



Blombach, F., Smollett, K. L., Grohmann, D. and Werner, F. (2016)
Molecular mechanisms of transcription initiation—structure, function, and
evolution of TFE/TFIIE-like factors and open complex formation. *Journal
of Molecular Biology*, 428(12), pp. 2592-2606.
(doi:[10.1016/j.jmb.2016.04.016](https://doi.org/10.1016/j.jmb.2016.04.016))

This is the author's final accepted version.

There may be differences between this version and the published version.
You are advised to consult the publisher's version if you wish to cite from
it.

<http://eprints.gla.ac.uk/150488/>

Deposited on: 29 November 2017

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Molecular Mechanisms of Transcription Initiation - structure, function and evolution of TFE/TFIIE-like factors and open complex formation

Fabian Blombach¹, Katherine L. Smollett¹, Dina Grohmann², and Finn
Werner^{1,*}

¹Institute of Structural and Molecular Biology, Division of Biosciences,
University College London, London WC1E 6BT, United Kingdom

²Institute of Microbiology, University of Regensburg, 93053 Regensburg,
Germany

*Correspondence:

f.werner@ucl.ac.uk

Abbreviations: PIC – pre-initiation complex, OC – open complex, CC –
closed complex, NTS - non-template strand, TSS – transcription start site, ZR
– Zinc ribbon, WH – winged helix-turn-helix, eWH – extended WH, FRET –
Förster Resonance Energy Transfer, smFRET – single molecule FRET, Mja –
Methanococcus jannaschii, Sso – *S. solfataricus*, Pfu – *Pyrococcus furiosus*

Abstract

Transcription initiation requires that the promoter DNA is melted and the template strand is loaded into the active site of the RNA polymerase, forming the open complex. The archaeal initiation factor TFE and its eukaryotic counterpart TFIIE facilitate this process. Recent structural and biophysical studies that have revealed the position of TFE/TFIIE within the pre-initiation complex and illuminated its role in open complex formation. TFE operates via allosteric and direct mechanisms. Firstly, it interacts with the RNAP and induces the opening of the flexible RNAP clamp domain concomitant with DNA melting and template loading. Secondly, TFE binds physically to single stranded DNA in the transcription bubble of the open complex and increases its stability. The identification of the β -subunit of archaeal TFE enabled us to reconstruct the evolutionary history of TFE/TFIIE-like factors, which is characterised by winged helix (WH) domain expansion in eukaryotes as well as loss of metal centres including Iron Sulphur clusters and Zinc ribbons. Open complex formation is an important target for the regulation of transcription in all domains of life. We propose that TFE and the bacterial general transcription factor CarD, though structurally and evolutionary unrelated, show interesting parallels in their mechanism to enhance open complex formation. We argue that open complex formation is used as a way

to regulate transcription in all domains of life, and these regulatory mechanisms co-evolved with the basal transcription machinery.

Introduction

Multisubunit RNA polymerases (RNAP) initiate transcription with the aid of general transcription factors. These factors can facilitate the recruitment of RNAP to the promoter and stimulate the local melting of DNA around the transcription start site (TSS), and subsequent loading of the template strand into the active site in a process called open complex formation. The transition from the closed (CC) to the open complex (OC) is accompanied by large-scale structural rearrangements of the transcription initiation complex, in particular of the RNAP clamp. In this article we review recent contributions to the structural and functional understanding of OC formation in RNAP systems in the different domains of life, chiefly focussed on the archaeal transcription apparatus.

Evolution of transcription initiation

The classical phylogenetic interpretation provided by Carl Woese in the 1970s stipulates that all cellular life belongs to either of one of three domains: bacteria, archaea and eukarya [1]. However, the discovery of novel archaeal phyla and improved methods for the reconstruction of ancient phylogenetic relationships are revolutionising our view on archaeal evolution

and eukaryogenesis [2-5]. The most recent scenarios have the evolution of eukaryotes beginning deep within the archaeal domain and hence archaea show great promise to illuminate the origin and evolution of the eukaryotic transcription machineries. To this end, it is essential to understand the diversity of the general transcription factors and RNAP within the archaeal domain.

According to their evolutionary conservation (Figure 1) only one RNAP-associated general transcription factor shares the deep ancestry of RNAP core subunits going back to the last universal common ancestor of life (LUCA): the transcription elongation factor Spt5 (NusG in bacteria). Despite several attempts of reconciliation there is no strong evidence that general transcription initiation factors in bacteria and archaea/eukaryotes share a common ancestor. Even though the mechanisms that govern transcription initiation appear to be functionally conserved, the factors that facilitate this process have likely evolved independently in bacteria and archaea prior to the rise of the eukaryotes.

Bacterial RNAPs utilise a range of sigma (σ) factors that control the transcription of distinct subsets of genes. There are several families of sigma factors that can be divided in two categories, the common σ^{70} -related factors and the more narrow phylogenetically distributed σ^{54} -type factors [6]. Sequence and structural alignments demonstrate that σ^{54} and σ^{70} are not

derived from a common ancestor but have evolved independently in the bacterial domain [7]. There are thus two functionally discrete mechanisms of transcription initiation in bacteria that are either (i) spontaneous (σ^{70}) or (ii) relying on ATP hydrolysis by bacterial enhancer binding proteins (bEBP) of the AAA+ family (σ^{54}). The combination of gene-specific bEBPs and σ^{54} orchestrates the nitrogen metabolism, and various stress responses including the phage shock response in *E. coli* [8, 9].

The archaeal RNAP and the three orthodox eukaryotic RNAPs (RNAPI, II and III) all depend on two homologous general transcription factors, the TATA binding protein (TBP) and transcription factor B (TFB). The latter is homologous to Taf1b, TFIIB and Brf1 in the eukaryotic RNAPI, II and III systems, respectively. RNAPII requires additional general transcription factors including TFIIH, and ATP hydrolysis for productive transcription initiation. Both accessory factor dependency and energy expenditure are reminiscent of the bacterial bEBP- σ^{54} ensemble but are evolutionary and mechanistically unrelated. The archaeal RNAP, RNAPII and -III employ a third conserved factor that stimulates OC formation, called transcription factor E (TFE), TFIIE and the RPC82/34 complex (hRPC62/39 in human), respectively [10]. Recent publications have provided intriguing insights into the structural organisation of these factors [11-17].

This review provides a perspective into OC formation during transcription initiation. We focus on the structure, function and evolution of TFE-related factors, draw parallels to the bacterial transcription machinery, and discuss the possibility of OC formation as a means to regulate transcription in bacteria, archaea and eukaryotes.

Architecture of archaeal RNAP

The conserved core of all multisubunit RNAPs adopts a crab claw shape with the lobe and clamp domains forming the two pincers. Lobe and clamp domains are constituted by the two largest subunits Rpo1 and Rpo2 (Rpb1 and Rpb2 in RNAPII, β' and β in bacterial RNAP) that also encompass all structural elements composing the active site. In most archaea Rpo1 is split into two subunits Rpo1' and Rpo1'' and similarly Rpo2 is split in methanogenic archaea. All archaeal and eukaryotic RNAPs include an Rpo4/7 stalk module that protrudes from the main body of the enzyme orthogonal to the direction of transcription; it binds to the nascent transcript via an OB fold thereby modulating both processivity and termination of transcription [18]. The stalk module is absent in bacterial RNAP. Of the twelve RNAPII subunits 10 are conserved within all archaea, whereas cren- and korarchaeota also contain an RPB8 homologue in line with the notion that these archaea are closer related to eukaryotes than euryarchaeota [2, 19]. The crenarchaeal *Sulfolobus* RNAP harbours an additional RNAP subunit, Rpo13, which is not

conserved in eukaryotic nor bacterial RNAPs [20]. Rpo13 is a largely disordered protein that interacts with the downstream DNA in a sequence-independent fashion and may contribute to RNAP-DNA interactions during transcription initiation and/or elongation [21].

All RNAPs are molecular machines that are made of rigid and flexible parts undergoing conformational changes during the transcription process. The most prominent flexible motif is the RNAP clamp (consisting of segments of the two largest subunits) that both adjusts the width of the DNA binding channel and has the ability to translate allosteric changes from the outside of the enzyme to the active centre, in particular to the bridge helix that is anchored to the inside of the clamp. It is thought that the opening and closing movements of the clamp over the DNA binding channel are integral to the process of OC formation during transcription initiation, and the initiation factor TFE binds to the tip of the clamp and changes its conformation [12].

