Archaeal promoter architecture and mechanism of gene activation

Nan Peng†, Xiang Ao*, Yun Xiang Liang* and Qunxin She†1
†State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, 430070 Wuhan, Hubei, China, and †Archaea Centre, Department of Biology, University of Copenhagen, Biocenter, Ole Maaloes Vej 5, DK-2200 Copenhagen N, Denmark

Abstract
Sulfolobus solfataricus and Sulfolobus islandicus contain several genes exhibiting D-arabinose-inducible expression and these systems are ideal for studying mechanisms of archaeal gene expression. At sequence level, only two highly conserved cis elements are present on the promoters: a regulatory element named ara box directing arabinose-inducible expression and the basal promoter element TATA, serving as the binding site for the TATA-binding protein. Strikingly, these promoters possess a modular structure that allows an essentially inactive basal promoter to be strongly activated. The invoked mechanisms include TFB (transcription factor B) recruitment by the ara-box-binding factor to activate gene expression and modulation of TFB recruitment efficiency to yield differential gene expression.

General feature of archaeal gene transcription
The archaeal transcriptional machinery represents a simplified version of the eukaryal counterpart both in terms of composition and structure [1]. As for the eukaryal RNAP (RNA polymerase) II, the archaeal RNAP is a multiple subunit complex comprising 13 subunits in Sulfolobales [2], whereas bacterial RNAP only has five. At the structural level, all cellular RNAPs exhibit strong similarity at the catalytic domain, but archaeal and eukaryal RNAPs have a protruding stalk that is absent from the bacterial system [3]. In fact, all archaeal information-processing machineries, including replication and translation apparatus, resemble the eukaryal counterparts more than the bacterial machineries. However, being unicellular prokaryotes with compactly organized genomes, archaea have many genes clustered into operons as for bacteria. Consequently, there is a great interest in understanding the mechanisms of gene expression in archaea, since different strategies are employed to regulate gene expression in bacteria and in eukarya.

In transcription initiation, promoter signals upstream of a gene are recognized by transcriptional machinery to form a transcription PIC (pre-initiation complex). The PIC-formation process is mechanistically different between the archaeal/eukaryal machinery and the bacterial RNAP. Whereas the latter uses an intrinsic basic transcription factor σ for promoter recognition, an archaeal/eukaryal RNAP employs extrinsic basal transcriptional factors for doing the same. As a consequence, archaeal core promoters resemble those of eukaryal RNAP II genes, but differ from bacterial ones. Bacterial core promoters possess two cis DNA elements (−10 and −35 boxes), serving as binding sites for the σ subunit of the bacterial RNAP holoenzyme, whereas archaeal and eukaryal promoters contain sequence motifs that interact with the extrinsic basal transcriptional factors TBP (TATA-box-binding protein) and TFB (transcriptional factor B) site-specifically. The promoter element that binds to TBP, termed the TATA box motif, is centred at −26 to −28 relative to TSS (transcription start site) in an archaeal promoter and the BRE (TFB recognition element) is located immediately upstream [4]. Since transcription is mainly controlled at the stage of initiation, transcriptional factors primarily regulate gene expression by modulating interactions between the transcriptional machinery and gene promoters. Thus studies on archaeal gene expression mechanisms are also focused on identifying transcriptional factors and the cis-DNA elements with which they interact. One of the archaeal systems attained a detailed study is the regulated expression of the genes involved in arabinose metabolism in the thermophilic crenarchaeon Sulfolobus.

D-Arabinose-responsive expression involves a global regulation in Sulfolobus
Sulfolobus solfataricus exhibits great versatility in utilizing carbohydrates and encodes several distinct types of transporters for substrate uptake, including glucose and arabinose ABC (ATP-binding cassette) transporters [5,6]. It can utilize D-arabinose as the sole carbon and energy source, and this involves an operon and four other genes in S. solfataricus. The expression of these genes is arabinose-inducible [7,8]. The promoter of the operon is located upstream of the first gene in the operon araS encoding an arabinose-binding protein and this promoter directs arabinose-inducible expression from a viral vector [9].

