

Type III CRISPR complexes from *Thermus thermophilus*

Marta Szychowska^{1,2,a,#}, Wojciech Siwek^{1,b,#}, Damian Pawolski^{1,c}, Asgar Abbas Kazrani¹, Krzysztof Pyrc^{3,4} and Matthias Bochtler^{1,5}✉

¹International Institute of Molecular and Cell Biology, Warsaw, Poland; ²Institute of Technical Biochemistry, Faculty of Biotechnology and Food Sciences, Technical University of Lodz, Łódź, Poland; ³Microbiology Department, Faculty of Biochemistry Biophysics and Biotechnology, Jagiellonian University, Kraków, Poland; ⁴Malopolska Centre of Biotechnology, Jagiellonian University, Kraków, Poland; ⁵Institute of Biochemistry and Biophysics PAS, Warsaw, Poland

Pathogen-specific acquired immunity in bacteria is mediated by the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas systems. *Thermus thermophilus* strain HB8 contains CRISPR systems of several major subtypes (type I, IIIA and IIIB), and has become a widely studied model for CRISPR biology. We have selected two highly expressed CRISPR spacers, crRNA 2.1 and crRNA 2.2, and have enriched endogenous *T. thermophilus* proteins that co-purify with these crRNAs. Mass spectroscopy indicates that the chromatography protocol enriches predominantly Csm complex subunits, but also Cmr subunits. After several chromatographic steps, size exclusion chromatography indicated a molecular mass of the crRNA associated complex of 265±69 kDa. In agreement with earlier work, crRNAs of different lengths (containing the selected spacers) were observed. Most of these were completely lost when several *T. thermophilus* csm genes were ablated.

Key words: CRISPR-Cas, endonuclease, Csm, Cmr, *Thermus thermophilus*, crRNA, mass spectrometry (MS), chromatography

Received: 05 February, 2016; **revised:** 24 May, 2016; **accepted:** 24 May, 2016; **available on-line:** 14 June, 2016

INTRODUCTION

Pathogen-specific, acquired immunity against invading nucleic acids, like bacteriophages or plasmids, is mediated in bacteria by CRISPR-Cas systems (Marraffini & Sontheimer, 2010a). These systems consist of the CRISPR loci and the CRISPR-associated (*cas*) genes (Wiedenheft *et al.*, 2012; van der Oost *et al.*, 2014). The CRISPR-loci store the immune memory in a form of pathogen-derived DNA sequences (known as spacers) separated by repeat sequences (Mojica *et al.*, 2005; Barrangou *et al.*, 2007). The Cas proteins cooperate with host factors to mediate spacer acquisition from invading DNA, crRNA biogenesis, and finally expression of immunity against invading DNA or RNA (Wiedenheft *et al.*, 2012; van der Oost *et al.*, 2014). Most bacteria harboring CRISPR-Cas systems share characteristic *cas* genes (typically *cas1-cas6*) (Jansen *et al.*, 2002a), which are complemented by additional, less widely distributed *cas* genes (Haft *et al.*, 2005). The gene composition and architecture of the *cas* operons as well as the phylogenies of the most conserved and prevalent *cas* genes (*cas1*, *cas2*) have been used to classify the CRISPR-Cas systems (Jansen *et al.*, 2002a; Jansen *et al.*, 2002b; Haft *et al.*, 2005; Makarova *et al.*, 2011). According to a recent review (Makarova *et al.*, 2015), CRISPR systems come in two main classes.

Class 1 systems are defined by the presence of multi-subunit crRNA effector complexes. They can be further classified into type I, type III and type IV, with characteristic signature genes (*cas3* for type I, *cas10* for type III and *csf1* for the poorly characterized type IV). Class 2 systems have a single subunit crRNA effector nuclease, either *Cas9* for type II systems, or *Cpf1* for type V systems. Type I and type II systems are already very well characterized (Jore *et al.*, 2013; Zhao *et al.*, 2014; Sinkunas *et al.*, 2013; Gasiunas *et al.*, 2012; Karvelis *et al.*, 2013; Mali *et al.*, 2013), not least because of the major role that *Cas9* is now playing for genetic engineering applications (Mali *et al.*, 2013; Jinek *et al.*, 2012).

Type III CRISPR systems are broadly divided into type IIIA and type IIIB CRISPR systems, with Csm and Cmr effector complexes, respectively (depending on whether the signature gene is *cmr1* or *cmr2*). *In vitro* experiments with purified complexes initially pointed to an RNA endonucleolytic activity for both type IIIA (Tamulaitis *et al.*, 2014; Staals *et al.*, 2014) and type IIIB (Hale *et al.*, 2009) complexes, but *in vivo* experiments on plasmid immunity were interpreted as evidence for an activity against invading DNA (Hatoum-Aslan *et al.*, 2014; Marraffini & Sontheimer, 2008). The puzzle was solved by the observation that plasmid immunity of *Sulfolobus islandicus* REY15A (a host of two different Cmr complexes) required transcription (Deng *et al.*, 2013). Subsequently, it was shown that the Csm complex from *Staphylococcus epidermidis* also harbored a strictly transcription dependent DNA endonuclease activity (Samai *et al.*, 2015). As the DNA and RNA endonuclease active sites were distinct, their roles could be genetically separated. Again using the *S. epidermidis* model, it was shown that the major contribution to plasmid and (DNA) virus immunity was made by the DNA endonucleolytic activity, with RNA endonucleolytic activity rather serving as a “backup” (Samai *et al.*, 2015). Proto-spacer adjacent motifs are not required in targets of type III CRISPR systems. Instead, base pairing between crRNA and target extending beyond the spacer (or protospacer) region

✉ e-mail: mbochtler@iimcb.gov.pl

^aCurrent address: Nencki Institute of Experimental Biology, Ludwika Pasteura 3, 02-093 Warsaw, Poland

^bCurrent address: Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal

^cCurrent address: Center for Molecular Bioengineering, Technische Universität Dresden, Arnoldstrasse 18, 01307 Dresden, Germany

[#]equal contribution

Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeats; Cas, CRISPR-associated; crRNA, CRISPR-RNA

and extending into the repeat region suppresses activity (Marraffini & Sontheimer, 2010b).

The molecular architecture of type III CRISPR systems is largely conserved between type IIIA and IIIB systems and also reminiscent of the architecture of type I CRISPR complexes. A hallmark of all type III complexes is the occurrence of multiple copies of some of their subunits. Cryo-electron microscopy data for the *Sulfolobus solfataricus* and *Thermus thermophilus* Csm complexes, and for the *Thermus thermophilus* Cmr complex suggest a two-filament structure (Wiedenheft *et al.*, 2011; Staals *et al.*, 2013; Staals *et al.*, 2014). The more prominent of these filaments serves as the primary binding site for the crRNA, the less prominent filament wraps around it. The remaining complex subunits “buttress” the filament pair at both ends. Biochemical data indicate substantial variability in the stoichiometry of the complexes, which appears to correlate with crRNA content.