A recruitment cascade nucleates transcription initiation

Both the archaeal RNAP and RNAPII only require two factors to facilitate promoter-directed transcription *in vitro* using strong promoters and negatively supercoiled templates [22, 23], even though both systems utilise additional factors to enhance this process. The sequential assembly of the archaeal pre-initiation complex (PIC) consisting of DNA, TBP, TFB and RNAP

(Figure 2) is congruent with the model for RNAPII PIC assembly first described by Steve Buratowski in 1989 [24]. First, TBP binds to the TATA box of the promoter and distorts this by bending the DNA roughly at a right angle [25]. The kinetics of archaeal TBP-TATA box interactions are significantly faster compared to their eukaryotic counterparts with a complex lifetime in the milliseconds range [26]. Second, or concomitant with TBP binding, TFB is recruited to TATA-TBP forming the ternary complex. Sequence-specific interactions of the TFB core C-terminal cyclin repeat with the promoter DNA immediately upstream of the TATA-box, the B recognition element (BRE) [27], are required for stable TBP-TFB-TATA-box complex formation; the BRE also provides the means to give the PIC the correct directionality on the archaeal promoter in lieu of additional core promoter elements fulfilling this role in eukaryotes [28]. The incorporation of TFB into the ternary complex stabilises the TBP-TATA interaction. This stabilisation appears to be more prevalent in some archaea (e.g. the crenarchaeon *Sulfolobus solfataricus*, Sso) where no TBP-TATA-box complexes in the absence of TFB are observable *in vitro*. TBP-induced bending of the DNA only in the presence of TFB suggesting that Sso TBP and Sso TFB bind concomitantly to the promoter DNA. In others (e.g. the euryarchaeon *Methanocaldococcus jannaschii*, Mja) TBP-TATA-box complexes are formed and the addition of TFB appears to have no influence

on their stability [26]. Sso TFB-TBP-DNA complexes exhibit a ten-fold increased complex lifetime compared to Mj TBP-DNA complexes supporting the notion that TFB stabilises the TBP-DNA complex. Third, RNAP is recruited to the ternary complex by a sophisticated network of interactions between TFB and RNAP. The N-terminal TFB Zn ribbon (ZR) domain binds to the RNAP dock domain [29]. The TFB B-reader and B-helix motifs, which connect the TFB ZR and -core domains, make intricate interactions with the inside of the RNAP clamp proximal to the active site. Mutational analysis of these structural elements in Mja and *Pyrococcus furiosus* (Pfu) TFB revealed conserved elements residing in the TFB linker region that are important for PIC stabilisation and synthesis of the initial phosphodiester bonds [29-31]. Finally, the N-terminal cyclin repeat of the TFB core domain makes contacts with the RNAP DNA binding channel. Initially the PIC assembles to form the CC with double stranded DNA likely hovering over the DNA binding cleft and RNAP making little or no contact with the promoter DNA. During OC formation the DNA strands of the promoter are separated and the template strand is loaded into the active site concomitant with conformational changes of the PIC. This mechanism of transcription initiation and the basal factors facilitating it, TBP and TFB/TFIIB, are conserved between the archaeal RNAP and eukaryotic RNAPII transcription systems. The archaeo-eukaryotic recruitment cascade is in contrast to bacterial transcription, where both σ^{70}

and σ^{54} form holoenzymes with the RNAP that are able to recognise the promoter.

Is there more to the archaeal promoter than TATA and BRE?

The eukaryotic core promoter is composed of multiple promoter elements including TATA-box, BRE_u and BRE_d (upstream and downstream BRE), Inr (initiator element), and DPE (downstream promoter element) that occur in various combinations [28, 32]. None of these promoter elements are strictly conserved and on many promoters they are absent altogether. The TATA-box is the most abundant promoter core element and TATA-like sequences are likely to be present in nearly all yeast promoters [33]. Likewise the bacterial promoter uses a modular architecture with -35, -10, extended -10 and UP elements being present in various combinations but all recognised by RNAP subunits or σ factors [34]. Archaeal promoters seem composed of mainly two elements, the TATA-box that binds TBP, and BRE that binds the TFB core domain (corresponding to the eukaryotic BRE_u). In addition, a sequence bias surrounding the TSS with the sequence (-1)T-A/G-T(+2) has been coined Inr element and is likely to interact with the RNAP itself rather than transcription factors like TBP-associated factors (TAFs) in the RNAPII system [35]. Pioneering work from Wolfram Zillig's laboratory identified an AT-rich sequence upstream of the TSS important for the promoter activity [36]. Permanganate foot-printing experiments on the Sso and Mja OC have

revealed similar boundaries for the initially melted region (IMR) of the promoter DNA extending up to position -12 relative to the TSS [11, 37, 38]. The propensity of DNA to melt depends on the energy required to disrupt base pair hydrogen bonding and base stacking interactions. Using chimeric archaeal promoter constructs we have shown that the IMR can change the promoter strength dramatically and independently of the TATA and BRE motifs of the promoter. The IMR is an important determinant for the strength of TFE stimulation [11]. However, it does not preclude additional sequence-specific interactions of this region with the transcription machinery, particularly RNAP, TFB and TFE.

Recent next generation sequencing approaches have enabled the genome-wide mapping of TSSs and thereby the sequence determinants of archaeal promoters for a number of organisms [39-42]. Sequence alignments centred around the TSSs for *Sso* and *Tko* confirm that BRE and TATA are the two canonical promoter elements in archaea and that the IMR is generally AT-rich (Figure 3). Sequence bias around the TSS representing the *Inr* is strong in *Sso*, partly due to the majority of transcripts being leaderless and coinciding with the translational, ATG, start site. The AT bias of the IMR shows some variation between archaea, with some species, such as *Tko*, only showing a preference at the upstream edge of the transcription bubble (positions -10 and -11) and others, such as *Sso*, showing a bias across the entire IMR, which

is more prominent at the upstream edge. Although this reflects a tendency for AT over GC base composition rather than a specific sequence element it is reminiscent of the bacterial -10 element which is also AT-rich and also forms the upstream edge of the transcription bubble in the *E. coli* σ^{70} -OC [43, 44].

In conclusion, the TATA-box and BRE are critical archaeal promoter elements specifically recruiting the general transcription factors TBP and TFB. The IMR upstream of the TSS contributes to promoter strength likely by enabling efficient OC formation catalysed by TFE.

Topology of the archaeal TATA-TBP-TFB-RNAP pre-initiation complex

Due to the high conservation of all involved components the archaeal PIC is likely to be near-identical to the RNAPII PIC in structural terms. However, despite heroic efforts it has yet not been possible to crystallise any complete archaeal or eukaryotic PIC, while structural information of partial complexes and biochemical- and biophysical proximity analyses have enabled structural models of PICs [13, 30, 38, 45]. We have recently prepared a solution model of the Mja OC, shown in Figure 4. A wholly recombinant RNAP system enabled us to incorporate fluorescent dye pairs into a range of strategically chosen surface-exposed locations on RNAP, TBP, TFB and TFE, and in the

template- and non-template DNA strands of the strong SSV1 T6 promoter [46]. OCs were assembled on immobilised promoter templates, and Förster Resonance Energy Transfer (FRET) between fluorescent donor-acceptor pairs was monitored at the single molecule level (smFRET) [38]. From the FRET efficiencies we calculated the interprobe distances, which in turn by triangulation and prior knowledge of partial structures of PIC components (including the RNAP, RNAP-TFIIB and TATA-TBP-TFB) were used to prepare the first structural model of the archaeal OC [38]. This analysis showed on one hand a stunning similarity between archaeal and eukaryotic RNAPII PICs, in addition to revealing subtle but intriguing differences, most notably an altered position of the TATA/TBP/TFB core ensemble relative to the upstream RNAP surface. A direct comparison of the archaeal solution OC model based on smFRET with a eukaryotic OC model reveals that TBP and the TFB core domain are located closer to the upstream RNAP surface and leaning towards to the RNAP stalk module. Since the downstream promoter DNA is fixed between the jaws of the RNAP and the upstream TATA and BRE promoter elements are anchored to RNAP via TBP/TFB core this could induce a torsional strain in the promoter DNA that results in a spontaneous localised DNA melting followed by the loading of the template strand into the active site cleft. This process can occur in eukaryotic PICs but is very

inefficient without the aid of TFIIIE and in particular without the translocase activity of TFIIH [47].

Enhancing the CC to OC transition – molecular mechanism of TFE

Even though the combination of TBP and TFB enable OC formation in archaea without requiring additional proteins, a third general transcription initiation factor, transcription factor E (TFE) enhances this process. Many euryarchaeota including *Mja* utilise monomeric TFE α . However, the prototypical TFE α/β exemplified by *Sso* is a heterodimer homologous to eukaryotic TFIIIE α/β . In *Sso* the gene coding for TFE β is essential while deletion of the gene coding for TFE α has not been attempted [11]. In the euryarchaeon *Methanococcus maripaludis* the gene coding for monomeric TFE α is essential [48]. TFE α interacts with the RNAP in a bidentate fashion: The N-terminal extended winged helix (eWH) domain interacts with the tip of the RNAP clamp coiled-coil, while the C-terminal Zinc ribbon (ZR) domain interacts with the base of the clamp and the RNAP stalk [12] (Figure 4). This archaeal binding mode is in agreement with the location of yeast and human TFIIIE in their cognate PICs based on biochemical crosslinking patterns and electron microscopy structures [13, 14]. The contributions of the β -subunit to TFE function are less understood. Like TFE α , TFE β is also a bipartite protein consisting of an N-terminal classic winged-helix (WH) domain and a C-

terminal domain containing a structural and seemingly redox inactive cubane [4Fe-4S] cluster. The former domain seems not to be required for TFE activity *in vitro*, while the latter domain is essential for heterodimerisation with the α -subunit and recruitment to the RNAP [11]. The gene encoding TFE β is essential in *Sulfolobus acidocaldarius* [11]. The recruitment of TFE α to the Mja RNAP and the stimulation of transcription are dependent on the RNAP stalk [49, 50], which suggests that interactions between the TFE α ZR and the stalk are required for TFE function. On the other hand, both TFE α and TFE α/β form stable complexes with a recombinant Sso RNAP clamp in the absence of the stalk [11]. Moreover, deletion of the ZR domain in the context of the Sso TFE α/β factor does not abolish its function, which suggests that the β -subunit can compensate for the contribution of the TFE α ZR domain. In summary, TFE interacts with the RNAP clamp and stalk domains that have been implicated in OC formation.

