The competition between viruses and hosts is played out in all branches of life. Many prokaryotes have an adaptive immune system termed ‘CRISPR’ (clustered regularly interspaced short palindromic repeats) which is based on the capture of short pieces of viral DNA. The captured DNA is integrated into the genomic DNA of the organism flanked by direct repeats, transcribed and processed to generate crRNA (CRISPR RNA) that is loaded into a variety of effector complexes. These complexes carry out sequence-specific detection and destruction of invading mobile genetic elements. In the present paper, we report the structure and activity of a Cas6 (CRISPR-associated 6) enzyme (Sso1437) from *Sulfolobus solfataricus* responsible for the generation of unit-length crRNA species. The crystal structure reveals an unusual dimeric organization that is important for the enzyme’s activity. In addition, the active site lacks the canonical catalytic histidine residue that has been viewed as an essential feature of the Cas6 family. Although several residues contribute towards catalysis, none is absolutely essential. Coupled with the very low catalytic rate constants of the Cas6 family and the plasticity of the active site, this suggests that the crRNA recognition and chaperone-like activities of the Cas6 family should be considered as equal to or even more important than their role as traditional enzymes.

Key words: antiviral defence, Cas6, clustered regularly interspaced short palindromic repeats (CRISPR), ribonuclease, *Sulfolobus*.

INTRODUCTION

The CRISPR (clustered regularly interspaced short palindromic repeats) system is an adaptive antiviral defence system found in many bacteria and most archaea. In organisms with an active CRISPR system, invading viral DNA can be captured and incorporated into the genome. This process, known as adaptation, requires the Cas (CRISPR-associated) proteins Cas1 and Cas2 and results in the insertion of a short piece (typically 25–50 bp) of foreign DNA into a region of the genome known as the CRISPR locus [1,2]. The inserted sequence, known as a ‘spacer’, is flanked by direct repeat sequences in an array; a locus can grow to comprise >100 repeat–spacer units. A promoter in the CRISPR locus drives transcription of the array, generating a pre-crRNA (crRNA is CRISPR RNA) that is processed into unit-length mature crRNA species [3–5]. These are loaded into effector complexes, which use the crRNA sequence to detect and degrade invading DNA or RNA during interference, providing immunity from infection (reviewed in [6]).

Three main types of CRISPR system (I, II and III) have been defined on the basis of the presence of the key protein subunits Cas3, Cas9 and Cas10 respectively [7]. These are defined further into specific subtypes, such as type I-A–I-F, depending on the particular proteins present. In type I and III systems, pre-crRNA is processed by the Cas6 endonuclease, whereas type II systems utilize an alternative method involving host RNase III [8]. Cas6 is thus a key component of the majority of CRISPR effector systems. Cas6 proteins typically consist of two tandem RRM (RNA-recognition motif)-like folds [also known as RAMP (repeat-associated mysterious protein) domains] [9] with a diagnostic glycine-rich loop near the C-terminus. Although this core fold is conserved, there are significant differences in Cas6 enzymes from different species. Many, such as TiCas6 (Thermus thermophilus Cas6, also known as Cse3) and PfuCas6 (Pyrococcus furiosus Cas6, also known as Csy4) recognize an RNA hairpin structure formed by the CRISPR repeat [10,11]. RNA cleavage typically occurs at the base of the hairpin, whereupon the enzyme remains tightly bound to the product and chaperones it to the effector complex. In most cases, these Cas6 enzymes are integral subunits of the effector complexes and catalyse only single-turnover cleavage of the pre-crRNA [12].

In many of the archaea, CRISPR repeats are not palindromic and are thus unable to form stable hairpin structures [13]. The structure of PfuCas6 (*Pyrococcus furiosus* Cas6) in complex with repeat RNA revealed that the RNA is wound around the outside of the enzyme between the two RAMP domains, analogous to the string around a yo-yo [14]. The pre-crRNA is engaged in a binding cleft on one side of the protein and cleaved in the active site on the other side of the enzyme. However, it is unclear whether this is a general mode for binding unstructured RNA in archaeal Cas6 homologues. For the archaeal type I-A and III-B systems, Cas6 appears to be more loosely associated with the effector complexes [15–17] and should, in theory, be capable of true multiple turnover to generate crRNAs for different clients.

All Cas6 variants studied to date are monomeric proteins with an active-site histidine side chain that is thought to act as a general acid or base [12] during the catalytic cycle. In the present paper, we report the crystal structure and accompanying biochemical data for Cas6 from the crenarchaeon *Sulfolobus solfataricus*. This enzyme has a novel dimeric arrangement that appears to be important for catalysis, as a monomeric variant is significantly less active. Furthermore: 5. *solfataricus* Cas6, in common with...
many other crenarchaeal enzymes, lacks the ‘essential’ histidine moiety. This suggests a different mechanism of catalysis, which is probed by site-directed mutagenesis.

