Promoter-proximal elongation regulates transcription in archaea

Fabian Blombach[1*], Thomas Fouqueau[1], Dorota Matelska[1], Katherine Smollett[1], Finn Werner[1*]

Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, United Kingdom

*Correspondence: f.blombach@ucl.ac.uk, f.werner@ucl.ac.uk

Abstract

Recruitment of RNA polymerase and initiation factors to the promoter is the only known mechanisms for transcription activation and repression in archaea. Whether any of the subsequent steps towards productive transcription elongation is involved in regulation is not known. We characterised how the basal transcription machinery is distributed along genes in the archaeon *Sulfolobus 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.

Introduction

Transcription output is determined by the frequency of transcription initiation and premature termination [1]. The recruitment stage of transcription initiation is the main target for regulation in yeast and bacteria [2,3], however, the initiation rate can also be affected indirectly by downstream events. In metazoan, promoter-proximal pausing of RNAP with slow turnover times blocks pre-initiation complex (PIC) formation for the following RNAP and is thereby widely rate-limiting for transcription initiation [4,5]. Promoter-proximal paused RNAPII can also be subject to premature termination providing an additional way of transcription regulation [6-10]. Promoter-proximal RNAP dynamics also limit gene expression in *E. coli* [11]. Well-established processes of post-recruitment regulation include Sigma70-dependent pausing and transcription attenuation mediated by premature termination [12](REFs). Another possible underlying molecular mechanism might be pausing during initial transcription [13,14], though its contribution to genome-wide gene regulation remains to be investigated [15].

Archaea form the 'third domain of life' next to bacteria and eukaryotes, with the latter likely originating from an archaeal ancestor [16]. The basal archaeal transcription machinery represents an evolutionarily ancient core of the RNAPII system encompassing RNAP subunits, basal transcription initiation and -elongation factors, and core promoter elements [17-19]. The mechanisms of initiation have been characterised in great detail *in vitro* (Figure 1A). 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 preinitiation complex (PIC) [20,21]. 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 [22-25]. TFE stimulates transcription initiation but is not strictly required *in vitro*. Like the archaeal PIC, the transcription elongation complex (TEC) corresponds to an evolutionarily ancient RNAPII TEC encompassing homologues to a subset of RNAPII elongation factors: Spt4/5 (DSIF in human) [26,27] and potentially the archaeal homologue of elongation factor Elf1 (Elof1 in humans) [28,29]. In addition, the transcript cleavage factor TFS (homologous to TFIIS) transiently associates with the TEC and reactivates arrested TECs [30]. Spt4/5 and TFE bind to RNAP in a mutually exclusive manner and the transition from transcription initiation to elongation requires factor switching between TFE and Spt4/5 [31]. Transcription termination

in archaea occurs via intrinsic or factor-dependent mechanisms. The latter involves termination factor aCPSF1 (or FttA) [32], a ribonuclease that is evolutionary 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 [33,34]. Likewise, our understanding of transcription regulation is limited to factors modulating the recruitment of PICs [35,36] where repression generally involves steric hindrance of RNAP or basal initiation factor binding and activation is achieved by enhancing their binding [37-40]. 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.

We analysed the genome-wide distribution of RNAP and transcription initiation and elongation factors in the crenarchaeon *Sulfolobus 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 Elf1) 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 TEC escape as an important feature 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 *Sulfolobus solfataricus*. We developed and adapted chromatin immuno-precipitation sequencing (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 [41].

Aggregate profiles of ChIP-exo data for a set of 298 transcriptionally active TUs with mapped TSS (see methods) showed a distinct footprint for RNAP and initiation factors TFB and TFEβ around the TSS (Figure 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 [5,42]. 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 *in vitro* exonuclease foot-printing experiments of archaeal PICs [43,44] (Figure 1B).

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 TSS (Figure 1c). Both TFB- and TFE occupancy correlated strongly with RNAP. (Spearman's r = 0.92 in both cases, Figure 1de). Since TFB binding is critically dependent on TBP binding, and TFE binding depends on RNAP, our results show 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 Figure 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 [45]. DNA scrunching results in downstream extension of the DNA bubble thereby making thymine basis 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 +10 [46]. 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 [47,48]. Aggregate plots showed that DNA melting occurred in the -12 to +3 region relative to the TSS, peaking at position -10, which is consistent with the *in vitro* permanganate foot printing of recombinant PICs [22,44,49]. 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.

RNAP escape limits productive transcription

How does archaeal RNAP progress from transcription initiation into productive elongation? To address these poorly understood processes, 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 TU, 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. Choosing transcription units with lengths of >500 bp provided us with a window within the TU body where the RNAP occupancy reflects the productive elongation phase well-separated from the PIC signal. 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 Figure 2). Crucially, some TUs showed a strong decrease in RNAP occupancy from the promoter towards the TU body revealing heterogeneity in how RNAPs progress into productive elongation (Figure 2a). For TUs encoding housekeeping genes such as *thsB*, encoding a subunit of the thermosome chaperone (Figure 2b), or *rps8E* (ribosomal protein S8e, Figure 2c) the decrease in RNAP occupancy was rather small. In contrast, for expample *dhg-1* (one of two glucose-1-dehydrogenase isoenzymes), shows a drastic decrease in RNAP occupancy (Figure 2d). Notably, all CRISPR loci showed strongly reduced RNAP escape into productive transcription suggesting that crRNA synthesis is regulated at this level (Figure 2e). If the transition into the productive transcription elongation phase is rate-limiting for RNA synthesis genome-wide, global mRNA levels should correlate better with RNAP occupancy within the TU body than RNAP occupancy at the promoter. Consistent with the rate-limiting role of RNAP escape, mRNA expression levels of the first cistron in the TU did correlate significantly better with the average RNAP occupancy within the TU body ($RNAP_{Bd}$, calculated over positions +251 to +500) than with RNAP promoter occupancy ($RNAP_{Pr}$. Spearman's r of 0.75 versus 0.44, Figure 2fg), or indeed TFB occupancy at the promoter (Spearman correlation statistically not significant, Figure 2h).

In summary, the ChIP-seq results reveal that the escape of RNAP into productive elongation varies greatly across different TUs.

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 $RNAP_{Bd}$ over $RNAP_{Pr}$. 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 (Figure 3ab). Both elongation factors accumulate in the promoter-proximal region of TUs with low EI alongside RNAP (Figure 3b). In support of this, escape index calculations for both elongation factors revealed strong correlations with

RNAP EI (Supplementary Figure 3). 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 (Figure 3ab). Henceforth we refer to the two TEC complexes as TEC_{Spt45} and $TEC_{Spt45\text{-}Elf1}$. 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 independently, we characterised the nascent RNAs synthesised at the 5′ end of the TUs referred to as TSS-RNAs. Short RNAs (20 to 200nt length) were isolated, enriched for triphosphorylated 5′-ends using the Cappable-seq method [50] and deep-sequenced. In quantitative terms, the occupancy of promoter-proximal elongation complexes (using $Spt4/5_{Pr}$ as proxy) correlated well with these TSS-RNA read counts (Spearman's r = 0.61, Figure 3c) and significantly better than mRNA counts from total RNA-seq (Spearman's r = 0.48, Supplementary Figure 4). This shows that they were enriched in 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 (< 50nt) consistent with promoter-proximal accumulation of TECs (Figure 3d).

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

aCPSF1 recruitment to the TEC correlates with reduced TEC escape

The balance between premature termination and antitermination in the 5′ 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 [32,51,52]. aCPSF1 is capable of inducing transcription termination on TECs stalled in the promoter-proximal region (+54) [32]. Intriguingly, aCPSF1 accumulated

in the promoter-proximal-region of most TUs (188 out of 212 TUs with peaks passing detection threshold) including *thsB*, *rps8E* and *dhg-1* (Figure 2b-d). In contrast, aCPSF1 does not form clearly defined peaks at 3'-ends of most TUs we predicted from RNA-seq data. Instead we observed a decrease in occupancy of aCPSF1 together with RNAP, Spt4/5 and Elf1 downstream of the predicted mRNA 3'-ends, and only in some cases well defined CPSF1 peaks (Supplementary Figure 5).

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 (Figure 4a). 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 [32]. In line with the *in vitro* observations, the position of promoter-proximal peak summits demonstrate that Spt4/5 is recruited to the TEC prior to aCPSF1 in vivo (Figure 4b).

Provided that aCPSF1 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 (Figure 4c). This anticorrelation holds true whether the aCPSF1 load is calculated as ratio of aCPSF1 to Elf1 (Figure 4c) or aCPSF1 to Spt4/5 promoter occupancy (Supplementary Figure 6). Importantly, a higher aCPSF1 load was correlated with lower mRNA levels (Figure 4d, Supplementary Figure 6).