The archaeal RNAP clamp exists in two states

OC formation requires significant rearrangements of the DNA template including DNA melting – a stepwise disruption of base pair interactions - and loading of the template strand into the active site. Once formed, the transcription bubble is prone to collapse and has to be stabilised to prevent this from happening. Securing the non-template (NT) strand on the outside

of the DNA binding channel of RNAP above the clamp provides one means to stabilise the OC [38]. Early crystal structures of eukaryotic RNAPII with and without the Rpo4/7 stalk [51-53] suggested that multisubunit RNAPs could adopt open and closed conformations. It appeared that in the presence of the stalk the RNAP clamp was closed over the DNA binding channel leading to a narrowed DNA cleft. More recent structures of archaeal RNAPs have shown the clamp in either open or closed conformations (Figure 5A), which indicates that the RNAP clamp conformation can change even in the presence of the stalk (Figure 5A). While the crenarchaeal Sso RNAP crystallised in a closed clamp conformation [54], the RNAP from the euryarchaeal organism *Thermococcus kodakarensis* exhibited an open clamp [55]. Movement of the clamp is accompanied by a shift in the position of the Rpo4/7 stalk, and the magnitude of the conformational change of the RNAP clamp (17 Å) is comparable to the changes deduced from comparisons of different eukaryotic RNAPII complexes (15 Å). However, these crystal structures are static snapshots of a flexible molecule and a key question in the field remained whether (i) alternative RNAP clamp states were relevant in solution, and (ii) whether they changed in response to recruitment of RNAP to the promoter, and (iii) during the transition between the closed and the OC. The combination of a biochemically tractable recombinant archaeal RNAP system with smFRET measurements allowed us to address these

questions [56]. We site-specifically introduced a fluorescent donor-acceptor dye (FRET) pair at the tip of the RNAP clamp coiled coil (subunit Rpo1) and into the lobe (subunit Rpo2) on the opposite side of the DNA cleft (Figure 5A). The double-labelled RNAPs were incorporated into DNA-TBP-TFB ternary complexes and changes in RNAP clamp conformation was assessed using the FRET efficiencies as proxy for inter-probe distances across the DNA binding channel (Figure 5B and C). Classification of the molecules according to their FRET efficiencies showed that the RNAP as part of the DNA-TBP-TFB-RNAP initiation complex indeed adopts two conformations, an open and a closed conformation similar to *E. coli* RNAP [57]. Notably, the conformation of the clamp differs from the conformations monitored for the RNAP not associated with transcription factors or DNA suggesting that the clamp adopts defined states in the initiation complex guided by the intricate network of interactions in the PIC.

TFE induces clamp opening

The recruitment of TFE into the PIC and the transition from the CC to OC occurs concurrently with the redistribution of the two conformational states of the clamp. In the CC, formed on double-stranded DNA, the clamp is preferentially in a closed state. Assembling the initiation complex on a synthetically pre-melted promoter template shifts the equilibrium towards an

open clamp state (Figure 5D), which suggests that DNA melting involves clamp opening. This process of clamp opening occurs spontaneously and does not require TFE, in agreement with the fact that TFE is not strictly required for transcription *in vitro*. However, the efficiency of clamp opening - and by inference template strand loading - is significantly stimulated in the presence of TFE α . While TFE α recruitment to RNAP depends on the presence of the Rpo4/7 stalk [12, 50, 58], the stalk itself does not influence the clamp conformation in the context of the PIC. These results suggest that the stimulatory effect of TFE on DNA melting and OC formation has an allosteric component: TFE binding to the RNAP leads to structural changes that result in the opening of the clamp. In the archaeal OC model register -12 of the NTS is juxtaposed to the RNAP clamp coiled-coil and the TFE α eWH domain [38]. There is a second component to TFE stimulation; TFE not only induces OC formation but helps to maintain the OC by securing the NTS at the upstream edge of the transcription bubble to the top of the clamp. This hypothesis is supported by cross-linking data and fluorescence quenching assays that showed a close proximity of the TFE α -eWH domain and the -12 NTS position [12, 59].

smFRET experiments show that changes in clamp conformation also occur during OC formation in the bacterial RNAP which does not utilise TFE-like factors [57]. Interestingly, the direction seems reversed in as much as the CC

has an open- and the OC a closed clamp, which could be due to the fact that different structural intermediates were captured. Once the archaeal RNAP has escaped the promoter and entered the processive elongation stage the clamp has closed again [56], and it is possible that the bacterial RNAP in the OC has proceeded one conformational step further than the archaeal RNAP. The structure of the human RNAPII PIC has recently been investigated using electron microscopy at intermediate resolution. Similar to the bacterial RNAP the CC to OC transition of the human RNAPII induces clamp closure [14]. The effect of TFIIE on the clamp conformation in RNAPII has not been directly tested. Similar to archaeal TFE, TFIIE can stimulate transcription *in vitro* from negatively supercoiled or partially pre-melted DNA templates [47, 60], suggesting that TFE and TFIIE use a conserved mechanism. Likewise, the clamp of RNAPIII adopts open and closed conformations [16]. However, the clamp movement is less pronounced (change of 9 Å) as compared to the archaeal RNAP and RNAPII (change of 17 Å). In the RNAPIII system, subunits C82/34 are an integral part of the RNAP and, similar to subunits TFIIE α/β (RNAPII), span the DNA cleft [13, 14, 16], which is likely to restrict the flexibility of the clamp.

Evolution of TFII E-like factors in archaea and eukaryotes

The identification of a bona fide homologue of TFE β in archaea allowed for reconstructing the evolutionary history of TFII E-like factors [11]. In terms of domain composition, the Sso TFE α - and β -subunits combine features of TFII E α and the human RNAPIII subunit hRPC39 (Figure 6A). Sso TFE α and TFII E α share the bipartite eWH [61] and ZR domain organisation. The eukaryote-specific C-terminus of TFII E α facilitates the recruitment of eukaryotic TFIIH to the PIC [62]. The Sso TFE β subunit is composed of one WH domain (WH) and one [4Fe-4S] cluster-containing domain that are homologous to RNAPIII subunit hRPC39 [11] (Figure 6A). The conservation of structural features suggests that TFII E and hRPC62/39 are derived from a common ancestor likely to be very similar to TFE α/β in archaea. Following duplication the TFE paralogues associated with RNAPII and RNAPIII transcription were reshaped by evolution (Figure 6B). TFII E β most likely retained its WH domain from its TFE β -like precursor, although no significant sequence homology can be detected between C34 (the yeast homologue of RPC39) and yeast Tfa2 [63]. The [4Fe-4S] cluster domain was eroded in TFII E β coinciding with the emergence of a new dimerisation interface. Several unicellular eukaryotes such as *Giardia lamblia* and *Leishmania major* appear to miss TFII E α and TFII E β homologues altogether, or the homologues are too divergent to be recognised as such [64, 65]. The C82 subunit of yeast

RNAPIII (homologous to hRPC62 in human RNAPII) diverged from its TFE α -like precursor by loss of the ZR domain and several duplications of the WH domain. In functional analogy to TFIIIE and TFE the C82/34 complex plays a role in OC formation of RNAPIII [66]. Furthermore, the C82/34 complex interacts with the RNAPIII-specific general transcription initiation factor Brf1, which possibly facilitated the partitioning of the transcription space of genes transcribed by RNAPIII [67, 68]. A common feature of TFIIIE β and RPC39 (Tfa2 and C34 in yeast) is the duplication of partially redundant WH domains. The WH domains in yeast TFIIIE form an array reaching over the DNA binding channel connecting the RNAP clamp and protrusion domains, and while deletion of one domain (Tfa2 WH1) has only a mild phenotype *in vivo* the deletion of both WH domains is lethal [13]. In this sense archaeal TFE mimics the minimal TFIIIE WH domain configuration. While the RNAPI system does not utilise any apparent TFIIIE homologue, the C-terminal tandem WH domains of subunit A49 might function similar to the tandem WH domains of Tfa2 and C34 [69]. The loss of [4Fe-4S] and ZR domains during evolution is paralleled in the evolution of TFE in archaea (Figure 7B). Both TFE α and TFE β are widely distributed within the archaeal phylum suggesting that they were present in the last common ancestor of archaea (and eukaryotes). However, several archaeal genomes lack recognisable TFE β genes, which might be the result of 'streamlining'. Members of the class Thermoplasmata lack both

TFE α and TFE β altogether. Halobacterial TFE β lack the conserved cysteine residues that are required for coordination of the [4Fe-4S] cluster indicating that the cluster has been lost [70]. While TFE α factors from species lacking a TFE β homologues, such as Mja and Pfu are fully functional for OC formation [12, 29, 49], the dimeric TFE of *Sulfolobus* is critically dependent on the TFE β subunit for its function [11].