EXPERIMENTAL

Cloning and site-directed mutagenesis

The sso1437 gene was amplified by PCR from S. solfataricus genomic DNA and cloned into the pET151-TOPO plasmid by TOPO cloning with a cleavable N-terminal His6-tag as described previously [18]. Site-directed mutagenesis was performed following standard protocols (QuickChange®, Stratagen). The sequences of oligonucleotides used for cloning and mutagenesis are available from M.F.W. upon request.

Expression and purification

Sso1437 [SsoCas6 (S. solfataricus Cas6)] was expressed in C43(DE3) Escherichia coli cells in LB (Luria–Bertani) medium at 37°C until reaching a D600 of 0.6, followed by induction with 1 mM IPTG (isopropyl β-d-thiogalactopyranoside) and overnight incubation at 25°C. Cell pellets were resuspended in PBS with 1 M NaCl, 10% glycerol, 9 mM 2-mercaptoethanol, 10 μg·ml⁻¹ lysozyme, 0.05 unit·ml⁻¹ DNase I and EDTA-free protease inhibitor tablets (Roche). The cells were lysed using a cell disruptor (Constant Systems) at 30 000 psi (1 psi = 6.9 kPa) and the lysate was cleared by centrifugation at 40 000 g for 45 min at 25°C. The soluble fraction was applied to an Ni-NTA (Ni²⁺-nitrilotriacetic acid) column (Qiagen), washed with 30 mM imidazole and eluted with 400 mM imidazole. The sample was dialysed into PBS with 1 M NaCl, 10% glycerol and the His6-tag cleaved with TEV (tobacco etch virus) protease overnight. The sample was reapplied and washed through the Ni-NTA column equilibrated in 20 mM Tris/HCl (pH 7.5), 1 M NaCl and 10% glycerol. The sequences of oligonucleotides used for cloning and mutagenesis are available from M.F.W. upon request.

Reductive methylation of surface lysine residues

Selenomethionine-labelled SsoCas6 was expressed in B834(DE3) E. coli cells in M9 medium supplemented with Selenomethionine Nutrient Mix (Molecular Dimensions) and 50 mg·l⁻¹ (L)-selenomethionine. The expression protocol was as above except that the cells were harvested after 30 h. The sample was purified as above except that 1 mM 2-mercaptoethanol was added to all buffers.

Structural biology

Optimized crystals of selenomethionine-labelled methylated SsoCas6 were grown under conditions of 0.05 M bicine (pH 9.1) and 28% PEG [poly(ethylene glycol)] 3350 using vapour diffusion with a protein/precipitant ratio of 1:1. The crystal was cryoprotected in mother liquor supplemented with 20% glycerol and flash-cooled in liquid nitrogen. A SAD (single-wavelength anomalous diffraction) dataset was collected on the ESRF (European Synchrotron Radiation Facility, Grenoble, France) ID14-4 beamline at the Se–K absorption edge using a single crystal cooled to 100 K. The data were processed with xia2 [19] using XDS [20] and SCALA [21] and the heavy atom sites located using Phenix AutoSol [22]. The map was improved further using Parrot [23] and the chains were traced using Buccaneer [24], both as part of the CCP4 suite [25]. The structure was refined with cycles of correction in COOT [26] and refinement with REFMAC version 5.6 [27] using TLS (Translation–Libration–Screw-rotation) parameters generated by the TLSMD server [28,29]. The side chains of residues in β₂ and the β₂–β₃ loop were disordered and real-space averaging, using AVE as part of the Uppsala Software Factory [30], was required to assign the residues correctly. Methyl groups were not modelled on to the lysine residues owing to insufficient density. The structure was validated using the MolProbity server [31]. Data collection and refinement statistics are presented in Table 1. The co-ordinates and data for the structure of SsoCas6 were deposited in the PDB under code 3ZFV.

Table 1 Data collection and refinement statistics for the structure of selenomethionine-labelled methylated SsoCas6

| Parameter                      | SsoCas6 |
|--------------------------------|---------|
| Data processing                |         |
| Wavelength (Å)                 | 0.979   |
| Space group                    | P2₁     |
| a, b, c (Å)                    | 71.7, 127.5, 83.6 |
| α, β, γ (°)                    | 90.0, 110.5, 90.0 |
| Resolution                     | 73.32–2.80 (2.87–2.80) |
| Rmerge                         | 0.07 (0.71) |
| I/σI                           | 23.3 (3.7) |
| Completeness                   | 99.9 (99.1) |
| Multiplicity                   | 10.4 (10.2) |
| Anomalous completeness         | 99.7 (98.5) |
| Anomalous multiplicity         | 5.3 (5.2) |
| Refinement                     |         |
| Rwork/Rfree                    | 0.20/0.24 |
| Mean B-value (Å²)              |         |
| All atoms                      | 33      |
| Protein                        | 33      |
| Water                          | 43      |
| Glycerol                       | 47      |
| RMSD                           |         |
| Bond lengths (Å)               | 0.01    |
| Angles (°)                     | 1.45    |