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 link between the termination factor aCPSF1 and RNA output and are consistent with a premature termination mechanism.

Oxidative stress decreases TEC escape genome-wide

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 treatment protocol has been partially characterised [53]. We expected that besides the induction of transcription for stress genes such as *dps-I* [53], 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 Figure 7) and a global attenuation of TFEαβ promoter occupancy reflected in the ChIP-seq analysis (Supplementary Figure 8).

To understand how oxidative stress affects TEC escape, we compared a set of transcriptionally active TUs (n=71) with little signal overlap from neighbouring TUs during both exponential growth and oxidative stress (see methods for data filtering). RNAP, Spt4/5 and Elf1 all showed globally reduced escape in response to oxidative stress (Figure 5a). In addition, the promoter-proximal recruitment of aCPSF1 and the negative correlation to TEC escape is reduced in response to oxidative stress (Supplementary Figure 9). The lower aCPSF1 signal cannot be explained by protein depletions because immunodetection revealed that the protein levels remained unaffected by oxidative stress (Supplementary Figure 7).

The comparison between exponential growth and oxidative stress provides us with an opportunity to unravel the changes occurring at promoters 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 (Figure 5b) to a similar extent as were changes in TFB occupancy. This suggests that TEC escape is an integral 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 (Figure 5b). The *rrn* promoter is one of the strongest promoters in *Sulfolobus*. 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 rRNA synthesis occurs in part at the level of TEC escape (Figure 5c). TFB accumulated at many promoters showing a strongly reduced TEC escape under oxidative stress conditions such as *gdha-4* and *NuoB* (Figure 5de). In contrast, promoters where TEC escape remained relatively unaffected did not show TFB accumulation as in the case of

SSO8549 (Figure 5f). These TUs were generally characterised by low TEC escape under both growth conditions.

Theoretically, the accumulation of TFB at the promoter could be due to ternary DNA-TBP-TFB complexes failing to recruit RNAP. To test this hypothesis, we compared TFB and RNAP ChIP-exo data regarding the fold-changes in PIC signal between exponential growth and oxidative stress (Supplementary Figure 10). 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. The accumulation of initiation factors when TEC escape is reduced thus offers a possible mechanistic explanation how TEC escape can affect productive transcription. Slow $TEC_{Spt4/5}$ and $TEC_{Spt4/5\text{-}Elf1}$ could block PICs from completing initiation, either during initial transcription or promoter clearance. 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 occupancy relative to TFB might be rather the result of TFEβ loss in stalled PICs resulting from low TEC escape, rather than promoters showing different affinities for TFEβ under oxidative stress. This does not preclude a broader, genome-wide effect of TFE depletion on transcription initiation during oxidative stress.

In order to test a second, independent, stress condition, we compared the EIs of RNAP and Spt4/5 during exponential growth and the stationary phase. Similar to oxidative stress, entering the stationary growth phase resulted in lowered EIs of RNAP and Spt4/5 that positively correlated with transcriptome changes (Supplementary Figure 11).

In summary, our results suggest that TEC escape can be modulated during different growth conditions, and that environmental insults lower the TEC escape and concomitant RNA output in archaea.

Stability of the upstream DNA duplex affects TEC escape

Ultimately, the differences in TEC escape are directly or indirectly dictated by the promoter sequence context. This includes the promoter-proximal accumulation of PICs and TECs observed under oxidative stress (Figure 5ab). 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 Figure 12). 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 (Figure 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 (Figure 6b) suggesting that the early steps of TEC assembly become sensitive under oxidative stress conditions resulting in the accumulation of $TEC_{Spt4/5}$ and 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 Figure 13). These results indicate that early transcription elongation and TEC assembly is enhanced by the stable reannealing of upstream DNA but only during the altered conditions of oxidative stress where TFEβ is for example limiting.

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 Figure 14). Furthermore, the models reproduced the growth condition-specific effects of aCPSF1 load (as ratio $aCPSF1_{Pr}$ to $Spt4/5_{Pr}$) and DNA duplex stability around the TSS. An increased aCPSF1 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 Figure 14).

The model for oxidative stress revealed new insight into the PIC composition associated with TEC escape. TFEβ accumulates at the promoter similar to TFB when TEC escape is low (Figure 5b). The model suggests that TUs with low TEC escape do have a lower fraction of PICs containing TFEβ. 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* [54]. 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.

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 (RNAP II), archaeal promoters show a lower apparent complexity in terms of promoter element composition [33,34]. What we know thus far is that archaeal transcription is regulated via enhancing or impairing the recruitment of PICs to the promoter [37-40]. To date, there are no known examples of transcription attenuation or antitermination mechanisms in archaea. How does the archaeal transcription machinery generate the diversity in promoter strength and regulation? Our results point to promoter-proximal elongation as key target in determining promoter strength (Figure 7). DNA duplex stability around the TSS is likely to constitute a conditional promoter element that plays an important role during oxidative stress where it affects the first stages of TEC escape. The underlying mechanism remains to be explored yet but is likely based on the reannealing of upstream DNA within the Inr region during promoter clearance towards early TEC progression. This is consistent with data from *E. coli*, where duplex stability of the initial transcribed region affects promoter escape *in vitro* [55].

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 [1]. Importantly, altered TEC dynamics will only affect productive transcription if they lead to a reduced initiation frequency as in the case of promoter-proximally pause RNAPII that blocks PIC formation in metazoans [4,5]. In *S. solfataricus*, promoter-proximal TEC accumulation coincides with accumulation rather than depletion of PICs. This indicates functional interaction between PICs and TECs and we propose that promoter-proximal TECs might prevent PICs from completing initiation and clear the promoter. Notably, our ChIP-exo data reveal broader PIC footprints than previously anticipated based on *in vitro* data. This could be due to wrapping of the downstream DNA around the PIC [56] or additional DNA-binding factors in the cross-linked complexes such as chromatin proteins [57]. But, independent of the mechanism, it indicates a larger spatial overlap and thereby interference between PICs and promoter-proximal TECs possibly forming the basis for their functional interaction.

The conserved transcription elongation factor Elf1

The function of Elf1 remains poorly understood and our data provide the first insight into archaeal Elf1. Yeast Elf1 is recruited after Spt4/5 to the TEC with a gradual increase in Elf1 occupancy towards the poly-adenylation site [58] and Elf1 recruitment depends on Spt4 [59]. In *S. solfataricus*, Elf1 is recruited subsequent to Spt4/5 in the promoter-proximal region. The role of Elf1 as an integral part of the TEC genome-wide makes it a likely target for regulation, possibly by post-translational modification of the N- and C-terminal tails of Elf1.

Evolution of promoter-proximal regulation in archaea and eukaryotes

The pivotal role of Spt4/5, Elf1 and aCPSF1 in the early TEC dynamics in *Sulfolobus* shows intriguing parallels to metazoans. Firstly, the early elongation phase of transcription is rate limiting for gene expression. Secondly, Spt4/5 (DSIF) is an integral component of promoter-proximal elongation complexes in humans as well as archaea. Thirdly, CPFS73-related RNases are likely to mediate premature termination of promoter-proximal RNAPs both eukaryotes [10,60] and archaea. Our discovery that early elongation complex dynamics

modulate transcription in *Sulfolobus* 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.

Perspective

Here 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 on the molecular mechanisms underlying TEC escape.

Acknowledgements

We would like to thank Amy Schmid (Duke University), Mark Williams (Birkbeck College) for their advice. We also thank Jonathan Chubb, Jürg Bähler, Alan Cheung (UCL) and Phil Robinson (Birkbeck College) for critical reading of the manuscript. Research in the RNAP laboratory at UCL is funded by a Wellcome Investigator Award in Science 'Mechanisms and Regulation of RNAP transcription' to FW (WT 207446/Z/17/Z).

Methods

ChIP-seq

Rabbit antisera against *S. solfataricus* TFB, TBP, TFEβ, TFEα and Rpo4/7 have been described previously [22]. Polyclonal rabbit antisera against recombinant Spt5, Elf1, and

aCPSF1 were produced at Davids Biotechnology (Regensburg, GER). All antibodies were purified from antiserum by Protein A-agarose affinity chromatography.

