Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in *Sulfolobus*

Ling Deng, Chandra S. Kenchappa, Xu Peng, Qunxin She* and Roger A. Garrett*

Archaea Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200N Copenhagen K, Denmark

Received September 8, 2011; Revised November 4, 2011; Accepted November 6, 2011

ABSTRACT

CRISPR loci are essential components of the adaptive immune system of archaea and bacteria. They consist of long arrays of repeats separated by DNA spacers encoding guide RNAs (crRNA), which target foreign genetic elements. Cbp1 (CRISPR DNA repeat binding protein) binds specifically to the multiple direct repeats of CRISPR loci of members of the acidothermophilic, crenarchaeal order *Sulfolobales*. cbp1 gene deletion from *Sulfolobus islandicus* REY15A produced a strong reduction in pre-crRNA yields from CRISPR loci but did not inhibit the foreign DNA targeting capacity of the CRISPR/Cas system. Conversely, overexpression of Cbp1 in *S. islandicus* generated an increase in pre-crRNA yields while the level of reverse strand transcripts from CRISPR loci remained unchanged. It is proposed that Cbp1 modulates production of longer pre-crRNA transcripts from CRISPR loci. A possible mechanism is that it minimizes interference from potential transcriptional signals carried on spacers deriving from A-T-rich genetic elements and, occasionally, on DNA repeats. Supporting evidence is provided by microarray and northern blotting analyses, and publicly available whole-transcriptome data for *S. solfataricus* P2.

INTRODUCTION

Archaeal CRISPR-based immune systems provide a defence against invading genetic elements, primarily viruses and conjugative plasmids, and they fall into three main types, the DNA-targeting CRISPR/Cas systems where CRISPR loci and cas genes are invariably linked on the genome, and the DNA-targeting CRISPR/Csm and RNA-targeting CRISPR/Cmr systems for which the csm and cmr gene cassettes are often uncoupled from CRISPR loci (1–4). Almost all archaea carry CRISPR-based defence systems and crenarchaea often exhibit a complex mixture of different types (5,6). The CRISPR locus is an essential functional component of all these systems and consists of a long leader region followed by up to about 100 spacer-repeat units. Spacer sequences originating from foreign genetic elements are about 30–40 bp long and the interspaced identical direct repeats are ~25–37 bp in length; both tend to be conserved in length for a given CRISPR locus (2,7–9).

All CRISPR-based immune systems are basically modular with three primary functions: (i) adaptation that involves excision of DNA from invading DNA genetic elements and integration of the DNA as a new spacer in a CRISPR locus at or near the leader, (ii) generation and processing of CRISPR transcripts to yield mature crRNAs, and (iii) interference of the genetic element by targeting and cleavage via a crRNA–protein complex (10). A few Cas, Cmr and Csm proteins have been assigned roles associated with each of these functional steps on the basis of predictions from bioinformatical or crystal structure analyses and, less commonly, experiments (2,4).

The crenarchaeal genus *Sulfolobus*, in particular, has yielded novel insights into these CRISPR-based systems. *Sulfolobus* species generally carry complex and diverse systems including DNA-targeting CRISPR/Cas and CRISPR/Csm and RNA targeting CRISPR/Cmr, sometimes encoded in multiple copies in a given species (5,7,11). Moreover, many novel viruses have been characterized for *Sulfolobus* and the related genus *Acidianus* which have recently been classified into seven new viral families with several remaining unclassified (12,13), and plasmids with an archaea-specific conjugative apparatus have also been identified (14). These provide a major advantage for studying the interplay between genetic elements and host CRISPR-based systems. For example, numerous virus and conjugative plasmid sequence matches to CRISPR spacers were used to demonstrate that the uptake of DNA from invading genetic elements was essentially a random process (15). Moreover,
it was recently shown that by employing vectors carrying viral genes or sequences matching CRISPR spacers under selection, one can induce different sized CRISPR deletions which all include the matching spacer (16), and these genetic systems were also used to study sequence stringency requirements for DNA targeting by crRNAs (16,17).