Key words: activator, archaea, transcriptional regulation, transcription factor B-recognition element (BRE), transcription factor B recruitment (TFB recruitment), upstream activator sequence (UAS).
Abbreviations used: BRE, transcription factor B-recognition element; PIC, pre-initiation complex; RNAP, RNA polymerase; TBP, TATA-box-binding protein; TFB, transcription factor B; TSS, transcription start site.
†To whom correspondence should be addressed (email qunxin@bio.ku.dk).
Four D-arabinose-metabolizing genes were identified in \textit{S. solfataricus} from a combined functional genomics approach in addition to the arabinose-transporter genes [8]. These include Sso1300 coding for a D-arabinose 1-dehydrogenase (araDH); Sso3124 for a D-arabinonate dehydratase (araD); Sso3117 for 2-oxo-3-deoxy-D-arabinonate dehydratase (dopDH) and Sso3118, 2,5-dioxopentanoate dehydrogenase (kdaD). It has been proposed that these enzymes work in concert to yield 2-oxoglutarate by oxidizing D-arabinose, as has been shown in bacteria [8].

Interestingly, the \textit{Sulfolobus} D-arabinose-metabolizing genes are differently organized compared with the bacterial counterparts: the bacterial genes are clustered into an operon, whereas the archaeal ones (\textit{araDH}, \textit{araD}, \textit{kdaD} and \textit{dopDH}) are dispersed at three different locations on the chromosomes of \textit{S. solfataricus} P2 [10] and in six of the \textit{S. islandicus} genomes published recently [11]. As the result, these archaeal genes have to be regulated globally to yield arabinose-inducible expression. Because the \textit{ara}-box motif identified in \textit{S. solfataricus} [8] is conserved in all \textit{Sulfolobus} species that encode the D-arabinose pathway, it has been implicated in arabinose-inducible expression for all of these organisms.

However, \textit{in vivo} characterization of these \textit{cis}-regulatory DNA elements presented on these arabinose-inducible promoters requires a genetic system that can determine low promoter activities precisely. Fortunately, a versatile host–vector system developed for the crenarchaeal model organism \textit{Sulfolobus islandicus} REY15A isolated from Iceland (reviewed in [12]) meets the criterion because the genetic host of the system carries a deletion in the \textit{lacS} gene coding for a \(\beta\)-galactosidase and it does not retain any detectable \(\beta\)-galactosidase activity [13].

### Architecture of D-arabinose-responsive promoters in \textit{Sulfolobus} species

A gene reporter system was then developed for \textit{S. islandicus} REY15A in which the \textit{Sulfolobus–Escherichia coli} shuttle vector pHZ2 [13] was used as the backbone for constructing the gene reporter plasmid with \textit{lacS}. Then the system was employed to analyse the \textit{S. solfataricus} \textit{araS} promoter, and this revealed the minimal active promoter as a 59-bp-long DNA fragment extending from −55 to +4 relative to the TSS [14]. The minimal active promoter contains the core promoter plus the \textit{ara}-box motif with 5′-AACAGTT-3′ as the consensus sequence identified from functional genomics approaches [8]. The only identifiable core promoter element is the AT-rich motif (TATA box) serving as the binding site for TBP, whereas BRE showed strong variation from the \textit{Sulfolobus} BRE consensus revealed from a transcriptomic analysis [15]. Importantly, \textit{in vivo} analysis of the \textit{araS} promoter shows that promoters containing a non-conserved BRE can be very weak or inactive [14], and gene expression from such promoters has to be activated from an upstream activation sequence such as the \textit{ara}-box element.

Furthermore, scanning mutagenesis of the \textit{araS} basic promoter indicates that several dinucleotides positioned outside the conserved TATA and \textit{ara}-box elements are of crucial importance to promoter activity. The promoter elements identified include the well-conserved TATA-box and the \textit{ara}-box motif, a BRE variant element immediately upstream of TATA, a PPE (proximal promoter element) widely conserved in the promoters in \textit{Sulfolobus} genes [15] as well as an initiator element (Inr) overlapping with the TSS that has been demonstrated in a separate experiment (N. Peng and Q. She, unpublished work).

We then studied how conserved the arabinose metabolism is among all known \textit{Sulfolobus} species. It is well known that neither \textit{Sulfolobus acidocaldarius} nor \textit{Sulfolobus tokodaii} utilizes D-arabinose and they do not encode any arabinose transporter. Nevertheless, one or two enzymes implicated in the metabolism are present: \textit{S. tokodaii} encodes the D-arabinonate dehydratase and \textit{S. acidocaldarius} encodes 2,5-dioxopentanoate dehydrogenase and 2-oxo-3-deoxy-D-arabinonate dehydratase. On the other hand, six of the seven \textit{S. islandicus} strains published recently [11] encode a complete set of arabinose-transporter and -metabolizing enzymes. The remaining one, Y.G. 57.14, apparently does not have the D-arabinose pathway owing to the lack of an arabinose transporter. Nevertheless three of the four arabinose-metabolizing enzymes are present; only D-arabinose 1-dehydrogenase is lacking.

