Promoter-proximal elongation regulates transcription in archaea

Fabian Blombach, Thomas Fouqueau, Dorota Matelska, Katherine Smollett & Finn Werner

Recruitment of RNA polymerase and initiation factors to the promoter is the only known target for transcription activation and repression in archaea. Whether any of the subsequent steps towards productive transcription elongation are involved in regulation is not known. We characterised how the basal transcription machinery is distributed along genes in the archaeon *Saccharolobus solfataricus*. We discovered a distinct early elongation phase where RNA polymerases sequentially recruit the elongation factors Spt4/5 and Elf1 to form the transcription elongation complex (TEC) before the TEC escapes into productive transcription. TEC escape is rate-limiting for transcription output during exponential growth. Oxidative stress causes changes in TEC escape that correlate with changes in the transcriptome. Our results thus establish that TEC escape contributes to the basal promoter strength and facilitates transcription regulation. Impaired TEC escape coincides with the accumulation of initiation factors at the promoter and recruitment of termination factor aCPSF1 to the early TEC. This suggests two possible mechanisms for how TEC escape limits transcription, physically blocking upstream RNA polymerases during transcription initiation and premature termination of early TECs.
The rate of RNA synthesis is determined by the frequency of transcription initiation and premature termination. The recruitment stage of transcription initiation is the main target for regulation in yeast and bacteria, however, the initiation rate can also be affected indirectly by downstream events. In metazoans, promoter-proximal pausing of RNA polymerase II with slow turnover times blocks pre-initiation complex (PIC) formation for the following RNAP and is thereby widely rate-limiting for transcription initiation. Promoter-proximal paused RNAPII can also be subject to premature termination providing an additional way of transcription regulation. Promoter-proximal RNA dynamics also limit gene expression in E. coli. Well-established processes of post-recruitment regulation include Sigma70-dependent pausing and transcription attenuation mediated by premature termination. Another possible underlying molecular mechanism might be pausing during initial transcription, though its contribution to genome-wide gene regulation remains to be investigated.

Archaea form the ‘third domain of life’ next to bacteria and eukaryotes, with the latter likely originating from an archaeal ancestor. The basal archaeal transcription machinery represents an evolutionarily ancient core of the RNAP II system encompassing RNAP subunits, basal transcription initiation and -elongation factors and core promoter elements. The mechanisms of initiation have been characterised in great detail in vitro. The basal transcription factors TBP and TFB bind to their cognate promoter elements (TATA box and BRE, respectively) and sequester RNAP to form the minimal PIC. A third transcription initiation factor TFE binds to RNAP to form the complete PIC and facilitates DNA melting leading to formation of the open complex. TFE stimulates transcription initiation but is not strictly required in vitro. Like the PIC, the transcription elongation complex (TEC) corresponds to an evolutionarily ancient RNAPII Tec encompassing homologues of a subset of RNAPII elongation factors: Spt4/5 (DSIF in human) and TFE (Elof1 in humans). In addition, the transcript cleavage factor TFS (homologous to TFIIIS) transiently associates with the TEC and reactivates arrested TECs. Spt4/5 and TFB bind to RNAP in a mutually exclusive manner and the transition from transcription initiation to elongation requires factor switching between TFE and Spt4/5. Transcription termination in archaea occurs via intrinsic or factor-dependent mechanisms. The latter involves termination factor aCPSF1 (or FtaA), a ribonuclease that is evolutionarily related to the RNAP II termination factor CPSF73 and the integrator subunit Ints11.

Archaeal promoters seem to comprise fewer promoter elements compared to their bacterial and eukaryotic counterparts, but it is possible that additional unknown sequence elements as well as the physicochemical properties of promoter DNA contribute to promoter strength. Likewise, our understanding of transcription regulation is limited to factors modulating the recruitment of PICs where repression generally involves steric hindrance of RNAP or basal initiation factor binding and activation is achieved by enhancing their binding. How archaeal RNA polymerase progresses further through the transcription cycle and whether subsequent stages beyond initiation are targeted for transcription regulation in archaea is currently poorly understood.

The crenarchaeon *Saccharolobus solfataricus* (formerly *Sulfolobus*) is a well-established model organism for archaeal transcription (e.g., refs. 24,32,43–45). Importantly, *S. solfataricus* harbours the full repertoire of known archaeal transcription initiation and elongation factors including Elof1 making it a good model to study mechanisms of transcription regulation beyond initiation.

We analysed the genome-wide distribution of RNAP and transcription initiation and elongation factors in *S. solfataricus* by using a multi-omics approach including chromatin immunoprecipitation–sequencing-based techniques (ChIP-seq) and transcriptomics. Our results provide evidence for a sequential recruitment cascade of elongation- (Spt4/5 and Elof1) and termination (aCPSF1) factors to RNAPs in the promoter-proximal region of the transcription unit. We show that escape of TECs from this region is rate-limiting for transcription and subject to regulation. Thereby we establish early elongation as an important checkpoint to set and regulate promoter strength in archaea.

**Results**

**Uniform PIC assembly during exponential growth.** We mapped the genome-wide occupancy of RNAP, initiation-, elongation- and termination factors to shed light on how the individual stages of transcription are subject to transcription regulation in *S. solfataricus*. We developed and adapted ChIP-seq using polyclonal antibodies raised against RNAP subunits Rpo4/7 and recombinant transcription factors. In order to obtain the resolution that separates PICs from promoter-proximal, early TECs, we adapted a ChIP-exo approach for RNAP and initiation factors TFB and TFEβ that includes 5'-3' exonuclease-trimming of the immunoprecipitated-DNA fragments.

We calculated aggregate profiles of ChIP-exo data for a set of 2587 TUs with mapped TSSs. These TUs were selected from a total set of 1054 mapped TSSs based on the corresponding ChIP-seq data for TFB and TFEβ to include only TUs displaying TFB and TFEβ occupancy as well as to exclude TUs with problematic regions for the mapping of sequencing reads (see Methods section). The profiles showed a distinct footprint for RNAP and initiation factors TFB and TFEβ around the TSS (Fig. 1b). The overall similarity of the RNAP, TFB and TFEβ profiles reflect the footprints of entire cross-linked PICs rather than the DNA-binding sites of the individual factors within the PIC. The main upstream border of the PIC is formed by a broad peak centred around position −12 to −14 that can be most likely attributed to the N-terminal cyclin fold of TFB interacting with the DNA downstream of the TATA-box, which is in good agreement with ChIP-exo mapping of RNAPII PICs. The downstream border for the PIC signal on the template strand was relatively broad and reached well beyond the ~20 bp downstream of the TSS protected in vitro exonuclease foot-printing experiments of archaeal PICs.

To investigate any heterogeneity in the recruitment of basal factors, we quantified the ChIP-exo signal on the non-template strand over a window from −30 to +20 relative to the TSS (Fig. 1c). Both TFB- and TFEβ occupancy correlated strongly with RNAP (Spearman’s r = 0.92 in both cases, Fig. 1d, e). Since TFB binding is critically dependent on TBP binding, and TFE binding depends on RNAP, our results suggest that all components of the archaeal PIC (TBP, TFB, RNAP and TFE) assemble on promoters in a homogenous, or uniform, fashion.