Our understanding of mechanisms of transcriptional regulation of CRISPR loci is still at an early stage. In enterobacteria transcription of CRISPR loci and associated cas genes is silenced by the H-NS regulator and activation of the system requires an anti-silencer (18–20). Moreover, a bacteriophage EPV1 was characterized in a metagenomic study encoding the H-NS protein which could inactivate the host defence system upon viral infection (21). For Sulfolobus, the complexity and diversity of the CRISPR systems, and the presence of putative transcriptional regulators associated with the different genetic modules, would require multiple regulatory mechanisms. For example, the putative provirus M164 was found integrated into the gene of the putative transcriptional regulator Csa3 associated with the gene cassette encoding proteins involved in adaptation (22).

The simplest way to inactivate the diverse systems would be to inhibit production of crRNAs. However, it has been shown for Sulfolobus, and other hyperthermophilic archaea, that pre-crRNAs are generated constitutively in the absence of invading foreign DNA elements (7,11,23–25) and, currently, there is no evidence to indicate that the level of pre-crRNA transcripts increases when genetic elements enter cells. Presumably, this reflects that the CRISPR immune system can respond rapidly to the continual exposure to a wide variety of foreign genetic elements that frequent these extreme natural environments (12). The regulatory difference between pre-crRNA regulation in the enterobacteria and Sulfolobus could also reflect that the diverse and complex CRISPR-based systems of Sulfolobus and other archaea are actively involved in maintaining relatively low levels of viruses intracellularly (12,13).

Only one protein to date has been shown to bind directly to a CRISPR locus. That is the Sulfolobus solfataricus P2 protein Sso0454 (formerly SRSR repeat-binding protein) that exhibits a triplex repeat repeat sequence and binds specifically to DNA repeats of CRISPR loci within a cell, in total 208 repeats of CRISPR loci present (11) and evidence for the formation of reverse-strand transcripts from the six CRISPR loci of S. solfataricus was also provided by a whole-transcriptome analysis (31). The functional significance, if any, of these reverse-strand transcripts remains unclear but potentially they can base pair with crRNAs and impede their interference reactions.

Sequence-specific DNA-binding proteins are often transcription factors but, since Cbp1 can potentially bind to all repeats of CRISPR loci within a cell, in total 208 repeats in S. islandicus REY15A and 423 in S. solfataricus P2 (11,16), it is not a typical transcriptional regulator. In order to gain some insight into its function(s), we exploited the recently developed genetic systems for S. islandicus REY15A and S. solfataricus P2 as well as a microarray for the latter organism. Cbp1 knockout and overexpression mutants were generated and transcriptional properties of the CRISPR loci were examined for the mutants (16,32).

**MATERIALS AND METHODS**

**Strains, media and constructs**

Experiments were performed with *S. islandicus* E233S carrying a deletion within the *pyrEF* genes and a complete *lacS* gene deletion. *S. solfataricus* InF1 was also employed carrying an inactivated *pyrF* gene. Cells were cultured at 75–78°C in the complex medium TYS or in the selective media SCVy, GCVy or ACVy (5,32). To arrest transcription, actinomycin D was added to the culture at 20 µg/ml and samples were taken at different time points (33).

Plasmid pK454 was used to generate the *cbp1* gene deletion in *S. islandicus* E233S. It was constructed by a triple ligation of pHZ1, the L arm (amplified using primers 5′-tgtatGCATTGACAAAACTCAATAATCTCCT-3′ and 5′-tttcgagAGGATTGCTAACCCGGAGAATGTA ACTT-3′) and the R arm (amplified using 5′-tttcgagTG CAAAAGACTCTAATTTCTACTA-3′ and 5′-tttcgagGAAGCTAGGACACCAATCCCTCATCAT-3′) before transforming into *S. islandicus* E233S to produce first pop-in transformants and then the deletion mutant (16,32). Primers 5′-AATCCCAACAGTAAC CACC-3′ and 5′-GCATGTCATGCCTTGGAGAAC G-3′ were used to confirm the recombination events.