In summary, TFIIIE-like factors show considerable variation in terms of domain composition, and are broadly phylogenetically distributed. Evolution has brought about an expansion of winged helix (WH) domains in eukaryotes, but also resulted in the loss of metal centres such as Iron sulphur clusters and Zinc ribbons.

Regulation of open complex formation

In bacteria the role of OC formation in the regulation of transcription is well documented (Figure 7). The most well characterised global regulation system that targets OC formation is the stringent response. Following amino acid starvation a subset of genes under the control of 'stringent' promoters are efficiently repressed by a destabilisation of the OC [71]. This mechanism relies on a sequence element, the discriminator, residing in the IMR of the promoter. Repression is mediated by the RNAP-associated regulator DksA and the guanosine nucleotide analogue ppGpp, which is synthesised by RelA

bound to stalled ribosomes in response to low aminoacylated tRNAs [71]. The binding site of ppGpp on *E. coli* RNAP is between two rigid modules of bacterial RNAP, the shelf and the core. Therefore binding of ppGpp is thought to restrict conformational changes in the RNAP [72, 73]. ppGpp binding reduces the stability of the OC, which collapses to the closed state and thus represses transcription from promoters with short-lived OCs such as the rRNA promoter in *E. coli* [74]. Another important example of OC limited transcription regulation is the σ^{54} transcription system. The σ^{54} -holo RNAP readily forms a CC, while OC formation requires the action of bEBPs and ATP-hydrolysis [7]. Class II transcription activators such as catabolite activator protein (CAP) also regulate transcription by enhancing OC formation [75].

The potential of OC formation as a regulator of gene expression has recently been emphasised by structural studies of the regulator CarD. The structure of the *Thermus thermophilus* RNAP-CarD initiation complex (Figure 9) shows that CarD stabilises the OC by interacting with the upstream edge of the transcription bubble [76] providing a mechanistic rationale for its role in activating transcription in *Mycobacterium* [77, 78]. This interaction is partially sequence specific due to the intercalation of a highly conserved tryptophan residue into the NT strand at position -12 [76]. The stabilisation of the OC through interaction with the upstream edge of the transcription bubble is reminiscent of TFE-like factors, whose eWH domains are interacting with the -

12 register of the NTS (Figure 8). CarD shows a wide but patchy phylogenetic distribution in sequenced bacterial species; it is present in *Mycobacterium* and *Thermus*, but absent in others such as *E. coli*. Interestingly, the OC formed by *Mycobacterium bovis* in absence of CarD is considerably less stable as compared to the *E. coli* OC (which does not utilise CarD) [77]. The incorporation of CarD into the *Mycobacterium* OC compensates for the lower stability. This suggests some degree of coevolution between the proteins forming the OC (RNAP and sigma factors) and CarD. *Mycobacterium smegmatis* CarD is induced by oxidative stress, DNA damage and starvation which altogether suggests that CarD is a regulator [79]. However, the whole genome occupancy of CarD implies that its bound to the majority of promoters [78] and should be considered a general transcription initiation factor [76].

The classical means of transcription regulation in eukaryotes include (i) improving the access of regulatory factors to their cognate DNA elements by chromatin remodelling, (ii) the cascade that results in the recruitment of RNAP to the promoter, and (iii) promoter-proximal pausing of early transcription elongation complexes. Recently, regulation of OC formation has emerged as a novel mechanism of global transcription activation in naïve lymphocytes [80]. Activation of naïve lymphocytes results in a massive

increase in global mRNA levels. Genome-wide mapping of single stranded DNA regions in resting lymphocytes revealed that while the bulk of RNAPII is promoter-bound, nearly all promoters rest in the closed state. This correlates with very low expression levels of TFIIH subunits including the translocase XPB. Upon lymphocyte activation TFIIH expression is induced, resulting in opening of RNAPII promoters, and global mRNA levels are amplified a hundredfold. In summary, OC formation as a means to regulate transcriptional output is used in both bacteria and eukaryotes.

Regulation of transcription in archaea is less well characterised than in bacteria or eukaryotes. Typically repression is achieved by promoter occlusion. Transcription activation is facilitated by enhancing the recruitment of TBP and TFB (reviewed in [81]). Recent findings suggest that similar to CarD, TFE could play a pivotal role in transcription regulation. Unlike TBP and TFB, TFE is non-essential for transcription initiation *in vitro*, therefore it could provide an opportunity for fine-tuning transcription *in vivo*. The expression levels of both Sso TFE subunits vary as a function of the growth rate and in response to stresses, unlike the other general transcription factors. The levels of TFE β decrease dramatically during oxidative stress and starvation [11], while TFE α is depleted upon heat shock; The steady-state levels of TBP, TFB, and RNAP remain unchanged under all conditions [82]. The amplitude of the TFE stimulation is dependent on the promoter context,

in particular on the sequence of the IMR [11]. As a consequence, the TFE depletion during stress is likely to affect distinct subsets of promoters in different ways. Finally, TFE could moderate the promoter activity by interacting differentially with alternative TFB paralogs since it has been shown that Pfu TFE stimulates transcription facilitated by the two paralogs TFB1 and TFB2 on different promoters to different extents [83].

Conclusions and outlook

Transcription initiation by the archaeal RNAP is closely related to that of eukaryotic RNAPII, in terms of promoter elements (BRE, TATA, IMR and Inr) and general transcription initiation factors (TBP, TFB and TFE), and it uses conserved molecular mechanisms during OC formation. However, there are interesting differences. OC formation is stimulated by TFE-like factors in both archaea and eukaryotes, but has become dependent on TFIIF and energy expenditure in the latter's RNAPII system. The reason for this remains opaque, but good arguments can be made that dependency of OC formation on exogenous factors provides a means for regulation of gene expression. There are multiple examples from bacteria, eukaryotes, and now also archaea that the transition between the CC and OC is a potent angle of attack for regulators including σ^{54} -bEBPs, ppGpp, TFIIF and TFE.

The current picture of transcription in archaea is relatively simple, which at least partly stems from our ignorance since most biochemical and structural

information is derived from a few exclusively (hyper-)thermophilic species. The recent discovery of TFE β and the RNAP subunit Rpo13 highlight the diversity in archaeal transcription that has yet to be explored. And there are still many open questions: Are TFE, TBP and TFB really general transcription factors in the sense that they are bound to and required for initiation on all promoters? Thus far only a single limited genome-wide occupancy study for TBP and TFB has been reported [84], whereas nothing is known about the distribution of TFE, and most importantly RNAP. Have we exhausted the general transcription factors constituting the archaeal PIC? Is it not possible that additional factors e.g. enable transcription of subsets of genes such as non-coding RNA genes like CRISPR or ribosomal RNA operons? TBP and TFB were originally identified from cell lysate and identified as essential transcription factors [85], while the identification of the two TFE genes was based on bioinformatics and only subsequently proven by experimental approaches. Experiments attempting the isolation of proteins associated with the archaeal PIC or with the RNAP from archaeal extracts may yet lead to the identification of novel archaea-specific transcription initiation factors.

Acknowledgements

We would like to thank all members of the RNAP laboratory at UCL - in particular Thomas Fouqueau - for inspiring discussions about transcription,

and Tine Arnvig for critical reading of the manuscript. Research in the RNAP laboratory at University College London is funded by Wellcome Trust Investigator Award 079351/Z/06/Z and BBSRC Grant BB/E008232/1 (to F.W.). F.B. acknowledges support from the German Research Foundation DFG (Research fellowship BL 1189/1-1). Research in the Grohmann laboratory is funded by the Fond der Chemischen Industrie, the Boehringer Ingelheim Fonds, and the DFG (Grants GR 3840/2-1 and SFB960).

References

- [1] Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*. 1977;74:5088-90.
- [2] Guy L, Ettema TJ. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol*. 2011;19:580-7.
- [3] Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*. 2013;504:231-6.
- [4] Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*. 2015;521:173-9.
- [5] Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci U S A*. 2015;112:6670-5.
- [6] Gruber TM, Gross CA. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol*. 2003;57:441-66.
- [7] Yang Y, Darbari VC, Zhang N, Lu D, Glyde R, Wang YP, et al. TRANSCRIPTION. Structures of the RNA polymerase-sigma54 reveal new and conserved regulatory strategies. *Science*. 2015;349:882-5.
- [8] Ghosh T, Bose D, Zhang XD. Mechanisms for activating bacterial RNA polymerase. *FEMS microbiology reviews*. 2010;34:611-27.
- [9] Buck M, Gallegos MT, Studholme DJ, Guo YL, Gralla JD. The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. *Journal of Bacteriology*. 2000;182:4129-36.
- [10] Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*. 2012;45:439-46.