We expected exceptions to this rule where transcription regulators would interfere with the recruitment of RNAP. E.g., the SSO8620 promoter shows strong TFB- but weak RNAP- and TFEβ signals, which indicates repression of RNAP recruitment to the TBP-TFB ternary complex. Consistent with this notion, the predominant TFB footprint on SSO8620 was significantly narrower compared to TFB footprints on promoters showing unimpaired RNAP recruitment to the PIC (Supplementary Fig. 1).
One possible explanation for why the ChIP-exo footprint of PICs was extended downstream could be that the PICs might be in a state of extended DNA scrunching where they ‘reel in’ downstream DNA during initial transcription. DNA scrunching results in downstream extension of the DNA bubble thereby making thymine bases within the melted region sensitive to permanganate. Promoter clearance by RNAP limits the extent of DNA scrunching and in vitro crosslinking data suggest that archaeal RNAP clears from the promoter approximately when it reaches position +1051. To test whether PICs undergo extended DNA scrunching in vivo beyond the anticipated position of promoter clearance, we mapped the melted DNA regions in the PIC genome wide by permanganate ChIP-seq using TFB as IP target. Aggregate plots showed that DNA melting occurred in the −12 to +3 region relative to the TSS, peaking at position −10 (Fig. 1f), which is consistent with the in vitro permanganate footprinting of reconstituted PICs. Importantly, the signal decreased to background levels beyond position +10, the expected point of promoter clearance. Thus, extended DNA scrunching is unlikely to explain the downstream border of PICs.

The discrepancy between in vitro exonuclease and in vivo ChIP-exo footprints suggest that additional, yet uncharacterised components associate with the PIC in the cell such as chromatin proteins or extended interaction of the PIC with downstream DNA.

RNAP escape limits productive transcription. How does archaeal RNAP progress from transcription initiation into productive elongation? To address this poorly understood process, we generated paired-end ChIP-seq data sampled to a mean fragment size of 120 bp. Because the S. solfataricus genome has very short intergenic regions with juxtaposed promoters of different TUs, ChIP-seq data with such short mean fragment size provide a good compromise between the requirement of good spatial resolution and the overall higher robustness of ChIP-seq compared to ChIP-exo. These data ensured unequivocal assignment of initiation factor peaks to specific promoters. At the promoters, the RNAP, TFB and TFEβ ChIP-seq data were in good agreement with ChIP-exo data, i.e., indicative of uniform PIC assembly (Supplementary Fig. 2). Choosing transcription

---

**Fig. 1 Uniform PIC assembly during exponential growth.** a Stages of transcription in archaea can be assessed by a combination of RNAP and basal transcription initiation and elongation factor occupancy. b Aggregate plots of ChIP-exo signal at the promoter for RNAP and initiation factors (n = 298 TUs). The average signal on the non-template and template strand is shown above and below the line, respectively. Data are pooled from three biological replicates. c Schematic showing the 50-nt window on the non-template strand that we used to quantify and correlate the ChIP-exo signal. Because RNAP and initiation factors TFB and TFEβ yield similar profiles on the non-template strand, this signal can be attributed to the PIC. d, e Scatter plots depicting correlation between the main ChIP-exo signal for TFB and RNAP (d) or TFEβ (e) within the PIC (n = 298). Data represent the average signal over a 50-nt window that we attributed to the PIC (see panel c). The geometric mean of three biological replicates is shown. f Aggregate plots of permanganate ChIP-seq signal at the promoter for initiation factor TFB (n = 298 TUs). The average signal for T-encoding positions on the non-template and template strand is shown above and below the line, respectively. Data are pooled from two biological replicates.
TECs accumulate in the promoter-proximal region. The observed accumulation of RNAP in the promoter-proximal region can be due to PICs or TECs. If TECs accumulate in this region, then the elongation factor Spt4/5 and possibly Elf1 should show similar promoter-proximal accumulation as RNAP. To test this, we classified TUs based on their RNAP escape by calculating an escape index (EI) for each TU defined as the log-transformed ratio of RNAP_{adj} over RNAP_{pp}. We divided the TUs into two subsets with a high (EI > −1) or low escape index (EI < −2.5) and compared the aggregate profiles for RNAP, Spt4/5 and Elf1 (Fig. 3a, b). Both elongation factors accumulate in the promoter-proximal region of TUs with low EI alongside RNAP (Fig. 3b). In support of this, escape index calculations for both elongation factors revealed strong correlations with RNAP EI (Supplementary Fig. 4). Escape indices for all three proteins (RNAP, Spt4/5 and Elf1) were strongly correlated with mRNA expression levels (Supplementary Fig. 4). Thus, the observed accumulation of RNAP in the promoter-proximal region appears to reflect reduced TEC escape into productive transcription.

Notably, Spt4/5 is consistently recruited to the TEC prior to Elf1, independent of whether escape is high or low (Fig. 3a, b). Consecutive recruitment of elongation factors is consistent with the observation that some TUs with low RNAP escape showed a much stronger Spt4/5 recruitment compared to Elf1 suggesting that TECs stall at an earlier stage, prior to Elf1 recruitment, on these TUs (Supplementary Fig. 5). Henceforth we refer to the two TEC complexes as TEC_{Spt4/5} and TEC_{Spt4/5-Elf1}. Our data thereby also provide the first experimental evidence that the archaeal Elf1 homologue is a general part of the archaeal TEC.

In order to corroborate the differences in TEC escape with a second, independent method, we characterised the nascent RNAs synthesised at the 5′ end of the TUs referred to as TSS-RNAs. Short RNAs (20–200 nt length) were isolated, enriched for triphosphorylated 5′-ends using the Cappable-seq method and deep-sequenced. In quantitative terms, the occupancy of promoter-proximal elongation complexes (using Spt4/5_{pp} as proxy) correlated well with TSS-RNA read counts (Spearman’s $r = 0.61$) (Fig. 3c) and significantly better than mRNA counts from total RNA-seq (Spearman’s $r = 0.48$) (Supplementary Fig. 6). This is in line with the isolated short 5′ triphosphorylated RNAs being predominantly composed of nascent RNAs synthesised by early TECs rather than degradation products of full-length RNAs. In qualitative terms, TUs with low RNAP escape were associated with the synthesis of shorter TSS-RNAs (<50 nt) consistent with promoter-proximal accumulation of TECs (Fig. 3d).

In summary, our results demonstrate that RNAP accumulates in the promoter-proximal region in the form of early TECs that incorporate Spt4/5 and Elf1.

In order to corroborate the differences in TEC escape with a first experimental evidence the that archaeal Elf1 homologue is a general part of the archaeal TEC.

aCPSF1 recruitment to the TEC correlates with reduced TEC escape. The balance between premature termination and anti-termination in the promoter-proximal regions of genes is a well characterised mode of transcription regulation in bacteriophages and bacteria, and has more recently also been reported for eukaryotic transcription systems. Premature termination could also contribute to the observed promoter-proximal enrichment of TECs that we observed in archaea. As we found no evidence for any significant sequence bias in the promoter-proximal region including uridine-stretches that could serve as intrinsic terminators, we considered factor-dependent termination mediated by the archaeal termination factor aCPSF1. aCPSF1 is capable of inducing transcription termination on TECs stalled in the promoter-proximal region (+54) and intriguingly, aCPSF1 accumulated in the promoter-proximal-region of most TUs (188 out of 212 TUs with peaks passing detection threshold) including the genes dhg-1 and rps8E (Fig. 2b). We also analysed aCPSF1 association with predicted TSSs from pairs of TUs in divergent orientation. These TSSs are thus isolated from termination sites at the 3′-end of TUs. aCPSF1 peaks showed strong association with these TSSs similar to initiation factor TFB that we used as control in line with the results above (Supplementary Fig. 6). In contrast to the promoter-proximal aCPSF1 peaks, aCPSF1 does not form clearly defined peaks at 3′-ends of most TUs predicted from RNA-seq data. Instead, we observed a decrease in occupancy of aCPSF1 together with RNAP downstream of the predicted mRNA 3′-ends, and only in some cases well-defined CPSF1 peaks (Supplementary Fig. 8).