Plasmid pC454 was used to complement the Cbp1 deletion mutant and was constructed by inserting the SOE-PCR product into pHZ1 (32). Briefly, primers CompL 5′-ttgcgtCCATTGACAAAACTCAATAATCTCCT-3′ and CompR 5′-CCTAGTATTTTCT TAAAATTCTCAAT-3′ were employed to produce fragment L, and fragment R was generated using

The possibility of a transcriptional role for the protein. CRISPR loci generally appear to be transcribed as single transcripts from a promoter within the leader region (7,11) followed by processing within repeat sequences by Cas6 to yield mature crRNAs, carrying part of the repeat and all or most of the spacer sequence (28–30). It was also shown for *S. acidocaldarius* that transcripts are produced from reverse CRISPR strands for each of the five CRISPR loci present (11) and evidence for the formation of reverse-strand transcripts from the six CRISPR loci of *S. solfataricus* was also provided by a whole-transcriptome analysis (31). The functional significance, if any, of these reverse-strand transcripts remains unclear but potentially they can base pair with crRNAs and impede their interference reactions.
Expression, purification and detection of Cbp1

The cbp1 gene of S. islandicus was amplified using primers SisCbpF  5'-TTTGGATATTCATCCATGGTAGAAGAAGAAAATGTTAAGAAATG-3' and SisCbpR  5'-TTTCCGCTGAGAGCAGATGTGTGAGGAGAGATTCACGAA-3'. The PCR product was inserted into a pET28a vector (Novagen, Darmstadt, Germany) using Xhol and NdeI (Fermentas, St Leon-Rot, Germany) and the stop codon was replaced by codons for a C-terminal hexameric His-tag. The resulting clone was inoculated into LB broth carrying kanamycin and chloramphenicol and cultured at 37°C to A600 = 0.6. The protein was expressed overnight at room temperature in BL21 (DE3) pLYSs cells (Promega, Madison, USA) and expression was induced by adding isopropylthio-β-D-galactoside (IPTG) at 0.5 mM.

The cbp1 gene of S. islandicus was amplified using primers XChbpFs  5'-tttggccgcaATGAGTGAAGAAGAAATAATTGAAAAAGT-3' and XChbpRs  5'-TTTCCGCTGAGAGCAGATGTGTGAGGAGAGATTCACGAA-3' for expression in S. islandicus. Primer XChbpSrS  5'-tttggccgcaATGAGTGAAGAAGAAATAATTGAAAAAGT-3' and XChbpRs were employed to amplify the cbp1 gene of S. solfataricus for expression in S. solfataricus. The modified pEXA vector pSeSD was used for protein overexpression (16,34).

Cbp1 protein were detected by western blots using antibodies derived from Escherichia coli-expressed Cbp1 (Innovagen AB, Lund, Sweden) and an alkaline phosphatase-conjugated goat anti-rabbit secondary antibody (Invitrogen, Paisley, UK), using the 5-bromo-4-chloro-3-indolylphosphate (BCIP)-nitroblue tetrazolium (NBT) reagent (Sigma-Aldrich, Munich, Germany).

DNA band-shift assay

A 157-bp DNA fragment carrying a repeat–spacer-repeat sequence and short-flanking regions was amplified from S. islandicus by PCR using oligonucleotide pairs Ssis2pF  5'-CTTTCCTCATCCATCTCTCCACCACC-3' and Ssis2pR  5'-CTTTCCTTTGCTACATCCCTCGCAG-3' and [32P]-5'-end labelled using T4 polynucleotide kinase (Fermentas). Cbp1 (35 nM) was incubated with 7 nM [32P]-5' end labelled CRISPR-2r substrate in DNA binding buffer (10 mM Tris–Cl, pH 7.6, 150 mM KCl, 2 mM DTT, 10% glycerol) for 20 min at 50°C before loading on an 8% polyacrylamide gel. To test for binding specificity, the unlabelled 157 bp CRISPR-2r substrate was used as specific competitor and a 179 bp recA gene PCR product was used as unspecific competitor. After cooling to room temperature, 3 μl loading buffer (10 mM Tris–Cl, pH 7.6, 1 mM EDTA, 50% glycerol, 0.5% bromophenol blue) was added and samples were fractionated in a prerun 8% polyacrylamide gel containing 89 mM Tris–Cl, 25 mM taurine, 0.5 mM EDTA, pH 8.9 and autoradiographed. The competition assay of S. solfataricus Cbp1 to different CRISPR repeats followed the same experimental conditions. For amplification of the 148 bp CRISPRs2-2r (A+B) DNA, primers Ssis2pF and Ssis2pR were employed to amplify S-CCGACACTGTATAAATACATGC-3' and S-CCGATCATTTGACTG-3', and the 151 bp CRISPRs2-2r (C+D) DNA was amplified using primers S-CCGACACTGTATAAATACATGC-3' and S-CCGATCATTTGACTG-3'.