To investigate whether the architecture revealed for the \textit{araS} promoter is conserved for all arabinose-responsive promoters, we retrieved sequences of the promoter regions for all genes or operons identified in the aforementioned experiment and analysed them. Consistent with inability to utilize D-arabinose, the identified \textit{S. tokodaii} and \textit{S. acidocaldarius} genes do not show the same promoter architecture as for arabinose-inducible promoters. The \textit{ara}-box element is completely lacking in two of them; in the third (Saci_1938), a perfect \textit{ara}-box motif is present, but the spacing between the \textit{ara} box and the TATA element is increased by two nucleotides and the BRE motif matches the consensus, a promoter architecture that diminishes \textit{ara}-box activation. Taking all these into consideration, the Saci_1938 promoter could by no means exhibit arabinose-inducible expression. On the other hand, the \textit{S. islandicus} Y.G. 57.14 strain does not possess an arabinose transporter and nor is \textit{araDH} encoding arabinose-1-dehydrogenase present, whereas the remaining three arabinose-metabolizing genes are identical with those of the \textit{S. islandicus} strains that encode the entire arabinose metabolic pathway in both gene and promoter sequences. This suggests that \textit{S. islandicus} Y.G. 57.14 has lost the arabinose-oxidation capacity only recently. Thus these \textit{ara}-box-containing promoters are likely to direct the inducible expression in the \textit{S. islandicus} reporter gene system. The conservation of the identified arabinose-responsive promoters is summarized in Figure 1.

Compiling the arabinose-inducible promoters identified reveals striking features in promoter architecture: whereas their BREs are very diverged from the \textit{Sulfolobus} consensus, the distance remains constant (10 bp) between TATA and the \textit{ara} box on the promoters (Figure 1). The implications
of the modular promoter structure are two-fold. First, the constant distance between the regulatory element ara box and the basal promoter elements must reflect the constraint of the interaction between the putative activator and the transcriptional machinery. Secondly, sequence divergence between the ara box and TATA provides a mechanism for directing differential gene expression. There are in vivo results supporting this assumption. Several S. solfataricus araS mutant promoters containing substitutions in the non-conserved nucleotides on the promoter show different capabilities for driving reporter gene expression [14], which can reflect the modulation of functional interaction between the activator and transcriptional machinery. This fashion of modulation was also seen for the promoters of araDH, araD, kdaD and dopDH of S. islandicus (X. Ao and Q. She, unpublished work).

Taken together, the genetic determinants at a promoter for arabinose-inducible expression include a well-conserved ara-box element, a weak BRE variant, a strong TATA element as well as stringent spacing between the ara box and TATA, whereas their expression levels are to be determined by nucleotides positioned between them as well as other TFB/RNAP-interaction sites of the promoter (Figure 1). In this regard, characterizing more arabinose-inducible promoters present in the recently published S. islandicus genomes [11] yield insights into the DNA–protein and protein–protein interactions on these promoters.

TFB recruitment as the activation mechanism in Sulfolobus arabinose-inducible expression

As has been revealed from biochemical experiments, the first step in the formation of the PIC is for TBP to bind the TATA-box element in an archaeal promoter. Then TFB interacts with TBP and binds specifically to BRE to form a TBP–TFB–promoter complex, which in turn recruits the archaeal RNAP to yield PIC, and any defect of, or interference with, PIC formation will inactivate gene expression.

The essentially inactive nature of the araS basal promoter suggests that it does not support the formation of a stable TFB–TBP–promoter complex because the araS basal promoter lacks an operator-like motif. Furthermore, the araS TATA motif appears to be a strong element both in sequence conservation and in vivo activity. This leaves a weak BRE as the only possible reason to yield an almost inactive araS basal promoter and this has been demonstrated experimentally by site-directed mutagenesis and BRE swapping [14]. Therefore the apparent mechanism for arabinose-responsive activation in Sulfolobus is that an ara-box-binding factor recruits TFB to the weak BRE element to activate gene expression.