Provided that the promoter-proximal occupancy reflects recruitment of aCPSF1 to TECs, the distribution of aCPSF1 in the promoter-proximal region should depend on the distribution of RNAP. Accordingly, the aCPSF1 peaks sharpened on TUs with low RNAP escape likely due to a lower elongation rate or processivity (Fig. 3a, b). aCPSF1-mediated transcription termination is stimulated in the presence of Spt4/5 in vitro suggesting that Spt4/5 might facilitate aCPSF1 recruitment to the TEC or modulate aCPSF1 activity. In line with the in vitro observations, TUs with low RNAP escape recruited aCPSF1 consecutive to Spt4/5 (Fig. 3b).

Provided that aCPSF1 recruitment results in the premature termination of elongation complexes, its recruitment to the promoter-proximal TECs should decrease TEC escape and RNA levels. The CPSF1 recruitment was indeed inversely associated with TEC escape (Fig. 3e). This anticorrelation holds true whether the aCPSF1 load is calculated as ratio of aCPSF1 to Elf1 (Fig. 3e) or aCPSF1 to Spt4/5 promoter occupancy (Supplementary Fig. 9). Importantly, a higher aCPSF1 load was correlated with lower mRNA levels (Fig. 3f and Supplementary Fig. 9).

In summary, our data show that promoters with high levels of promoter-proximal aCPSF1 recruitment show decreased TEC escape and low mRNA levels. These observations demonstrate the
Fig. 2 Productive transcription is limited by RNAP escape. 

a Heatmap of ChIP-seq data for RNAP and the basal transcription machinery on a selected set of 212 TUs for exponential growth phase. The corresponding RNA-seq data for the plus strand are depicted on the right. Data are based on one representative of two biological replicates.
b–e ChIP-seq occupancy plots on *thsB* coding for a subunit of the thermosome chaperone complex (b), *rps8e* (c), *dhg-1* coding for a glucose-1-dehydrogenase (d) and CRISPR C (e). Traces show mean occupancy for two biological replicates with the range depicted as semi-transparent ribbon.
f–h Correlation of steady-state mRNA levels with RNAP occupancy at the promoter (RNAP$_p$, f), the TU body (RNAP$_b$, g) and TFB promoter occupancy (h), $n = 211$ TUs. ChIP-seq data represent the geometric mean of two biological replicates. Rockhopper estimates of mRNA levels are based on two biological replicates.
The link between the termination factor aCPSF1 and RNA output and are consistent with a premature termination mechanism.

The **CRISPR C** promoter shows pausing in the promoter-proximal region in vitro. To test whether promoter-proximal pausing of RNAP can be observed in vitro for low TEC escape promoters, we developed a synchronised in vitro transcription assay to monitor promoter-proximal transcription elongation dynamics. We used *S. solfataricus* cell lysates generated from cells in exponential growth phase to provide a full set of auxiliary factors.

To generate templates for in vitro transcription, we inserted target promoter regions encompassing −50 to +100 relative to the TSS into plasmids that were subsequently linearised downstream of the insert to allow for run-off transcripts of 115 nt length. Synchronisation of transcription was achieved by a transcriptionally inhibitory variant of TFB termed TFBc that comprises only the C-terminal cyclin fold domains. Pre-formed PICs are able to initiate a single round of transcription, but subsequent PIC assembly and transcription re-initiation is blocked by an excess of TFBc outcompeting the inherent TFB for recruitment to TBP-bound promoters (Fig. 4a). The generated transcripts were affinity purified using immobilised 25 nt antisense oligonucleotides. We tested the assay on two promoters showing high TEC escape (*thsB* and *rps8e*) and two promoters showing low TEC escape (*dhg-1* and **CRISPR C**), the same promoters with ChIP-seq profiles depicted in Fig. 2b–e. All four promoters gave raise to run-off transcripts within 30 s under the experimental conditions (Fig. 4b). Notably, the **CRISPR C** promoter displayed some level of early, broad pausing at 30–40 nt transcript length, about the shortest length that is reliably detectable with the assay. Thus, the in vitro transcription data for the **CRISPR C** promoter are in line with early pausing of TECs in the promoter-proximal region. The absence of a corresponding pausing pattern for the *dhg-1* promoter may suggest that it is difficult to establish the proper context such as chromatinization of the DNA templates that reflects the in vivo situation.

**TEC escape regulation contributes to oxidative stress response.** Our results demonstrate that TEC escape is an important factor for determining promoter strength and RNA levels. In order to investigate whether cells can modulate TEC escape to regulate transcription, we tested how TEC escape changed in response to environmental changes such as oxidative stress. The impact of oxidative stress on *S. solfataricus* using a hydrogen peroxide...
S. solfataricus cell lysate preparation
NTP degradation by alkaline phosphatase & phosphatase inactivation
Addition of linearised DNA template & PIC formation
Single-round transcription
Antisense oligonucleotide affinity purification

Fig. 4 The CRISPR C promoter shows early pausing in vitro. a Schematic overview of the cell lysate-based synchronised in vitro transcription system for S. solfataricus. Cell lysates were treated with shrimp alkaline phosphatase for NTP degradation before heat inactivation of the phosphatase. Linearised plasmid DNA containing a S. solfataricus promoter were added to allow PIC formation on the templates with inherent initiation factor TFB. Simultaneously with the addition of ribonucleotides to allow the PICs to initiate a single round of transcription, we added an excess of a recombinant TFB variant termed TFBc that blocks subsequent rounds of PIC formation. The generated transcripts were purified by affinity purification using immobilised 25 nt antisense oligonucleotides. b Synchronised in vitro transcription assay with two promoters showing high TEC escape (thsB and rps8e) and two promoters showing low TEC escape in vivo (dhg-1 and CRISPR C). Samples were withdrawn 15 s, 30 s and 45 s after simultaneous addition of 50 µM rNTPs including [α-32P]-UTP and TFBc. Purified radiolabelled transcripts were resolved on a denaturing polyacrylamide gel. The position of run-off transcripts and transcripts resulting from pausing in the promoter-proximal region is indicated. A representative experiment of three technical replicates is shown.

The comparison between exponential growth and oxidative stress treatment protocol has been partially characterised. We expected that besides the induction of transcription for stress genes such as dps-1, oxidative stress would cause a broader, global transcriptional response such as a widespread attenuation of the transcriptome. Relevant to transcription initiation, the peroxide treatment results in the depletion of the TFEβ subunit from the cytoplasm (Supplementary Fig. 10). TFEβ depletion coincided with globally decreased promoter occupancies for both TFEα and TFEβ subunits in ChIP-seq implying that TFEα recruitment to the promoter is TFEβ-dependent (Supplementary Fig. 11).

To understand how oxidative stress affects TEC escape, we generated ChIP-seq data for RNAP, initiation- and elongation factors and we applied the same data filtering as for exponential growth phase data to obtain a set of 118 TUs with EI estimates. We then analysed the intersection of 71 TUs from both exponential growth and oxidative stress data sets to compare TEC escape between the two conditions. TUs displaying high TEC escape under exponential growth conditions showed overall reduced RNAP and Spt4/5 escape in response to oxidative stress (Fig. 5a). In addition, the promoter-proximal recruitment of aCPSF1 and the negative correlation to TEC escape are reduced in response to oxidative stress (Supplementary Fig. 12). The lower aCPSF1 signal cannot be explained by protein depletion because immunodetection revealed that the protein levels remained unaffected by oxidative stress (Supplementary Fig. 10).

The comparison between exponential growth and oxidative stress provides us with an opportunity to unravel the changes in PIC occupancy when TEC escape is affected. The changes in TEC escape — in particular Elf1 EI — were positively correlated with changes in the transcriptome between the two conditions (Fig. 5b) to a similar extent as were changes in TFB occupancy. This suggests that TEC escape is an integral part of the transcriptional stress response.