RNA preparation and northern blotting

Total cellular RNA was prepared as described earlier (20). For northern blotting, 15μg RNA was mixed with the same volume of Gel Loading Buffer II or NorthernMax-Gly Sample Loading Buffer (Applied Biosystems/Ambion, Austin, USA) and fractionated in a 6% polyacrylamide gel containing 7 M urea and 90 Mm Tris-borate, 2 mM EDTA, pH 8.3, or a 1.5% agarose gel in 10 mM PIPES, 30 mM Bis-Tris, 1 mM EDTA, pH 8.0 with 0.1–2.0 or 0.5–10 kb RNA ladders (Invitrogen) or a 50- to 1000-nt RNA ladder (New England Biolabs, Boston, USA).

Procedures for transferring and immobilizing RNA on nylon membranes, prehybridizing, end-labelling of complementary nucleotides, hybridization and film exposure followed earlier protocols (11). Probes were stripped from hybridized membranes with 0.5% SDS for 1 h and membranes were reused for hybridization when no residual radioactivity was detected. Oligonucleotide probes used were as follows: repeat of loci 1 and 2 of S. islandicus pre-crRNA—5'-TTTCCATATTTATAGTAGATTAGC-3'; spacer 10, locus 2—5'-GGCCCGAG TAGATTAGC-3'; spacer 1, locus C—5'-TTGAAAGATTTGAACGTTAGCGAG-3'; repeat of loci A+B—5'-TTTGGAATTCCATATGGTGAGTGAAG-3'; repeat of loci A+B—5'-TTTGGAATTCCATATGGTGAGTGAAG-3'; repeat of loci C+D—5'-TTTCCATATTTATAGTAGATTAGC-3'; and repeat of loci A+B—5'-TTTCCATATTTATAGTAGATTAGC-3'.

Localizing CRISPR locus deletions

PCR products were obtained across CRISPR loci 1 and 2 of S. islandicus REY15A using premixed Ex Taq (Takara, Otsu, Japan) following the manufacturer’s protocol with 75–100 ng genomic DNA in 10 μl. To amplify S. islandicus CRISPR loci 1 and 2, respectively, we used C1F  5'-AGCTTGTTAACCTTAAGGTACTTACGT-3' and C1R  5'-TATAAACAACGAGTTTTCTCCCTGAT-3', and C2F  5'-AGGATACGGAATGCGACTGTTGGTGAT-3' and C2R  5'-TAACGCACGTTATGGAAATTTCCTC6-3'. Purified PCR products were sequenced either directly (Eurofins MWG Operon, Ebersberg, Germany) or after cloning using CloneJET™ PCR Cloning Kit (Fermentas).
Analysis of microarray and transcriptome data for *S. solfataricus* P2

*S. solfataricus* P2 microarrays were designed by the *Sulfolobus* genome chips consortium and manufactured by Oicimun Biosolutions (Hyderabad, India). They carried 3042 gene probes and several sets of probes against crenarchaeal viral genomes and plasmids (35). Microarray hybridizations were performed as described (35) except that the CyScribe Post-Labeling Kit (GE Healthcare) was used for labelling. Data analyses were conducted by ImaGene (BioDiscovery, CA, USA) using default settings. A dye swap was performed for each time point and values were averaged. CLC Genomic Workbench (Aarhus, Denmark) was employed for analysing the raw CRISPR transcriptome sequencing data (http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?study=SRP001461). Strand-specific sequences from three different cDNA libraries were analysed and the parameters were set such that the only perfect matches of the 36 nt reads to the genome were taken.