RNA oligonucleotide purification, end labelling and marker ladder

The RNA sequence corresponding to the repeat of the C and D CRISPR loci of S. solfataricus P2 was ordered from IDT, with the sequence 5′-GAUAAUCUUAAUGAAGUGAAAG-3′. The oligonucleotide was gel-purified and end-labelled with γ-³²P]ATP as described previously [15]. The RNA marker ladder was generated by alkaline hydrolysis as described previously [16].

Single-turnover endonuclease assays

SsoCas6 protein (0.5 μM) was incubated with 1–5 nM labelled RNA in nuclease assay buffer [20 mM sodium phosphate...
(pH 7.5), 100 mM potassium glutamate, 5 mM EDTA and 0.5 mM DTT (dithiothreitol) at the temperature indicated. At relevant time points, 10 μl samples were removed from the main reaction volume and quenched by addition to 30 μl of pre-distributed acid phenol/chloroform (Ambion), vortex-mixed for ∼10 s and centrifuged at 15 000 g for 1 min. Then, 5 μl of the upper aqueous phase was removed and mixed 1:1 with formamide. Samples were heated at 95 °C for 2 min immediately before loading on to a prewarmed 20% denaturing polyacrylamide gel [20% acrylamide, 8 M urea and 1 × TBE (45 mM Tris/borate and 1 mM EDTA)] and the products were separated by electrophoresis at 80 W for 90 min in 1 × TBE running buffer. Following electrophoresis, gels were scanned by phosphorimaging and analysed using Fuji Imagegauge software as described previously [32].

Thermofluor assay for protein stability
Protein (SsoCas6 wild-type or SsoCas6-L170D/V202D) (5 μM) was incubated in nuclease assay buffer supplemented with 2× SyproOrange® (Invitrogen). Volumes of 100 μl were accommodated in a 96-well polypropylene plate (Agilent Technologies) and covered with an optically clear adhesive film (Molecular Dimensions). Plates were spun briefly at 1500 g before commencing the assay. The temperature was raised from 25 to 95 °C in 0.5 °C increments, 1 min cycles and the fluorescence levels were monitored in a QPCR System (Stratagene® MX3005P™) using a FFROX filter set with excitation and emission wavelengths of 492 and 610 nm respectively. Post-assay manipulation of data was undertaken using the DFS analysis (version 2.5) tool developed (and kindly provided) by Niesen et al. [33] (downloadable at ftp://ftp.sgc.ox.ac.uk/pub/biophysics). The inflection point of the sigmoidal monophasic thermal profile of the protein was taken to be the melting temperature (T_m) of the protein [33].

Calibrated gel filtration
SsoCas6 (the K28H variant, which has the same elution profile as that of wild-type protein) and SsoCas6-L170D/V202D were applied to a HiLoad 26/600 Superdex 200™ gel-filtration column (GE Healthcare) equilibrated in 20 mM Tris/HCl (pH 7.5) and 1 M NaCl. The column was calibrated using gel-filtration standards (Bio-Rad Laboratories) and the molecular masses were estimated as described previously [34].

RESULTS
Sso1437: an authentic Cas6 enzyme
The genome of S. solfataricus contains five Cas6 paralogues, one of which (Sso2004) was shown previously to be an active ribonuclease by Lintner et al. [15]. Sso2004 shares 90% sequence identity with its parologue Sso1437, hereafter called SsoCas6. This parologue was cloned and expressed in E. coli with an N-terminal polyhistidine tag. SsoCas6 was purified to homogeneity by a combination of immobilized metal-affinity chromatography and gel filtration, and the polyhistidine tag was removed using TEV protease as described previously [35]. The purified protein was assayed for ribonuclease activity using a 5′-32P-end-labelled oligonucleotide corresponding to the CRISPR RNA repeat sequence, 5′-GAUAUACUCUUAUAGAUAUAGAG-3′, which matches the CRISPR repeat found in the C and D loci of the S. solfataricus P2 genome [36]. Cleavage occurred specifically 8 nucleotides from the 3′ end, generating the conserved ‘5′ handle’ motif [15]. The single-turnover catalytic rate constant for Sso1437 was measured at a range of temperatures from 20 to 80 °C (Figure 1). As expected for an enzyme from a thermophile, maximal activity was observed at 70 °C, close to the optimum growth temperature of the organism of 80 °C. Reaction rates for thermobstable enzymes from S. solfataricus typically increase 2-fold for each 10 °C rise in incubation temperature [37]. This pattern was followed by SsoCas6 up to 60 °C, with a modest increase at 70 °C and a significant fall in activity thereafter, suggesting that the protein is heat-denatured under these conditions in vitro at temperatures above 70 °C. The first-order reaction rate of the enzyme in vivo at 80 °C may thus approach 3–4 min⁻¹. SsoCas6 was studied further using a combination of crystallography and site-directed mutagenesis to elucidate its structure and catalytic mechanism.