Further insight into the possible mechanisms of the gene activation is derived from studying interactions between archaeal TFBs and core promoters. In Pyrococcus furiosus, a hyperthermophilic euryarchaeon, TFBs binds to the entire basal promoter region of a glutamate dehydrogenase promoter and at several positions, both TFB and RNAP make strong footprints, including the −9 region relative to the TSS [16,17]. This is in good agreement with in vivo results where structurally sensitive positions (−7 and −8 relative to the TSS) are present in the S. solfataricus araS promoter. Taken together, the TFB and RNAP interactions between TATA and the TSS are very important for gene expression. It is interesting to see how widespread this manner of gene regulation is in Sulfolobus and other archaea. Another striking point is to investigate the minimal requirement for a functional TFB–promoter interaction to yield gene expression and how the interaction is modulated to yield differential gene expression.

Figure 1 | Conservation of the promoters of Sulfolobus arabinose-inducible genes

The promoter elements experimentally characterized in the S. solfataricus araS gene are marked with bars underneath the sequence logo. BREup, upstream BRE; BREdn, downstream BRE. Promoter sequences were retrieved from the GenBank® public databases using arabinose-inducible promoters encoded in S. islandicus REY15A as probes. Promoters used for this analysis include those of araS (arabinose-binding protein), araDH (o-arabinose 1-dehydrogenase), araD (o-arabinonate dehydratase), dopDH (2-oxo-3-deoxy-o-arabinonate dehydratase) and kdaD (2,5-dioxopentanoate dehydrogenase) and they were retrieved from the genomes of S. solfataricus P2 [10] and S. islandicus L.D.S.8, L.S.2.15, M.14.25, M16.4, M.16.27, Y.G.15.51 and Y.N.14.57 [11].

Figure 1: Sequence logos of the arabinose-inducible promoters. The sequence logos were generated using the logo.igv tool [14], which are then shown alongside a multiple sequence alignment of the promoters. The promoter elements experimentally characterized in the S. solfataricus araS gene are marked with bars underneath the sequence logo. BREup, upstream BRE; BREdn, downstream BRE. Promoter sequences were retrieved from the GenBank® public databases using arabinose-inducible promoters encoded in S. islandicus REY15A as probes. Promoters used for this analysis include those of araS (arabinose-binding protein), araDH (o-arabinose 1-dehydrogenase), araD (o-arabinonate dehydratase), dopDH (2-oxo-3-deoxy-o-arabinonate dehydratase) and kdaD (2,5-dioxopentanoate dehydrogenase) and they were retrieved from the genomes of S. solfataricus P2 [10] and S. islandicus L.D.S.8, L.S.2.15, M.14.25, M16.4, M.16.27, Y.G.15.51 and Y.N.14.57 [11].
Concluding remarks
Many archaeal repressors have been shown to compete with TBP and/or TFB for binding at the core promoter elements, or blocking the binding of archaeal RNAP on the promoters (reviewed in [18,19]) and this mimics the bacterial repression where repressors compete for binding sites of the promoter recognizing factor σ and/or RNAP, a phenomenon that was attributable to the fact that neither bacteria nor archaea possess homologues of the multiple protein complexes that remodel eukaryal chromatin.

However, mechanisms responsible for activation of gene expression in archaea remain largely to be revealed. Only a few archaeal activation systems have attained extensive studies thus far and these include the Methanocaldococcus jannaschii activator Ptr2 and Halobacterium salinarum vesicle synthesis activator GvpE. These activators were found to interact with TBP and stimulate transcription from RNAP in vitro [20,21]. This establishes that TBP recruitment as one of the mechanisms in archaeal activation. Several other activators are implicated in activating gene expression by recruiting TBP/TFB including S. solfataricus LysM [22], Sta1 [23,24], BldR [25,26] and Ss-LrpB [27], Thermococcus kodakaraensis Tgr [28] and P. furiosus SurR [29], but the bona fide mechanisms of gene activation remain to be illustrated for all of these systems.

On the other hand, genetic characterization of arabinose-inducible expression in Sulfolobus suggests that recruiting TFB on to inactive core promoters constitutes the activation mechanism for ara-box-containing promoters. But the interaction between the putative activator and the transcriptional machinery needs to be demonstrated. Unfortunately, several attempts failed to isolate a protein factor that specifically binds to the ara-box sequence in an affinity binding experiment (N. Peng, unpublished work). Once obtained, investigating the interaction between and the transcriptional machinery on different ara-box containing promoters will yield insights into the mechanistic details of the proposed TFB recruitment and the mechanism of differential gene expression from arabinose-inducible promoters.

Funding
This research is supported by the Danish Research Council of Technology and Production [grant number 274-07-0116] and by State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, China.

References


Received 16 September 2010
doi:10.1042/BST0390099