JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

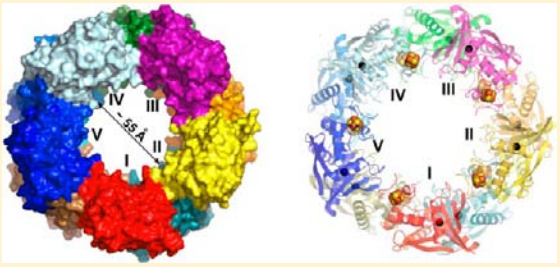
Toroidal Structure and DNA Cleavage by the CRISPR-Associated [4Fe-4S] Cluster Containing Cas4 Nuclease SSO0001 from *Sulfolobus solfataricus*

Sofia Lemak,[†,§] Natalia Beloglazova,[†,§] Boguslaw Nocek,[‡] Tatiana Skarina,[†] Robert Flick,[†] Greg Brown,[†] Ana Popovic,[†] Andrzej Joachimiak,[‡] Alexei Savchenko,[†] and Alexander F. Yakunin*[,†]

[†]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario M5S 3E5, Canada
[‡]Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, United States

Supporting Information

ABSTRACT: Cas4 proteins, a core protein family associated with the microbial system of adaptive immunity CRISPR, are predicted to function in the adaptation step of the CRISPR mechanism. Here we show that the Cas4 protein SSO0001 from the archaeon *Sulfolobus solfataricus* has metal-dependent endonuclease and 5′→3′ exonuclease activities against single-stranded DNA, as well as ATP-independent DNA unwinding activity toward double-stranded DNA. The crystal structure of SSO0001 revealed a decameric toroid formed by five dimers with each protomer containing one [4Fe-4S] cluster and one Mn^{2+} ion bound in the active site located inside the internal tunnel. The conserved RecB motif and four Cys residues are important for DNA binding and cleavage activities, whereas DNA unwinding depends on several residues located near the [4Fe-4S] cluster. Our results suggest that Cas4 proteins might contribute to the addition of novel CRISPR spacers through the formation of 3′-DNA overhangs and to the degradation of foreign DNA.



INTRODUCTION

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and associated proteins (Cas) comprise a novel prokaryotic immune system, which resembles eukaryotic RNA interference and protects microbial cells from invading viruses and plasmids.[1−3] In the first step of CRISPR mechanism ("adaptation"), Cas proteins recognize foreign DNA and incorporate short sequences (25−50 nt) of this DNA (protospacers) into the host chromosome as new spacers. During the "expression" step, these viral-specific spacers are transcribed and processed by other Cas proteins to produce short CRISPR RNAs (crRNAs). In the last step ("interference"), the crRNAs in complex with different Cas proteins specifically recognize and degrade foreign DNA or RNA.[4−8]

Cas proteins can be organized into as many as 45 protein families with predicted nuclease, helicase, nucleic acid binding, or unknown activities.[1,9] Six families of Cas proteins (Cas1 to Cas6) form the core group, which are present in most CRISPR-harboring genomes. Based on the composition of Cas proteins in microbial genomes, the CRISPR-Cas systems have been classified into three major types: I, II, and III. The type I system is the most widespread and is divided into six subtypes (I-A to I-F) on the basis of subtype-specific genes.[2] The type I system includes the Cas3 protein and complexes of subtype-specific Cas proteins involved in crRNA processing and recognition of DNA targets.[6,10] The CRISPR type II system contains the large multidomain protein Cas9 instead of Cas3, whereas the type III system includes the Cas6-RAMP module (Cmr or Csm).

The least understood step of CRISPR interference is adaptation, which has been proposed to involve the recognition of foreign DNA (self/non-self) and PAM sequences, excision of protospacers from the invader DNA, and incorporation of them into CRISPR loci performed by the core proteins Cas1, Cas2, Cas4, and the subtype II-A-specific Csn2.[1,3,5,11,12] Cas4 proteins have been predicted to function in the adaptation step, because their genes are usually associated and sometimes fused to *cas1* genes, and the *Thermoproteus tenax* Cas4 protein forms a complex with Cas1 and Cas2 proteins.[13] The sequences of Cas4 proteins contain a RecB-like nuclease motif and four conserved Cys residues, which are predicted to coordinate an [Fe-S] cluster. Based on sequence, Cas4 proteins belong to RecB-like nucleases, which are related to the λ exonuclease and other enzymes from the restriction endonuclease-like superfamily.[1,14] The multifunctional *E. coli* RecBCD and *Bacillus subtilis* AddAB helicase−nuclease complexes represent the main engine of DNA recombination and repair, which involve unwinding of DNA duplex and formation of 3′-ssDNA tails.[15−17] In addition, the RecBCD complex contributes to the degradation of foreign DNA entering the microbial cell.[15]