RESULTS

*Sulfolobus islandicus* Cbp1 binds specifically to CRISPR repeats

Cbp1 of *S. islandicus* REY15A shows 93% sequence identity to Sso0454 of *S. solfataricus* and was expressed in *E. coli* and purified (Figure 1A). DNA repeat binding activity was assayed using a [32P]-end labelled 157 bp repeat–spacer–repeat construct with short flanking sequences (CRISPR-2r). Electrophoretic band-shift assays showed the formation of two retarded bands consistent with Cbp1 binding to one or both repeats of the substrate (Figure 1B, lane 2). Competition experiments were performed using either unlabelled CRISPR-2r DNA as specific competitor or a 179-bp DNA fragment amplified from a recA gene of *S. islandicus* as an unspecific competitor. Only the unlabelled CRISPR-2r competed strongly over the range 2.5- to 20-fold molar excess (lanes 3–6 compared with lane 2) showing a progressive transition from the upper to the lower bands thereby providing support for a specific Cbp1–DNA interaction (Figure 1B).

Generation of a deletion mutant, a complemented strain and overexpression vectors for Cbp1

In order to test for Cbp1 function, we first generated a Cbp1-minus mutant of *S. islandicus* REY15A using a ‘pop in-pop out’ gene targeting method (32). Plasmid pK454 was constructed and then transformed to produce pop in transformants and then Cbp1 deletion mutants (see ‘Materials and Methods’ section). To minimize possible detrimental effects on expression of flanking genes, part of the *cbp1* gene, encoding the N-terminal 38 amino acids, was retained after the knockout. Deletion mutants were identified by colony PCR for four of the eight colonies that formed on counter-selective plates (Figure 2A). Growth rates and morphologies of the Cbp1-minus cells were indistinguishable from those of wild-type cells indicating that Cbp1 is not essential for cell viability (data not shown).

Next, a complemented strain for the deletion mutant was produced using the same strategy as for the deletion mutant (Figure 2B). It carried a single A-G mutation, confirmed by PCR amplification and sequencing that resulted in conversion of the lysine-86 codon from AAA to AAG (data not shown). Both codons occur frequently such that the translation efficiency should not be affected. A western blot analysis was employed to verify that Cbp1 was only expressed from the wild-type and Cbp1 complemented strains, and it was also demonstrated that the expression level from the complemented strain was similar to the wild-type (Figure 2C).

The *cbp1* gene from *S. islandicus* was cloned into the pSeSD overexpression vector and transformed into *S. islandicus* E233S. A similar overexpression construct was also generated for the *cbp1* gene of *S. solfataricus* (sso0454) and was transformed into *S. solfataricus* lnF1. The *cbp1* genes were preceded by an araS promoter such that Cbp1 expression could be controlled in both species by using different carbon sources (16,34).

Active foreign DNA interference in the *S. islandicus* Cbp1-minus mutant

First, we tested whether foreign DNA interference by the CRISPR/Cas system was inhibited by the absence of Cbp1. A vector was employed carrying spacer 45 of the CRISPR locus 2, a CC protospacer adjacent motif (PAM) and *pyrE/F* genes and it was transformed into the uracil auxotrophic Cbp1-minus *S. islandicus* E233S cells. A very low transformation efficiency was observed compared with non-targeted plasmids that is consistent with active DNA targeting (16). Ten transformants were cultured and PCR products were generated for CRISPR locus 2 carrying the matching spacer 45 and for...
non-targeted CRISPR locus 1 as a control. The results demonstrate that deletions occurred in locus 2 in up to eight transformants, including the apparent loss of CRISPR loci 1 and 2 in transformants 1 and 4 (Figure 3). PCR products from transformants 5 and 6 were sequenced and showed deletions from repeats 1–62 and 7–56, respectively, in locus 2 which both included the matching spacer 45. These results indicate that the DNA interference system is still active in the absence of Cbp1 (16).

Reduced pre-crRNA levels in the Cbp1-minus mutant

To test for a possible transcriptional role, we estimated the pre-crRNA and mature crRNA transcript levels for the Cbp1-minus mutant. Northern blots were obtained using a probe specific for the identical repeat sequences of CRISPR loci 1 and 2 and another for spacer 10 of CRISPR locus 2. In each experiment, processed intermediates generated a typical Sulfolobus pre-crRNA ladder corresponding to multiples of two to three spacer-repeat units (24) (Figure 4).