The Csm complex from *Staphylococcus epidermidis* (which lacks other CRISPR-Cas systems simplifying the interpretation of *in vivo* results) (Marraffini & Sontheimer, 2008; Marraffini & Sontheimer, 2010b; Hatoum-Aslan *et al.*, 2011; Hatoum-Aslan *et al.*, 2014) has been described as a complex of 331 kDa (Hatoum-Aslan *et al.*, 2013). Formation of the complex was impeded in the absence of Csm1 (also known as Cas10), Csm3 or Csm4. However, subcomplexes containing all other members still formed when subunits Csm2 or Csm5 were missing (Hatoum-Aslan *et al.*, 2014). The Csm complex from *S. solfataricus* (which has several Csm3 paralogues) was described as a complex of 427 kDa, but several subcomplexes, most likely representing fragments of the full complex, were also observed (Rouillon *et al.*, 2013). For the *Streptococcus thermophilus* Csm complex, masses of 345 kDa and 486 kDa were found, depending on whether the complexes were associated with short (40 nt) or long (72 nt) crRNAs (Tamulaitis *et al.*, 2014). A Cmr complex from *T. thermophilus* was also described as having variable stoichiometry (molecular masses about 310 and about 350 kDa, bound crRNA of 40 or 46 nt in length) (Staals *et al.*, 2013).

T. thermophilus strain HB8 contains three *cas* operons of the type IE, IIIA and IIIB and twelve CRISPR loci, located both on the pTT27 megaplasmid and on the chromosomal DNA (see Fig. 1 in Staals *et al.*, 2014). All *cas* operons are expressed, and the level of their transcription increases upon phage infection (Agari *et al.*, 2009). In addition all CRISPR loci, except the CRISPR-8 locus, are constitutively transcribed and processed, and the pre-crRNAs are primarily cleaved 8 nt upstream of the spacer (Juraneck *et al.*, 2012). In this work, we have chosen two crRNAs (crRNA 2.1 and crRNA 2.2) which were previously reported to be highly abundant in *T. thermophilus* HB8 (Juraneck *et al.*, 2012) and purified protein complexes associated with them. We find that the mix of these two crRNAs is predominantly associated with Csm, but also Cmr complexes. We determine the approximate mass of the complexes by size exclusion chromatography, and we demonstrate that the biogenesis or stability of these crRNAs is dependent on all tested Csm subunits and not only on Csm3 (the subunit recently shown to harbor the RNase activity).

MATERIALS AND METHODS

Bacterial strains and growth conditions. Wild type *T. thermophilus* strain HB8 (WT) (Oshima & Imahori, 1974) and HB27 (lacking the CRISPR-2 array) (Henne *et al.*, 2004) and Δ csm1, Δ csm3, Δ csm5 strains were grown

at 65°C in *Thermus* medium (TM, 0.8% polypeptone, 0.4% yeast extract, 0.2% NaCl, pH 7.0) liquid culture shaken at 180 rpm or on *Thermus* medium agar plates containing 1.5% (w/v) agar. When required, TM was supplemented with kanamycin (20 µg/ml).

T. thermophilus transformation and gene deletion.

T. thermophilus overnight culture was supplemented with CaCl₂ and MgCl₂ to the final concentration of 0.4 mM and shaken at 65°C for 2 h. 50 ml of such culture was centrifuged and the pellet was suspended in 0.5 ml of TM with 0.4 mM of CaCl₂ and MgCl₂. The Δ csm1, Δ csm3, Δ csm5 strains were made according to the protocol from the *Thermus* consortium (Hashimoto *et al.*, 2001). One µg of deletion vector (commercially distributed by RIKEN consortium) was added to the suspension and the mixture was shaken at 65°C for 2 h. Culture was then spread on TM agar plates supplemented with kanamycin and grown for 48 h at 65°C.

Confirmation of gene deletion by Southern blot analysis.

T. thermophilus megaplasmid DNA was isolated from *T. thermophilus* WT and Δ csm1, Δ csm3, Δ csm5 strains liquid cultures using Plasmid Midi AX kit (A&A Biotechnology). Two µg of megaplasmid DNA from Δ csm1 strain was digested with R.SacI, digestion of megaplasmid DNA from Δ csm3 and Δ csm5 strains was performed with R.SmaI enzyme. The same restriction endonucleases were used to digest megaplasmid DNA from the WT strain as a negative control for the assay. Fragmented DNA samples were separated on 1% agarose gels alongside with a [³²P]-5'-end radiolabeled GeneRuler DNA Ladder Mix (Thermo-Scientific). Gels were covered with denaturation buffer (1.5 M NaCl, 0.5 M NaOH) and incubated at RT for 30 minutes with gentle shaking. After removal of denaturation buffer gels were gently shaken with neutralization buffer (1 M Tris-Cl, 1.5 M NaCl, pH 7.4) at RT for 20 minutes. DNAs were blotted to the positively charged nylon membrane (Ambion) by capillary transfer performed in 10× SSC for approximately 15 h and bound to the membrane by UV cross-linking. Prehybridization and hybridization was done at 42°C in the presence of prehybridization buffer (4× SSC, 0.5% SDS, 1× Denhardt's solution, 0.1 mg/ml herring sperm ssDNA). After 2 h of prehybridization, [³²P]-5'-end radiolabeled probe specific for kanamycin resistance cassette (5'ATATATAGTGGATGTGTCAAACGCATACCATTTTGAACGATGACCTCTAATAATTTGTTAATCATGTTGGT-TACGCTG-3') was added to the prehybridization buffer and hybridization was performed for at least 2 h. Subsequently, the membrane was washed twice with washing solution (4× SSC, 1% SDS) at 42°C for 20 minutes in order to remove unbound probe. The blots were overnight exposed to the phosphor screen and scanned using a STORM imager.

Confirmation of gene deletion by sequencing.

T. thermophilus genomic DNA was isolated from *T. thermophilus* Δ csm1, Δ csm3, Δ csm5 strains liquid cultures using Genomic Mini kit (A&A Biotechnology). Amplification of *csm* loci was carried using genomic DNA from particular *T. thermophilus* deletion strains as a template. PCR reactions were performed in 50 µl of 1× DreamTaq green buffer containing 2 mM MgCl₂, 0.2 mM of each dNTP, 0.5 µM of relevant forward and reverse primers (Table 3), 1.25U DreamTaq DNA Polymerase, 50 ng of genomic DNA. Touchdown PCR was performed with an initial denaturation step at 95°C for 3 min followed by 30 cycles, which involved a denaturation step at 95°C for 30 s, annealing at 65°C for 30 s in the initial cycle and at decreasing temperatures by 0.5°C/cycle until

a temperature of 50°C was reached, extension at 72°C for 1 min. Final extension was performed at 72°C for 10 min. PCR products were resolved on 1% agarose gels alongside with GeneRuler DNA Ladder Mix (Thermo-Scientific) and isolated from a gel using Gel/PCR Mini Kit (Syngen). Sequencing of PCR products was carried by Genomed as a commercial service. PCR products were sequenced using the same primers as for amplification of particular *csm* loci (Table 3).