Received: August 22, 2013
Published: October 30, 2013

   dx.doi.org/10.1021/ja408729b | J. Am. Chem. Soc. 2013, 135, 17476−17487

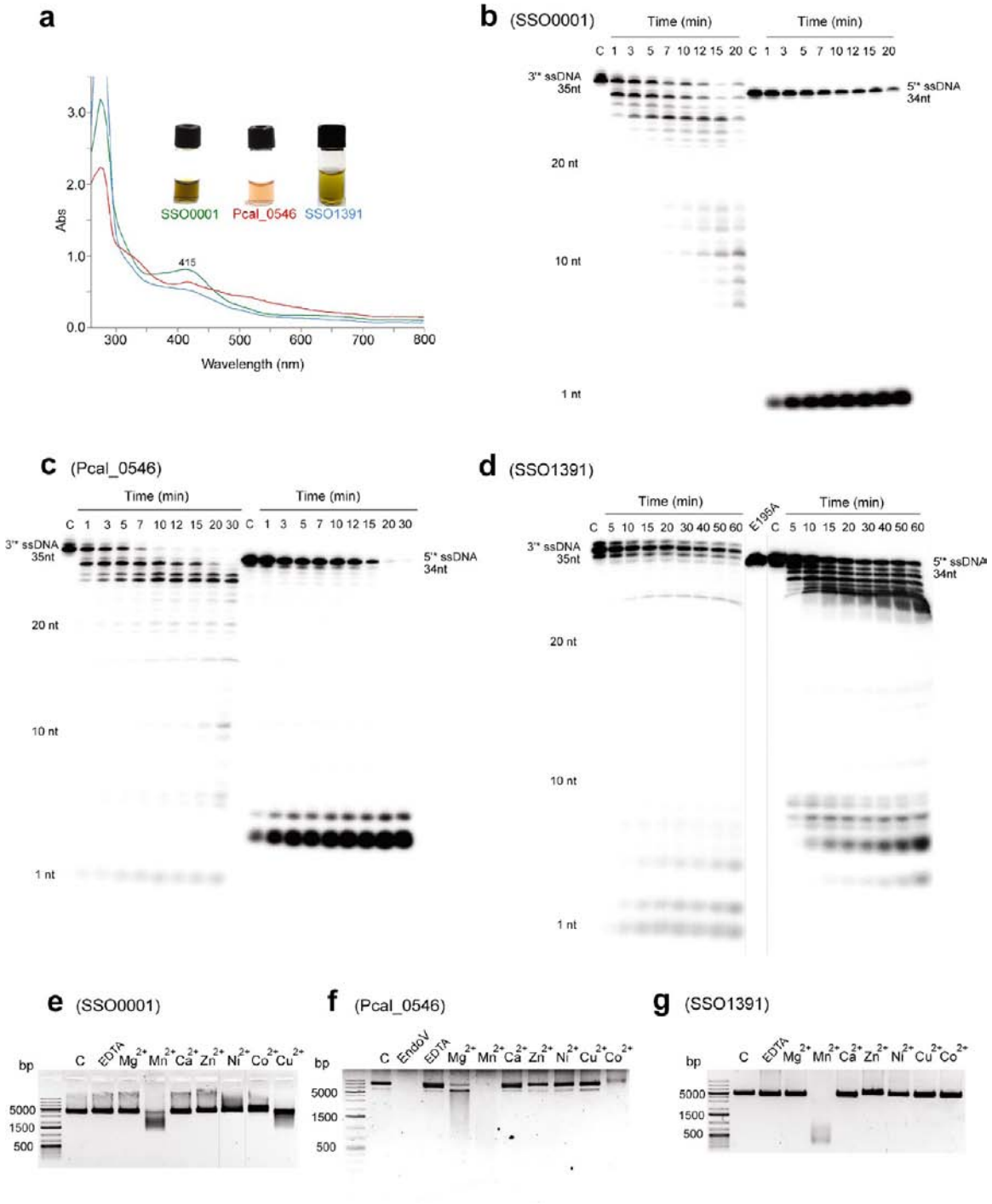


Figure 1. Absorption spectra and nuclease activity of purified Cas4 proteins. (a) Absorption spectra of SSO0001 (green line), Pcal_0546 (red line), and SSO1391 (blue line). Insets are photos of the vials with the concentrated purified protein. Cleavage of ssDNA: (b) SSO0001, (c) Pcal_0546, and (d) SSO1391. The 3′- or 5′-[^{32}P]-labeled ssDNA was incubated without protein (lane C) or with 8 nM SSO0001 (b), 493 nM Pcal_0546 (c), or 250 nM SSO1391 (wt or E195A mutant) (d) for the indicated time (60 min for mutant) at 45 °C and analyzed by denaturing gel electrophoresis. (e−g) Endonuclease assays: cleavage of the circular ssDNA of M13mp18. The M13mp18 ssDNA (5 nM) was incubated at 45 °C without protein addition (lane C) or in the presence of (e) 8 μM SSO0001 (45 min), (f) 0.4 μM Pcal_0546 (20 min), or (g) 0.5 μM SSO1391 (45 min), and the reaction products were analyzed by agarose gel electrophoresis and SYBR Green staining. Lanes EndoV and EDTA represent control experiments with endonuclease V or 5 mM EDTA.

The crenarchaeon *Sulfolobus solfataricus* contains a complex CRISPR system with two subsystems (types I-A and III-B) organized into six CRISPR loci (A-F) with 52 Cas genes.[18,19] The *S. solfataricus* genome encodes five Cas4-like proteins which share low sequence similarity (15−30% sequence identity) and belong to DUF83 (SSO0001, SSO1392, SSO1449) and DUF911 (SSO1391, SSO1451) (Figure S1). Recently, the purified SSO0001 has been shown to possess a 5′-to-3′ exonuclease activity.[20]

Here, we demonstrate that the *S. solfataricus* Cas4 protein SSO0001 exhibits metal-dependent endonuclease and 5′→3′ exonuclease activities against ssDNA, as well as ATP-