The results with the repeat probe revealed a strong reduction in yields of pre-crRNA products for two isolated Cbp1-minus mutant clones (M1 and M2) compared to wild-type and Cbp1 complemented strains, consistent with an overall reduction in the level of longer CRISPR transcripts (Figure 4A). The results with the spacer 10 probe also showed the presence of relatively weaker intermediate bands at ~180, 300 and 540 nt but strong mature crRNA bands were still present consistent with the demonstration that foreign DNA targeting remains active (Figure 4A). The uppermost bands >600 nt may result partially from cross-hybridization effects with the spacer probe.
These results suggest that in the absence of Cbp1 there is a reduced overall level of longer pre-crRNA transcripts generated from loci 1 and 2, but the reduction could also reflect that the absence of Cbp1 leads to a destabilization of pre-crRNAs. Therefore, we tested for pre-crRNA stability employing actinomycin D that actively blocks transcription in *Sulfolobus* cells for up to 2 h (33). The relative yields of pre-crRNA and mature crRNAs in the Cbp1-minus mutant and wild-type cells were monitored by northern blotting using a repeat probe. Yields of the RNA products remained essentially constant over a 2-h period for both mutant and wild-type (Supplementary Figure S1) indicating that the absence of Cbp1 did not affect the stability of the pre-crRNAs. The experiment was also repeated by probing spacer 1 of CRISPR locus 2 for the wild-type and Cbp1-minus mutant. This showed that the yields of the mature crRNAs were also unchanged over the 2-h period (Supplementary Figure S1). These combined results support the conclusion that production of longer pre-crRNAs is reduced in the absence of Cbp1 but they also establish that the pre- and mature crRNAs have relatively long half-lives. For the mature crRNAs, this may reflect that they are stabilized by complexing with the Csa2 (Cas7) protein or Cmr proteins (2,25,34).

**Overexpression of Cbp1 produces increased levels of pre-crRNAs**

The preceding results suggest that the presence of repeat-bound Cbp1 enhances production of longer CRISPR transcripts from the leader because transcripts initiating and terminating within spacers of a CRISPR locus would tend to generate additional irregularly sized intermediate pre-crRNA products. Since it was estimated for *S. solfataricus* P2 that there are insufficient Cbp1 copies to saturate all the CRISPR repeats (26), we reasoned that overexpression of Cbp1 should increase levels of longer pre-crRNA transcripts. Therefore, we introduced the Cbp1 overexpression vector into *S. islandicus* E233S and examined the yields of larger pre-crRNA products relative to those of wild-type cells by northern blot analysis. The results demonstrated a significant increase in the yields of the larger (>600 nt) pre-crRNA products (Figure 4B) consistent with increased coverage of CRISPR repeats by Cbp1 leading to higher yields of longer pre-crRNAs. Since unprocessed CRISPR transcripts could be 6–7000 nt in length, the approximate size range of pre-crRNAs was estimated by separating RNAs from the overexpressed Cbp1 strain in agarose gels prior to northern blotting. Most pre-crRNAs fell within the size range 60–1000 nt (Figure 4C).

**Microarray analysis of selected crRNAs in *S. solfataricus***

Since overexpression of Cbp1 produced higher yields of larger pre-crRNAs (Figure 4B), we exploited an available microarray carrying probes against all predicted ORFs and selected spacers of CRISPR loci A to F of the closely related species *S. solfataricus* P2 (11). No gene knockout system is available for this strain but we could...
overexpress Cbp1. Samples were collected at 0, 4 and 20 h after changing to an arabinose medium when transformants carrying empty arabinose-inducible expression vectors, or Cbp1 overexpression vectors, were selectively cultured (Figure 5A). Western blots performed on transformants from each culture showed enhanced Cbp1 expression at 0 h that increased strongly after 4 h and remained at a similar level after 4 h (Figure 5B). Enhanced expression at the first time-point reflects leakiness of the strong arabinose promoter (36).