Purification of crRNA-protein complexes. *T. thermophilus* HB8 cells were harvested by centrifugation. The pellet was suspended in 30 ml of buffer A (5% (v/v) glycerol, 50 mM Hepes (pH 8.0), 50 mM NaCl, 10 mM β -mercaptoethanol). Cells were disrupted in a French press at 40000 psi and the lysate was cleared by ultracentrifugation (20 min; 40000 $\times g$). crRNA-protein complexes were purified from the cell extract by a combination of ion-exchange, affinity and size-exclusion chromatographic steps. After each step the presence of crRNA was tested by the Northern blot analysis and fractions containing crRNA were subjected to further purification steps. First, cell extract was applied on a HiTrap SP HP column (GE Healthcare Life Sciences) equilibrated with buffer A and flow-through was collected. The flow-through was then loaded on a HiTrap Q HP column (GE Healthcare Life Sciences) equilibrated with buffer A. The column was then washed with buffer A and proteins were eluted using 10–100% buffer B (5% (v/v) glycerol, 50 mM Hepes (pH 8.0), 1M NaCl, 10 mM β -mercaptoethanol) gradient in buffer A. Fractions that eluted at a conductivity between 14–23 mS/cm were collected and buffer A was added to the sample to bring the final volume to 50 ml. Subsequently, sample was applied on HiTrap Heparin HP column equilibrated with buffer A. The column was then washed with buffer A. For the elution of proteins, a mix of 80% buffer A with 20% buffer B was used, and the contribution of buffer B to the elution buffer was then gradually increased to 45%. Fractions eluted at conductivity between 15–30 mS/cm were further purified with a flow rate of 0.5 ml/min on a HiPrep 16/60 Sephacryl S-400 HR column (GE Healthcare Life Sciences) equilibrated with buffer A. The volume of collected fractions was 2 ml. Proteins with retention volume in the range of 73–85 ml were pooled together and used for further analysis.

Preparation of samples for Northern blotting. *T. thermophilus* WT and $\Delta csm1$, $\Delta csm3$, $\Delta csm5$ cells were harvested by centrifugation. The total RNA was isolated using TRIzol reagent (ThermoFisher) from approximately 100 mg of the pellet. The RNA isolation was performed according to the protocol provided by the manufacturer. The concentration of RNA was measured on NanoDrop Spectrophotometer (Thermo-Scientific) and the amount of RNA in the sample was standardized (30 μ g of RNA/sample). The samples were mixed with equal amount of formamide loading dye (95% deionized formamide, 0.025% bromophenol blue, 5 mM EDTA pH 8.0) and incubated at 95°C for 10 minutes. In the case of the fractions collected during purification of crRNA-protein complexes from *T. thermophilus* HB8 WT (including cell extract and samples collected after purification on HiTrap SP HP, HiTrap Q HP, HiTrap Heparin HP, HiPrep 16/60 Sephacryl S-400 HR columns) sample preparation was performed as follows. The extracts were concentrated on Vivaspin centrifugal concentrators (Sartorius) with 100 kDa cut-off, standardized with regard to the total protein amount (2000 μ g, 1500 μ g, 200 μ g, 50 μ g and 30 μ g, respectively), mixed with

15 μ l of formamide loading dye and incubated in 95°C for 10 minutes.

Northern blot detection of crRNAs. The RNA containing samples were separated alongside with a [³²P]-5'-end radiolabeled 10 bp DNA Ladder (Invitrogen) on 15% polyacrylamide 7 M urea gels, transferred onto positively charged nylon membrane (Ambion) and immobilized on the membrane by UV cross-linking. Pre-hybridization and hybridization was performed at 42°C in the presence of pre-hybridization buffer (4 \times SSC buffer, 0.5% SDS, 1 \times Denhardt's solution, 0.1 mg/ml herring sperm ssDNA). After 2 h of pre-hybridization, 4 μ l of [³²P]-5'-end radiolabeled probe (20 pmol of crRNA_2.1_probe: 5'-GAAAGAACATGCCGAAGCCGTCGTCGTTGGACT-3' and 20 pmol of crRNA_2.2_probe: 5'-CTTCAGTCCCGGCCCGAGGCCGCTCACGCCCTC-3') or in case of the loading control with the 2 μ l of [³²P]-5'-end radiolabeled probe against 5S rRNA (20 pmol of 5'-AGTACCATCGGCGCTGGCGCGTTTCACTT-C-3' probe) or Leu-rRNA (20 pmol of 5'-TCAAGATGGCGTGTCTACCA-GTTCCACCAC -3') was added to the pre-hybridization buffer and hybridization was performed for at least 2 h. Subsequently, the membrane was washed twice with washing solution (4 \times SSC, 1% SDS) at 42°C for 20 minutes in order to remove unbound probe. The blot was overnight exposed to the phosphor screen and scanned using the STORM instrument.

Estimation of the molecular mass of the crRNA protein complex. The HiPrep 16/60 Sephacryl S-400 HR column was calibrated using thyroglobulin (670 kDa), γ -globulin (158 kDa), albumin fraction V (69 kDa) and myoglobin (17 kDa) as the molecular mass markers. Column calibration was done at a flow rate of 0.5 ml. The retention volume of the crRNA-protein complex (after the HiTrap SP HP, HiTrap Q HP, HiTrap Heparin HP, HiPrep 16/60 Sephacryl S-400 HR purification steps) was verified by Northern blotting against crRNA 2.1 and crRNA 2.2 probe and used to estimate the molecular mass of the complex. Its error was deduced from the error of its logarithm, which in turn was estimated from the error of the retention volume and uncertainty of the calibration curve.

Proteomic analyses – MS and protein identification. Proteins co-purifying with crRNAs were analyzed by liquid chromatography coupled to the mass spectrometer in the Laboratory of Mass Spectrometry, IBB PAS (Warsaw, Poland). Samples were subjected to standard procedure of trypsin digestion, during which proteins were reduced with 10 mM DTT for 30 minutes at 56°C and alkylated with iodoacetamide in darkness for 45 minutes at room temperature and digested overnight with 10 ng/ μ l trypsin. The resulting peptide mixtures were concentrated and desalted on a RP-C18 pre-column (Waters), and further peptide separation was achieved on a nano-UPLC RP-C18 column (Waters, BEH130 C18 column, 75 μ m i.d., 250 mm long) of a nanoACQUITY UPLC system, using a 45 minutes linear acetonitrile gradient. Column outlet was directly coupled to the Electrospray ionization (ESI) ion source of the Orbitrap Velos type mass spectrometer (Thermo), working in the regime of data dependent MS to MS/MS switch. An electrospray voltage of 1.5 kV was used. Raw data files were pre-processed with Mascot Distiller software (version 2.4.2.0, MatrixScience). The obtained peptide masses and fragmentation spectra were matched to the NCBI non-redundant database (57412064 sequences/20591031683 residues), with a *Thermus thermophilus* filter (10650 se-

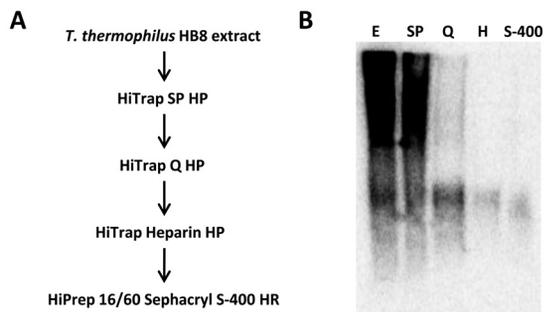


Figure 1. Purification scheme.