S. solfataricus P2 cells were grown in Brock medium [61] at 76 °C in a Thermotron air incubator (Infors) to mid-exponential growth phase ($O.D._{600}$ 021 to 0.29). To get cells into the late stationary phase, we grew cells for 96 hrs at which time a small decline in $O.D._{600}$ was observed (final $O.D._{600}$ 1.8 - 2.0). For oxidative stress, cells were grown overnight in modified Brock medium without $FeCl_2$ supplemented with 0.2% tryptone to mid-exponential growth phase ($O.D._{600}$ 0.11 to 0.24) before the addition of 30 µM H_2O_2 similar to what has been described before [62]. 105 min after H_2O_2 addition cells were cross-linked. 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, 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. Debris was removed by centrifugation before freezing in liquid nitrogen and storage at -80 °C. For ChIP, 500 µl lysate diluted to a DNA content of 20 ng/µl (based on A260 measurements) was supplemented with antibody (2 µg of TFB or Rpo4/7 antibodies, 4 µg for TFEα or TFEβ) and incubated overnight in an end-over-end rotator at 4 °C. After addition of 50 µl of sheep anti-rabbit IgG Dynabeads M-280 (Thermo Scientific) were added and incubation was continued for 1 hr. Beads were washed twice with 1 ml lysis buffer, once with 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, 0.5% Nonidet P-40) and TE buffer. Immuno-precipitated material was eluted from beads by the addition of 200 µl ChIP-elution buffer (50 mM Tris/HCl pH 8.0, 10 mM EDTA, 1% SDS), de-cross-linked overnight at 65 degree in the presence of 10 µg RNase A and 40 µg Proteinase K. DNA was purified using Qiaquick PCR purification kit (Qiagen). For Spt4/5, Elf1 and CPSF1 ChIP experiments, the lysate volume was increased to 1 ml and 8 µg of antibody were used in combination with Protein G Dynabeads M-280 (Thermo Scientific) as described above. The yield of the ChIP experiments was determined using Qubit dsDNA

HS (Thermo Scientific). Libraries were prepared using the NEBNext UltraII 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 Hi-seq data from our ChIP-seq data (two biological replicates). Reads were trimmed to 50 nt and aligned to the *S. solfataricus* P2 genome (NC_002754.1) using bowtie [63] (parameters -v 2 -m 1 --fr) allowing only for a single match and proper read and converted to bam file format using SAMtools [64]. To ensure comparability between different ChIP-seq samples, we sequenced all samples to high genomic coverage and adjusted the fragment size distribution using a computational approach by subsampling the reads. The fraction size range in a ChIP-seq experiment is shaped in different ways including the chromatin shearing, size selection method and conditions during the library preparation, and gating (minimal and maximal fragment sizes) during read alignment and calculation of genomic coverage.

In order to adjust the fragment size distribution computationally, bam files were converted to bedpe format using BEDTools bamtobed [65] and imported into R. We adjusted the fragment size distribution to fit a normal distribution with mean 120 and standard deviation of 18. To this end, then read pairs were binned into 200 bins according to fragment size in the range of 51 to 250 bp randomly drawn from each bin using the sample() function in R. The number of reads to be drawn was determined by their relative frequency in the target distribution (dnorm()) multiplied by maximum total number of reads possible without exhausting any of the bins. This procedure retained on average 46% of the read pairs. Data were exported as bed files with the fragment coordinates and subsequently converted back to bam format for any downstream analysis. DeepTools bamCoverage and multiBamSummary [66] with bin size 50 was used to calculated read coverage. Data were imported into R and pairwise correlation between unfiltered and sampled data as well as and the pairwise correlation between biological replicates was assessed using the cor.test() function (see Supplementary Item).

Peak calling

Peaks were identified with MACS2 [67] in BEDPE mode, q 0.01 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 [65]. For the consistency analysis of the peaks between replicates based on p-values [68], we set a global IDR threshold of 0.05 using the Cran IDR package in R [69]. A number of reproducible TFEα and TFEβ peaks within the rRNA 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 pairwise correlations were calculated using the spearman.ci() function from the RVAideMemoire 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 width) [66,70] and converted to bigwig format. For individual gene plots normalised bigwig files were imported into R via the rtracklayer package [71] and plotted using ggplot2 [72]. Heatmaps were generated using deepTools computeMatrix and plotHeatmap functions [66].

TU selection and escape index calculation

In order to generate a list of TUs with robust signal quantification, we used a previously published map of 2229 TUs (after filtering for consistency) based on RNA-seq data [73]. The dataset contained 1035 primary TSSs with unique assignment to TU starts. We also included two previously determined transcription start sites for 5S and 16S/23S rRNA genes [74]. For the remaining 1192 TUs no experimentally verified TSS data were available, but based on

the generally short 5′ UTR length of Sso mRNAs (69% with length of 4bp or shorter) [73], we used the start codon position as rough indicator for TSS position with additional adjustments based on our ChIP-exo data (see below). The start codon for genes Sso0845 (TU 565) and Sso1077 / *fumC* (TU 710) were reassigned to the third and second ATG codon 33 bp and 21 bp downstream, respectively, as TFB and TFE ChIP-seq, TFB ChIP-exo and permanganate ChIP-seq signals as well as RNA-seq data all consistently suggested the reassignment.

To identify transcriptionally active TUs, we filtered for those TUs showing TFB occupancy at the promoter with bijective correspondence and pairing of TFB and TFEβ peaks with bijective correspondence. To this end, TFB ChIP-seq peaks were matched to TFEβ peaks 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 over 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 500bp relative to TSS to ensure that RNAP and Spt4/5 occupancy is not influenced by any TU internal promoters.

Divergent promoters in *S. solfataricus* are often tightly spaced causing the RNAP and Spt4/5 ChIP-seq signal 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.

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 TSSs [73] were used to build a training set. A peak with median position -14 relative to TSS (Inter-Quartile Range -16 to -12) was identified.

For nine TUs out of 82 TUs without experimentally mapped TSS in total, adjustment of TSS positions by 10 to 33 bp were suggested. For TU 1536 no consistent peak could be identified and it was removed from the data set. Lastly, all suggested adjustments of TSS positions were checked manually for their consistence with ChIP-seq, ChIP-exo, permanganate ChIP-seq and RNA-seq data and TSS positions were adjusted further by max 3 bp to match the (C/T)(A/G) consensus of the Inr promoter element.

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 2 [75] output of the RNA-seq data (see below). Predicted RNA 3'-ends were filtered by the following criteria: (i) a TU length of > 300 nt, (ii) no TFB peaks within the surrounding 600 bp to filter out aCPSF1 peaks resulting from promoter-proximal recruitment, (iii) continuous input coverage of > 20 reads in the surrounding 500 bp to ensure reliable input normalisation (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 window 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 [76] in 'oops' mode with 8-15 bp 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 [77]. Cell growth, cross-linking 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. Immuno-precipitation was carried out 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 hr before following the manufacturer's recommendations for washing of the beads and library preparation. Library quality was assessed using Agilent High Sensitivity DNA it (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* strain P2 genome using bowtie v1.2.2 [63] (parameters -v 2 -m 1 –best –strata -S), converted to bam file format SAMtools [64] and converted to strand-specific 5'-end coverage data using MACE v1.2 pre-processor function with default settings normalised to 1x genome coverage [78]. For pooling of replicates we calculated the geometric mean.

DNA duplex stability

DNA duplex stability was estimated using a nearest-neighbour model [79]. Gibbs free energy values for each dinucleotide were extrapolated to the optimal growth temperature for *Sulfolobus* of 76 °C. Initially, windows with sizes of 3 to 20 bp were tested for correlation of duplex stability with escape indices and a 7 bp from position -3 to +4 relative to TSS yielded the highest Spearman 's r. In the nearest neighbour model, the salt concentration is included in the Gibbs free energy calculation as offset. While the internal salt concentration of *Sulfolobus* cells is not known and could only be roughly estimated, the Spearman correlation is not influenced by offsets.