- [11] Blombach F, Salvadori E, Fouqueau T, Yan J, Reimann J, Sheppard C, et al. Archaeal TFE α /beta is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *Elife*. 2015;4:e08378.
- [12] Grohmann D, Nagy J, Chakraborty A, Klose D, Fielden D, Ebright RH, et al. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell*. 2011;43:263-74.
- [13] Grünberg S, Warfield L, Hahn S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol*. 2012;19:788-96.
- [14] He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. *Nature*. 2013;495:481-6.
- [15] Grunberg S, Hahn S. Structural insights into transcription initiation by RNA polymerase II. *Trends in biochemical sciences*. 2013;38:603-11.
- [16] Hoffmann NA, Jakobi AJ, Moreno-Morcillo M, Glatt S, Kosinski J, Hagen WJ, et al. Molecular structures of unbound and transcribing RNA polymerase III. *Nature*. 2015;528:231-6.
- [17] Lefèvre S, Dumay-Odelot H, El-Ayoubi L, Budd A, Legrand P, Pinaud N, et al. Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation. *Nat Struct Mol Biol*. 2011;18:352-8.
- [18] Hirtreiter A, Grohmann D, Werner F. Molecular mechanisms of RNA polymerase--the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic acids research*. 2010;38:585-96.
- [19] Koonin EV, Makarova KS, Elkins JG. Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biol Direct*. 2007;2:38.
- [20] Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, et al. Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS biology*. 2009;7:e102.
- [21] Wojtas MN, Moggi M, Millet O, Bell SD, Abrescia NG. Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic Acids Res*. 2012;40:9941-52.
- [22] Parvin JD, Sharp PA. DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell*. 1993;73:533-40.
- [23] Qureshi SA, Bell SD, Jackson SP. Factor requirements for transcription in the Archaeon *Sulfolobus shibatae*. *EMBO J*. 1997;16:2927-36.
- [24] Buratowski S, Hahn S, Guarente L, Sharp PA. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*. 1989;56:549-61.
- [25] Littlefield O, Korkhin Y, Sigler PB. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A*. 1999;96:13668-73.

- [26] Gietl A, Holzmeister P, Blombach F, Schulz S, von Voithenberg LV, Lamb DC, et al. Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res.* 2014;42:6219-31.
- [27] Bell SD, Kosa PL, Sigler PB, Jackson SP. Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A.* 1999;96:13662-7.
- [28] Kadonaga JT. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol.* 2012;1:40-51.
- [29] Werner F, Weinzierl RO. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol Cell Biol.* 2005;25:8344-55.
- [30] Kostrewa D, Zeller M, Armache K, Seizl M, Leike K, Thomm M, et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature.* 2009;462:323-30.
- [31] Wiesler SC, Weinzierl RO. The linker domain of basal transcription factor TFIIB controls distinct recruitment and transcription stimulation functions. *Nucleic Acids Res.* 2011;39:464-74.
- [32] Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol.* 2008;20:253-9.
- [33] Rhee HS, Pugh BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature.* 2012;483:295-301.
- [34] Decker KB, Hinton DM. Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annu Rev Microbiol.* 2013;67:113-39.
- [35] Ao X, Li Y, Wang F, Feng M, Lin Y, Zhao S, et al. The *Sulfolobus* initiator element is an important contributor to promoter strength. *J Bacteriol.* 2013;195:5216-22.
- [36] Hain J, Reiter WD, Hudepohl U, Zillig W. Elements of an archaeal promoter defined by mutational analysis. *Nucleic Acids Res.* 1992;20:5423-8.
- [37] Bell SD, Jaxel C, Nadal M, Kosa PF, Jackson SP. Temperature, template topology, and factor requirements of archaeal transcription. *Proc Natl Acad Sci U S A.* 1998;95:15218-22.
- [38] Nagy J, Grohmann D, Cheung AC, Schulz S, Smollett K, Werner F, et al. Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nature communications.* 2015;6:6161.
- [39] Jager D, Forstner KU, Sharma CM, Santangelo TJ, Reeve JN. Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics.* 2014;15:684.
- [40] Jager D, Sharma CM, Thomsen J, Ehlers C, Vogel J, Schmitz RA. Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc Natl Acad Sci U S A.* 2009;106:21878-82.
- [41] Li J, Qi L, Guo Y, Yue L, Li Y, Ge W, et al. Global mapping transcriptional start sites revealed both transcriptional and post-transcriptional regulation of

- cold adaptation in the methanogenic archaeon *Methanolobus psychrophilus*. *Sci Rep*. 2015;5:9209.
- [42] Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R. A single-base resolution map of an archaeal transcriptome. *Genome research*. 2010;20:133-41.
- [43] Sasse-Dwight S, Gralla JD. KMnO₄ as a probe for lac promoter DNA melting and mechanism in vivo. *J Biol Chem*. 1989;264:8074-81.
- [44] Zuo YH, Steitz TA. Crystal Structures of the *E. coli* Transcription Initiation Complexes with a Complete Bubble. *Molecular Cell*. 2015;58:534-40.
- [45] Murakami K, Tsai KL, Kalisman N, Bushnell DA, Asturias FJ, Kornberg RD. Structure of an RNA polymerase II preinitiation complex. *Proc Natl Acad Sci U S A*. 2015;112:13543-8.
- [46] Schulz S, Kramm K, Werner F, Grohmann D. Fluorescently labeled recombinant RNAP system to probe archaeal transcription initiation. *Methods*. 2015.
- [47] Holstege FC, van der Vliet PC, Timmers H. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO journal*. 1996;15:1666.
- [48] Sarmiento F, Mrazek J, Whitman WB. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A*. 2013;110:4726-31.
- [49] 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*. 2007;282:11047-57.
- [50] Ouhammouch M, Werner F, Weinzierl RO, Geiduschek EP. A fully recombinant system for activator-dependent archaeal transcription. *J Biol Chem*. 2004;279:51719-21.
- [51] Armache KJ, Kettenberger H, Cramer P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100:6964-8.
- [52] Bushnell DA, Kornberg RD. Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A*. 2003;100:6969-73.
- [53] Cramer P, Bushnell DA, Kornberg RD. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science*. 2001;292:1863-76.
- [54] Hirata A, Klein BJ, Murakami KS. The X-ray crystal structure of RNA polymerase from Archaea. *Nature*. 2008;451:851-4.
- [55] Jun SH, Hirata A, Kanai T, Santangelo TJ, Imanaka T, Murakami KS. The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nature communications*. 2014;5:5132.

- [56] Schulz S, Gietl A, Smollett K, Tinnefeld P, Werner F, Grohmann D. TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc Natl Acad Sci U S A*. 2016;in press.
- [57] Chakraborty A, Wang DY, Ebright YW, Korlann Y, Kortkhonjia E, Kim T, et al. Opening and Closing of the Bacterial RNA Polymerase Clamp. *Science*. 2012;337:591-5.
- [58] Hirata A, Kanai T, Santangelo TJ, Tajiri M, Manabe K, Reeve JN, et al. Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol Microbiol*. 2008;70:623-33.
- [59] Grünberg S, Bartlett MS, Naji S, Thomm M. Transcription factor E is a part of transcription elongation complexes. *J Biol Chem*. 2007;282:35482-90.
- [60] Holstege FC, Tantin D, Carey M, van der Vliet PC, Timmers HT. The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO journal*. 1995;14:810-9.
- [61] Meinhardt A, Blobel J, Cramer P. An extended winged helix domain in general transcription factor E/IIE alpha. *J Biol Chem*. 2003;278:48267-74.
- [62] Ohkuma Y, Hashimoto S, Wang CK, Horikoshi M, Roeder RG. Analysis of the Role of Tfiie in Basal Transcription and Tfiih-Mediated Carboxy-Terminal Domain Phosphorylation through Structure-Function Studies of Tfiie-Alpha. *Molecular and Cellular Biology*. 1995;15:4856-66.
- [63] Carter R, Drouin G. The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol Biol Evol*. 2010;27:1035-43.
- [64] Best AA, Morrison HG, McArthur AG, Sogin ML, Olsen GJ. Evolution of eukaryotic transcription: Insights from the genome of *Giardia lamblia*. *Genome research*. 2004;14:1537-47.
- [65] Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science*. 2005;309:436-42.
- [66] Brun I, Sentenac A, Werner M. Dual role of the C34 subunit of RNA polymerase III in transcription initiation. *EMBO J*. 1997;16:5730-41.
- [67] Werner M, Chaussivert N, Willis IM, Sentenac A. Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. *J Biol Chem*. 1993;268:20721-4.
- [68] Khoo B, Brophy B, Jackson SP. Conserved functional domains of the RNA polymerase III general transcription factor BRF. *Genes Dev*. 1994;8:2879-90.
- [69] Geiger SR, Lorenzen K, Schrieck A, Hanecker P, Kostrewa D, Heck AJ, et al. RNA polymerase I contains a TFIIIF-related DNA-binding subcomplex. *Mol Cell*. 2010;39:583-94.
- [70] Blombach F, Makarova KS, Marrero J, Siebers B, Koonin EV, van der Oost J. Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA

