (gene: Saci\_0816), *Metallosphaera sedula* (gene: Msed\_0052), and *Staphylothermus marinus F1* (gene: Smar\_1004).

Found at doi:10.1371/journal.pbio.1000102.sg005 (920 KB DOC).

**Figure S6.** Stereo Views of 2Fo-Fc Electron Density Maps of the Two Independent Rpo13 Subunits of the Crystal in  $P2_12_12$  Space Group

Stereo views of the two independent Rpo13 subunits (as orange C $\alpha$  trace) within the asymmetric unit of the crystal in  $P2_12_12$  space group with correspondent 2Fo-Fc maps (contoured at  $0.9\sigma$ ) calculated from the current model (see Crystal\_1 refinement in Table 1). Both Rpo13 molecules slightly differ by having less ordered N- and C- terminal extensions than the Rpo13 modelled in the  $P2_12_12$  crystal, a situation contributed by the different packing environment constraints that supports the prediction of a more flexible N- and C-terminal domains (Figure 5A).

Found at doi:10.1371/journal.pbio.1000102.sg006 (1.03 MB DOC).

**Figure S7.** Stereo View of 2Fo-Fc and Fo-Fc Electron Density Maps of Rpo13 Corresponding Regions in the RNAP from *S. solfataricus*

Stereo view of 2Fo-Fc and Fo-Fc electron density maps calculated via EDS in Coot [53] from the PDB entry 2PMZ corresponding to the structure of the RNAP from *Sulfolobus solfataricus* [9] and contoured respectively at  $0.9\sigma$  (blue) and  $2.3\sigma$  (positive green; negative red) of the corresponding region where we have found Rpo13 in our RNAP. Two side-by-side rod-like shaped densities are clearly visible packing against Rpo5 (as yellow ribbon). Density was also observed for subunit Rpo8 in *Sulfolobus solfataricus* RNAP structure [9] (unpublished data).

Found at doi:10.1371/journal.pbio.1000102.sg007 (1.50 MB DOC).

**Figure S8.** Mascot Data Results for Rpo13

Mascot data results for Rpo13 (*S. shibatae*) from LC MS/MS experiment. On top, the sequence coverage (63%; red) obtained by identification of the fingerprinted peptides generated by trypsinization of gel fractions.

Found at doi:10.1371/journal.pbio.1000102.sg008 (1.17 MB DOC).

**Protocol S1.** Modes from 1 to 15 Correspond to DCD Files Labelled ca#.dcd Where # Corresponds to the Mode Number

To visualize the motions of the individual normal modes, the C $\alpha$  pdb coordinate file (labelled ca.pdb) and the mode of interest can be uploaded in the VMD program (<http://www.ks.uiuc.edu/Research/vmd/>).

Found at doi:10.1371/journal.pbio.1000102.sd001 (5.79 MB ZIP).

**Protocol S2.** Modes from 16 to 30 Correspond to DCD Files Labelled ca#.dcd Where # Corresponds to the Mode Number

To visualize the motions of the individual normal modes, the C $\alpha$  pdb coordinate file (labelled ca.pdb) and the mode of interest can be uploaded in the VMD program (<http://www.ks.uiuc.edu/Research/vmd/>).

Found at doi:10.1371/journal.pbio.1000102.sd002 (5.76 MB ZIP).

**Protocol S3.** Modes from 31 to 50 Correspond to DCD Files Labelled ca#.dcd Where # Corresponds to the Mode Number

To visualize the motions of the individual normal modes, the C $\alpha$  pdb coordinate file (labelled ca.pdb) and the mode of interest can be uploaded in the VMD program (<http://www.ks.uiuc.edu/Research/vmd/>).

Found at doi:10.1371/journal.pbio.1000102.sd003 (7.68 MB ZIP).

**Table S1.** Primers for Sequencing of *S. shibatae* RNAP Subunits  
Found at doi:10.1371/journal.pbio.1000102.st001 (39 KB DOC)

**Table S2.** Sequence Identity to *S. solfataricus* Ortholog  
Found at doi:10.1371/journal.pbio.1000102.st002 (44 KB DOC)

**Text S1.** Supplementary Methods: Model Building and Refinement Protocol  
Found at doi:10.1371/journal.pbio.1000102.sd004 (41 KB DOC).

**Text S2.** Supplementary Methods: Modelling of the Preinitiation Complex  
Found at doi:10.1371/journal.pbio.1000102.sd005 (32 KB DOC).

## Acknowledgments

NGAA is indebted to S. Harrison for his support throughout the project and for his valuable comments on the manuscript. We thank K. Severinov and J. Brown for suggestions on the text, B. Poon and S.