A reduction in TEC escape (RNAP and Spt4/5 EI, but not Elf1 EI) was generally associated with the accumulation of TFB and TFEβ at the promoter (Fig. 5b). The rrr promoter is one of the strongest promoters in Saccharolobus. RNAP and Spt4/5 accumulated at the promoter in response to oxidative stress (RNAP EI from −0.8 to −2.6) suggesting that the control of RNA synthesis occurs in part at the level of TEC escape (Fig. 5c). TFB accumulated at many promoters showing a strongly reduced TEC escape under oxidative stress conditions such as gdha-4 and NuoB (Fig. 5d, e). In contrast, promoters where TEC escape remained relatively unaffected did not show significant changes in TFB accumulation as in the case of SSO8549 (Fig. 5f). These promoters were generally characterised by low TEC escape under both growth conditions.

The link between increased TFB accumulation and reduced TEC escape thus offers a possible mechanistic explanation how TEC escape can affect productive transcription. The accumulation of TFB at the promoter in ChIP-seq experiments could reflect a slower progression from the initial formation of ternary DNA-TBP-TFB complexes towards dissociation of TFB from RNAP during promoter clearance. To test whether RNAP recruitment to DNA-TBP-TFB is impaired, we compared TFB

---

**NATURE COMMUNICATIONS** | **(2021) 12:5524 | https://doi.org/10.1038/s41467-021-25669-2 | www.nature.com/naturecommunications

---
Traces show mean occupancy for two biological replicates with the range depicted as semi-transparent ribbon. Machinery for four different promoters during exponential growth (Expon.) and oxidative stress (Oxid.):

- Reduced TEC escape could block PICs during initial transcription, which does not support the notion that lowered TEC escape is associated with impaired RNAP recruitment to DNA-TBP-TFB complexes. Alternatively, reduced TEC escape could block PICs during initial transcription or promoter clearance.

**Fig. 5 TEC escape changes during oxidative stress response.**

(a) High escape TUs show reduced TEC escape under oxidative stress. Scatter plots comparing escape indices under exponential growth and oxidative stress conditions for 71 TUs accessible for analysis in both conditions. 

(b) Heatmap showing correlated changes in initiation factor occupancy, escape indices and RNA output between exponential growth and oxidative stress. Spearman rank correlations were calculated for protein-encoding TUs accessible for analysis in both conditions (n = 70). Correlations were calculated for the mean escape index and the geometric mean of all other values for two biological replicates. * denotes an adjusted p-value < 0.05 after multiple testing correction (Benjamini-Hochberg) and p < 0.05 for two combinations of individual biological replicates tested. 

c-d ChIP-seq profiles of the basal transcription machinery for four different promoters during exponential growth (Expon.), and oxidative stress (Oxid.): *rn* (c), *gdhA-4* (d), *NuoB* (e), and *SSO8549* (f). Traces show mean occupancy for two biological replicates with the range depicted as semi-transparent ribbon.

and RNAP ChIP-exo data regarding the fold-changes in PIC signal between exponential growth and oxidative stress (Supplementary Fig. 13). The TFB and RNAP ChIP-exo signals showed equal changes between the two growth conditions independent of the co-occurring changes in TEC escape, which does not support the notion that lowered TEC escape is associated with impaired RNAP recruitment to DNA-TBP-TFB complexes. Alternatively, reduced TEC escape could block PICs during initial transcription or promoter clearance. The finding that PICs accumulate to higher levels when TEC escape is reduced indicates that changes in RNAP escape indices (which integrate both PIC and promoter-proximal TEC signal) reflect both cause and effect of TEC escape regulation. Curiously, changes in RNAP EIs do show a weaker correlation to transcriptome changes compared to Spt4/5 and Elf1.

In contrast to TFB, changes in TFEβ occupancy did not show any significant correlation to transcriptome changes. We reasoned that the observed heterogeneity in TFEβ promoter
Stability of the upstream DNA duplex affects TEC escape. Ultimately, the differences in TEC escape are directly or indirectly dictated by the promoter and template sequence context. This includes the promoter-proximal accumulation of PICs and TECs observed under oxidative stress (Fig. 5a, b). Neither the strength of the BRE-TATA promoter element nor its spacing relative to the TSS showed any significant correlation to any feature of TEC escape (Supplementary Fig. 14). In contrast, we found that oxidative stress-specific accumulation of RNAP and Spt4/5 in the promoter region appeared to be influenced by the stability of the DNA duplex across the TSS. To this end, DNA duplex stability of individual promoters was calculated as the inverse of the predicted Gibbs free energy for a 7-bp sliding window within the promoter-proximal region. Under oxidative stress conditions, DNA duplex stability across the TSS showed a robust correlation with RNAP and Spt4/5 escape indices, but much less so with Elf1 (mean of two biological replicates) under oxidative stress conditions (Fig. 6a) (maximum spearman’s $r = 0.51$, $r = 0.57$ and $r = 0.35$, respectively). Notably, under exponential growth conditions, TSS duplex stability showed a weaker, non-significant positive correlation with TEC escape (Fig. 6b) suggesting that the early steps of TEC assembly become sensitive under oxidative stress conditions resulting in the accumulation of TEC in stalled PICs. Consistent with that notion, TSS DNA duplex stability is directly correlated with the ratio of Elf1 to Spt4/5 promoter occupancy under oxidative stress conditions (Supplementary Fig. 15). These results indicate that early transcription elongation and TEC assembly are enhanced by the stable reannealing of upstream DNA but only during the altered conditions of oxidative stress when TFE shows any significant correlation to any feature of TEC escape.

Discussion

Transcription regulation in archaea. Transcription in all domains of life has to be fine-tuned over a wide range of synthesis rates that can respond to environmental cues. Compared to Bacteria as well as Eukaryotes, archaeal promoters show a lower apparent complexity in terms of promoter element composition. What we know thus far is that archaeal transcription is regulated via enhancing or impairing the recruitment of PICs to the promoter similar to TFB when TEC escape is low (Fig. 5b). The model suggests that TUs with low TEC escape do have a lower fraction of PICs containing TFE. The causative relationship of this relative change in PIC composition could work in either direction. Slow TEC escape could impair PICs assembled on the promoter from completing initiation. This retention could lead to the loss of TFE similar to stalled open PICs of yeast RNAPII that have been shown to lose TFIIE in vitro. Alternatively, low TFE occupancy and slow TEC escape could both be a result of slower transcription initiation. In summary, the multiple regression analysis proposes a link between PIC composition and TEC escape.

Multiple regression analysis reveals changes in PIC composition with low TEC escape. Two features are interfering with a quantitative analysis of the relationships between different components of the basal transcription machinery: the non-normal distribution of ChIP-seq occupancy data and the widespread collinearity between occupancy data for different factors. To provide a more comprehensive view of the changes at promoters with high or low TEC escape, we performed a multiple regression analysis for TEC escape (represented by Spt4/5 occupancy data, see methods) under exponential growth and oxidative stress conditions using negative binomial generalised linear models. The models reproduced the observed accumulation of TFB when TEC escape is low (indicated by the negative coefficient for TFB in the model Supplementary Fig. 16). Furthermore, the models reproduced the growth condition-specific effects of aCPSF1 load (as ratio aCPSF1 to Spt4/5) and DNA duplex stability around the TSS. An increased aCPSF1 load was associated with lower TEC escape specifically under exponential growth conditions. TSS DNA duplex stability was associated with increased TEC escape specifically under oxidative stress conditions (Supplementary Fig. 16).

The model for oxidative stress revealed new insight into the PIC composition associated with TEC escape. TFB accumulates at the promoter similar to TFB when TEC escape is low (Fig. 5b). The model suggests that TUs with low TEC escape do have a lower fraction of PICs containing TFE. The causative relationship of this relative change in PIC composition could work in either direction. Slow TEC escape could impair PICs assembled on the promoter from completing initiation. This retention could lead to the loss of TFE similar to stalled open PICs of yeast RNAPII that have been shown to lose TFIIE in vitro. Alternatively, low TFE occupancy and slow TEC escape could both be a result of slower transcription initiation. In summary, the multiple regression analysis proposes a link between PIC composition and TEC escape.