(A) Enrichment scheme for crRNAs from the CRISPR-2 locus and bound proteins. After each chromatography step, eluted fractions were analyzed by Northern blotting. A mixture of [³²P]-5'-end labeled oligonucleotides serving as probes for crRNA 2.1 and crRNA 2.2 was used. Fractions with a strong Northern signal were pooled and used in the subsequent purification step. (B) Northern blot signal for the crRNAs present in *T. thermophilus* HB8 cell extract (E) and in the fractions collected after purification on the HiTrap SP HP (SP), HiTrap Q HP (Q), HiTrap Heparin HP (H) and HiPrep 16/60 Sephacryl S-400 HR column (S-400).

quences) using the Mascot search engine (Mascot Daemon v. 2.4.0, Mascot Server v. 2.4.1, MatrixScience). The following search parameters were applied: enzyme specificity was set to trypsin, peptide mass tolerance to ± 30 ppm and fragment mass tolerance to ± 0.6 Da. The protein mass was left as unrestricted, and mass values as monoisotopic with one missed cleavage being allowed. Alkylation of cysteine by carbamidomethylation was set as fixed, and oxidation of methionine and carboxymethylation of lysine were set as variable modifications.

Protein identification was performed using the Mascot search engine (MatrixScience), with the probability based algorithm. Data were searched with automatic decoy database and were filtered to obtain a false discovery rate below 1%.

DNA cleavage assays. The DNase activity of the sample obtained after all chromatographic steps was tested on a set of 5' radio-labeled oligonucleotides (Table 2). The oligonucleotides were used at a final concentration of 0.005 pmol/ μ l. The cleavage reactions contained 4 μ l of the purified sample with a protein concentration of 0.2 mg/ml. The control samples contained water instead of the purified cell extract. The assays were performed for 1 h at 65°C in 20 μ l reaction volume with the standard FastDigest buffer (ThermoScientific) supplemented with 0.5 mM ATP and 0.5 mM GTP. Reactions were stopped by addition of 15 μ l of formamide loading dye (95% deionized formamide, 0.025% bromophenol blue, 5 mM EDTA, pH 8.0) and incubated at 95°C for 10 minutes. Reaction products were resolved on 15% polyacrylamide 7 M urea gels alongside with a [³²P]-5'end radiolabeled GeneRuler Ultra Low Range DNA Ladder (ThermoScientific), transferred onto positively charged nylon membrane (Ambion) and UV cross-linked. The blot was exposed overnight to a phosphor screen and scanned using a STORM instrument.

RESULTS

Chromatographic purification of crRNA-protein complexes

crRNAs 2.1 and 2.2 (from the cluster termed NC_006462_3 in the CRISPRdb (Grissa *et al.*, 2007)

Table 1. Mass spectrometry (MS) results of Cas proteins co-purifying with mature crRNAs from the CRISPR-2 locus.

Fractions collected after Heparin and Sephacryl S-400 chromatographic steps were subjected to mass spectrometry. Cas proteins in the samples were ordered according to their Mascot scores, which indicate the quality of identification (a high score corresponds a confident identification).

Name	Position on the list	Score
Results after the heparin column		
Csm1	57	261
Csm3	58	261
Csm4	135	58
Csm5	141	52
Csm2	143	50
Results after the S-400 column		
Csm3	11	1018
Csm5	24	655
Csm1	32	362
Csm4	37	316
Cmr4	42	264
Csm2	48	195
Cmr3	56	148
Cmr6	65	111
Cmr2	66	108
Cas7	72	85
Cmr1	76	80
Cmr5	77	79

nomenclature) are abundant in *T. thermophilus*, and their abundance increases upon phage infection (Agari *et al.*, 2009). We purified macromolecular complexes of these crRNAs, using a protocol similar to that previously reported for the Cmr complex from *P. furiosus* (Hale *et al.*, 2009). The cell extracts from the WT *T. thermophilus* HB8 strain were fractionated on SP, Q, Heparin and Sephacryl S-400 columns. After each chromatographic step we monitored the presence of crRNA 2.1 and crRNA 2.2 by Northern blot analysis (Fig. 1A). To increase sensitivity of the Northern blot analysis we used probes against crRNAs 2.1 and 2.2 simultaneously. Fractions in which we detected mature crRNAs were subjected to further purification. The Northern blot results show that throughout the purification the mature crRNAs 2.1 and 2.2 were separated from the larger crRNA species (Fig. 1B).

Determination of the complex composition by mass spectrometry

In order to identify proteins which co-migrate with mature crRNAs we analyzed the samples obtained after Heparin and Sephacryl S-400 chromatographic steps by mass spectrometry (Table 1). The analysis performed after Heparin chromatographic step identified 187 proteins including Csm1-5. After the Sephacryl S-400 chromatographic step 117 proteins were detected, including the Csm1-5 and Cmr1-6 proteins. Importantly, the Csm proteins were substantially enriched in the sample collected after chromatography on Sephacryl S-400 column when

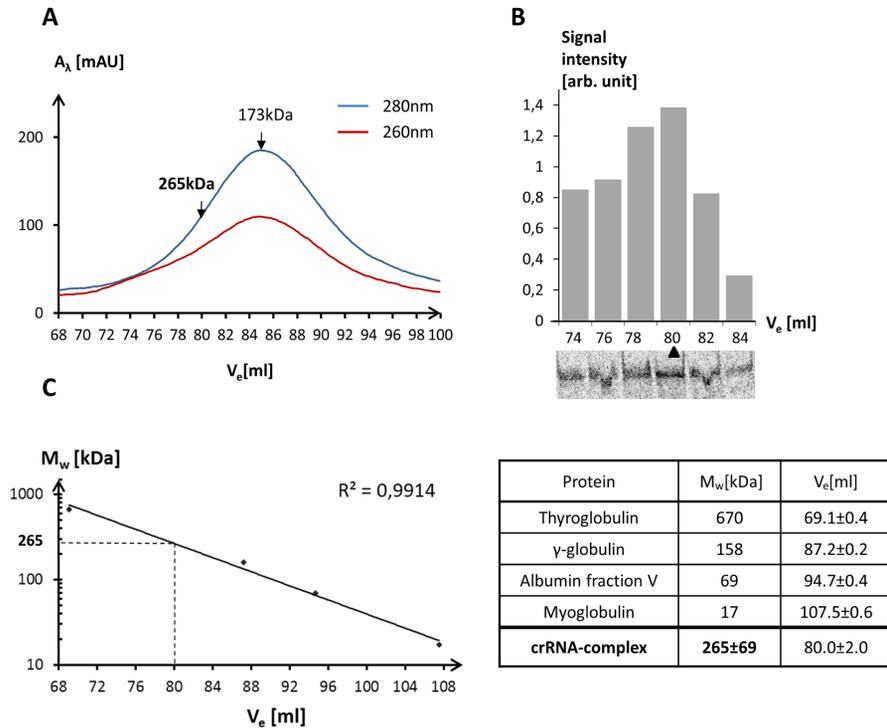


Figure 2. Molecular sizing chromatography.

(A) The elution profile from the Sephacryl S-400 chromatographic column. Molecular mass estimates for the main peak (173 kDa) and for the peak of crRNA elution (265 kDa) are indicated. (B) crRNA elution profile analyzed by Northern blotting with the crRNA 2.1 and crRNA 2.2 probes. As the purification of the endogenous complex was only partial, the main peak of UV absorption (173 kDa) did not coincide with the peak for crRNA elution (265 kDa). The elution volume of mature crRNAs is indicated by an arrow. (C) Estimation of the crRNA-complex molecular mass. The column was calibrated with the indicated set of molecular mass standards and the mass of the crRNA associated complex was obtained assuming a linear relationship between the molecular mass and the logarithm of the retention time. Based on the uncertainty of calibration line and the uncertainty of the retention time, the error of the mass is estimated to be ± 69 kDa. The molecular mass standards and the estimated mass of the crRNA-binding complex are indicated in the table.

compared to the sample obtained after purification on Heparin column (Table 1).