The crystal structure of SsoCas6
Although native SsoCas6 crystallized readily, reductive methylation of the surface lysine residues was required to produce well-diffracting crystals. A selenomethionine derivative of the methylated protein was used to solve the structure with a SAD dataset collected to 2.8 Å (1 Å resolution). Data collection and refinement statistics are presented in Table 1. Residues 43–50 and 90–93 are disordered in each of the four monomers in the crystallographic asymmetric unit.

Structural homology with other Cas6 proteins was determined with the PDBeFold server [38], which identified the non-catalytic P. furiosus Cas6 (PfuCas6nc) as the most structurally similar [PDB code 3UFC, Cα RMSD (root mean square deviation) of 3.4 Å over 187 residues], followed by its catalytic parologue PfuCas6 (PDB code 3I4H, RMSD 3.4 Å over 179 residues). The monomer consists of eleven β-strands and eight helices arranged as two tandem RAMP folds linked by a single helix, consistent with the canonical Cas6 fold [3] (Figures 2A and 2B). RAMP domains are similar to ferredoxin-like and RRM folds, comprising a central four-stranded antiparallel β-sheet (βαβα), which forms a characteristic ‘RAMP turn’ motif [20]. The RAMP domain of Cas6 deviates from the standard motif in that the second RAMP helix (α2) that normally precedes β2 is missing and replaced by a loop with an extended conformation. In addition to the standard RAMP elements, the N-terminal domain contains a short β-stand located after the missing RAMP helix (α2), which forms a small three-stranded antiparallel β-sheet with two strands from the...
central \( \beta \)-sheet. In the C-terminal RAMP domain, there are three additional helices located on the helical face of the domain as well as a \( \beta \)-hairpin that connects two strands of the central \( \beta \)-sheet. The characteristic glycine-rich loop of the RAMP superfamily is found only in the C-terminal domain and is in the same conformation seen in the other Cas6 homologues.

**Dimerization of SsoCas6**

Analysis with the PDBePISA server [39] indicates that the four SsoCas6 monomers in the asymmetric unit assemble into two identical dimers (complexation significance score of 1), consistent with gel-filtration data that also suggested a dimer (Figures 2C and 3B). The dimer interface was formed between the C-terminal domains of both proteins with an average buried surface area of 1349 Å\(^2\). The three non-conserved helices of the C-terminal domain are positioned at the interface and may be responsible for causing dimerization in a Cas6 protein that is typically monomeric. Introduction of charged residues at the centre of the interface (SsoCas6-L170D/V202D, Figure 3A) resulted in a protein that ran on gel filtration as a monomer (Figure 3B). The same dimeric arrangement was seen in low-resolution structures of methylated and non-methylated SsoCas6 crystallized under different conditions (Supplementary Figure S1 at http://www.biochemj.org/bj/452/bj4520223add.htm). This is the first time that a Cas6 protein has been confirmed to be a dimer in the absence of RNA.

The dimeric wild-type enzyme was observed to undergo thermal melting with a melting temperature \((T_m)\) of 75 °C. For the monomeric variant, the equivalent \(T_m\) was 68 °C. Although 7 °C lower than the wild-type enzyme, the monomer remained heat-stable under the assay conditions. The monomeric Cas6 variant had a single-turnover catalytic rate constant of 0.034 min\(^{-1}\) at 60 °C, equivalent to 4 % of the wild-type rate. A similar reduction in rate for the monomeric variant was observed at lower reaction temperatures, suggesting that the loss of activity is not simply due to protein instability (results not shown).