Model of productive transcription elongation under oxidative stress

We generated negative binomial generalised linear models with log link function using the glm.nb() function in the MASS package in R. Raw Spt4/5 TU body coverage was used as the dependent variable in the models while log-transformed raw input coverage over the same region was included as fixed offset effectively performing the input normalisation. We calculated the total occupancy signal within the TU body (Bd, +250 to +500 bp relative to

596 TSS) for Spt4/5 ChIP-seq data as well as the chromatin input for the subset of TUs with
597 mapped TSSs. These values were corrected for the average fragment length (120 nt) to
598 obtain an estimate of overlapping read pair counts. Log-transformed input-normalised
599 Spt4/5$_{Pr}$ occupancy was included as a second offset term. Thereby, the models effectively
600 identified variables predictive for TEC escape efficiency. As explanatory variables, we
601 tested TFB and TFEβ promoter occupancy (log-transformed), aCPSF1 load on promoter-
602 proximal TECs (log-transformed aCPSF1$_{Pr}$ to Spt4/5$_{Pr}$ ratio), and TSS DNA duplex stability.
603 The full model was structured as follows:
604
605 $$log(\text{Spt4/5}_{Bd}\ coverage) \sim\ log(input_{Bd}\ coverage) + log(\text{Spt4/5}_{Pr}) + \beta_0 + \beta_1 * log(\text{TFB})$$
606 $$+\ \beta_2 * log(\text{TFE}\beta) + \beta_3 * log\left(\frac{\text{aCPSF1}_{Pr}}{\text{Spt4/5}_{Pr}}\right) + \beta_4 * TSS\ DNA\ duplex\ stability$$
607
608 TUs passing a Cook's distance of 0.5 for a model of all explanatory variables were removed
609 from the dataset. To identify the optimal model, we used the step() function in the MASS
610 package for automatic model building using default settings for AIC minimisation. The null
611 model was used as lower boundary and the full model including all variables as upper
612 boundary. Searches were initiated with the null model (including only the raw input
613 coverage and Spt4/5$_{Pr}$ offsets) as well as all explanatory variables added separately.
614 Statistical significance of the included variables in the model was confirmed using the
615 Likelihood-ratio chi-squared test implemented in anova.negbin() (car package) by
616 comparing the optimal models against models excluding single variables ($p < 0.05$). Only
617 statistically significant explanatory variables consistent between replicates were retained in
618 the final model. To validate the model, bootstrapped 95% confidence intervals were
619 calculated for each coefficient and checked whether they were consistently positive or
620 negative using the Boot() function from the car package.
621
622 ### Permanganate ChIP-seq
623 For permanganate ChIP-seq, we adapted our ChIP-exo protocol analogously to the method
624 described by Gilmour, Pugh and co-workers [47,48] as follows. After cross-linking, cells were
625 washed once in 25 ml ice-cold PBS and resuspended in 25 ml room temperature PBS. 833

µl 300 mM $KMnO_4$ was added to a final concentration of 10 mM and incubated at room temperature. After 1 min, 25 ml of stop solution (PBS supplemented with 0.8M 2-mercaptoethanol and 40 mM EDTA) was added and cells were washed three more times in PBS before freezing in liquid nitrogen. Cells were further processed as for ChIP-exo, but λ-exonuclease and $RecJ_F$ digestion steps were omitted. After reversal of crosslinks and ethanol precipitation, the purified DNA was dissolved in 10% piperidine and incubated at 90 °C for 30 min. Piperidine was removed by three steps of 1-butanol extraction followed by chloroform extraction before ethanol precipitation. P7 primer extension, P5 adaptor ligation and PCR amplification were carried out according to the ChIP-exo protocol. Reads were mapped onto the *S. solfataricus* strain P2 genome as described above and converted to strand-specific 5'-end coverage data using BEDTools genomecov. Coverage data were loaded into R and corrected for the position by -1 nt so that it corresponds to the modified nucleotides eliminated during permanganate/piperidine cleavage. Coverage data were filtered for genomic positions with Ts. While regions around TSSs are generally characterised by strongly increased signal on T positions over non-T positions (expected for potassium permanganate/piperidine treatment of DNA), we found that these regions were often flanked by broader regions of lower coverage signal that did not show any specificity for Ts and possibly results from incomplete P7 adaptor ligation during the library preparation. For this reason, we performed a background correction on the data. Background signal was calculated as from four neighbouring non-T positions for each T (the two closest non-T residues on either side). The median background signal was subtracted and the resulting T-specific signal was normalised to 1x genome coverage of Ts for both strands.

RNA-seq

RNA was isolated by mixing samples directly with three volumes ice-cold TRIzol LS (Thermo-Fisher). RNA was isolated according to manufacturer's protocol and remaining genomic DNA was removed using the TURBO DNA-free kit (Thermo-Fisher). RNA was quantified using the Qubit RNA BR Assay kit (Thermo Scientific) and quality was assessed using the RNA ScreenTape system (Agilent). Libraries were prepared at Edinburgh Genomics using the TruSeq® Stranded Total RNA Library Prep kit (Illumina) including partial ribosomal RNA

657 depletion with the Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina). 75 bp paired-end reads
658 were generated on a HiSeq 4000 system (Illumina). Coverage tracks were produced using
659 the Rockhopper 2 software package in --rf mode [75]. Rockhopper 2 quantifies transcripts by
660 applying an upper quartile normalisation. The coverage tracks with raw coverage were
661 corrected for sequencing depth and the fraction of reads mapping sense to protein
662 encoding genes (mRNA) yielding normalised coverage in counts per million (cpm).
663

664 ## Cappable-seq short RNA sequencing

665 15 ml cell culture was rapidly mixed with 30 ml pre-cooled RNAprotect Bacteria Reagent
666 (Qiagen) placed in an ice bath. Cells were harvested by centrifugation (5 min at 4000 x g at
667 4 °C). Pellets were immediately subjected to RNA isolation using the mirVana miRNA
668 isolation kit (Ambion) with an initial resuspension buffer volume of 200 µl following the
669 protocol for small RNAs (20-200 nt length). Library preparation and deep sequencing was
670 conducted at Vertis Biotechnologie (Germany). In brief, 5′-triphosporylated RNA was
671 capped with 3′-desthiobiotin-TEG-GTP (NEB)) using the Vaccinia virus Capping enzyme
672 (NEB) and biotinylated RNA species were subsequently enriched by affinity purification
673 using streptavidin beads yielding 0.6 to 1.3% of the sRNA preparation. The eluted RNA was
674 poly-adenylated using *E. coli* Poly(A) polymerase and 5′-ends were converted to mono-
675 phosphates by incubation with RNA 5′ Pyrophosphohydrolase (NEB). Subsequently, an
676 RNA adapter (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') was ligated to the
677 newly formed 5'-monophosphate structures. First-strand cDNA synthesis was performed
678 using an oligo(dT)-adapter primer and the M-MLV reverse transcriptase at 42 °C. The
679 resulting cDNA was finally PCR-amplified (12 cycles) with TruSeq Dual Index sequencing
680 primers (Illumina) and Herculase II Fusion DNA Polymerase (Agilent). The libraries were
681 sequenced on an llumina NextSeq 500 system with 75 bp read length. In order to remove
682 poly(A)-tails and adaptors, we trimmed the reads using Cutadapt [80] in two rounds with the
683 following settings to prevent trimming of naturally occurring A-rich RNAs due to the low
684 GC-content of the *S. solfataricus* genome: (i) -a "{A15}" -e 0 -m 15 to remove all poly(A)
685 stretches of at least 15 nt length plus downstream regions and (ii) -a "A{15}"X -e 0 -O 5 to
686 terminal shorter poly(A) stretches of minimum 5 nt length. Trimmed and untrimmed reads
687 were split into separate fastq files using awk. Both fastq files were aligned to the *S.*

solfataricus genome using bowtie v1.2.2 [63] (parameters -v 1 -m 1 --best --strata -S) with untrimmed 75 nt reads shortened to 71 nt (-3 4). The bam file output was merged, sorted and indexed using SAMtools [64]. Bam files were imported into the R environment using the rsamtools package and filtered to ensure a unique sequence of the initial 20 bp within the *S. solfataricus* genome required to map the reads. TSS-RNA 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 for 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.

Immuno-detection

Cell lysates were resolved on 12% Tris-tricine SDS gels and blotted onto nitrocellulose membranes. All immune-detections were carried out using polyclonal antisera (see above) in combination with donkey anti-rabbit IgG Dylight680 (Bethyl Laboratories). Dps-I antiserum was a kind gift of Mark Young (Montana State University, USA). As loading control, we used sheep Alba antiserum (kind gift of Malcolm White, University of St. Andrews, UK) in combination with donkey anti-sheep IgG Alexa488 (Thermo Fisher). Blots were scanned on a Typhoon FLA 9500 scanner (GE Lifesciences).

Data availability

We deposited all sequencing data have at NCBI GEO under superseries GSE141290.

Code availability

The code for TSS-RNA analysis, ChIP-seq read sampling, permanganate ChIP-seq background correction as well as the multiregression analysis is available at github.com/fblombach/ChIP-seq.

References

1 Ehrensberger, A. H., Kelly, G. P. & Svejstrup, J. Q. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* **154**, 713-715, doi:10.1016/j.cell.2013.07.032 (2013).

2 Browning, D. F. & Busby, S. J. Local and global regulation of transcription initiation in bacteria. *Nat Rev Microbiol* **14**, 638-650, doi:10.1038/nrmicro.2016.103 (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, doi:10.1534/genetics.111.127019 (2011).

4 Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* **6**, doi:10.7554/eLife.29736 (2017).

5 Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet*, doi:10.1038/ng.3867 (2017).

6 Krebs, A. R. *et al.* Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol Cell* **67**, 411-422 e414, doi:10.1016/j.molcel.2017.06.027 (2017).