Molecular mass estimation

Next, we attempted to quantify the molecular mass of the crRNA associated complexes by quantifying the data from the size exclusion step (Fig. 2). As the native Csm and Cmr complexes appear to be present in *T. thermophilus* in low abundance, the peak of the UV absorption profile (presumably due to non-CRISPR associated proteins) was found at a lower mass than the maximum signal from Northern blotting, which was determined to correspond to a molecular mass of (265 ± 69) kDa.

Activity assay

Biochemical data available now demonstrate that the Csm complex targets DNA in a strictly transcription dependent manner (Koneremann *et al.*, 2015). Moreover, it can also cleave RNA (Jamulaitis *et al.*, 2014, Staats *et al.*, 2014). A protospacer adjacent motif (PAM) is not required (Marraffini & Sontheimer, 2010b). At the time when the experiments described below were carried out, there was only genetic evidence for DNA targeting (Marraffini & Sontheimer, 2008), but no *in vitro* demonstration of DNA or RNA endonuclease activity. In order to maximize the chances of observing cleavage in assays with single or double stranded DNA as the target, we chose flanking sequences for the protospacers that were previously shown to be present in an *S. epider-*

midis Csm DNA substrate (in the two possible orientations, oligonucleotides assigned as A and B, Fig. 3). The control dsDNA and ssDNA oligonucleotides resembled

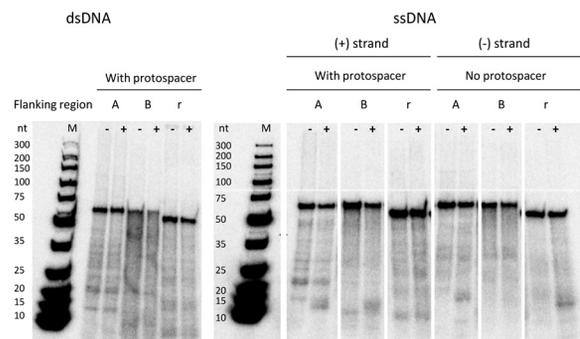


Figure 3. Autoradiogram demonstrating the inactivity of crRNA 2.1 associated complexes against DNA substrates.

The activity assay was performed on 5'-radiolabeled double stranded (dsDNA) and single stranded DNA molecules (ssDNA). Throughout, the extracts were tested after all four chromatography steps (+), in comparison with the addition of no extract (-). Oligonucleotides designated with A and B letters are non-complementary outside the protospacer region and should be cleaved if the crRNA 2.1 associated complexes targeted DNA. All other DNA molecules were negative controls and should not be cleaved, either because they were complementary also to the repeat region (oligonucleotides designated with letter "r"), or because their insert was identical to the crRNA spacer region rather than to the protospacer (no protospacer substrates). The full sequences of test substrates are given in Table 2.

Table 2. DNA molecules used in the activity assay. The (+) strands of all oligodeoxynucleotides contain the region complementary to the crRNA 2.1 spacer sequence, the (-) strands contain regions with sequence identical to the spacer (except for the obvious U to T change).

The spacer and protospacer sequences are indicated by upper case letters. In the control substrate protospacer flanking regions are complementary to the crRNA repeat region (DNA_r). The DNA_A and DNA_B substrates contain non-matching flanks. Although it was not expected that type III systems would require a PAM, we chose these regions identical to *S. epidermidis* protospacer flanking regions in two possible orientations (DNA_A and DNA_B). For comparison, the sequence of crRNA 2.1 is also given at the top of the table.

	Strand	Sequence	Length
crRNA 2.1		5' -auugcgacAUGUGGCAGUCCAACGACGACGGCUUCGGCAUGUUCUUUCGguugc-3'	54 nt
DNA_A	(+)	5' -gattccttaccCGAAAAGAACATGCCGAAGCCGTCGTCGTTGGACTGCCACATtctctaata-3'	61 nt
	(-)	5' -tatttagagaATGTGGCAGTCCAACGACGACGGCTTCGGCATGTTCTTTCGggtagaatc-3'	61 nt
DNA_B	(+)	5' -tatttagagaCGAAAAGAACATGCCGAAGCCGTCGTCGTTGGACTGCCACATggtagaatc-3'	61 nt
	(-)	5' -gattccttaccATGTGGCAGTCCAACGACGACGGCTTCGGCATGTTCTTTCGtctctaata-3'	61 nt
DNA_r	(+)	5' -gcaacCGAAAAGAACATGCCGAAGCCGTCGTCGTTGGACTGCCACATgtcgcaat-3'	54 nt
	(-)	5' -attgcgacATGTGGCAGTCCAACGACGACGGCTTCGGCATGTTCTTTCGggttgc-3'	54 nt

the *T. thermophilus* genomic CRISPR array and contained the protospacer flanked by a part of the genomic repeat sequence (oligonucleotides assigned as “r”, Fig. 3). In the cleavage assay on the ssDNA targets we used the oligonucleotides with sequence identical to the guide crRNA spacer as an additional control (oligonucleotides assigned as “(-) strand”). Since the CRISPR-Cas effector complexes mediate cleavage only if the target is complementary to the guide crRNA such oligonucleotides should not be cleaved. All DNA molecules were then end-labeled radioactively and then incubated with “+” or without “-” crRNA 2.1 associated complexes. No difference was observed (Fig. 3), in agreement with the current view that type IIIA and most likely also type IIIB complexes strictly require transcription of their DNA substrates (Samai *et al.*, 2015).

Gene disruption studies

Next, we tested the effect of *csm* gene deletions on the content of crRNA in *T. thermophilus* extracts. Single *csm* gene deletions were carried out using gene disruption plasmids available from RIKEN (for *csm1*, *csm3* and *csm5*). These plasmids carry kanamycin resistance genes for selection in *T. thermophilus*. They can replicate in *E. coli*, but not *T. thermophilus*, and therefore require an insertion in the latter bacteria to mediate resistance. Flanking regions target the resistance gene to the desired locus. As the resistance gene insertion is not “scar-less”, only single gene mutations were constructed. Kanamycin resistance cassette insertion into the *csm1*, *csm3* and *csm5* was confirmed by Southern blotting (Fig. 4) and DNA sequencing (Table 3). We then probed the crRNA content of the cell extracts (and of extracts of *T. thermophilus* HB27, which lacks the CRISPR-2 array) by Northern blotting with the crRNA 2.1 and 2.2 probes. In contrast to the protein purification experiment, RNAs were pre-purified by the TRIzol method, which led to much sharper RNA bands, but also to a loss of hybridization signal already in the wild-type (Fig. 5).