Correlation between DNA duplex stability and TEC escape. The model suggests that TUs with low TEC escape do have a lower fraction of PICs containing TFE. The causative relationship of this relative change in PIC composition could work in either direction. Slow TEC escape could impair PICs assembled on the promoter from completing initiation. This retention could lead to the loss of TFE similar to stalled open PICs of yeast RNAPII that have been shown to lose TFIIE in vitro. Alternatively, low TFE occupancy and slow TEC escape could both be a result of slower transcription initiation. In summary, the multiple regression analysis proposes a link between PIC composition and TEC escape.
Mechanisms underlying promoter-proximal TEC dynamics. Besides premature termination, the promoter-proximal accumulation of TECs can be explained by slow elongation or pausing that might result in backtracking. Bacterial cleavage factor GreB facilitates the release of E. coli RNAP from Sigma70-dependent pause sites. Likewise, cleavage factor TFII S facilitates the release of promoter-proximally paused RNAPII in Drosophila and human cell lines. It is tempting to speculate that TFS, the archaeal TFII S homologue, might play a role in controlling promoter-proximal TEC dynamics in Saccharolobus. Unfortunately, our TFS ChIP-experiments were not successful and TFS association with promoter-proximal TECs remains to be tested.

Functional interactions between initiation and elongation complexes. Promoter-proximal enrichment of TECs can be caused by altered dynamics such as slower elongation rates or pausing, or premature termination. Importantly, altered TEC dynamics will only affect productive transcription if it leads to a wrapping of the downstream DNA around the PIC68 or addi-

Evolution of promoter-proximal regulation in archaea and eukaryotes. The pivotal role of Sp45/5, Elf1 and aCPSF1 in the early TEC dynamics in Saccharolobus shows intriguing parallels to metazoans. Firstly, the early elongation phase of transcription is rate limiting for gene expression. Secondly, Sp45/5 (DSIF) is an integral component of promoter-proximal elongation complexes in humans as well as archaea. Thirdly, CPSF53-related RNases are likely to mediate premature termination of promoter-proximal RNAPs in both eukaryotes and archaea. Our discovery that early elongation complex dynamics modulate transcription in Saccharolobus suggests that promoter-proximal regulation is an ancient feature of the archaeo-eukaryotic transcription machinery. Control of promoter-proximal TEC escape efficiency could provide a simple primordial mechanism for gene regulation, from which a more complex process evolved that involves the stable pausing of elongation complexes, and a tightly controlled pause-release by factors including NELF and P-TEFb.

In Sum, we provide evidence for widespread promoter-proximal transcription regulation in archaea. Our data suggest that the archaeal transcription cycle involves at least two major regulatory checkpoints: (i) recruitment of RNAP and initiation factors to the promoter and (ii) TEC escape into productive elongation likely involving a negative feedback effect on initiating RNAPs upstream as well as premature termination. Together they create a dynamic mosaic of mechanisms that determines the transcription output in archaea. The relative simplicity and biochemical tractability of archaeal transcription complexes provides for the development of in vitro models to elucidate the molecular mechanisms underlying TEC escape.

Methods

ChIP-seq. Rabbit antisera against S. solfataricus TFB, TBP, TFEβ, TFEα and Rpo4/7 have been described previously. Polyclonal rabbit antisera against recombinant Sp5, Elf1, and aCPSF1 were produced at Davids Biotechnology (Regensburg, GER). All antibodies were purified from antisera by Protein A-agarose affinity chromatography.

S. solfataricus P2 cells were grown in Brock medium at 76 °C in a Thermotron air incubator (Infors) to mid-exponential growth phase (O.D.600 0.29). Oxidative stress cells were grown overnight in modified Brock medium without FeCl3 supplemented with 0.2% tryptophan to mid-exponential growth phase (O.D.600 0.11 to 0.24) before the addition of 30 mM H2O2 similar to what has been described before. Cultures were cross-linked for 10 min after H2O2 addition. All cultures were cross-linked by the addition of 0.4% formaldehyde for 1 min before quenching with 100 mM Tris/HCl pH 8.0.

Cells were washed three times in PBS buffer before freezing in liquid nitrogen and storage at −80 °C. To prepare lysates for ChIP experiments, cells were resuspended in lysis buffer (50 mM HEPES/NaOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.1% Na-deoxycholate and 1% Triton X-100) supplemented with complete protease inhibitor cocktail (Roche). DNA was sheared in polystyrene tubes in a Q700 cup sonicator (Qsonica) at 4 °C to an average fragment size of 150 bp as judged by agarose gel electrophoresis. Cells were sonicated using a Branson sonicator (300 W) and resuspended in lysis buffer containing 500 mM NaCl, wash buffer (10 mM Tris/HCl pH 8.0, 100 mM LiCl, 1 mM EDTA, 0.5% Na-deoxycholate and 0.5% Nonidet P-40) and TE buffer.

Intracytoplasmic material was eluted from beads by the addition of 200 µl ChIP-elution buffer (50 mM Tris/HEPES/NaOH pH 8.0, 10 mM EDTA and 1% SDS), de-cross-linked overnight at 65 °C in the presence of 10 µg RNase A and 40 µg Proteinase K. DNA was purified using Qiaquick PCR purification kit (Qiagen). For Sp45/5, Elf1 and CPSF1 ChIP experiments, the lysate volume was increased to 1 ml and 8 µg of antibody was used in combination with Protein G Dynabeads M-280 (Thermo Scientific). Libraries were prepared using the NEBNext Ultra II DNA library prep kit for Illumina (NEB) according to the manufacturer’s protocol. Library quality and quantity was assessed using
Agilent High Sensitivity DNA kit (Agilent Technologies) and Qubit dsDNA HS assay kit (Thermo Scientific).

ChIP-seq read mapping and fragment size distribution adjustment. We generated paired-end ChIP-seq data on Illumina HiSeq platforms with two biological replicates per condition. 75 or 125 nt reads were trimmed from the 3′-ends to 50 nt read length and the read pairs were aligned to the S. solfataricus P2 genome (NC_002754.1) using bowtie2 by scanning promoters with mapped TSSs and Spt4/5 occupancy within the TU body (+251 to 500 bp relative to TSS) for the promoter region (−50 to +100) was calculated.

Positional adjustment of TSS prediction based on ChIP-exo data. For TSS positions that were initially estimated from start codon positions only, we used TFB ChIP-exo signal on the non-template strand to get a more precise estimate of TSS position. The ChIP-exo TFB signals for genes with experimentally mapped TSS were used to build a training set. A peak with median position −14 relative to TSS (Inter-Quartile Range −16 to −12) was identified. TUs were assigned to TSSs using BEDTools window with the peak summit position and then assigned to TSSs using BEDTools window with the peak summit position within 40 bp maximal distance from the TSS.

In order to ensure reliable data normalisation and quantification of the TU body, we only considered TUs with 500 bp minimum length and a minimum coverage of 20 reads in the chromatin input sample within the −250 to +500 interval relative to the TSS. TUs were further filtered against internal TFB peaks within +40 to 500 bp relative to TSS to ensure that RNAP and Spt4/5 occupancy is not influenced by any TU internal promoters.

Different promoters in S. solfataricus are often tightly spaced causing the RNAP and Spt4/5 ChIP-seq signals from these promoter pairs to be convoluted to considerable extent. To address this problem, we filtered our TU set further for those where the input-normalised Spt4/5 occupancy at TSS position was at least 1.5x increased compared to occupancy at position 150 relative to TSS.

The resulting set of TUs was checked for consistent positions of the ChIP-exo data as additional control for TSS position (see below).

To determine the escape index, the log2 ratio of input-normalised RNAP occupancy within the TU body (+251 to 500 bp relative to TSS) to the promoter region (−50 to +100) was calculated.