7 Steurer, B. *et al.* Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc Natl Acad Sci U S A* **115**, E4368-E4376, doi:10.1073/pnas.1717920115 (2018).

8 Erickson, B., Sheridan, R. M., Cortazar, M. & Bentley, D. L. Dynamic turnover of paused Pol II complexes at human promoters. *Genes & Development* **32**, 1215-1225, doi:10.1101/gad.316810.118 (2018).

9 Nilson, K. A. *et al.* Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic Acids Research* **45**, 11088-11105, doi:10.1093/nar/gkx724 (2017).

10 Elrod, N. D. *et al.* The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. *Molecular Cell* **76**, 738-752.e737, doi:https://doi.org/10.1016/j.molcel.2019.10.034 (2019).

11 Reppas, N. B., Wade, J. T., Church, G. M. & Struhl, K. The Transition between Transcriptional Initiation and Elongation in E. coli Is Highly Variable and Often Rate Limiting. *Molecular Cell* **24**, 747-757, doi:10.1016/j.molcel.2006.10.030 (2006).

12 Ring, B. Z., Yarnell, W. S. & Roberts, J. W. Function of E. coli RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. *Cell* **86**, 485-493 (1996).

13 Lerner, E. *et al.* Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A* **113**, E6562-E6571, doi:10.1073/pnas.1605038113 (2016).

14 Duchi, D. *et al.* RNA Polymerase Pausing during Initial Transcription. *Mol Cell* **63**, 939-950, doi:10.1016/j.molcel.2016.08.011 (2016).

15 Petushkov, I., Esyunina, D. & Kulbachinskiy, A. Possible roles of sigma-dependent RNA polymerase pausing in transcription regulation. *RNA Biol* **14**, 1678-1682, doi:10.1080/15476286.2017.1356568 (2017).

16 Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nature Reviews Microbiology* **15**, 711-723, doi:10.1038/nrmicro.2017.133 (2017).

17 Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* **9**, 85-98, doi:10.1038/nrmicro2507 (2011).
18 Fouqueau, T., Blombach, F. & Werner, F. Evolutionary Origins of Two-Barrel RNA Polymerases and Site-Specific Transcription Initiation. *Annu Rev Microbiol* **71**, 331-348, doi:10.1146/annurev-micro-091014-104145 (2017).
19 Korkhin, Y. *et al.* Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS biology* **7**, e102, doi:10.1371/journal.pbio.1000102 (2009).
20 Bell, S. D., Kosa, P. L., Sigler, P. B. & Jackson, S. P. Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A* **96**, 13662-13667 (1999).
21 Gietl, A. *et al.* Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res* **42**, 6219-6231, doi:10.1093/nar/gku273 (2014).
22 Blombach, F. *et al.* Archaeal TFEalpha/beta is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *Elife* **4**, e08378, doi:10.7554/eLife.08378 (2015).
23 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, doi:10.1074/jbc.M611674200 (2007).
24 Schulz, S. *et al.* TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc Natl Acad Sci U S A* **in press** (2016).
25 Werner, F. & Weinzierl, R. O. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol Cell Biol* **25**, 8344-8355, doi:25/18/8344 [pii] 10.1128/MCB.25.18.8344-8355.2005 (2005).
26 Hirtreiter, A. *et al.* Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res* **38**, 4040-4051, doi:10.1093/nar/gkq135 (2010).
27 Smollett, K., Blombach, F., Reichelt, R., Thomm, M. & Werner, F. A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat Microbiol* **2**, 17021, doi:10.1038/nmicrobiol.2017.21 (2017).
28 Fouqueau, T. *et al.* The cutting edge of archaeal transcription. *Emerging Topics in Life Sciences* **2**, 517, doi:10.1042/ETLS20180014 (2018).
29 Daniels, J. P., Kelly, S., Wickstead, B. & Gull, K. Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol Direct* **4**, 24, doi:10.1186/1745-6150-4-24 (2009).
30 Fouqueau, T. *et al.* The transcript cleavage factor paralogue TFS4 is a potent RNA polymerase inhibitor. *Nat Commun* **8**, 1914, doi:10.1038/s41467-017-02081-3 (2017).
31 Grohmann, D. *et al.* The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell* **43**, 263-274, doi:10.1016/j.molcel.2011.05.030 (2011).
32 Sanders, T. J. *et al.* FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat Microbiol*, doi:10.1038/s41564-020-0667-3 (2020).
33 Blombach, F., Matelska, D., Fouqueau, T., Cackett, G. & Werner, F. Key Concepts and Challenges in Archaeal Transcription. *J Mol Biol*, doi:10.1016/j.jmb.2019.06.020 (2019).

34 Blombach, F. & Grohmann, D. Same same but different: The evolution of TBP in archaea and their eukaryotic offspring. *Transcription* **8**, 162-168, doi:10.1080/21541264.2017.1289879 (2017).
35 Lemmens, L., Maklad, H. R., Bervoets, I. & Peeters, E. Transcription Regulators in Archaea: Homologies and Differences with Bacterial Regulators. *J Mol Biol*, doi:10.1016/j.jmb.2019.05.045 (2019).
36 Peeters, E., Peixeiro, N. & Sezonov, G. Cis-regulatory logic in archaeal transcription. *Biochem Soc Trans* **41**, 326-331, doi:10.1042/BST20120312 (2013).
37 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, doi:10.1146/annurev-genet-120116-023413 (2017).
38 Ochs, S. M. *et al.* Activation of archaeal transcription mediated by recruitment of transcription factor B. *J Biol Chem* **287**, 18863-18871, doi:10.1074/jbc.M112.365742 (2012).
39 Ouhammouch, M., Werner, F., Weinzierl, R. O. & Geiduschek, E. P. A fully recombinant system for activator-dependent archaeal transcription. *J Biol Chem* **279**, 51719-51721, doi:C400446200 [pii]
10.1074/jbc.C400446200 (2004).
40 Bell, S. D. & Jackson, S. P. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**, 222-228 (1998).
41 Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408-1419, doi:10.1016/j.cell.2011.11.013 (2011).
42 Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295-301, doi:10.1038/nature10799 (2012).
43 Nagy, J. *et al.* Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat Commun* **6**, 6161, doi:10.1038/ncomms7161 (2015).
44 Spitalny, 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, doi:10.1074/jbc.M303633200 (2003).
45 Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139-1143, doi:10.1126/science.1131398 (2006).
46 Dexl, S. *et al.* Displacement of the transcription factor B reader domain during transcription initiation. *Nucleic Acids Res*, doi:10.1093/nar/gky699 (2018).
47 Lai, W. K. & Pugh, B. F. Genome-wide uniformity of human 'open' pre-initiation complexes. *Genome Res* **27**, 15-26, doi:10.1101/gr.210955.116 (2017).
48 Li, J. *et al.* Kinetic Competition between Elongation Rate and Binding of NELF Controls Promoter-Proximal Pausing. *Molecular Cell* **50**, 711-722, doi:10.1016/j.molcel.2013.05.016 (2013).
49 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 U S A* **95**, 15218-15222 (1998).
50 Ettwiller, L., Buswell, J., Yigit, E. & Schildkraut, I. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in

a model prokaryote and the gut microbiome. *BMC Genomics* **17**, 199, doi:10.1186/s12864-016-2539-z (2016).

51 Phung, D. K. *et al.* Archaeal beta-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res* **41**, 1091-1103, doi:10.1093/nar/gks1237 (2013).

52 Yue, L. *et al.* The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3'-end cleavage mode. *Nucleic Acids Res*, doi:10.1093/nar/gkaa702 (2020).

53 Maaty, W. S. *et al.* Something old, something new, something borrowed; how the thermoacidophilic archaeon Sulfolobus solfataricus responds to oxidative stress. *PLoS One* **4**, e6964, doi:10.1371/journal.pone.0006964 (2009).

54 Cabart, P. & Luse, D. S. Inactivated RNA polymerase II open complexes can be reactivated with TFIIE. *J Biol Chem* **287**, 961-967, doi:10.1074/jbc.M111.297572 (2012).

55 Heyduk, E. & Heyduk, T. DNA template sequence control of bacterial RNA polymerase escape from the promoter. *Nucleic Acids Res* **46**, 4469-4486, doi:10.1093/nar/gky172 (2018).

56 Rivetti, C., Guthold, M. & Bustamante, C. Wrapping of DNA around the E.coli RNA polymerase open promoter complex. *EMBO J* **18**, 4464-4475, doi:10.1093/emboj/18.16.4464 (1999).

57 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, doi:10.1038/nrmicro3467 (2015).

58 Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17**, 1272-1278, doi:10.1038/nsmb.1903 (2010).