In the wild-type, the majority of crRNAs 2.1 and 2.2 consisted of between ~45 and ~70 nucleotides, but there were also bands corresponding to smaller RNA molecules (Fig. 5). Our data do not have sufficient

resolution to determine the precise size difference between these bands, the differences may correspond to the 6 nucleotide spacing between crRNA lengths found earlier for the crRNAs of *S. epidermidis* (Hatoum-Aslan *et al.*, 2013) (and the 6 nucleotide spacing in RNA substrates reported recently) (Staals *et al.*, 2014; Gasiunas *et al.*

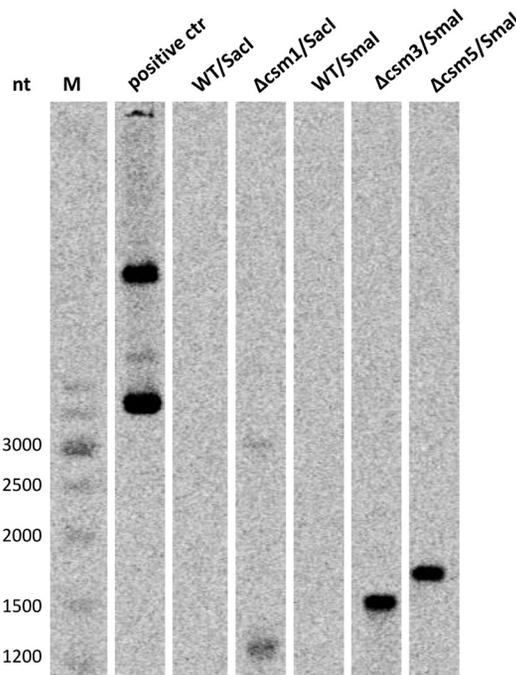


Figure 4. Confirmation of *csm1*, *csm3*, *csm5* gene deletion by Southern blot analysis.

Single *csm* gene deletions were carried out by insertion of the kanamycin resistance cassette into targeted genes. The megaplasmid DNA isolated from *T. thermophilus* strains was digested with either R.SacI or with R.SmaI enzyme and tested for the presence of kanamycin resistance cassette. The kanamycin carrying genome fragments were expected to be 1344 (for the $\Delta csm1$ strain), 1533 ($\Delta csm3$ strain) and 1715 nt ($\Delta csm5$ strain) base pairs long. Megaplasmid DNA from the wild type strain was used as a negative control. Positive ctr – *csm1* disruption plasmid; nt – number of nucleotides; M – molecular mass marker.

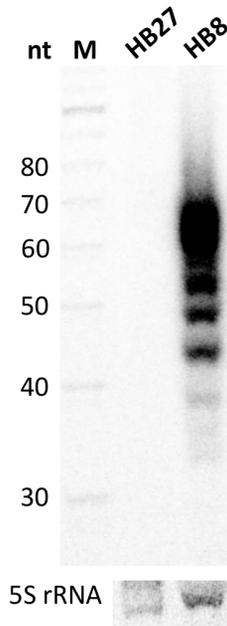


Figure 5. Expression of crRNAs from the CRISPR-2 locus in *T. thermophilus*.
Total RNA was isolated from the *T. thermophilus* HB8 and HB27 (which lacks the CRISPR-2 locus) using TRIzol reagent. The crRNA content was then visualized by Northern blotting using a mixture of the crRNA 2.1 and crRNA 2.2 probes. Two probes were used simultaneously to increase sensitivity of the analysis. Detection of 5S rRNA is shown as loading control. nt – number of nucleotides; M – molecular mass marker.

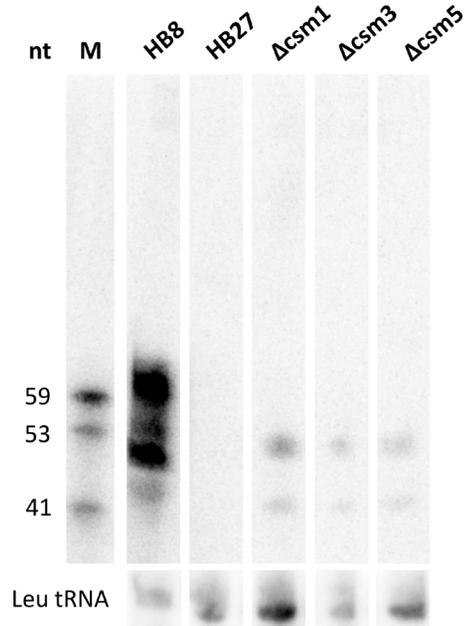


Figure 6. Effect of *csm* gene deletion on crRNAs from the CRISPR-2 locus in *T. thermophilus*.
Total RNA was isolated from the *T. thermophilus* HB8, HB27 and the otherwise isogenic Δ*csm*1, Δ*csm*3, Δ*csm*5 strains using TRIzol reagent. The crRNA content was then visualized by Northern blotting using a mixture of the crRNA 2.1 and crRNA 2.2 probes. Two probes were used simultaneously to increase sensitivity of the analysis. Detection of Leu-tRNA is shown as loading control. nt – number of nucleotides; M – molecular mass marker.

Table 3. Confirmation of *csm*1, *csm*3, *csm*5 gene disruption by sequencing of relevant *csm* loci.

Disrupted gene	Primer sequence	Sequence of the relevant <i>csm</i> locus in the mutant strain
<i>csm</i> 1	F 5'GGGAGGTG-CAGGTGCTG-TGG-3'	5' ATGCTTGGGGACGGCCTGAGCGTtctagacgttgacggcggatgatggtacaatgaacatggagacatg cagggttttctctgcatttttctttgtgcactcatcaagcgattttcggtaaagctaggaggaatga
	R 5'GGGCCGCT-CAAAGACCT-TGG-3'	ggatgtgtcaaaaacgcataccattttgaacgatgacctctaataattgttaatcatgttggttacgctgc agTTTACCGGAAAGGAGGGTGTGA-3'
-----897 nucleotides-----		
<i>csm</i> 3	F 5'AGGCG-GTGCTCGCC-TACTTCTAC-3'	5' ATGAAGCTCAAGAAGGTGATCggtaccggttgacggcggatgatggtacaatgaacatggagacatgca ggggttttctctgcatttttctttgtgcactcatcaagcgattttcggtaaagctaggaggaatgatg
	R 5'GGGCCT-TGGAAGTA-GAGGCGGA-AGAG-3'	atgtgtcaaaaacgcataccattttgaacgatgacctctaataattgttaatcatgttggttacgctgcag GTGGAGGAAGTGGTCTTTAG-3'
-----897 nucleotides-----		
<i>csm</i> 5	F 5'CACCGC-CTGCTGGAG-GTGAC-3'	5' GTGGGAGGCCCTCGTGGTCTTTTctctagacgttgacggcggatgatggtacaatgaacatggagacatg cagggttttctctgcatttttctttgtgcactcatcaagcgattttcggtaaagctaggaggaatga
	R 5'CCCGGTAC-CGCTCCA-AAGGG-3'	ggatgtgtcaaaaacgcataccattttgaacgatgacctctaataattgttaatcatgttggttacgctgc agTGGGGAGGCTAGAGCCCTCTAG3'

Confirmation of *csm*1, *csm*3, *csm*5 gene disruption by sequencing of the relevant *csm* loci. Single *csm* gene deletions were carried out by insertion of the kanamycin resistance cassette into targeted genes. The *csm*1, *csm*3, *csm*5 loci were amplified from genomic DNA of mutant *Thermus thermophilus* HB8 strains (Δ*csm*1, Δ*csm*3, Δ*csm*5, respectively) and sequenced. The primers used for amplification and sequencing of *csm* loci were designed to anneal in the regions flanking relevant *csm* genes. F – forward primer; R – reverse primer; upper case letters – fragments of the targeted gene left after gene disruption; lower case letters – kanamycin resistance cassette; black – region covered by sequencing; underlined region not covered by sequencing. Note that the *csm*5 open reading frame starts with GTG. The first few nucleotides in the kanamycin resistance gene in Δ*csm*3 strains are different from those in the Δ*csm*1 and Δ*csm*5 strains, the reminders of the resistance genes are identical.

al., 2012). As a result of the gene deletions, we observe a massive reduction in the amounts of the crRNA species (Fig. 6).