- polymerases for coding and non-coding RNAs in Archaea. *Biol Direct*. 2009;4:39.
- [71] Potrykus K, Cashel M. (p)ppGpp: still magical? *Annu Rev Microbiol*. 2008;62:35-51.
- [72] Ross W, Vrentas CE, Sanchez-Vazquez P, Gaal T, Gourse RL. The magic spot: a ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. *Mol Cell*. 2013;50:420-9.
- [73] Zuo Y, Wang Y, Steitz TA. The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Mol Cell*. 2013;50:430-6.
- [74] Barker MM, Gaal T, Josaitis CA, Gourse RL. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *J Mol Biol*. 2001;305:673-88.
- [75] Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). *Journal of Molecular Biology*. 1999;293:199-213.
- [76] Bae B, Chen J, Davis E, Leon K, Darst SA, Campbell EA. CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *eLife*2015.
- [77] Davis E, Chen J, Leon K, Darst SA, Campbell EA. Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res*2015. p. 433-45.
- [78] Srivastava DB, Leon K, Osmundson J, Garner AL, Weiss LA, Westblade LF, et al. Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci U S A*. 2013;110:12619-24.
- [79] Stallings CL, Stephanou NC, Chu L, Hochschild A, Nickels BE, Glickman MS. CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. *Cell*2009. p. 146-59.
- [80] Kouzine F, Wojtowicz D, Yamane A, Resch W, Kieffer-Kwon KR, Bandle R, et al. Global regulation of promoter melting in naive lymphocytes. *Cell*. 2013;153:988-99.
- [81] Peeters E, Driessen RP, Werner F, Dame RT. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol*. 2015;13:333-41.
- [82] Iqbal J, Qureshi SA. Selective Depletion of *Sulfolobus solfataricus* Transcription Factor E under Heat Shock Conditions. *Journal of Bacteriology*. 2010;192:2887-91.
- [83] Micorescu M, Grunberg S, Franke A, Cramer P, Thomm M, Bartlett M. Archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus*. *J Bacteriol*. 2008;190:157-67.
- [84] Facciotti MT, Reiss DJ, Pan M, Kaur A, Vuthoori M, Bonneau R, et al. General transcription factor specified global gene regulation in archaea. *Proc Natl Acad Sci U S A*. 2007;104:4630-5.

- [85] Hausner W, Wettach J, Hethke C, Thomm M. Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *J Biol Chem.* 1996;271:30144-8.
- [86] Andrecka J, Treutlein B, Arcusa MA, Muschielok A, Lewis R, Cheung AC, et al. Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic acids research.* 2009;37:5803-9.

Figures

Figure 1. Evolution of the transcription initiation machinery after LUCA.

The universally conserved core RNAP and the transcription elongation factor Spt5 are the only components of the transcription machinery that predate the last universal common ancestor (LUCA) of bacteria, archaea, and eukaryotes. The general transcription factors required for transcription initiation emerged later independently in bacteria and archaea, the two primary domains of life. The emergence of eukaryotes from their "archaeal parent" led to the evolution of additional general transcription factors and those belonging to the RNAPII system are listed.

Figure 2. Transcription initiation in archaea is a recruitment cascade.

Sequential or concomitant binding of TBP and TFB to the TATA-box and BRE nucleates the formation of the ternary complex. RNAP is recruited to this platform to form the preinitiation complex (PIC) in the closed form (closed complex, CC). The third factor TFE binds to the CC and assists conformational changes that facilitate DNA melting resulting in the open complex (OC).

Figure 3. The archaeal core promoter structure. Alignment of the DNA sequences upstream of TSSs identified individual promoter elements including BRE, the TATA box, the initially melted region (IMR) and the initiator (Inr) surrounding the TSS (+1). The strong Inr signal in *Sso* is due to the fact that the ATG start codon on most genes coincides with the TSS. Alignment of TSS identified by whole transcriptome sequencing from *S. solfataricus* [42] and *T. kodakarensis* [39]. The inserts show the TATA box motifs identified by the program MEME (<http://meme-suite.org>) in the same dataset. Alignment was performed using WebLogo3 (<http://weblogo.threeplusone.com>) adjusting to the background GC content for each organism.

Figure 4. Architecture of the complete archaeal open complex. The OC model encompasses the three archaeal general transcription factors TBP (green), TFB (blue) and TFE (magenta), the RNAP (grey) and TS (dark blue) and NTS (cyan). The overall topology of the archaeal OC is very similar to the human OC structure determined by electron microscopy [14]. The relative orientation of the TFE α eWH domain is somewhat uncertain. This model is based on distance constraints derived from smFRET measurements between fluorescent dye pairs introduced at strategic locations in components of the OC [38]. Interprobe distances were calculated from smFRET measurements

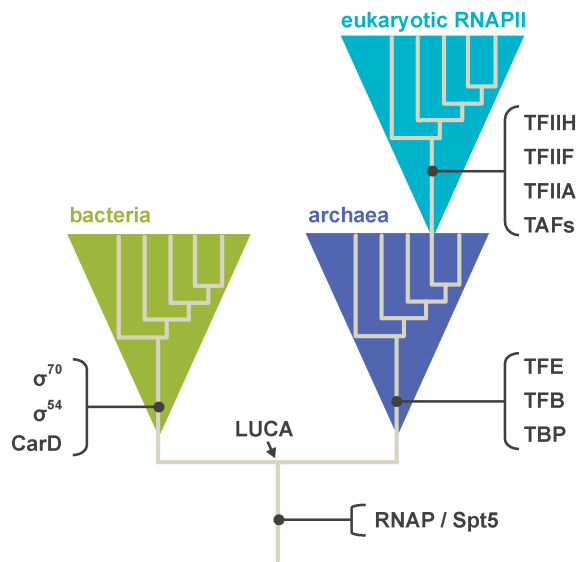
and processed using the NPS system [86]. Structural parents to the model included the archaeal Sso RNAP (pdb: 2WAQ), ternary complex DNA-TBP-TFB_{core} from *Pyrococcus woesei* (pdb: 1D3U), Sso TFE α eWH domain (pdb: 1Q1H), and yeast RNAPII-TFIIB (pdb: 4BBR) and human TFIIE α ZR domain (pdb: 1VD4).

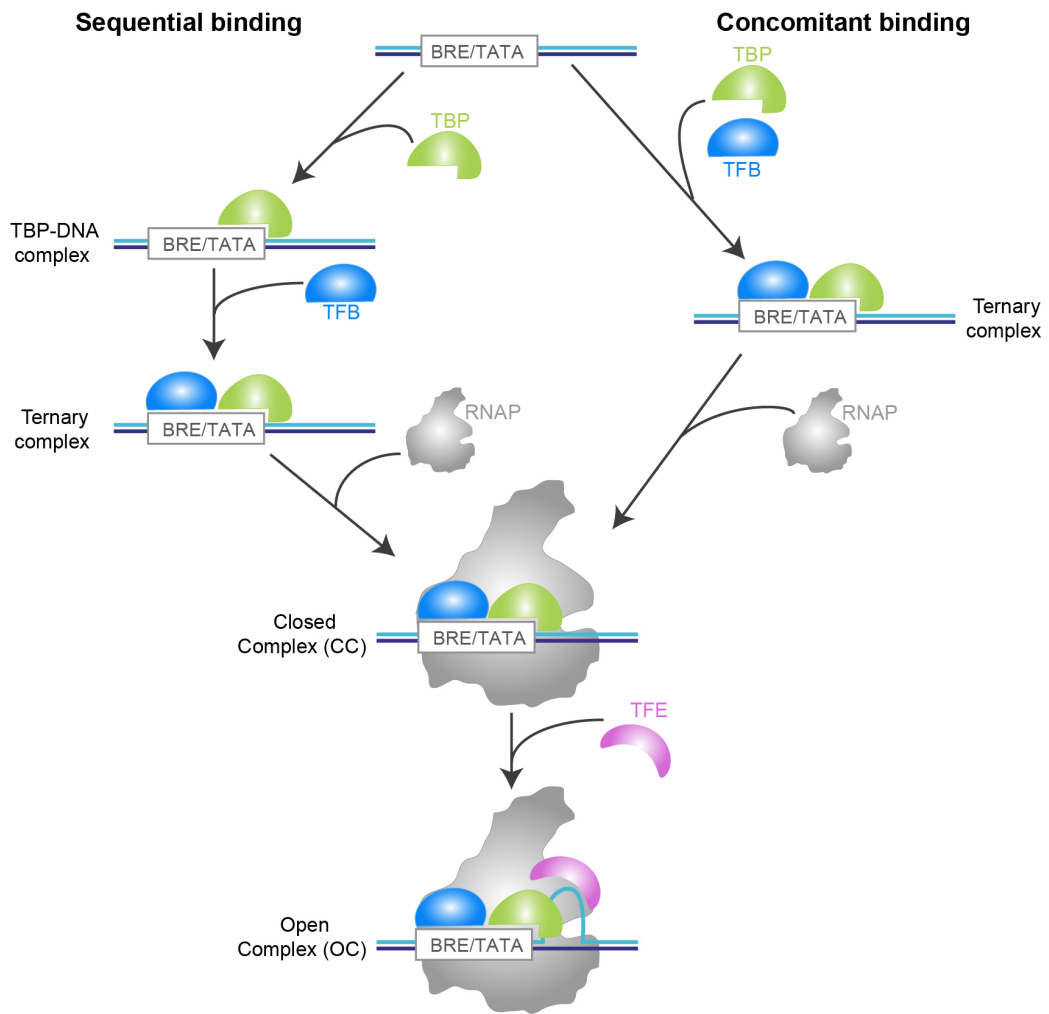
Figure 5. Conformation of the RNAP clamp in archaea. (A) Structural alignment of the crenarchaeal (*S. solfataricus*, pdb: 2PMZ) and euryarchaeal RNAP (*T. kodakarensis*, pdb: 4QIW) that adopt a closed or open clamp, respectively. (B) Single-molecule FRET measurements on immobilised initiation complexes assembled in the presence of a donor-acceptor-labelled RNAP inform about the conformation of the archaeal RNAP clamp. (C) FRET serves as molecular ruler with high sensitivity in the nanometer-range providing information about the width of the DNA cleft and the conformation of the clamp. (D) Opening and closing of the archaeal RNAP clamp during open formation as revealed by smFRET (colour coding as in A and B; non-template strand in cyan, template strand in blue) [56]. TFE binding and OC formation stimulate opening of the RNAP clamp.