## References

- Kwapisz M, Beckouet F, Thuriaux P (2008) Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet* 24: 211–215.
- Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292: 1863–1876.
- Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292: 1876–1882.
- Kettenberger H, Armache KJ, Cramer P (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* 114: 347–357.
- Bushnell DA, Westover KD, Davis RE, Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocystal at 4.5 Angstroms. *Science* 303: 983–988.
- Bell SD, Jaxel C, Nadal M, Kosa PF, Jackson SP (1998) Temperature, template topology, and factor requirements of archaeal transcription. *Proc Natl Acad Sci U S A* 95: 15218–15222.
- Geiduschek EP, Ouhammouch M (2005) Archaeal transcription and its regulators. *Mol Microbiol* 56: 1397–1407.
- Werner F (2007) Structure and function of archaeal RNA polymerases. *Mol Microbiol* 65: 1395–1404.
- Hirata A, Klein BJ, Murakami KS (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451: 851–854.
- Kusser AG, Bertero MG, Naji S, et al. (2008) Structure of an archaeal RNA polymerase. *J Mol Biol* 376: 303–307.
- Davis IW, Murray LW, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35: W375–W383.
- Lanzendorfer M, Langer D, Hain J, et al. (1994) Structure and function of the DNA-dependent RNA polymerase of *Sulfolobus*. *Syst Appl Microbiol* 16: 656–664.
- Iyer LM, Koonin EV, Aravind L (2003) Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol* 3: 1.
- Grunberg S, Bartlett MS, Naji S, Thomm M (2007) Transcription factor E is a part of transcription elongation complexes. *J Biol Chem* 282: 35482–35490.
- Todone F, Brick P, Werner F, Weinzierl RO, Onesti S (2001) Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol Cell* 8: 1137–1143.
- Ouhammouch M, Geiduschek EP (2005) An expanding family of archaeal transcriptional activators. *Proc Natl Acad Sci U S A* 102: 15423–15428.
- Werner F (2008) Structural evolution of multisubunit RNA polymerases. *Trends Microbiol* 16: 247–250.
- Yang ZR, Thomson R, McNeil P, Esnouf RM (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21: 3369–3376.
- Rost B, Yachdav G, Liu J (2004) The PredictProtein server. *Nucleic Acids Res* 32: W321–326.
- Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* 29: 231–262.
- Campbell EA, Korzheva N, Mustaev A, et al. (2001) Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104: 901–912.
- Vassilyev DG, Sekine S, Laptchenko O, et al. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417: 712–719.
- Cramer P, Bushnell DA, Fu J, et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288: 640–649.
- Graham for discussions, R. Esnouf for bioinformatics support, and B. Kessler and M. Edlmann for expert assistance with sample preparation and analysis by tandem mass spectrometry.
- Author contributions.** YK, PBS, SDB, and NGAA conceived and designed the experiments. YK, UMU, OL, PJJ, SDB, and NGAA performed the experiments. YK, PBS, DIS, SDB, and NGAA analyzed the data. DIS, SDB, and NGAA wrote the paper.
- Funding.** This investigation was supported in part by the Howard Hughes Medical Institute and the National Institutes of Health (PBS), Medical Research Council, UK (DIS and NGAA) and Edward Penley Abraham Trust (SDB). YK was a Eugene V. Weissman-Irvington Institute for Immunological Research fellow. NGAA is supported by a senior research contract by the Basque Foundation for Science-Ikerbasque and acknowledges travel grants from the Generalitat de Catalunya and the Crystallography Fund-University of Oxford. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
- Competing interests. The authors have declared that no competing interests exist.
- Todone F, Weinzierl ROJ, Brick P, Onesti S (2000) Crystal structure of RPB5, a universal eukaryotic RNA polymerase subunit and transcription factor interaction target. *Proc Natl Acad Sci U S A* 97: 6306–6310.
- Poon BK, Chen X, Lu M, et al. (2007) Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-Å crystallographic resolution. *Proc Natl Acad Sci U S A* 104: 7869–7874.
- Bushnell DA, Kornberg RD (2003) Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* 100: 6969–6973.
- Westover KD, Bushnell DA, Kornberg RD (2004) Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* 303: 1014–1016.
- Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* 438: 460–465.
- Westover KD, Bushnell DA, Kornberg RD (2004) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119: 481–489.
- Littlefield O, Korkhin Y, Sigler PB (1999) The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A* 96: 13668–13673.
- Parvin JD, Sharp PA (1993) DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* 73: 533–540.
- Schneider KL, Pollard KS, Baertsch R, Pohl A, Lowe TM (2006) The UCSC Archaeal Genome Browser. *Nucleic Acids Res* 34: D407–D410.
- Renfrow MB, Naryshkin N, Lewis LM, et al. (2004) Transcription factor B contacts promoter DNA near the transcription start site of the archaeal transcription initiation complex. *J Biol Chem* 279: 2825–2831.
- Korkhin Y, Littlefield O, Nelson PJ, Bell SD, Sigler PB (2001) Preparation of components of archaeal transcription preinitiation complex. *Methods Enzymol* 334: 227–239.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276: 307–326.
- Tong L, Rossmann MG (1997) Rotation function calculations with GLRF program. *Methods Enzymol* 276: 594–611.
- Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* 50: 760–763.
- Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D* 55: 849–861.
- Meinhart A, Blobel J, Cramer P (2003) An extended winged helix domain in general transcription factor E/II $\alpha$ . *J Biol Chem* 278: 48267–48274.
- Harrison SC, Aggarwal AK (1990) DNA recognition by proteins with the helix-turn-helix motif. *Annu Rev Biochem* 59: 933–969.
- She Q, Singh RK, Confalonieri F, et al. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* 98: 7835–7840.
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20: 3551–3567.
- Langer D, Hain J, Thuriaux P, Zillig W (1995) Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* 92: 5768–5772.
- Stuart DI, Levine M, Muirhead H, Stammers DK (1979) Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. *J Mol Biol* 134: 109–142.
- DeLano WL (2008) The PyMOL molecular graphics system. Palo Alto (California): DeLano Scientific.
- Kettenberger H, Armache KJ, Cramer P (2004) Complete RNA polymerase II elongation complex structure and its interaction with NTP and TFIIS. *Mol Cell* 16: 955–965.
- Bamford DH, Grimes JM, Stuart DI (2005) What does structure tell us about virus evolution? *Curr Opin Struct Biol* 15: 655–663.

48. Felsenstein J (1989) PHYLIP: Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164–166.
49. Kang X, Hu Y, Li Y, et al. (2006) Structural, biochemical, and dynamic characterizations of the hRPB8 subunit of human RNA polymerases. *J Biol Chem* 281: 18216–18226.
50. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32: W665–667.
51. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
52. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15: 305–308.
53. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60: 2126–2132.