**Delineation of catalytically important residues**

The PfuCas6 active site contains a triad of Tyr\(^{31}\), His\(^{46}\), and Lys\(^{52}\) [40]. The Y31A and H46A variants had no detectable enzyme activity, whereas the K52A variant resulted in a 40-fold reduction in nuclease activity. In SsoCas6, the catalytic lysine residue is conserved at position 51 (main chain ordered only in protomer D), whereas Tyr\(^{31}\) is not conserved (Supplementary Figure S2 at http://www.biochemj.org/bj/452/bj4520223add.htm). The essential histidine residue, which is also found in more divergent Cas6 orthologues such as Cas6e and Cas6f, is not present in SsoCas6. In fact, there were no histidine residues located around the putative active site. This is also true for the majority of crenarchaeal Cas6 orthologues, suggesting that there may be considerable mechanistic diversity within the archaeal Cas6 family (Supplementary Figure S3 at http://www.biochemj.org/bj/452/bj4520223add.htm).

In order to identify residues important for catalysis, 12 conserved residues present around the putative active site of SsoCas6 were targeted by site-directed mutagenesis (Supplementary Figure S3). The variant proteins were all purified as for the wild-type enzyme and were assayed using a...
Figure 3  Dimerization of SsoCas6

(A) View of Leu170 and Val202 at the dimer interface. These residues were each changed to aspartate to disrupt the interface. (B) Gel-filtration elution profiles of dimeric SsoCas6 and the SsoCas6-L170D/V202D variant, which elutes with a retention volume consistent with a monomeric composition (expected molecular mass of 33 kDa). (C) ThermoFluor analysis of heat-induced denaturation of wild-type (WT) and monomeric SsoCas6, showing the 7 °C difference in melting temperatures (Tm). (D) Single-turnover kinetic comparison of wild-type and monomeric SsoCas6. The catalytic activity of the monomer is reduced by 95% compared with the wild-type (WT) enzyme.

The active site of crenarchaeal Cas6

All Cas6 proteins studied in detail to date have an essential histidine residue in the active site: for example, His46 in PfuCas6 and His45 in Pyrococcus horikoshii Cas6. In contrast, the other Cas6 enzymes studied to date are all monomeric. This is likely to be a required property for the I-E and I-F systems, where only a single copy of Cas6 is present in the interference complex [41]. A monomeric protein would also ensure the delivery of a single crRNA to the complex. However, for PfuCas6, which is not strongly associated with effector complexes, the paradigm still holds [3]. Dimerization of both the Cas6 of Pyrococcus horikoshii and the Cas6b of Methanococcus maripaludis has been noted, but only in the presence of pre-crRNA [42,43]. Nevertheless, whereas the monomeric form of SsoCas6 generated in the present study is soluble, the enzyme is much less active than the wild-type dimeric protein. This suggests that the dimeric organization may have some bearing on catalysis, perhaps due to allosteric communication between the two active sites. It is worth bearing in mind that the substrate of Cas6 is a large pre-crRNA with many cleavage sites and that two RNA-cleavage events are required to generate one unit-length crRNA. The dimeric SsoCas6 structure has a broad area of positive charge in the region spanning the dimer interface and the two active sites (Figure 2D). It is possible that pre-crRNA spans this region, allowing both active sites to engage and cleave consecutive recognition sites. This would represent a departure from the simple crRNA wrapping proposed for PfuCas6 [14]. Alternatively, the reduced activity of the monomer may simply be a consequence of its reduced stability or of perturbation of the active site geometry.
Figure 4  Delineating the SsoCas6 active site

(A) Phosphorimage of a denaturing polyacrylamide gel showing the reaction products of repeat RNA incubated with wild-type (WT) and selected variant Cas6 enzymes. Lane m shows an RNA ladder generated by alkaline hydrolysis and lane c shows RNA incubated for 60 min in the absence of protein. Time points correspond to 1, 5 and 50 min incubations at 60°C. The cleavage site is indicated in the RNA sequence below the image. (B) Plot of the reaction kinetics for selected variants of SsoCas6 (red, wild-type; WT; black, S268A; green, K51A; blue, K28A). All data points were measured in triplicate and are means ± S.E.M. (C) Structure of the SsoCas6 monomer. The glycine-rich loop is shown in yellow to highlight the approximate position of the active site. The positions of selected side chains targeted by site-directed mutagenesis are shown as magenta sticks. Of these residues, it was not possible to define the absolute conformation of the side chains of Lys25, Lys28 and Lys51 from the electron density. Chains B and D are superimposed on chain A to include all of the desired residues. (D) First-order rate constants for wild-type and selected variant SsoCas6 enzymes. Relative activity (Rel. act.) is expressed as a percentage of wild-type (WT) activity.