DISCUSSION

Length distribution of crRNAs detected by crRNA 2.1 and 2.2 probes

Sequencing studies of crRNAs from *T. thermophilus* extracts and crRNA fractions isolated from Csm and Cmr complexes show consistently that the precursor crRNAs are processed between the -8 and -9 positions in the repeat, so that mature crRNAs with a repeat derived 8-nucleotide handle are formed. Despite the generality of this rule even for repeats of different types (which differ slightly in the repeat sequence even in this "handle" region), crRNAs bound to Csm complexes were found to have a variety of lengths, with some lengths (45 nucleotides, 53 nucleotides) represented more often than others (Staals *et al.*, 2014). crRNAs bound to the Cmr protein complex were less diverse (Staals *et al.*, 2013). In line with the latter result, we observe a defined set of crRNA lengths in our experiments that only probe crRNAs with two specific spacers from a single cluster, in agreement with earlier studies using a single spacer probe (see Fig. 6 of Juranek and coworkers (2012), the CRISPR cluster pTT27_3, referred to as cluster 2 in Staals and coworkers (2014)).

The CRISPR cluster 2 (pTT27_3) contains three repeats with two spacers only. The repeats are 36 nucleotides long, the spacers in between them have lengths of 41 and 40 nucleotides. A single cleavage at the -8 position of the repeat should therefore result in crRNAs of 77 and 76 nucleotides length. Complete removal of the repeat portion at the 3'-end of the crRNA would take off 28 nucleotides (36-8 nt for the 5'-handle), and hence result in crRNA lengths of 49 or 50 nucleotides. Any smaller crRNA must not only lack the repeat derived sequence at the 3'-end, but also part of the spacer, or alternatively, must have an unusual 5'-end. With the pTT27_3.1 probe, Juranek and coworkers (2012) detect a faint crRNA Northern blot band at around 70 nt, and several additional bands between 40 and 60 nucleotides (Juranek *et al.*, 2012). As in the previous work, we observe distinct bands, rather than a continuous size distribution. Assuming 8 nucleotides of repeat at the 5'-end, the main crRNA bands in the HB8 wild-type strain could result from slight to complete trimming of the repeat 3'-ends. Some trimming into the spacer may occur as well, in line with conclusions for other type III systems (van der Oost *et al.*, 2014) (Fig. 5).

Effect of *csm* gene disruption on crRNAs 2.1 and 2.2

The crRNA pool of *T. thermophilus* has been extensively characterized, but the proteins involved in their biogenesis have not been identified. In contrast, extensive work on crRNA biogenesis has been carried out for the type IIIA CRISPR system from *Staphylococcus epidermidis* (Hatoum-Aslan *et al.*, 2011; Hatoum-Aslan *et al.*, 2013; Hatoum-Aslan *et al.*, 2014). These studies have shown that Cas6 is responsible for the primary cleavage of crRNA but may require activation by Csm1 or Csm4 (Hatoum-Aslan *et al.*, 2011; Hatoum-Aslan *et al.*, 2014). The studies further showed that the primary processing product of crRNA is then trimmed further in a manner that requires a crRNA length measurement from

the 5'-end primary processing site (Hatoum-Aslan *et al.*, 2011). Gene deletion experiments in *S. epidermidis* suggest that this step requires the *csm2*, *csm3* and *csm5* genes, because their ablation leads to accumulation of the primary processing product (Hatoum-Aslan *et al.*, 2011). Further studies, again using *S. epidermidis* as the model, showed that the crRNA molecules generated by the primary cleavage were then processed further so that a series of RNA molecules differing between each other by six nucleotides in length were formed (Hatoum-Aslan *et al.*, 2013). The correlation between the length of crRNA and the number of bound Csm3 molecules further suggested that a Csm3 "helix", now known to lie at the core of the Csm complex (Staals *et al.*, 2014), binds crRNA in such a manner that each Csm3 subunit associates with 6 nucleotides of crRNA, which could explain the characteristic 6 nucleotide steps in crRNA sizes (irrespective of whether Csm3, now known to cleave target crRNA, or another nuclease is responsible for the post-processing of Cas6 cleaved crRNAs) (Hatoum-Aslan *et al.*, 2013).

In *T. thermophilus* HB8, we observe a massive reduction in the amounts of all but two smaller RNA species upon deletion of the *csm1*, *csm3* or *csm5* genes. We suspect that type III CRISPR systems are not only involved in the post-processing of crRNAs, but also protect them from other RNases. Alternatively, it is also possible that insertion of the resistance cassette in some *csm* genes might affect the expression of the operon and thus of other *csm* genes. Two bands in the Northern blot persist at least to some extent in the absence of the *csm* genes, perhaps due to the presence of other CRISPR systems in *T. thermophilus* (which is not the case in *S. epidermidis*) (Fig. 6).

Molecular mass of crRNAs 2.1 and 2.2 associated complexes

In this work, we used an unbiased approach to purify proteins or protein complexes that are associated with the two crRNAs from a particular crRNA cluster (cluster 2 in the nomenclature of Staals *et al.*), which had previously been shown to be abundantly expressed. After three initial purification steps, we were able to detect by mass spectrometry all five Csm subunits (Csm1-5) with significant scores, and after an additional gel filtration step, we could detect both Csm and Cmr subunits in our preparations. The large improvement in the rank score of the peptides as well as the inclusion of the Cmr proteins most likely reflects the substantial enrichment by a chromatography step that excludes most *T. thermophilus* proteins that are smaller than the purified complexes. When this work was originally carried out, the crRNA bound to Csm and Cmr complexes of *T. thermophilus* had not yet been sequenced, and it was surprising that the crRNA was found to be associated with both complexes. However, it is now clear that crRNAs with spacers that are highly represented in Csm complexes are also abundant in Cmr complexes and *vice versa*. Hence, the detection of the crRNA 2.1 and 2.2 with both complexes is now expected. As Csm complexes seem to dominate over Cmr complexes in our enrichment scheme, the molecular mass determined by gel filtration should more closely match the *T. thermophilus* Csm mass than the reported Cmr masses.

The most recent mass estimate for the full Csm complex determined by mass spectrometry is about 427 kDa, but several smaller fragments of the complex were also detected (Staals *et al.*, 2014). Our result, 265 ± 69 kDa, is smaller than the previous value for the full complex

determined by mass spectrometry. However, within error margin, our mass estimate for the *T. thermophilus* complex agrees with the mass of the Csm complex (331 kDa) (Hatoum-Aslan *et al.*, 2013) from *S. epidermidis*, which also has one paralogue of the *csm* genes in its genome (unlike *S. solfataricus*). As in our case, the estimate for the *S. epidermidis* complex was made by gel filtration. It is therefore possible that gel filtration leads to systematically too small masses for the Csm complexes. However, we consider this unlikely because the Csm complexes are known to be elongated and therefore their apparent mass should exceed their theoretical mass. An alternative explanation could be that during the purification procedure, some more peripheral subunits may be lost. This explanation appears plausible in the light of the new structural data for the Csm complex (Staals *et al.*, 2014) and need not contradict the observation that all subunits are detectable by the mass spectrometry, if the population of subcomplexes is heterogeneous. Another explanation for the smaller mass of the crRNA 2.1 and 2.2 associated complex compared to the reported mass for the *T. thermophilus* Csm complex could be a larger contribution from the Cmr complex than mass spectrometry suggests. The Cmr complex from *T. thermophilus* was previously found to be heterogeneous with molecular masses of 310 and 350 kDa, depending on the species of bound crRNA (Staals *et al.*, 2013). At least the lower mass is compatible with the mass reported in this work.