dx.doi.org/10.1021/ja408729b | J. Am. Chem. Soc. 2013, 135, 17476−17487

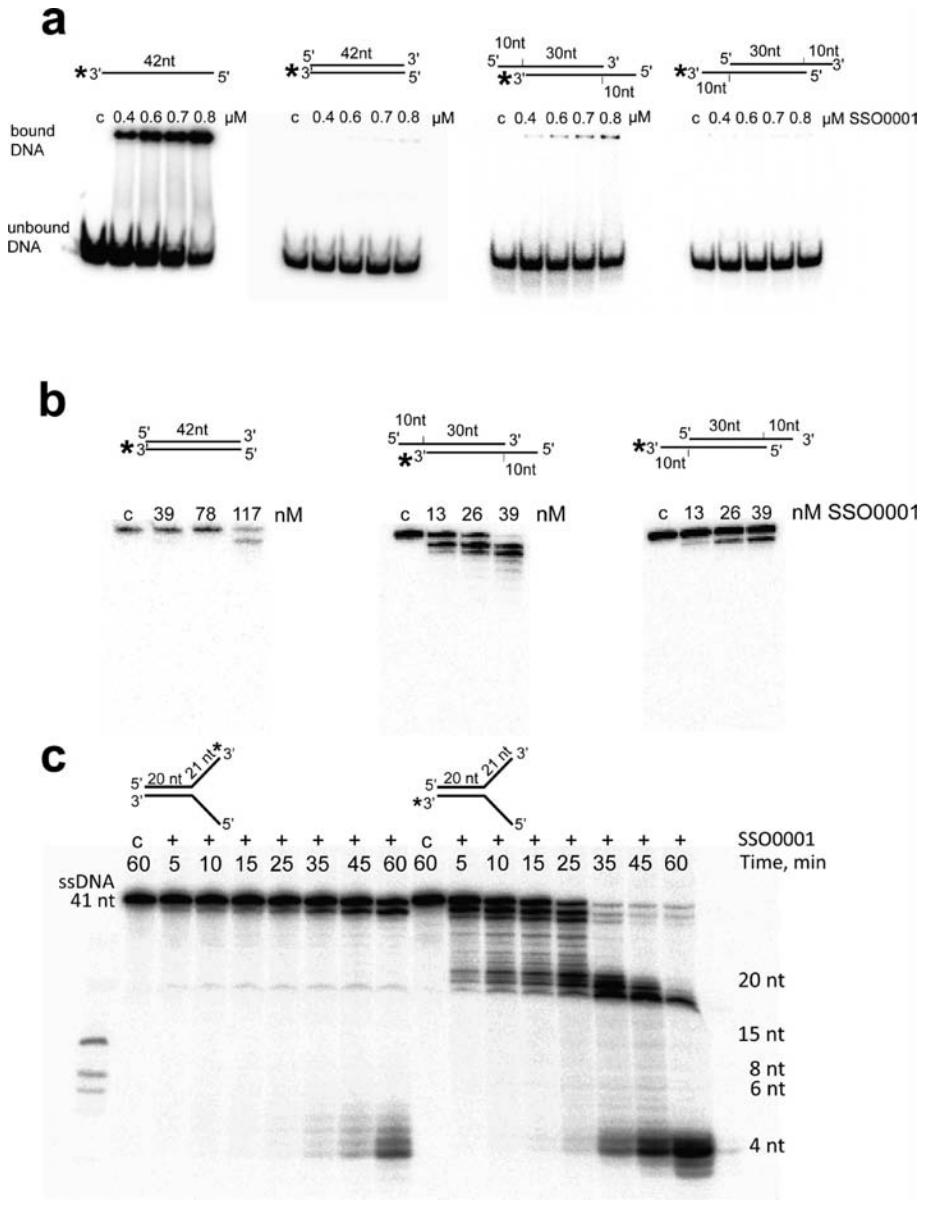


Figure 2. Binding and cleavage of dsDNA by SSO0001. (a) Gel-shift assays. The indicated amounts of SSO0001 were incubated with ssDNA, dsDNA with blunt ends, dsDNA with 5′-overhangs, or dsDNA with 3′-overhangs, and the reactions were analyzed by native PAGE. (b,c), Cleavage of dsDNA with blunt ends, 5′-overhangs, 3′-overhangs, or the splayed arm substrate (18 nM SSO0001) incubated for 15 min (b) or as indicated (c) at 45 °C. In all substrates, one strand is [^{32}P]-labeled at the 3′-end (indicated by asterisks), and the length of ssDNA overhangs is indicated on the substrate models (nt). Lanes C show incubation without SSO0001 addition.

independent unwinding activity toward dsDNA. Crystal structure of SSO0001 revealed a toroidal decameric complex with each protomer containing a [4Fe-4S] cluster and a Mn^{2+} ion. The DNA binding and cleavage activities are dependent on the conserved Cys and RecB motif residues, whereas the DNA unwinding activity is associated with residues located near the [Fe-S] cluster.

RESULTS

Nuclease Activity of Cas4 Proteins. The purified Cas4 proteins SSO0001 and SSO1391 from *S. solfataricus* and Pcal_0546 from *Pyrobaculum calidifontis* exhibited a brown color indicating the presence of functional [Fe-S] clusters (Figure 1a). Their absorption spectra showed the presence of a broad shoulder at 380−420 nm typical of [4Fe-4S] cluster-containing proteins (Figure 1a). Nuclease assays using the 5′- and 3′-[^{32}P]-labeled linear ssDNA revealed the presence of high metal-dependent nuclease activity in all three proteins (Figure 1b−d), including SSO1391, which was reported to be catalytically inactive in a previous work.[20] The nuclease activity of SSO0001 was maximal in the presence of Mg^{2+} or Mn^{2+} (5 mM), KCl (20−100 mM), at alkaline pH (9−10), and at high temperatures (65−70 °C) (Figure S2). In contrast to the HNH-like and GIY-YIG-like nucleases, Ca^{2+} did not support the nuclease activity of SSO0001 (Figure S2a). The rates of cleavage of ssRNA and dsDNA with blunt ends by SSO0001 were approximately 20 times and 200 times lower, respectively, compared to those of ssDNA (Figures 2b and S2d).