Random cDNAs were generated from crRNAs of the six CRISPR loci and hybridized to the microarray. Relative hybridization yields for the overexpression mutant and wild-type were estimated in duplicate samples probing for sequences of 22 individual crRNAs. We focused on the time points 0 and 4 h because strong growth retardation occurred after 4 h. Seven crRNA transcripts showed significant changes, six with increased expression in the overexpression mutant and one with decreased expression while the remaining eight changes were considered insignificant (Figure 5C). The results from loci E and F which carry a defect CRISPR promoter and lack a leader region, respectively (11), exhibited low transcript levels presumed to derive from internal promoters (Figure 5C). Northern blot analyses of selected crRNAs confirmed the enhanced levels of mature crRNAs in the Cbp1-overexpressed strain for locus A/spacer 28, locus B/spacer 24 and locus C/spacer 1 (Figure 5C).

Furthermore, northern blotting using probes against the repeats, revealed that pre-crRNAs of loci A and B were more strongly induced than those of C and D (Figure 6A). We tested for the relative binding strength of Cbp1 to the two repeats, which differ in sequence at five positions and in length by 1 bp, in a competition experiment (Figure 6B). The results showed that the C + D repeat displaced Cbp1 from the A + B repeat consistent with the protein binding more strongly to the former repeat (Figure 6B). This in turn provides a rationale for the result seen in Figure 6A. Under normal cellular conditions, Cbp1 will preferentially saturate CRISPR loci C + D while Cbp1 overexpression will lead to the additional saturation of loci A + B, consistent with strongly enhanced crRNA yields observed for these loci.

The major decrease in the level of transcripts observed for locus D/spacer 89 (Figure 5C) correlates with an exceptionally high level of transcripts observed in this region for the wild-type strain probably due to a strong promoter located in spacer 88 (Figure 7A). The northern blot result confirms that for the overexpression strain

![Figure 5](https://academic.oup.com/nar/article-abstract/40/6/2470/2409225)
there is a decrease in the level of crRNA transcripts downstream from this site (Figure 7B). These results suggest that Cbp1 has a modulating effect on pre-crRNA transcription and does not simply enhance transcription but can also reduce transcription at abnormally highly transcribed internal CRISPR sites.

In addition to affecting transcription of pre-crRNAs, overexpression of Cbp1 also produced other minor transcriptional changes at time points T1 and T2 (Figure 5A) for a few genes most of which encode conserved hypothetical proteins (data not shown). The single exception was the sso1101 gene encoding a putative transcription regulator. Transcription was repressed ~3-fold and 2-fold, respectively, at time points T1 and T2. Sso1101 is one of the few proteins that exhibits enhanced expression on biofilm formation for diverse Sulfolobus species (37).

**Reverse-strand transcripts from CRISPR loci of S. solfataricus P2**

In an earlier study, reverse-strand transcripts covering a range of sizes were detected from the five CRISPR loci of *S. acidocaldarius* by northern blot analyses (7,11). More recently, a high-coverage sequencing study of the
transcriptome of *S. solfataricus* P2 was performed (31) and the raw data are publicly available (see ‘Materials and Methods’ section). 5'-ends of both leader and reverse-CRISPR transcripts, constituting either start or cleavage sites, were identified for the six CRISPR loci. Loci A–D carry strong leader promoters and most 5'-ends were located on the leader strand and correspond to cleavage sites within repeats. Loci E and F exhibit defective transcription from the leader and while locus E (seven spacers) showed very few 5'-ends, for the larger locus F (88 spacers) 5'-ends were distributed fairly evenly along both strands. Representative results for these two groups of loci are shown for loci C and F, respectively, in Figure 8A. Corresponding results for both leader and reverse strands of the other loci A, B and D are given in Supplementary Figure S2.

5'-end locations and the number of reads with identical 5'-ends are shown for the two main peaks of reverse-strand transcripts, located distal from the leader region. They are associated with repeat–spacer 29 of locus C and with repeat–spacer 84 of locus F (Figure 8A). In order to gain some insight into the size of the reverse-strand transcripts generated and whether they are influenced by Cbp1 overexpression, northern blot analyses were performed by probing the transcripts in loci C and F adjacent to the illustrated start sites (Figure 8A). The results show discrete RNA bands of ~100 and 170 nt for locus C and strong larger bands, while a range of band sizes >100 nt was observed for locus F (Figure 8B). Thus the increased expression of Cbp1 had no apparent influence on the size or yields of these reverse-strand transcripts.