RNA 3′-end selection. To test aCPSF1 association with putative transcription termination sites, we used an initial dataset of 1727 predicted RNA 3′-ends based on the Rockhopper 250 bp genome (output of the RNA-seq data) to identify overlapping peaks. Predicted RNA 3′-ends were filtered by the following criteria: (i) a TU length >500 nt, (ii) no TFB peaks within the surrounding 600 bp, (iii) continuous input coverage of >20 reads in the surrounding 500 bp to ensure reliable input normalisation, and (iv) a two-fold increase in average Spt4/5 occupancy in the 250 bp downstream of the predicted RNA 3′-end compared to the 250 bp upstream. The final data set comprised 41 predicted RNA 3′-ends.

TATA-box assignment. The S. solfataricus BRE-TATA box motif was determined by scanning promoters with mapped TSS within a 24-bp window (positions −42 bp to −19 relative to TSS) using MEME in ‘ops’ mode with 8−15 motif width.

ChIP-exo. For ChIP-exo analysis, we used the ChIP-exo Kit (Active Motif) according to manufacturer’s specifications with the following modifications. This kit is based on the modified ChIP-exo protocol adapted for Illumina sequencing. Cell growth, crosslinking and DNA shearing were carried out as described for ChIP-seq samples, but DNA was sheared to a range of >200 bp to be suitable for ChIP-exo. Immunoprecipitation was carried out as for ChIP-seq samples by incubating 1 ml lysate with 8 µg antibody overnight. The lysates were transferred to a new tube with 50 µl Protein-G Dynabeads (Active Motif) and further incubated for 1 h before following the manufacturer’s recommendations for washing of the beads and library preparation. Library quality was assessed using Agilent High Sensitivity DNA kit (Agilent Technologies) and Qubit dsDNA HS assay kit (Thermo Scientific). Libraries were sequenced on the Illumina HiSeq platform with 50 cycles read-length. Reads were aligned to the S. solfataricus genome using bowtie2 v2.2.7. Heatmaps were generated using deepTools computeMatrix and plotHeatmap functions.

Peak calling. Peaks were identified with MACS2 2.0.1 and with the call-summit sub-function in order to identify overlapping peaks. MACS2 output provides summit coordinates and quality scores for each peak, but the coordinates for each enriched region are not split between the overlapping peaks. For this reason, we used the peak summit positions to merge peaks from replicates with 40 bp max distance which should correspond to more than 50% overlap between the peaks using BEDTools window function. For the reproducibility analysis of the peaks replicates based on p-values, we set a global IDR threshold of 0.05 using the Cran IDR package in R. A number of reproducible TEFs and TFB peaks within the RNA operon were removed as this region exhibited an overall strongly increased background. Finally, for reproducible peaks the average position and fold-enrichment between the replicates was calculated. All spearman correlation coefficients were calculated using the spearman.cor function from the RvaidaMemento package in R with confidence intervals calculated using bootstrapping (n = 1000). Spearman correlation estimates were considered to be significantly different at significance level α when the confidence intervals calculated from the bootstrapped data set for the same significance level were non-overlapping for both biological replicates.

Occupancy data plotting. Bam files were normalised against input using deepTools bamCompare using the SES method for scaling (10,000 bins, 200 bp bin size) and then assigned to TSSs using BEDTools window with the peak summit position within 40 bp maximal distance from the TSS.

Positional adjustment of TSS prediction based on ChIP-exo data. For TSS positions that were initially estimated from start codon positions only, we used TFB ChIP-exo signal on the non-template strand to get a more precise estimate of TSS position. The ChIP-exo TFB signals for genes with experimentally mapped TSS were used to build a training set. A peak with median position −14 relative to TSS (Inter-Quartile Range −16 to −12) was identified. TUs were assigned to TSSs using BEDTools window with the peak summit position and then assigned to TSSs using BEDTools window with the peak summit position within 40 bp maximal distance from the TSS.

In order to ensure reliable data normalisation and quantification of the TU body, we only considered TUs with 500 bp minimum length and a minimum coverage of 20 reads in the chromatin input sample within the −250 to +500 interval relative to the TSS. TUs were further filtered against internal TFB peaks within +40 to 500 bp relative to TSS to ensure that RNAP and Spt4/5 occupancy is not influenced by any TU internal promoters.

Different promoters in S. solfataricus are often tightly spaced causing the RNAP and Spt4/5 ChIP-seq signals from these promoter pairs to be convoluted to considerable extent. To address this problem, we filtered our TU set further for those where the input-normalised Spt4/5 occupancy at TSS position was at least 1.5x increased compared to occupancy at position 150 relative to TSS.

The resulting set of TUs was checked for consistent positions of the ChIP-exo data as additional control for TSS position (see below).

To determine the escape index, the log2 ratio of input-normalised RNAP occupancy within the TU body (+251 to 500 bp relative to TSS) to the promoter region (−50 to +100) was calculated.
paired-end reads were generated on a HiSeq 4000 system (Illumina). Coverage kit (Thermo-Fisher). RNA was quantified. The median background signal was subtracted and the resulting T-specific neighbour non-T positions for each T (the two closest non-T residues on either side) of the TU was included as a second offset term. Thereby, the models effectively identified variables predictive for TCE escape efficiency. As explanatory variables, we tested TFB and TFE promoter occupancy (log-transformed), acPSF1, to Sp4/5p, and TSS DNA duplex stability. The full model was structured as follows:

$$\log(\text{Sp4/5p coverage}) = \log(\text{input-normalised coverage}) + \log(\text{Sp4/5p})$$
$$+ \beta_1 + \beta_2 \log(\text{TFB})$$
$$+ \beta_3 + \log(\text{TFE}) + \beta_4 \log(\text{acPSF1}) + \beta_5 \log(\text{Sp4/5p})$$
$$+ \beta_6 \times \text{TSS DNA duplex stability}$$

TU s passing a Cook’s distance of 0.5 for a model of all explanatory variables were removed from the dataset. To identify the optimal model, we used the step() function in the MASS package for automatic model building using default settings for AIC minimisation. The null model was used as lower boundary and the full model including all variables as upper boundary. Searches were initiated with the null model (including only the raw input coverage and Sp4/5p offsets) as well as all explanatory variables added in a separate model. The significance of each added variable in the model was confirmed using the Likelihood-ratio chi-squared test implemented in anova.negbin() (car package) by comparing the optimal models with the input normalisation. We calculated the total occupancy signal within the TU region and 50 nt upstream and downstream regions. The resulting cDNA was finally amplified (12 cycles) with TruSeq Dual Index sequencing primers (Illunia) and Herculase II Fusion DNA Polymerase (Agilent).

The libraries were sequenced on an Illumina NextSeq 500 system with 75 bp read length. In order to include poly(A)-tails and adaptors, we trimmed the reads using Cutadapt93 in two rounds with the following settings to prevent trimming of non-coding A-rich RNA due to the low GC-content of the S. solfataricus genome: (i) -a “[A15]” -e 0 -m 15 to remove all poly(A) stretches of at least 15 nt length plus downstream regions and (ii) -a “[A15]”e -x 0 -O 5 terminal shorter poly(A) stretches of minimum 5 nt length. Trimmmed and untrimmed reads were split into separate fastq files using awk. Both fastq files were aligned to the S. solfataricus genome using bowtie v1.2.276 (parameters -v 1 -m 1 -best) with the S. solfataricus genome (version bowtie v1.2.276) as reference. Primers were trimmed 75 nt reads shorted to 71 nt (−3 4). The bam file output was merged, sorted and indexed using SAMtools77. Bam files were imported into the R environment using the rsamtools and GenomicRanges packages and filtered to ensure a unique sequence of the initial 20 bp within the S. solfataricus genome required to map the reads in the Biostrings packages. The rNA reads were defined as RNAs with a 5′-end within 20 nt of a mapped or predicted TSS. The two biological replicates showed good reproducibility of TSS-RNA occupancy with a Spearman correlation of 0.98 for 438 mappable promoters.