With the 5′-labeled ssDNA as substrate, SSO0001 produced mononucleotides as products, whereas the 3′-labeled substrate was cleaved with the formation of a series of oligonucleotides, with the shortest product containing four nucleotides (Figure

dx.doi.org/10.1021/ja408729b | J. Am. Chem. Soc. 2013, 135, 17476−17487

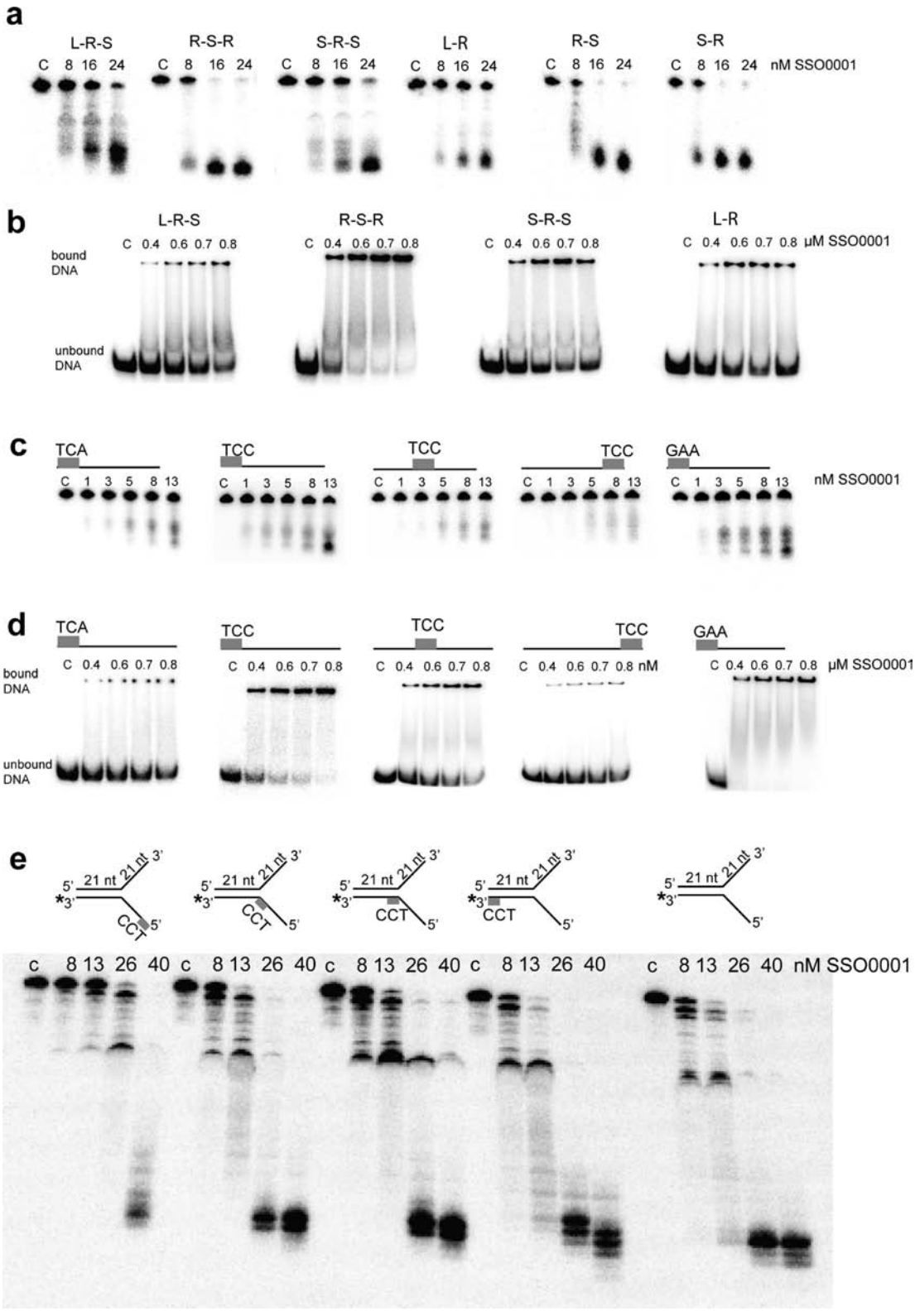


Figure 3. Cleavage and binding of CRISPR-related substrates by SSO0001. (a) Cleavage and (b) binding of ssDNA containing various combinations of sequences from the *S. solfataricus* CRISPR-B: L-R-S, leader (L; 18 nt), repeat (R; 25 nt), and spacer-1 (S; 17 nt); R-S-R, the spacer-1 fragment (10 nt) flanked by the repeat-1 sequences (25 nt each); S-R-S, repeat-1 (25 nt) flanked by the spacer-1 fragments (17 nt and 18 nt); L-R, the leader fragment (35 nt) with repeat-1 (25 nt); R-S, repeat-1 (25 nt) with spacer-1 (38 nt); S-R, spacer-1 (38 nt) with repeat-1 (25 nt). (c) Cleavage and (d) binding of ssDNA containing different *S. solfataricus* PAM (TCA, TCC) or non-PAM (GAA) sequences at the 5′-end, center, or 3′-end of the *S. solfataricus* CRISPR-A spacer-1. The sequence and position of PAMs are indicated by the letters on the substrate models. (e) Cleavage of splayed arm substrates containing the *S. solfataricus* CRISPR-B leader sequence with varying positions of the TCC PAM on the labeled strand. In all experiments, the 3′-[^{32}P]-labeled substrates were incubated with the indicated amounts of SSO0001 at 45 °C for 15 min. Lanes C show incubation without SSO0001 addition.

dx.doi.org/10.1021/ja408729b | *J. Am. Chem. Soc.* 2013, 135, 17476−17487

1b). These results confirm that SSO0001 is a 5′→3′ exonuclease that cleaves one nucleotide at a time from the 5′-end of the substrate. The short oligonucleotides (17 nt) were cleaved by SSO0001 faster than longer substrates (62 nt) (Figure S2j), whereas phosphorylation or biotinylation of substrates at the 5′-terminus had no effect on activity (Figure S2h,i). The 5′→3′ exonuclease activity was also observed in the Cas4 protein Pcal_0546 (also DUF83), whereas SSO1391 (DUF911) produced a range of cleavage products with both substrates, indicating that it cleaves ssDNA in both directions (5′→3′ and 3′→5′) (Figure 1c,d). The three proteins also showed Mn^{2+}-dependent endonuclease activity with circular ssDNA of the M13mp18 phage as a substrate (Figure 1e−g), but no activity against circular dsDNA (Figure S3a). Interestingly, the *E. coli* RecB nuclease domain is able to cleave ssDNA with dual polarity and endonucleolytically, whereas the RecB nuclease domain of the *B. subtilis* AddB is a 5′→3′ exonuclease.[17,21] Thus, our work together with the previous report on SSO0001[20] demonstrates that the CRISPR-associated RecB-like proteins represent a diverse group of exonucleases, which can cleave ssDNA in one or both directions and also exhibit endonuclease activity against ssDNA.

Most known nucleases cleave the phosphodiester bond with the formation of 5′-phosphates and 3′-OH groups.[22] To determine the phosphodiester bond cleavage site of SSO0001, we carried out two reactions using the unlabeled ssDNA as substrate and T4 polynucleotide kinase (PNK): direct phosphorylation of cleavage products at the 5′-end (requires a 5′-hydroxyl) and phosphate exchange reaction at the 5′-end (requires a 5′-phosphate end). The results show high product labeling in the PNK phosphorylation reaction and no labeling in the PNK exchange reaction, suggesting the absence of phosphate at the product 5′-end (Figure S3d). These results were confirmed using an unlabeled poly-dC_{10} ssDNA substrate and liquid chromatography−mass spectrometry analysis (LC-MS) of the cleavage products of the wild-type SSO0001 and its inactive mutant D99A (Figure S3e). Thus, SSO0001 cleaves ssDNA substrates with the formation of products containing 5′-hydroxyls and 3′-phosphates. In contrast, the characterized RecB-like nucleases and λ exonuclease have been shown to generate the products with 5′-phosphates and 3′-hydroxyls,[23,24] suggesting that these enzymes and SSO0001 might use different catalytic mechanisms.