59 Prather, D., Krogan, N. J., Emili, A., Greenblatt, J. F. & Winston, F. Identification and characterization of Elf1, a conserved transcription elongation factor in Saccharomyces cerevisiae. *Mol Cell Biol* **25**, 10122-10135, doi:10.1128/MCB.25.22.10122-10135.2005 (2005).

60 Lykke-Andersen, S. *et al.* Integrator is a genome-wide attenuator of non-productive transcription. *bioRxiv*, 2020.2007.2017.208702, doi:10.1101/2020.07.17.208702 (2020).

61 Zaparty, M. *et al.* "Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus. *Extremophiles* **14**, 119-142, doi:10.1007/s00792-009-0280-0 (2010).

62 Wiedenheft, B. *et al.* An archaeal antioxidant: characterization of a Dps-like protein from Sulfolobus solfataricus. *Proc Natl Acad Sci U S A* **102**, 10551-10556, doi:10.1073/pnas.0501497102 (2005).

63 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

64 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

65 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

66 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:10.1093/nar/gku365 (2014).

67 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

68 Li, Q. H., Brown, J. B., Huang, H. Y. & Bickel, P. J. Measuring Reproducibility of High-Throughput Experiments. *Annals of Applied Statistics* **5**, 1752-1779, doi:10.1214/11-Aoas466 (2011).

69 R Core Team. R: A language and environment for statistical computing. . (2014).

70 Diaz, A., Park, K., Lim, D. A. & Song, J. S. Normalization, bias correction, and peak calling for ChIP-seq. *Stat Appl Genet Mol Biol* **11**, Article 9, doi:10.1515/1544-6115.1750 (2012).

71 Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12**, 115-121, doi:10.1038/nmeth.3252 (2015).

72 Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Ggplot2: Elegant Graphics for Data Analysis*, 1-212, doi:10.1007/978-0-387-98141-3 (2009).

73 Wurtzel, O. *et al.* A single-base resolution map of an archaeal transcriptome. *Genome Res* **20**, 133-141, doi:10.1101/gr.100396.109 (2010).

74 Reiter, W. D. *et al.* Putative promoter elements for the ribosomal RNA genes of the thermoacidophilic archaebacterium Sulfolobus sp. strain B12. *Nucleic Acids Res* **15**, 5581-5595 (1987).

75 Tjaden, B. De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biology* **16**, 1, doi:10.1186/s13059-014-0572-2 (2015).

76 Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).

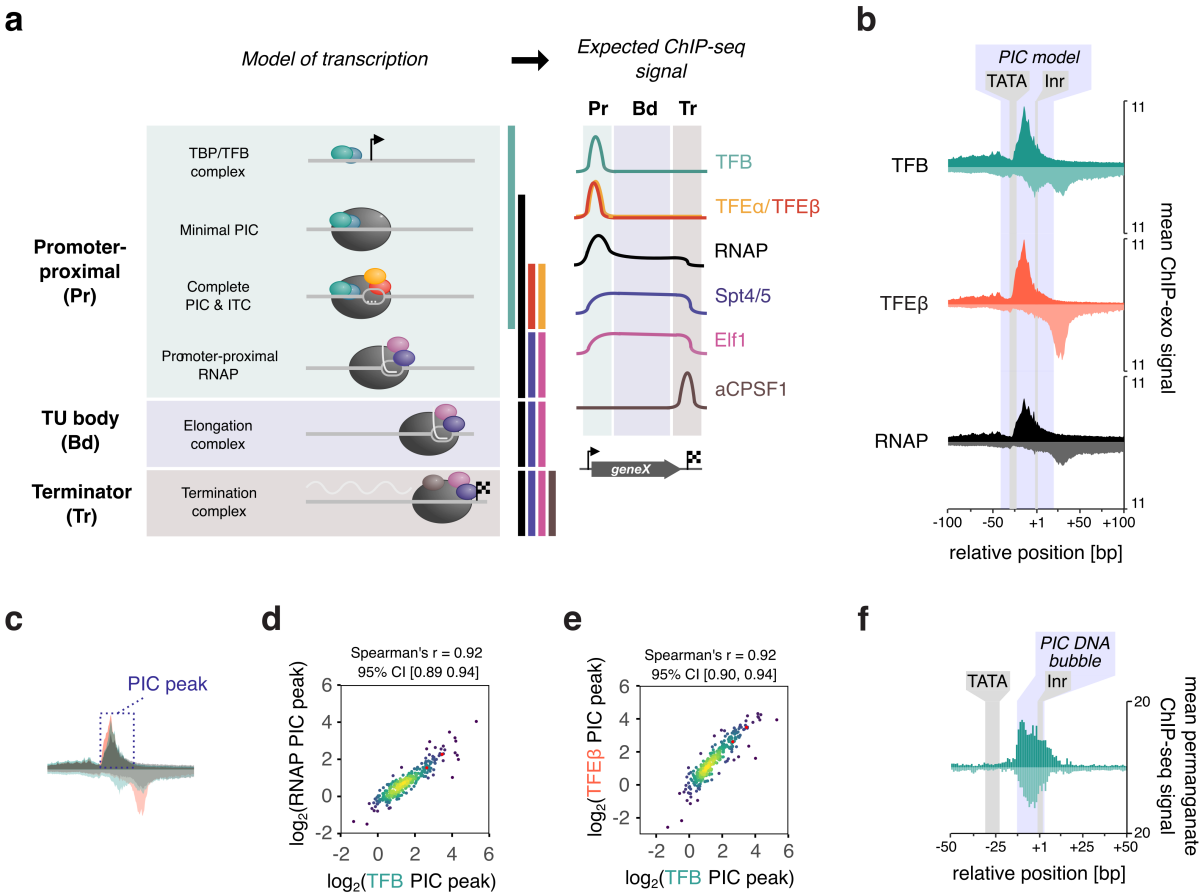
77 Serandour, A. A., Brown, G. D., Cohen, J. D. & Carroll, J. S. Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. *Genome Biol* **14**, R147, doi:10.1186/gb-2013-14-12-r147 (2013).

78 Wang, L. *et al.* MACE: model based analysis of ChIP-exo. *Nucleic Acids Res* **42**, e156, doi:10.1093/nar/gku846 (2014).

79 SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* **95**, 1460-1465, doi:10.1073/pnas.95.4.1460 (1998).

80 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data AnalysisDO - 10.14806/ej.17.1.200* (2011).

938 Figures



939

940 **Figure 1: Uniform PIC assembly during exponential growth.**

941 (a) Stages of transcription in archaea can be assessed by a combination of RNAP and basal
942 transcription initiation and elongation factor occupancy.

943 (b) Aggregate plots of ChIP-exo signal at the promoter for RNAP and initiation factors
944 (n=298 TUs). The average signal on the non-template and template strand is shown above
945 and below the line, respectively. Data are pooled from three biological replicates.

946 (c) Schematic showing the 50 nt window on the non-template strand that we used to
947 quantify and correlate the ChIP-exo signal. Because RNAP and initiation factors TFB and
948 TFEβ yield similar profiles on the non-template strand, this signal can be attributed to the
949 PIC.

950 (d-e) Scatter plots depicting correlation between the main ChIP-exo signal for TFB and
951 RNAP (d) or TFEβ (e) within the PIC (n = 298). Data represent the average signal over a 50
952 nt window that we attributed to the PIC (see panel c). The geometric mean of three
953 biological replicates is shown.

954 (f) Aggregate plots of permanganate ChIP-seq signal at the promoter for RNAP and
955 initiation factors TFB (n=298 TUs). The average signal for T-encoding positions on the non-
956 template and template strand is shown above and below the line, respectively. Data are
957 pooled from two biological replicates.
958
959
960