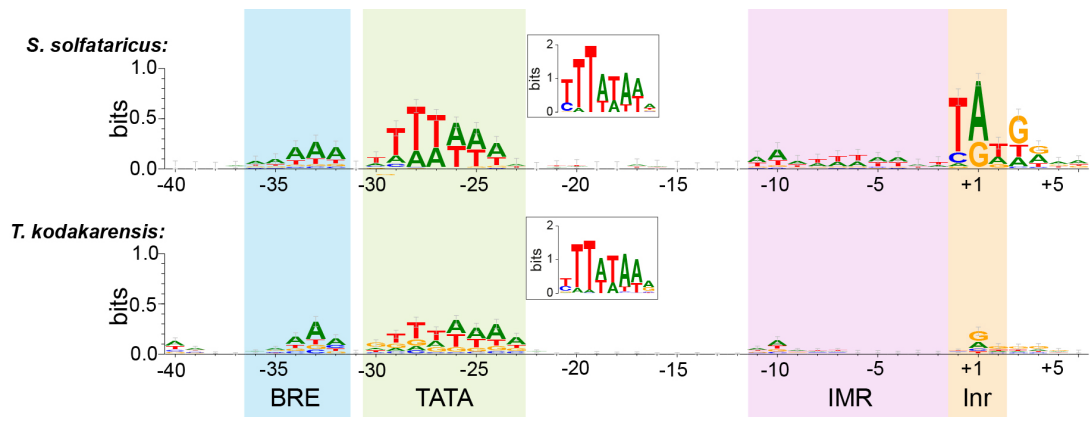
Figure 6 Diversity of TFIIIE-like factors in archaea and eukaryotes. (A) The domain composition of dimeric *Sulfolobus* TFE combines features of TFIIIE and RNAPIII subunits RPC62/39. The additional eWH and WH domains resulting from duplications are depicted in light blue and green, respectively. (B) The gene loss and loss of the [4Fe-4S] cluster and Zn-ribbon domains in TFIIIE-related factors is depicted on an updated archaeal phylogeny placing eukaryotes within the archaeal domain [5]. Different archaeal taxonomic groups belonging to the euryarchaeota or the 'TACK' superphylum [2] as well as the three classes of eukaryotic RNAP systems are included. In order to depict variation within the eukaryotic domain, *S. cerevisiae* (y) and human (h) counterparts were included separately. The prediction of conservation, or loss, of metal centres is based on the presence of the conserved cysteine residues required for coordination of Zn ions and [4Fe-4S] clusters.

Figure 7. OC formation as a mean to regulate transcription. Examples of factors and molecules regulating OC formation from different organisms are shown alongside the proposed mechanism of activation or repression (see text for discussion).

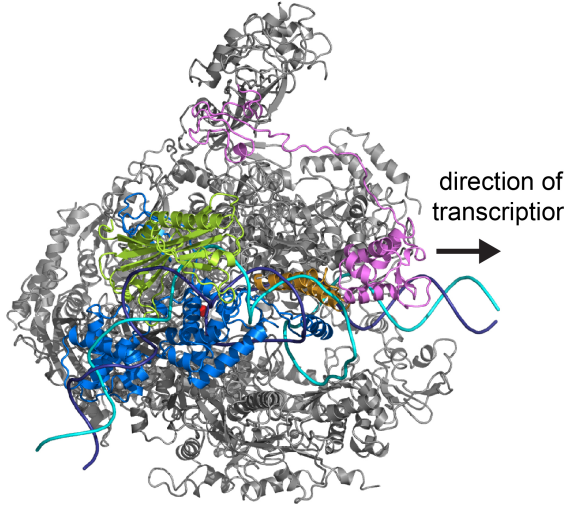
Figure 8. TFE-like factors and CarD both activate transcription by stabilising the OC. (A) Model of the *Methanocaldococcus* OC with TFE α shown in a magenta semitransparent surface representation. The TFE α eWH domain is perched on the tip of the RNAP clamp coiled coil (orange) in close contact with the NTS of the promoter. (B) Structure of the bacterial OC from *Thermus* with CarD shown in magenta semitransparent surface representation (pdb: 4XLR) [76]. A conserved tryptophan residue (Trp86) wedges into the minor groove of the upstream DNA thereby stabilising the transcription bubble (see close up).



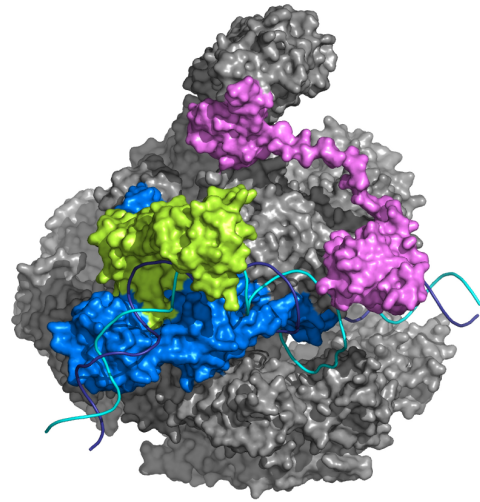




top view

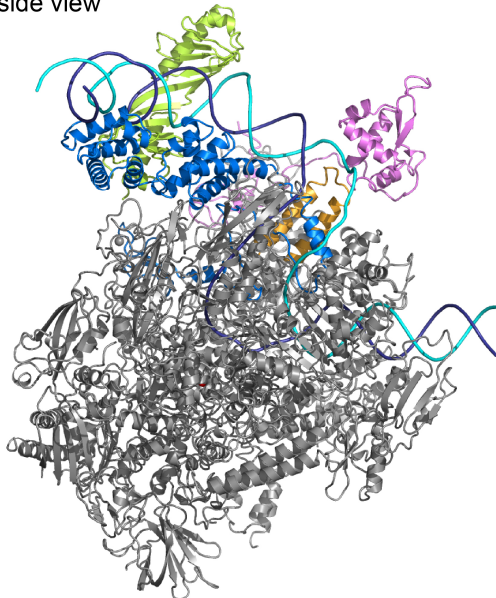


top view (surface representation)