To calculate the fraction of TSS-RNA with a length shorter than 50 nt each TU, a minimum read count of 10 TSS-RNAs per TU per replicate was used and values were averaged between the two biological replicates.

Immunodetection. Cell lysates were resolved on 12% Tris-tricine SDS gels and blotted onto nitrocellulose membranes. All immuno-detections were carried out using polyclonal antisera (see above) in combination with donkey anti-rabbit IgG (DyLight680 (Bethyl Laboratories). Dps-l antisera was a kind gift from Mark Young (Montana State University, USA). As loading control, we used sheep Alp anti- serum (University of Cambridge, UK) in combination with donkey anti-sheep IgG Alexa488 (Thermo Fisher). Blots were scanned on a Typhoon FLA 9500 scanner (GE Lifesciences).

In vitro transcription. In order to trace promoter-proximal transcription elongation dynamics in vitro, we developed a cell lysate-based synchronised transcription assay. The transcription was achieved by supplementing an inactive variant of initiation factor TFB comprising only the C-terminal cyclin folds (TFBc) simultaneously with rNTPs. TFBc forms ternary complexes with TBP and DNA containing TATA/BRE promoter motifs, but it fails to recruit RNA polymerase and does not facilitate transcription initiation60,64.

To generate cell lysates for the in vitro transcription, 4 litre S. solfataricus P2 cultures in exponential growth phase were transferred to an ice-water bath for cooling. Cells were harvested by centrifugation for 20 min at 4000 × g at 4 °C. Pellets were washed twice in ice-cold 20 mM sucrose solution before flash freezing in liquid nitrogen and storage at −80 °C. Cells were resuspended in 6 ml 10 M MOPS pH 6.5, 10 mM MgCl2, 1 mM DTT, 1 mM EDTA, 0.1% Triton X-100 and incubated for 1 h on ice for cell lysis. Insoluble material was removed by centrifugation for 20 min at 20000 × g at 4 °C and the supernatant was aliquoted and stored at −80 °C.

Templates for in vitro transcription were generated by PCR-amplifying promoter regions −50 to −100 relative to the TSS for thiB, rpsl, dhp-1 and CRISPR C from S. solfataricus P2 genomic DNA and ligating the products into the vector pGM-T (Promega) (see Supplementary Table 1 for details).

Before using the cell lysates in vitro transcription reactions, nucleotides were degraded by treating the lysate with 100 unit/ml recombinant shrimp alkaline phosphatase (New England Biolabs) for 20 min at 37 °C. Cells were harvested and DNA was removed using the phenol:chloroform extraction method described by Gilmour, Pugh and co-workers65. RNA was isolated according to the ChIP-exo protocol. Reads were mapped onto the genome using bowtie v1.2.276 (parameters -v 1 -m 1 -best) with the S. solfataricus genome (version bowtie v1.2.276) as reference. Primers were trimmed 75 nt reads shorted to 71 nt (−3 4). The bam file output was merged, sorted and indexed using SAMtools77. Bam files were imported into the R environment using the rsamtools and GenomicRanges packages and filtered to ensure a unique sequence of the initial 20 bp within the S. solfataricus genome required to map the reads in the Biostrings packages. The rNA reads were defined as RNAs with a 5′-end within 20 nt of a mapped or predicted TSS. The two biological replicates showed good reproducibility of TSS-RNA occupancy with a Spearman correlation of 0.98 for 438 mappable promoters.

To calculate the fraction of TSS-RNAs with a length shorter than 50 nt each TU, a minimum read count of 10 TSS-RNAs per TU per replicate was used and values were averaged between the two biological replicates.
[α-32P]-UTP (Perkin Elmer) were provided to initiate transcription together with 5 µM Tfac blocking further pre-initiation complex assembly. At given time points, 30 µl sample were withdrawn and mixed rapidly with 200 µl stop mix (20 mM EDTA pH 8.0, 200 mM NaCl, 1% SDS, 250 ng/µl torula yeast RNA and 0.1 mg/ml Proteinase K). After 5 min incubation, samples were purified by a single acidic phenol/chloroform extraction step. Transcripts were subsequently enriched by affinity purification following a protocol based on58. 200 µl supernatant from the phenol extraction step was mixed with 100 µl salt adjust mix (30 mM Tris/HCl pH 8.0, 50 mM NaCl) and 1.6 pmol α'-biotinylated antisense oligonucleotide matching the first 25 nt of the transcripts (based on the predicted TSS) (Supplementary Table 1). α'-Biotinylation of the oligonucleotides was carried out with biotin-14-dATP (Jena Bioscience) and Terminal transferase (New England Biolabs). Transcripts and biotinylated oligonucleotides were allowed to hybridise overnight at room temperature. Samples were mixed with 120 µl of 1.25 mg/ml Dynabeads MyOne Streptavidin C1 (ThermoFisher) pre-equilibrated in 2X B/W buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA, 2 M NaCl) for 15 min. Beads were washed twice with 300 µl washing buffer (10 mM Tris/HCl pH 8.0, 5 mM EDTA, 10 mM NaCl, 100 ng/µl torula yeast RNA) and finally eluted in 15 µl formamide sample buffer (95% deionised formamide, 18 mM EDTA, 0.025% SDS) for 5 min at 95°C. 10 µl of the samples were resolved on 12% polyacrylamide, 7 M Urea, 1x TBE sequencing gel. Transcripts were detected by phosphor imagery and quantification of bands was performed using the ImageQuant TL software (GE Life Sciences).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support this study are available from the corresponding author upon reasonable request. All sequencing data files (ChIP seq, ChIP-exo, permanganate ChIP-seq, RNA-seq and Capable-seq) and the processed data were deposited at NCBI GEO under accession code GSE141290.

Code availability
The analysis code is available on Zenodo.96 [https://doi.org/10.5281/zenodo.5196117].

Received: 25 August 2020; Accepted: 25 August 2021;
Published online: 17 September 2021