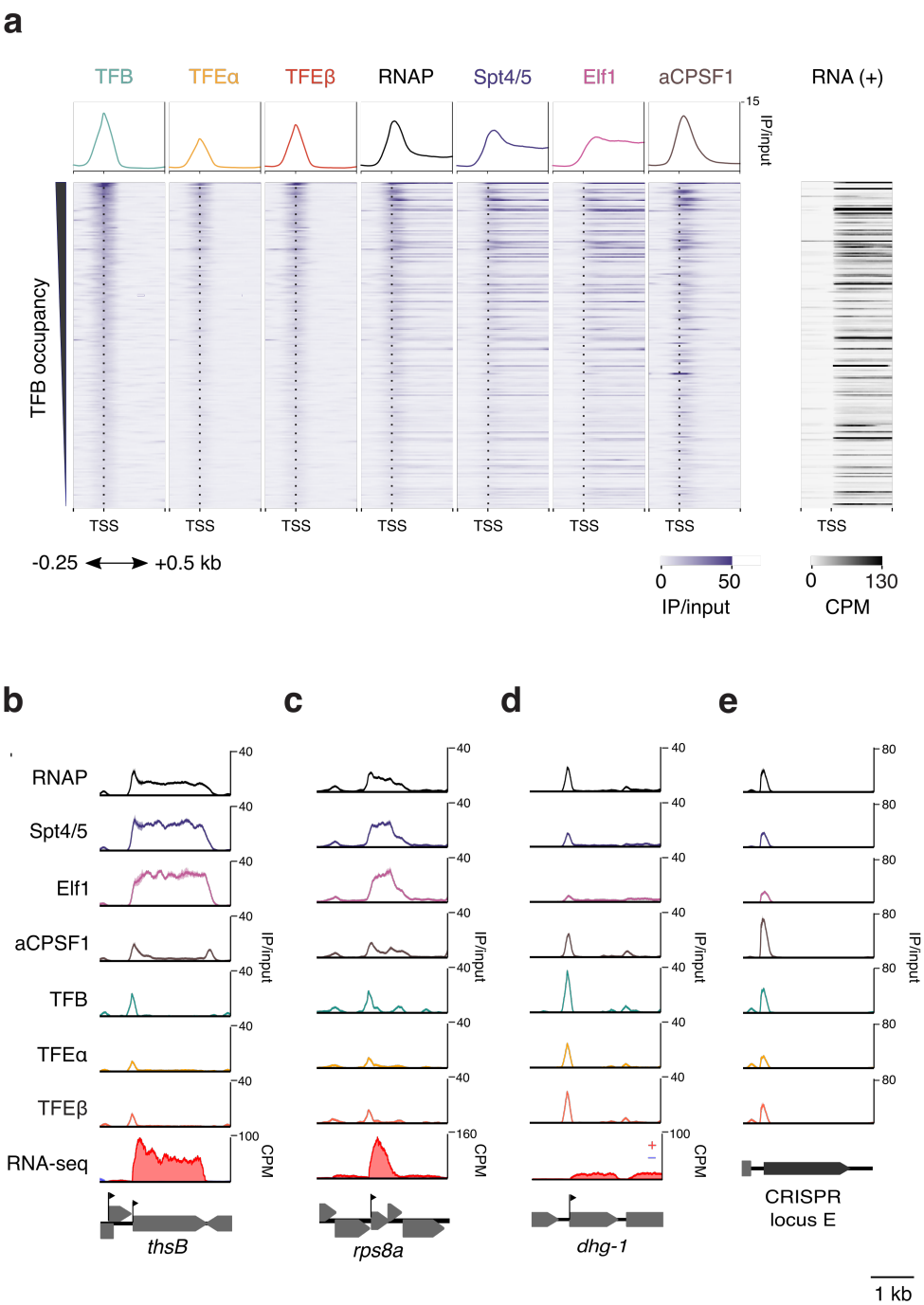


Figure 2: Productive transcription is limited by RNAP escape.

963 (a) Heatmap of RNAP and the basal transcription machinery on a selected set of 212 TUs
964 for exponential growth phase. The corresponding RNA-seq data for the plus strand are
965 depicted on the right. Data represent one representative of two biological replicates.
966 (b-e) ChIP-seq occupancy plots on *thsB* coding for a subunit of the thermosome chaperone
967 complex (b), rps8e (c), *dhg-1* coding for a putative glucose-1-dehydrogenase (d), and
968 CRISPR locus E (e). Traces show mean occupancy for two biological replicates with the
969 range depicted as semi-transparent ribbon.
970 (f-h) Correlation of steady-state mRNA levels with RNAP occupancy at the promoter
971 ($RNAP_{Pr,}$, f), the TU body ($RNAP_{Bd}$, g), and TFB promoter occupancy (h), n = 211 TUs. ChIP-
972 seq data represent the geometric mean of two biological replicates. Rockhopper [75]
973 estimates of mRNA levels are based on two biological replicates.
974
975
976
977
978

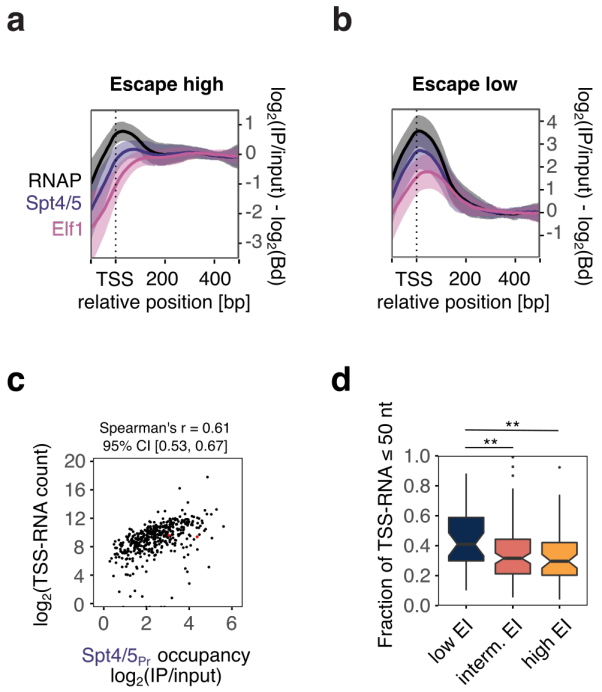


Figure 3: Promoter-proximal elongation determines RNAP escape.

(a-b) Elongation factors Spt4/5 and Elf1 are sequentially recruited to the TEC following TFE release. Aggregate plots for TUs with high (mean EI RNAP> -1, n= 58) (a) and low escape indices (< -2.5, n= 41) (b). RNAP, Spt4/5 and Elf1 ChIP-seq occupancy was scaled to their respective average occupancy within the body of each specific TU ($RNAP_{Bd}$, $Spt4/5_{Bd}$, $Elf1_{Bd}$). The profiles thus represent the relative recruitment of the factors to the TEC. Lines represent mean occupancy values with 1x standard deviation shown as semi-transparent ribbon. Data are from a single biological replicate.

(c) Scatter plot depicting the correlation between $Spt4/5_{Pr}$ and TSS-RNA (n=438). Data represent the mean of two biological replicates.

(d) TSS-RNA length distribution correlates with TEC escape. Boxplots depicting the fraction of TSS-RNAs smaller than 50 nt for TUs with low, intermediate, and high RNAP escape. Statistical significance of the observed differences was tested using a one-sidedwilcoxon rank sum test. ** denotes $p < 0.01$. The number of TUs in each EI category was 35 (low EI), 109 (intermediate EI), 56 (high EI). Data represent the mean of two biological replicates.

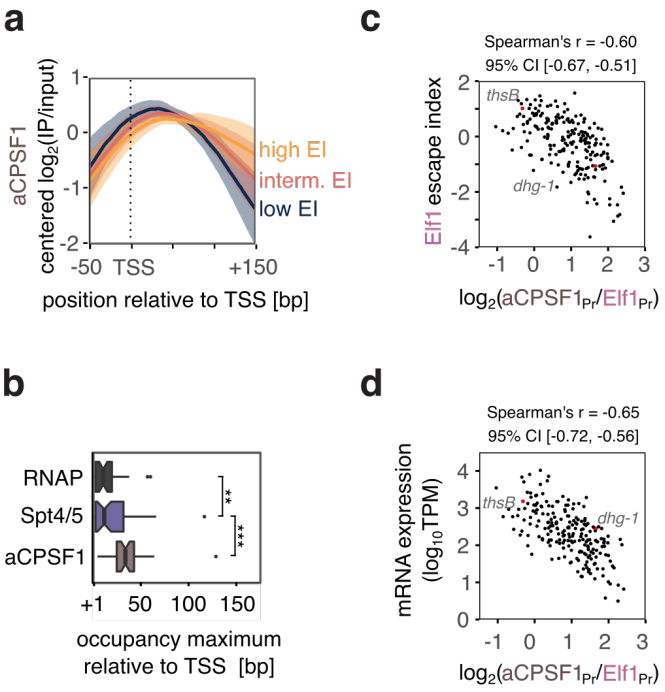


Figure 4: aCPSF1 promoter occupancy correlates with reduced TEC escape

(a) Promoter-proximal aCPSF1 occupancy pattern changes with TEC escape. aCPSF1 aggregate occupancy plots showing the averaged of centered log_2-transformed occupancy data for TUs with high (n=58), intermediate (n=113) and low escape (n=41). 1-fold standard deviation is depicted as semi-transparent ribbon. All ChIP-seq data in this figure are from a single, representative biological replicate.

(b) Boxplots depicting the position of maximum ChIP occupany within the first 150 bp of TUs with low RNAP EI (n=41) for RNAP, Spt4/5 and aCPSF1. Significance of the observed differences was tested using one-sided paired wilcoxon rank sum test. *** denotes $p < 0.001$, ** $p < 0.01$. Data for high EI are difficult to plot due to the strong but even (flat) occupancy profiles.