**DISCUSSION**

We provide evidence for Cbp1 modulating transcription of CRISPR loci from the leader. We still know little about the stepwise processing of large CRISPR transcripts, and the mechanisms may vary for different CRISPR-based systems (11,23,24). Moreover, the yields of individual mature crRNAs are non-uniform (see, for example, Figure 7A). Nevertheless, the data presented are consistent with a model in which Cbp1 inhibits internal initiation and termination at putative archaeal transcriptional motifs located within spacers, and occasionally within repeats, of CRISPR loci. Such a function could be especially important for the acidothermophilic Sulfolobales and their genetic elements which carry A+T-rich (~65%) genomes. The frequency of possible archaeal-specific promoter motifs (hexameric TATA-like sequences) and archaeal terminator motifs (T-rich pyrimidine sequences) is likely to be high (38,39). An estimate of the number of such motifs in the 4800 spacers contained in sequenced genomes of the Sulfolobales indicated that a large fraction carried potential promoter motifs and a smaller fraction of terminator motifs (3) and there is an additional promoter motif (ATTAAT) within the repeats of loci A and B of *S. solfataricus* P2. Even if most of these
motifs are ineffective or weakly effective, collectively they could severely impede generation of long CRISPR transcripts of up to about 7000 nt. The demonstration by footprinting studies that Cbp1 partially protects the ends of the spacer regions (26) provides a potential mechanism for the transcriptional modulation. To test this model further, we plan to generate knockout mutants of the CRISPR pre-crRNA processing enzyme Cas6 and study the effect of Cbp1 on the formation of primary CRISPR transcripts.

There are no obvious precedents for this type of transcriptional modulation but there could be a mechanistic link to the eukaryotic THO complex consisting of four protein subunits. This complex has been implicated in inhibiting formation of aberrant DNA structures during transcription of genes containing repeat sequences which might otherwise impede polymerase progression or lead to increased recombination (40). This potential mechanistic similarity also suggests a possible secondary role for Cbp1. Repeat-bound Cbp1 could reduce the likelihood of recombination occurring between CRISPR repeats which might lead to deletions within CRISPR loci that do occur periodically (7,11) and can be induced by vectors carrying matching protospacers (16,41).

The cbp1 gene is not linked directly to the CRISPR-based gene cassettes in Sulfolobus chromosomes which suggests that it has other cellular functions. There is a precedent for this with the bacteria-specific RNase III endonuclease. It contributes to important cellular RNA processing functions but is also essential for processing of bacteria-specific type II CRISPR RNAs (42). A more general role for Cbp1 in inhibiting recombination between repeat sequences mentioned earlier is one possibility. Another is the potential link of Cbp1 to biofilm formation. Enhanced Cbp1 transcription produced a significant reduction in Sso1101 expression and this is one of very few proteins that are overexpressed during biofilm formation in Sulfolobus species (37). If Cbp1 were to be overexpressed on viral infection biofilm formation might be inhibited in order to reduce mixing of uninfected with infected cells.

Cbp1 binds specifically to a range of similar repeat sequences associated with different CRISPR loci of Sulfolobus and it is likely that the more conserved binding specificity (26, our unpublished data). Our results show that Cbp1 binds more strongly to the most common family I repeats of CRISPR loci C + D that dominate in the Sulfolobales and in other crenarchaea than to the less common family II repeats of loci A + B (11). Given that CRISPR/Cas and Cmr modules have been shown to exchange intercellularly for S. islandicus species (6), Cbp1 could influence which types of CRISPR loci are retained in the cell and also, explain the predominance of the family I repeats amongst the Sulfolobales (11).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures S1 and S2.

ACKNOWLEDGEMENTS

We thank Shiraz A. Shah for helpful discussions on genome analyses and the referees for their critical insights.

FUNDING

Danish Natural Science Research Council (grant no. 272-08-0391); Danish Research Council for Technology and Production (grant no. 274-07-0116); Danish National Research Foundation. Funding for open access charge: Danish Natural Science Research Council.

Conflict of interest statement. None declared.

REFERENCES