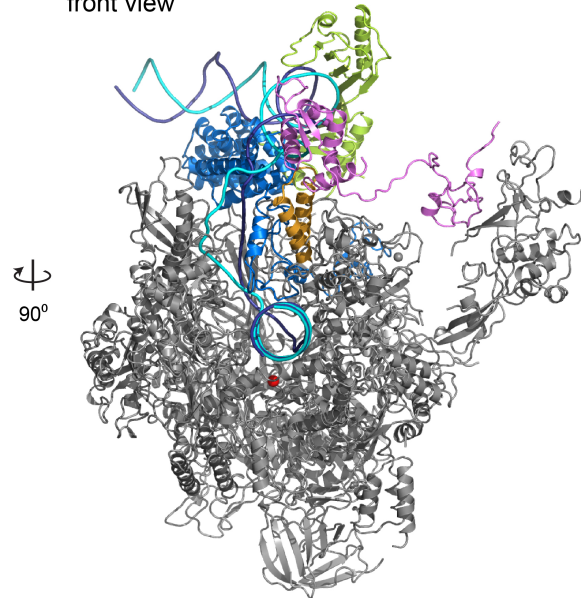


⊕
90°

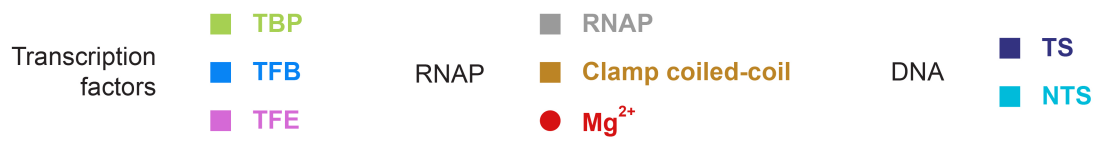
side view

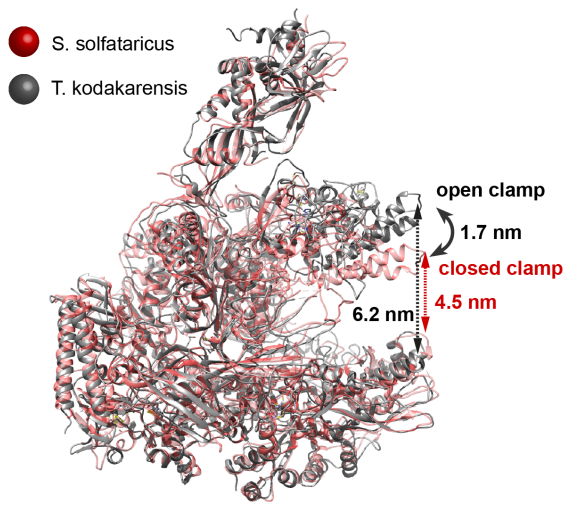
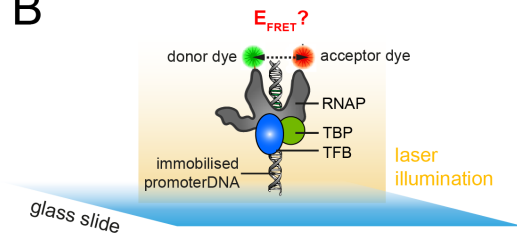
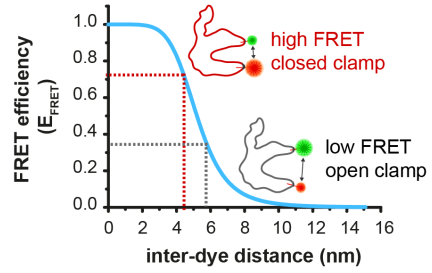


front view

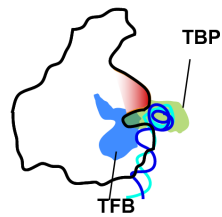


⊕
90°

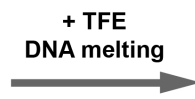


A**B****C****D**

Closed Complex
RNAP-dsDNA-TBP-TFB



closed clamp

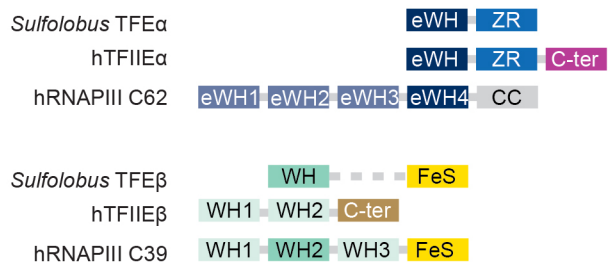


Open Complex
RNAP-open DNA-TBP-TFB-TFE

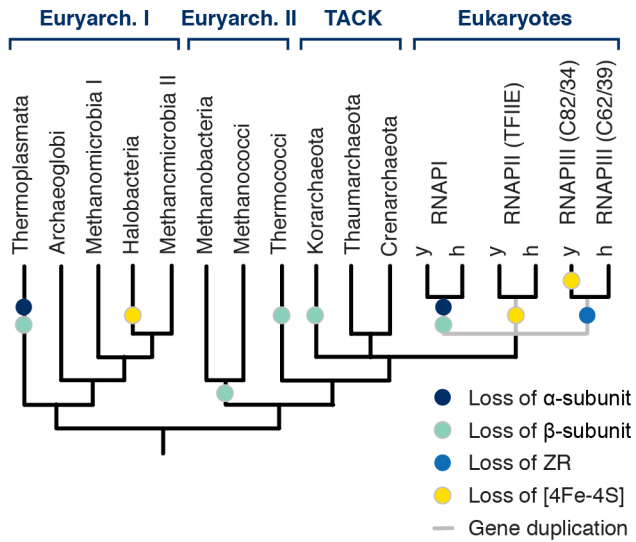


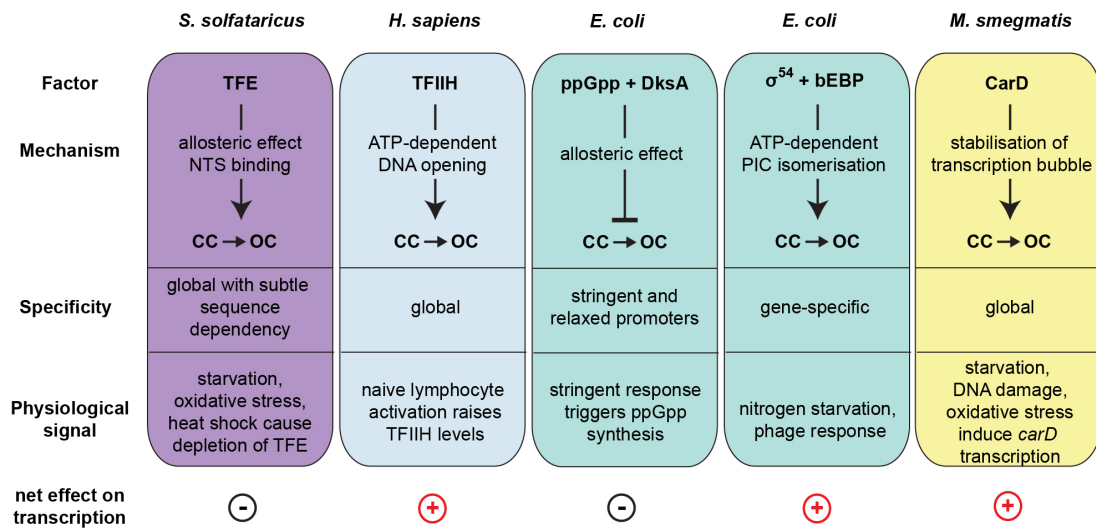
open clamp

A



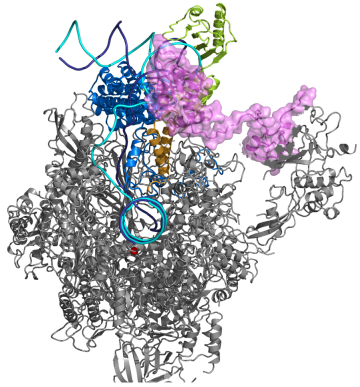
B



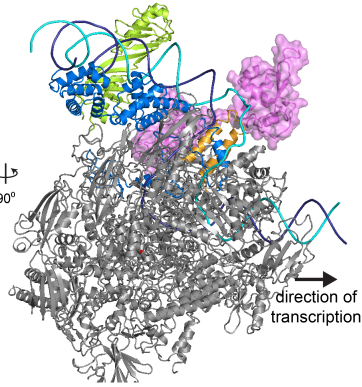


A *Methanocaldococcus* OC model with TFE

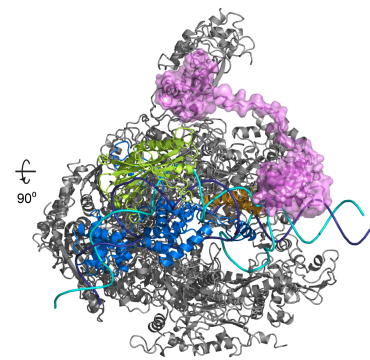
front view



side view

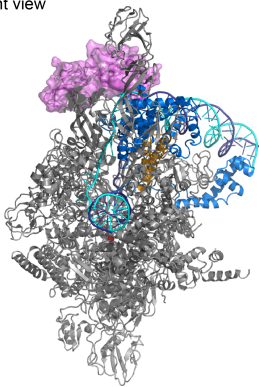


top view

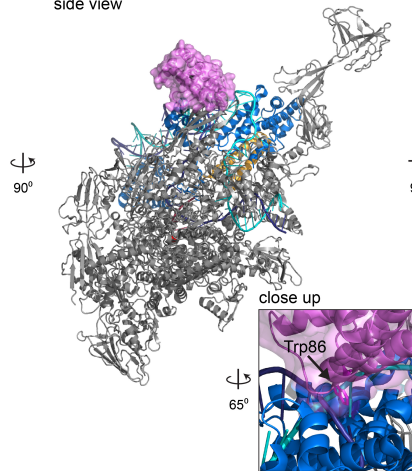


B *Thermus* OC with CarD

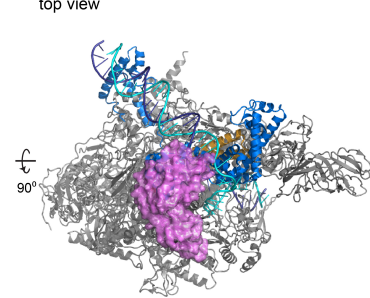
front view



side view



top view



Transcription factors
■ TBP
■ TFB and σ^A
■ TFE and CarD

RNAP
■ RNAP
■ Clamp coiled-coil
● Mg^{2+}

DNA/RNA
■ TS
■ NTS
■ RNA