(c) Scatter plots depicting the anticorrelation between Elf1 EI (mean of two biological replicates) and the relative load of aCPSF1 on the Elf1-bound TEC calculated as $aCPSF1_{Pr}$ to $Elf1_{Pr}$ ratio (geometric mean of two biological replicates, n=212 genes).

(d) Correlation of aCPSF1 load ($aCPSF_{Pr}/Elf1_{Pr}$) to mRNA expression levels (mean of two biological replicates, n=211 genes).

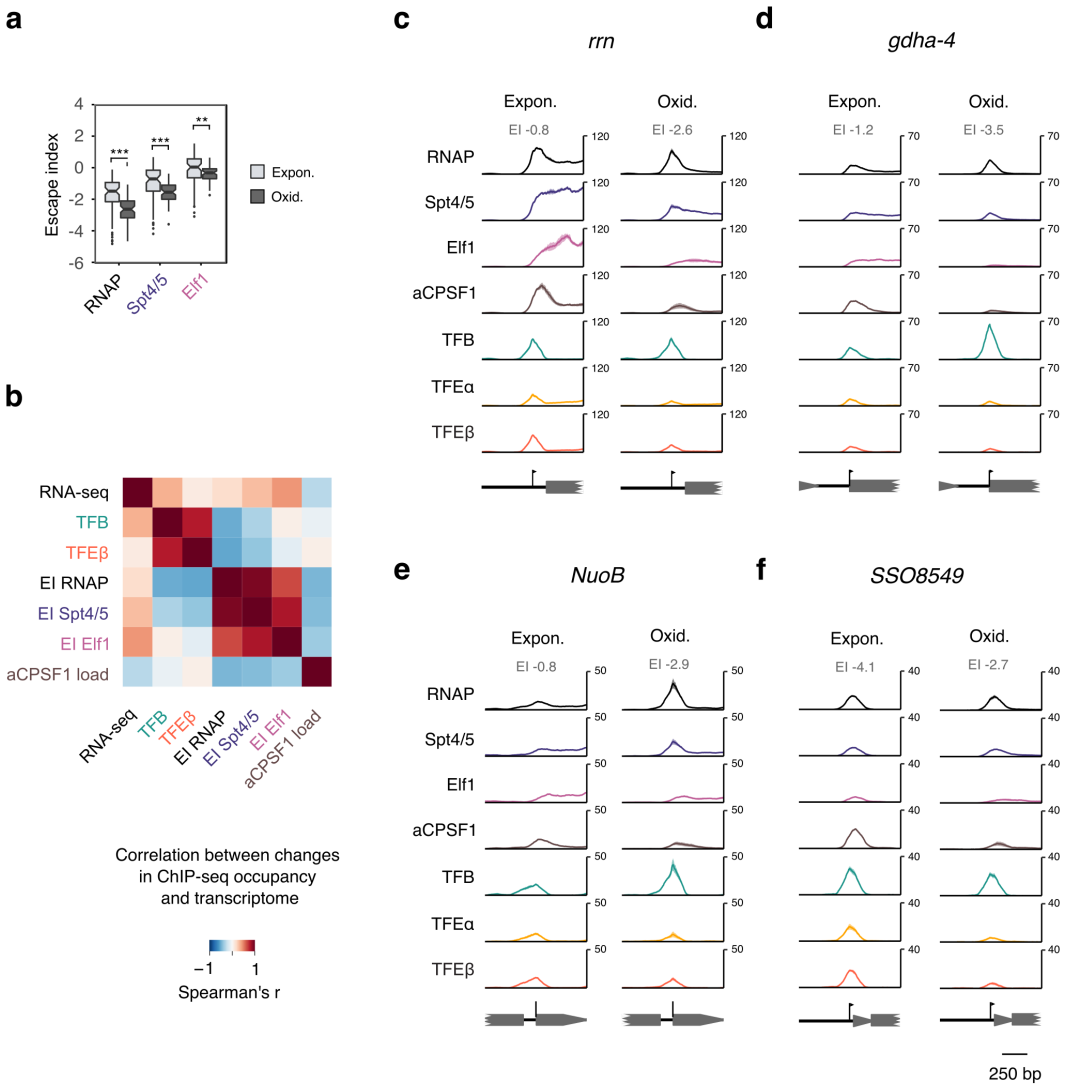


Figure 5: Oxidative stress reduces RNAP escape

(a) RNAP and elongation factors show reduced escape under oxidative stress. Boxplots comparing escape indices under exponential growth and oxidative stress conditions for TUs accessible for analysis in both conditions (n=71). Differences in escape index distribution were assessed using one-sided paired wilcoxon rank-sum test, *** denotes $p < 0.001$, ** $p < 0.01$.

(b) Heatmap showing correlated changes in initiation factor occupancy, escape indices and RNA output between exponential growth and oxidative stress. Spearman rank correlation was calculated for protein-encoding TUs accessible for analysis in both conditions (n=70).

1025 Correlations were calculated for the mean escape index and the geometric mean of all
1026 other values for two biological replicates.
1027 (c-f) ChIP-seq profiles of the basal transcription machinery for four different promoters
1028 during exponential growth (Expon.), and oxidative stress (Oxid.): *rrn* (c), *gdha-4* (d), *NuoB*
1029 (e), and *SSO8549* (f). Traces show mean occupancy for two biological replicates with the
1030 range depicted as semi-transparent ribbon.

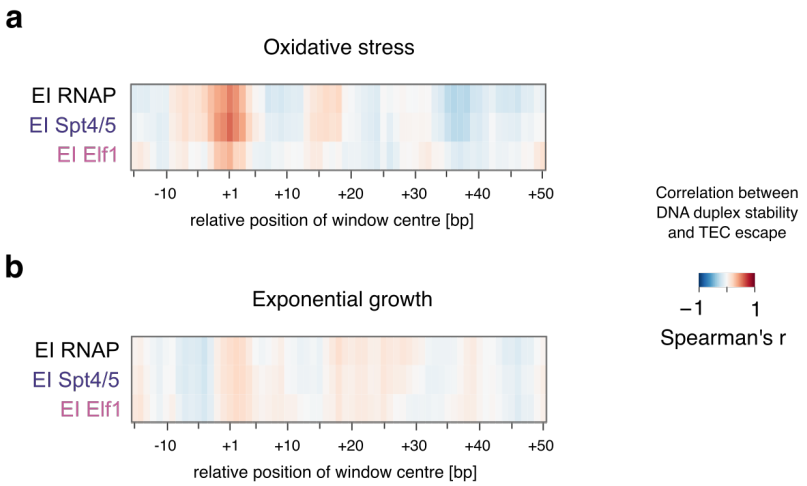


Figure 6: DNA duplex stability around the TSS is linked to TEC escape

(a-b) TEC escape is sensitive to DNA duplex stability around the TSS under oxidative stress. DNA duplex stability was calculated over a 7 bp sliding window for individual promoters and correlated with the escape indices for RNAP, Spt4/5, and Elf1 (mean of two biological replicates) under oxidative stress conditions (a) and during exponential growth (b). Selected TUs with mapped TSS were included (n=93 for oxidative stress and n=140 for exponential growth).

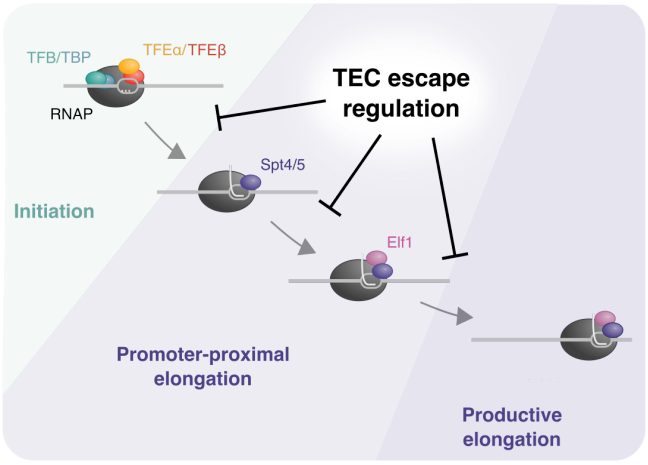


Figure 7: A model for the promoter-proximal elongation phase

Schematic overview of the promoter-proximal elongation phase and the effect of TEC escape regulation on individual steps leading towards productive transcription. Low TEC escape is associated with the accumulation of PICs and the two different TEC intermediates $TEC_{Spt4/5}$ and $TEC_{Spt4/5\text{-}Elf1}$.