[40], His39 in PaCas6f [12] and His26 in TcCas6e [44]. The exact location of the histidine residue differs in each of the available crystal structures, but is located either in the N-terminal RAMP α1 (TcCas6e and PaCas6f) or a non-conserved helix before α1 (PfuCas6) (Supplementary Figure S2). In M. maripaludis Cas6b, two histidine residues have also been implicated in the catalytic cycle, one of which probably corresponds to His46 of PfuCas6 [45]. The role of the catalytic histidine residue was originally thought to be that of a general acid, donating a proton to the leaving group during catalysis [40]. However, in Cas6f, the histidine is deprotonated and acts as a general base, abstracting a proton from the 2′-hydroxy nucleophile during the catalytic cycle [12]. Regardless, the crenarchaeal Cas6 enzymes lack any histidine residues in proximity to the presumed active site, suggesting that these enzymes employ another mechanism. Site-directed mutagenesis suggests that Lys25, Lys28, Lys31, and Arg31, which are conserved in crenarchaeal Cas6s (Supplementary Figure S3), are important for catalysis. These residues may play a role in stabilizing the pentacovalent phosphate transition state. Of the four, Lys28 appears to be the closest to an essential catalytic residue.

Other residues targeted by mutation had rather modest effects on catalysis. Tyr179 is well conserved in the crenarchaeal Cas6 family members and is suitably positioned to participate in catalysis, situated on the right-hand side of the active-site cleft. However, the fact that a phenylalanine residue is well-tolerated at this position suggests that its role may relate to positioning of other side chains or an RNA base. The only conserved acidic residue near the active site, Glu192, is clearly not involved in the catalytic cycle as the E192A variant retained full activity. The catalytic site of SsoCas6 thus appears to be unusually resistant to inactivation by targeted mutagenesis, with many highly conserved residues apparently not essential for catalysis. It is possible that many of these residues play a role in crRNA binding, as the single-turnover assay employed in the present study focuses purely on the chemical step of catalysis.

The single-turnover rate constant of SsoCas6, approximately 1 min⁻¹ at 60°C, is rather low, although comparable with those of the other Cas6 family enzymes that have been studied [12]. These rates are more akin to those observed for ribozymes than for ribonucleases [46]. However, they are presumably high enough to
fulfil the proteins’ functions in pre-crRNA processing in vivo, where it is probably more important to be a highly specific ribonuclease than a highly active one. The ease with which crRNA cleavage sites can evolve is exemplified by the type I-C CRISPR system. This subtype lacks a cas6 gene and instead the Cas5d protein is catalytic, cleaving pre-crRNA in vitro [47,48]. A putative catalytic triad consisting of Tyr116, Lys116 and His117 has been identified [47], but these residues are located on a different part of the RAMP domain compared with the active site of Cas6 proteins, suggesting that the active site has evolved independently. The crystal structure of S. solfataricus Cas6 has revealed a dimeric organization that may be relevant for its in vivo function as a stand-alone crRNA-processing endonuclease. The paradigm of an ‘active-site histidine residue’ for this enzyme family seems untenable, as no such residue exists within range of the active site. Instead, site-directed mutagenesis has revealed a network of basic residues that each contribute towards catalytic activity without being absolutely essential. The catalytic rate constants of Cas6 family enzymes are so low that they could be considered as much as RNA chaperones as true enzymes. This is particularly relevant for the enzymes that bind hairpin RNA structures very tightly and do not support multiple-turnover catalysis.

While this paper was in revision, Shao and Li [49] reported the crystal structure of a closely related SsoCas6 enzyme in complex with crRNA. This structure is also dimeric and site-directed mutagenesis confirmed the importance of Lys31, Lys53 and Arg52 for catalysis.

AUTHOR CONTRIBUTION
Judith Reeks, Richard Sokolowski, Shirley Graham and Huanting Liu generated the data. Judith Reeks, Richard Sokolowski, James Naismith and Malcolm White analysed the data and wrote the paper.

FUNDING
This work was funded by the Biotechnology and Biological Sciences Research Council (grant numbers BB/G011400/1 and BB/K000314/1 (to M.F.W. and J.H.N.)), a Biotechnology and Biological Sciences Research Council-funded studentship to J.R. and a Medical Research Council-funded studentship to R.D.S.