References
1. Ehrensberger, A. H., Kelly, G. P. & Svejstrup, J. Q. Mechanistic interpretation of promoter-proximal peaks and RNA PolI density maps. Cell 154, 713–715 (2013).
2. Browning, D. F. & Busby, S. J. Local and global regulation of transcription initiation in bacteria. Nat. Rev. Microbiol. 14, 638–650 (2016).
3. Hahn, S. & Young, E. T. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics 189, 705–736 (2011).
4. Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife 6, e29736 (2017).
5. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new promoter DNA bending pathways. Nucleic Acids Res. 42, 6219–6231 (2014).
6. Blombach, F. et al. Archaeal TTFalp/a/beta is a hybrid of TFIIIE and the RNA polymerase III subcomplex hRPc23/39. Elife 4, e03875 (2015).
7. Naji, S., Grünberg, 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).
8. Schulz, S. et al. TFE and Spt4/S5 open and close the RNA polymerase clamp during the transcription cycle. Proc. Natl Acad. Sci. USA 113, E1816–E1826 (2016).
9. Werner, F. & Weinzierl, R. O. Direct modulation of RNA polymerase core functions by basal transcription factors. Mol. Cell Biol. 25, 8344–8355 (2005).
10. Hirttreiter, A. et al. Spt4/S5 stimulates transcription elongation through the RNA polymerase clamp-coiled-coil motif. Nucleic Acids Res. 38, 4040–4051 (2010).
11. Smollett, K., Blombach, F., Reichelt, R., Thomm, M. & Werner, F. A global analysis of transcription reveals two modes of Spt4/S5 recruitment to archaeal RNA polymerase. Nat. Microbiol. 2, 17021 (2017).
12. Fouqueau, T. et al. The cutting edge of archaeal transcription. Emerg. Top. Life Sci. 2, 517 (2018).
13. Daniels, J. P., Kelly, S., Wickstead, B. & Gull, K. Identification of a crenarchaeal orthologue of EIIA: implications for chromatin and transcription in Archaea. Biol. Direct 4, 24 (2009).
14. Fouqueau, T. et al. The transcript cleavage factor paralogue TFS4 is a potent RNA polymerase inhibitor. Nat. Commun. 8, 1914 (2017).
15. Grohmann, D. et al. The initiation factor TFE and the elongation factor Spt4/S5 compete for the KANP clamp during transcription initiation and elongation. Mol. Cell 43, 263–274 (2011).
16. Sanders, T. J. et al. FtaA is a CPSF73 homologue that terminates transcription in Archaea. Nat. Microbiol. 5, 545–553 (2020).
17. Blombach, F., Matelska D., Fouqueau T., Cackett G. & Werner F. Key concepts in archaeal transcription. Biochim. Biophys. Acta 189, 18871–18871 (2012).
18. Blombach, F. & Grohmann, D. Same same but different: the evolution of TBP in archaea and their eukaryotic offspring. Transcripton 8, 162–168 (2017).
19. Lemmens L., Maklad H. R., Bervoets I. & Peeters E. Transcription regulators in archaea: homologies and differences with bacterial regulators. J. Mol. Biol. 431, 431–4146 (2019).
20. Peeters, E., Petriero, N. & Sezonov, G. Cis-regulatory logic in archaeal transcription. Biochim. Biophys. Acta 41, 326–331 (2013).
21. Martinez-Pastor, M., Tonner, P. D., Darnell, C. L. & Schmid, A. K. Transcriptional regulation in archaea: from individual genes to global regulatory networks. Annu. Rev. Genet. 51, 143–170 (2017).
22. Ochs, S. M. et al. Activation of archaeal transcription mediated by recruitment of transcription factor B. J. Biol. Chem. 287, 18863–18871 (2012).
23. Ochsmanmouch, M., Werner, F., Weinzierl, R. O. & Geiduscheck, E. P. A fully recombinant system for activator-dependent archaeal transcription. J. Biol. Chem. 279, 51719–51721 (2004).
24. Bell, S. D. & Jackson, S. P. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends Microbiol. 6, 222–228 (1998).
25. Hirata, A., Klein, B. J. & Murakami, K. S. The X-ray crystal structure of RNA polymerase from Archaea. Nature 451, 851–854 (2008).
Qureshi, S. A., Bell, S. D. & Jackson, S. P. Factor requirements for transcription in the Archaeon Sulfolobus shibatae. EMBO J. 16, 2927–2936 (1997).

Sheppard, C. et al. Repression of RNA polymerase by the archaeo-viral regulator ORF145/RIP. Nat. Commun. 7, 15395 (2016).

Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147, 1408–1419 (2011).

Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483, 295–301 (2012).

Nagy, J. et al. Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. Nat. Commun. 6, 6161 (2015).

Spatny, P. & Thomm, M. Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation. J. Biol. Chem. 278, 30497–30505 (2003).

Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science 314, 1339–1343 (2006).

Dex, S. et al. Displacement of the transcription factor B reader domain during transcription initiation. Nucleic Acids Res. 46, 10066–10081 (2018).

Lai, W. K. & Pugh, B. F. Genome-wide uniformity of human ‘open’ pre-initiation complexes. Genome Res. 27, 15–26 (2017).

Li, J. et al. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. Mol. Cell 50, 711–722 (2013).

Bell, S. D., Jaxel, C., Nadal, M., Kosa, P. F. & Jackson, S. P. Temperature, template topology, and factor requirements of archaeal transcription. Proc. Natl Acad. Sci. USA 102, 5212–5217 (2005).

Ettwiller, L., Buswell, J., Yi, G., Ehrlich, J. & Schildkraut, I. A novel enrichment strategy identifies flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191 (2014).

Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).

Li, Q. H., Brown, J. B., Huang, H. Y. & Buckel, P. J. Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5, 1752–1777 (2011).

Dex, S. et al. Displacement of the transcription factor B reader domain during transcription initiation. Nucleic Acids Res. 35, 553–564 (2019).

Ely, J. M. et al. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. Mol. Cell 50, 711–722 (2013).

Bell, S. D., Jaxel, C., Nadal, M., Kosa, P. F. & Jackson, S. P. Temperature, template topology, and factor requirements of archaeal transcription. Proc. Natl Acad. Sci. USA 102, 5212–5217 (2005).

Ettwiller, L., Buswell, J., Yi, G., Ehrlich, J. & Schildkraut, I. A novel enrichment strategy identifies flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191 (2014).

Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).

Li, Q. H., Brown, J. B., Huang, H. Y. & Buckel, P. J. Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5, 1752–1777 (2011).

Dex, S. et al. Displacement of the transcription factor B reader domain during transcription initiation. Nucleic Acids Res. 35, 553–564 (2019).

Phung, D. K. et al. Archaeal beta-CASP ribonucleases of the aCSF1 family are orthologs of the eukaryal CPSF73 factor. Nucleic Acids Res. 41, 1091–1103 (2013).

Yu, L. et al. The conserved ribonuclease aCSF11 triggers genome-wide transcription resolution of Archaea via a 3'-end cleavage mode. Nucleic Acids Res. 48, 9589–9605 (2020).

Bell, S. D. & Jackson, S. P. The role of transcription factor B in transcription initiation and promoter clearance in the archaeon Sulfolobus acidocaldarius. J. Biol. Chem. 275, 12934–12940 (2000).

Maita, W. S. et al. Something old, something new, something borrowed; how the thermoacidophilic archaeon Sulfolobus solfataricus responds to oxidative stress. PLoS ONE 4, e6964 (2009).

Cabrera, M. J. & Pugues, D. S. Inactivated RNA polymerase II open complexes can be reactivated with TFIIIE. J. Biol. Chem. 287, 961–967 (2012).

Heyduk, E. & Heyduk, T. DNA template sequence control of bacterial RNA polymerase escape from the promoter. Nucleic Acids Res. 46, 4469–4486 (2018).

Weinstein-Fischer, D., Elgraby-Weiss, M., Fau - Altuvia, S. & Altuvia, S. Escherichia coli response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. Mol. Microbiol. 35, 1413–1420 (2002).

Adelman, K. & Lif, J. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoa. Nat. Rev. Genet. 13, 720–731 (2012).

Adelman, K. et al. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIH. Mol. Cell 17, 103–112 (2005).

Sherrard, R. M., Fong, N., D’Alessandro, A. & Bentley, D. L. Widespread backtracking by RNA Pol II is a major effector of gene activation, 5’ pause release, termination, and transcription elongation rate Mol. Cell 73, 107–118.e4 (2019).

Rivetti, C., Guthold, M. & Bustamante, C. Wrapping of DNA around the E.coli RNA polymerase open promoter complex. EMBO J. 18, 4464–4475 (1999).

Peeters, E., Driessen, R. P., Werner, F. & Dame, R. T. The interplay between nucleoid organization and transcription in archaeal genomes. Nat. Rev. Microbiol. 13, 333–341 (2015).

Mayer, A. et al. Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol. 17, 1272–1278 (2010).

Prather, D., Krogan, N. J., Emili, A., Greenblatt, J. F. & Winston, F. Identification and characterization of ELL1, a conserved transcription elongation factor in Saccharomyces cerevisiae. Mol. Cell Biol. 25, 10122–10135 (2005).
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-25669-2.

Correspondence and requests for materials should be addressed to Fabian Blombach or Finn Werner

Peer review information Nature Communications thanks Sergei Borukhov, Katsuhiko Murakami and Evgeny Nudler for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© Crown 2021