Activity of the complexes associated with crRNA 2.1

The lack of activity of the purified fraction against DNA with crRNA 2.1 protospacers was surprising when the experiments were performed because genetic data suggested that the Csm complexes should target DNA. However, with the benefit of hindsight, the lack of activity against DNA substrates is now fully consistent with the recent discovery of strictly transcription dependent DNA cleavage activity of type III CRISPR systems (Koneremann *et al.*, 2015; Deng *et al.*, 2013). The easiest explanation for transcription dependence of DNA cleavage by Csm (and possibly Cmr) complexes appears to be that transcription generates single stranded DNA, which should anneal more easily with crRNA than double stranded DNA. However, experimental data by Samai and colleagues show that surprisingly, single stranded DNA is also not cleaved in the absence of transcription (Koneremann *et al.*, 2015), in agreement with the observations in this work. The reasons why single stranded DNA must be transcribed to be a substrate for type III CRISPR complexes remain to be understood.

Acknowledgements

We thank the Faculty of Biotechnology and Food Sciences of the Technical University of Łódź for allowing M.S. to carry out her M.Sc. thesis work at the International Institute of Molecular and Cell Biology. The RIKEN consortium is gratefully acknowledged for the *csm* gene deletion vectors. Prof. Michal Dadlez and his colleagues performed the mass spectrometry work as a commercial service. Dr. Ewa Sitkiewicz helped with the interpretation of the mass spectrometry results. We are grateful to Dr. Honorata Czapinska for critical reading of the manuscript.

This work was supported by grants N N302 654640 and UMO-2011/02/A/NZ1/00052 from the Polish National Science Centre to MB and UMO-2011/01/D/NZ6/00269 to KP.

REFERENCES

- Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A (2009) Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J Mol Biol* **395**: 270–281. doi: 10.1016/j.jmb.2009.10.057.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–1712. doi: 10.1126/science.1138140.
- Deng L, Garrett RA, Shah SA, Peng X, She Q (2013) A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Mol Microbiol* **87**: 1088–1099. doi: 10.1111/mmi.12152.
- Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109**: E2579–E2586. doi: 10.1073/pnas.1208507109.
- Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172. doi: 10.1186/1471-2105-8-172.
- Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**: e60. doi: 10.1371/journal.pcbi.0010060.
- Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945–956. doi: 10.1016/j.cell.2009.07.040.
- Hashimoto Y, Yano T, Kuramitsu S, Kagamiyama H (2001) Disruption of *Thermus thermophilus* genes by homologous recombination using a thermostable kanamycin-resistant marker. *FEBS Lett* **506**: 231–234. doi: 10.1016/S0014-5793(01)02926-X.
- Hatoum-Aslan A, Maniv I, Marraffini LA (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc Natl Acad Sci USA* **108**: 21218–21222. doi: 10.1073/pnas.1112832108.
- Hatoum-Aslan A, Samai P, Maniv I, Jiang W, Marraffini LA (2013) A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J Biol Chem* **288**: 27888–27897. doi: 10.1074/jbc.M113.499244.
- Hatoum-Aslan A, Maniv I, Samai P, Marraffini LA (2014) Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *J Bacteriol* **196**: 310–317. doi: 10.1128/JB.01130-13.
- Henne A, Bruggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martinez-Arias R, Jacobi C, Starkuviene V, Schlenczek S, Dencker S, Huber R, Klenk HP, Kramer W, Merkl R, Gottschalk G, Fritz HJ (2004) The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat Biotechnol* **22**: 547–553. doi: 10.1038/nbt956.
- Jansen R, Embden JD, Gaastera W, Schouls LM (2002a) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **43**: 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.
- Jansen R, van Embden JD, Gaastera W, Schouls LM (2002b) Identification of a novel family of sequence repeats among prokaryotes. *OMICS* **6**: 23–33. doi: 10.1089/15362310252780816.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–821. doi: 10.1126/science.1225829.
- Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Wagmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ (2013) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**: 529–536. doi: 10.1038/nsmb.2019.
- Juranek S, Eban T, Altuvia Y, Brown M, Morozov P, Tuschl T, Margalit H (2012) A genome-wide view of the expression and processing patterns of *Thermus thermophilus* HB8 CRISPR RNAs. *RNA* **18**: 783–794. doi: 10.1261/rna.031468.111.
- Karvelis T, Gasiunas G, Miksys A, Barrangou R, Horvath P, Siksnys V (2013) crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol* **10**: 841–851. doi: 10.4161/rna.24203.
- Koneremann S, Brigham MD, Trevino AE, Joung J, Abudayeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**: 583–588. doi: 10.1038/nature14136.
- Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**: 467–477. doi: 10.1038/nrmicro2577.

- Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**: 722–736. doi: 10.1038/nrmicro3569.
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823–826. doi: 10.1126/science.1232033.
- Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843–1845. doi: 10.1126/science.1165771.
- Marraffini LA, Sontheimer EJ (2010a) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**: 181–190. doi: 10.1038/nrg2749.
- Marraffini LA, Sontheimer EJ (2010b) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**: 568–571. doi: 10.1038/nature08703.
- Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174–182. doi: 10.1007/s00239-004-0046-3
- Oshima T, Imahori K (1974) Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. *Int J Syst Bacteriol* **24**: 102–112. doi: 10.1099/00207713-24-1-102.
- Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, Cannone G, Graham S, Robinson CV, Spagnolo L, White MF (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell* **52**: 124–134. doi: 10.1016/j.molcel.2013.08.020.
- Samai P, Pyenson N, Jiang W, Goldberg GW, Hatoum-Aslan A, Marraffini LA (2015) Co-transcriptional DNA and RNA cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**: 1164–1174. doi: 10.1016/j.cell.2015.04.027.
- Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P, Siksnys V (2013) *In vitro* reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J* **32**: 385–394. doi: 10.1038/emboj.2012.352.
- Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A, Dohmae N, Schaap PJ, Doudna JA, Heck AJ, Yonekura K, van der Oost J, Shinkai A (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* **52**: 135–145. doi: 10.1016/j.molcel.2013.09.013.
- Staals RH, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, Koehorst JJ, Vlot M, Neupane N, Varossieau K, Sakamoto K, Suzuki T, Dohmae N, Yokoyama S, Schaap PJ, Urlaub H, Heck AJ, Nogales E, Doudna JA, Shinkai A, van der Oost J (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol Cell* **56**: 518–530. doi: 10.1016/j.molcel.2014.10.005.
- Tamulaitis G, Kazlauskienė M, Manakova E, Venclovas C, Nwokeoji AO, Dickman MJ, Horvath P, Siksnys V (2014) Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Mol Cell* **56**: 506–517. doi: 10.1016/j.molcel.2014.09.027.
- van der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **12**: 479–492. doi: 10.1038/nrmicro3279.
- Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**: 486–489. doi: 10.1038/nature10402.
- Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331–338. doi: 10.1038/nature10886.
- Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, Wei Z, Wang Y (2014) Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* **515**: 147–150. doi: 10.1038/nature13733.