REFERENCES
1 Mojica, F. J., Diez-Villanueva, C., Garcia-Martinez, J. and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174–182
2 Bolotin, A., Quinquis, B., Sorek, A. and Ehrlich, S. D. (2005) Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561
3 Carte, J., Wang, R., Li, H., Terns, R. M. and Terns, M. P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev. 22, 3489–3496
4 Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Duff, M. O., Gravelle, B. R., Wells, L., Terns, R. M. and Terns, M. P. (2009) RNA-guided crRNA cleavage by a CRISPR RNA–Cas protein complex. Cell 139, 945–956
5 Payfuli, S., McMahon, S. A., Graham, S., Liu, H., Botting, C. H., Makarova, K. S., Koonin, E. V., Naismith, J. H. and White, M. F. (2012) Displacement of the canonical single-stranded DNA-binding protein in the Thermoproteales. Proc. Natl. Acad. Sci. U.S.A. 109, E389–E405
6 Winter, G. (2010) Xia2: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. 43, 106–109
7 Kabsch, W. (2010) XDS. Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 125–132
8 Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr., Sect. D: Biol. Crystallogr. 62, 72–82
9 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 213–221
10 Zhang, K. Y., Cowtan, K. and Mann, P. (1997) Combining constraints for electron-density modification. Methods Enzymol. 277, 53–64
11 Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr., Sect. D: Biol. Crystallogr. 62, 1002–1011
12 Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for macromolecular structure determination. Acta Crystallogr., Sect. D: Biol. Crystallogr. 50, 760–763
13 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr., Sect. D: Biol. Crystallogr. 60, 2126–2132
14 Mursudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr., Sect. D: Biol. Crystallogr. 53, 240–255
15 Painter, J. and Merritt, E. A. (2006) Optimal description of a protein structure in terms of main-chain and side-chain conformations and sites of anomalous scattering. Acta Crystallogr., Sect. D: Biol. Crystallogr. 62, 439–450
16 Winn, M. D., Murchow, G. N. and Papiz, M. Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol. 374, 300–321
17 Kleywegt, G. J. and Read, R. J. (1997) Not your average density. Structure 5, 1557–1569
18 Chen, V. B., Arenadell, 3rd, W. B., Head, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 12–21
19 Richter, J. J., Bell, S. D. and White, M. F. (2004) Physical and functional interaction of the archaean single-stranded DNA binding protein SSB with RNA polymerase. Nucleic Acids Res. 32, 1055–1074
20 Niesen, F. H., Berglund, H. and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. 2, 2212–2221
21 McRobbie, A. M., Meyer, B., Rousseau, C., Petrovic-Stojanovska, B., Liu, H. and White, M. F. (2012) Staphylococcus aureus DihE, a helicase that has evolved into a nuclease. Biochem. J. 442, 77–84
22 Oke, M., Carter, L. G., Johnson, K. A., Liu, H., McMahon, S. A., Yan, X., Keroa, M., Weikart, N. D., Kadi, N., Sheikh, M. A. et al. (2010) The Scottish Structural Proteomics Facility: targets, methods and outputs. J. Struct. Funct. Genomics 11, 167–180

© The Authors Journal compilation © 2013 Biochemical Society
36 Lillestol, R. K., Shah, S. A., Brügger, K., Redder, P., Phan, H., Christiansen, J. and Garrett, R. A. (2009) CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol. Microbiol. 72, 259–272
37 Parker, J. L. and White, M. F. (2005) The endonuclease Hje catalyses rapid, multiple turnover resolution of Holliday junctions. J. Mol. Biol. 350, 1–6
38 Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr., Sect. D: Biol. Crystallogr. 60, 2256–2268
39 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797
40 Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M. and Terns, M. P. (2010) Binding and cleavage of CRISPR RNA by Cas6. RNA 16, 2181–2188
41 Jore, M. M., Lundgren, M., van Duijn, E., Buttnera, J. B., Westra, E. R., Waghrmae, S. P., Wiedenhett, B., Pul, U., Wurm, R., Wagner, R. et al. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. 18, 529–536
42 Richter, H., Lange, S. J., Backofen, R. and Randau, L. (2013) SF CRISPR: comparative analysis of Cas6b-processing and CRISPR RNA stability. RNA Biol. 10, 405–417
43 Wang, R., Zheng, H., Preamplume, G., Shao, Y. and Li, H. (2012) The impact of CRISPR repeat sequence on structures of a Cas6 protein–RNA complex. Protein Sci. 21, 405–417
44 Sashital, D. G., Jinek, M. and Doudna, J. A. (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat. Struct. Mol. Biol. 18, 680–687
45 Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R. and Randau, L. (2012) Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. Nucleic Acids Res. 40, 9887–9896
46 Lilley, D. M. (2011) Catalysis by the nucleolytic ribozymes. Biochem. Soc. Trans. 39, 641–646
47 Nam, K. H., Hatjema, C., Liu, X., Ding, F., Wang, H., Delisa, M. P. and Ke, A. (2012) Cas5d protein processes pre-crRNA and assembles into a Cascade-like interference complex in subtype I-C/Dvulg CRISPR–Cas system. Structure 20, 1574–1584
48 Garside, E. L., Schellenberg, M. J., Gesner, E. M., Bonanno, J. B., Sauder, J. M., Burley, S. K., Almo, S. C., Mehta, G. and Macmillan, A. M. (2012) Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. RNA 18, 2020–2028
49 Shao, Y. and Li, H. (2013) Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. Structure, doi: 10.1016/j.str.2013.01.010
50 McNicholas, S., Potterton, E., Wilson, K. S. and Noble, M. E. (2011) Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr., Sect. D: Biol. Crystallogr. 67, 386–394

Received 22 February 2013; accepted 25 March 2013.
Published as BJ Immediate Publication 25 March 2013, doi:10.1042/BJ20130269
SUPPLEMENTARY ONLINE DATA
Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing

Judith REEKS¹, Richard D. SOKOLOWSKI¹, Shirley GRAHAM, Huanting LIU, James H. NAISMITH² and Malcolm F. WHITE²
Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, U.K.

Figure S1  Crystal packing in multiple crystal forms of SsoCas6 showing the conservation of the dimer

Initial molecular replacement (MR) solutions of low-resolution datasets collected from three non-related crystal forms. The data were processed using xia2 [1] and MR was performed using Phaser [2] in the Phenix suite [3] using a monomer of methylated selenomethionine-labelled SsoCas6 as the search model. (A) Crystals of SsoCas6 were grown in buffer (20 mM Tris/HCl, pH 7.5, 1 M NaCl and 10 % glycerol) without the use of a precipitant. The crystals were cryoprotected in 75 mM Tris/HCl (pH 7.5), 3 M NaCl and 20 % glycerol. Data were collected on I04-1 (Diamond Light Source) and processed to 4.5 Å resolution with a space group of F4132. The two molecules of the ASU (asymmetric unit) formed the dimer. (B) SsoCas6K28A (20 mM Tris/HCl, pH 7.5, 0.5 M NaCl and 10 % glycerol) was incubated with repeat RNA (sequence 5′-GAUAAUCUCUUAUAGAAUUGAAAG-3′, purchased from IDT) at a molar ratio of 1:1.1 and crystallized in 0.1 M Mes (pH 6.8), 0.05 M caesium chloride and 27.5 % Jeffamine M-600. The crystals were cryocooled directly from the drop and diffracted to 3.0 Å resolution on I24 (Diamond Light Source) with a space group of I222. No electron density was visible for the RNA. One molecule was present in the ASU, but formed a dimer with a symmetry-related molecule. (C) Methylated SsoCas6K28A (20 mM Tris/HCl, pH 7.5, 0.15 M NaCl and 10 % glycerol) was incubated with RNA at a molar ratio of 1:2 and crystallized in 0.1 M Tris/HCl (pH 8.8), 0.2 M MgCl₂ and 33 % PEG (polyethylene glycol) 4000 with the drops supplemented with 10 % glycerol. The crystals were cryoprotected with perfluoropolyether. Data were collected on I24 (Diamond Light Source) and processed to 6 Å resolution with a space group of P4₂;2₂. Again, no electron density was observed for the RNA. One molecule was present in the ASU, but formed a dimer with a symmetry mate.

¹ These authors contributed equally to this work.
² Correspondence may be addressed to either of these authors (email jhn@st-andrews.ac.uk or mfw2@st-andrews.ac.uk).
The co-ordinates and data for the structure of Sulfolobus solfataricus Cas6 have been deposited in the PDB under code 3ZFV.
Figure S2  Comparison of Cas6 sequences and structures

(A) Structures of (left to right) SsoCas6, PfuCas6 (PDB code 3PKM) and TtCas6e (PDB code 2Y8W) highlighting the positions of the catalytic residues (magenta sticks) and the glycine-rich loop (yellow). (B) Structure-based sequence alignment of PfuCas6 and SsoCas6. Sequence similarity is shown in black. Secondary-structure elements are shown above and below the alignment for PfuCas6 and SsoCas6 respectively and are coloured as in Figure S1. Gaps in the connections reflect disordered residues. Putative catalytic residues are indicated with purple stars. The Figure was created with Aline [4].
Structure of crenarchaeal Cas6

The residues targeted by site-directed mutagenesis are indicated. Representative Cas6 orthologues from *S. solfataricus* (Sso1437 and Sso2004), *Sulfolobus islandicus* (C3MMW1), *Sulfolobus tokodaii* (F9VPJ5), *Metallosphaera sedula* (A4YFU6), *Sulfolobus acidocaldarius* (Q4J7R1), *Ignisphaera aggregans* (E0SS14) and *Staphylothermus marinus* (A3DLD1) are shown. The alignment was generated using T-COFFEE [5].

**REFERENCES**

1. Winter, G. (2010) *xia2*: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. 43, 186–190
2. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674
3. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 213–221
4. Bond, C. S. and Schuttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr., Sect. D: Biol. Crystallogr. 65, 510–512
5. Notredame, C., Higgins, D. G. and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217

Received 22 February 2013; accepted 25 March 2013
Published as BJ Immediate Publication 25 March 2013, doi:10.1042/BJ20130269