SSO0001 Exhibits ATP-Independent DNA Unwinding Activity. Further biochemical characterization of Cas4 proteins was focused on SSO0001, since we solved its crystal structure. In electrophoretic mobility shift assays, the purified SSO0001 showed strong binding to ssDNA (42 nt) (Figure 2a). With dsDNA substrates, detectable binding was observed only for substrates containing 5′-overhangs (10 nt), whereas no significant binding was observed to dsDNA with 3′-overhangs or blunt ends. This is consistent with the 5′→3′ polarity of ssDNA cleavage by SSO0001 (Figure 1b). In nuclease assays, SSO0001 was able to cleave these dsDNA substrates, although with lower rates than ssDNA (Figure 2b). With dsDNA substrates, the highest activity was observed against the dsDNA with 5′-overhangs and the lowest activity toward dsDNA with blunt ends. These results also suggest the presence of DNA unwinding activity in SSO0001. This was confirmed using the splayed arm duplex substrates, which are the intermediates of DNA recombination and integration and also might represent a potential intermediate of CRISPR spacer acquisition. Splayed arm structures are known substrates for RecB-like nucleases and are cleaved by them through a combination of exonuclease, endonuclease, and ATP-dependent helicase activities.[15] Two types of splayed arm substrates were used: containing both dsDNA (20 nt) and ssDNA (21 nt) areas with a [^{32}P]-label located at the 3′-end of the bottom (in the dsDNA area) or top (in the ssDNA area) strands (Figures 2c and S3b). The time point experiments with the first substrate revealed the formation of a series of products with different lengths, with the transitional accumulation of the 20 nt long products, suggesting that the cleavage of labeled strand by SSO0001 slowed upon reaching the substrate duplex (Figure 2c). Longer incubation times (25 min) resulted in the formation of short products (1−4 nt), which can be produced only after complete unwinding of DNA duplex. The formation of short products was also observed with the second substrate containing the label at the 3′-end of the top strand, but this required longer incubation times, consistent with the requirement to unwind and degrade the unlabeled strand first (Figure 2c). With a series of splayed arm substrates with different arm lengths (5−30 nt), SSO0001 showed higher activity toward substrates with short arms (5−15 nt) (Figure S2k). ATP addition had no effect on the unwinding and cleavage of splayed arm substrates by SSO0001, suggesting that SSO0001 is an ATP-independent DNA-unwinding enzyme (Figure S2l). Thus, our results indicate that, in addition to endonuclease and 5′→3′ exonuclease activities toward ssDNA, SSO0001 also exhibits the unwinding activity against dsDNA.

SSO0001 Activity against CRISPR-Related Substrates. The CRISPR adaptation step has been proposed to involve the recognition of the leader, repeat, and PAM sequences and the formation of dsDNA breaks and 3′ ssDNA overhangs.[3,5,12,20] To determine if SSO0001 exhibits a preference for substrates containing the *S. solfataricus* CRISPR sequences, we analyzed its activity against 14 ssDNA substrates (60−63 nt) containing various combinations of the repeat, spacer-1, and leader sequences from the *S. solfataricus* CRISPR locus B, which is the biggest (and probably most active) CRISPR locus in this organism (Table S1). SSO0001 cleaved all these substrates with some preference for the repeat-spacer-repeat substrate, which also supported higher activity of SSO0001 in DNA binding assay, suggesting that this protein might exhibit some level of sequence preference (Figure 3a,b). It has been shown that the *Sulfolobus* CRISPR systems preferentially target protospacers located at the 3′-end of the PAM sequences CCN, TCN, and GTN or at the 5′-end of the PAM NGG.[19,25] We designed 32 ssDNA substrates (CRISPR-A spacer-1, 41 nt) containing these PAMs located at the substrate 5′-end, in the middle, or at the 3′-end, as well as 27 substrates with non-PAM sequences (ACA, CAA, GAA) (Table S1). SSO0001 showed similar cleavage and binding activities with both PAM and non-PAM substrates, suggesting that the purified SSO0001 is not strictly specific to the PAM-containing ssDNA substrates *in vitro* (Figures 3c,d, S4, and S5). The effect of PAMs on the DNA unwinding activity of SSO0001 was tested using several splayed arm substrates (42 nt) containing the sequences of the *S. solfataricus* CRISPR-B leader or spacer-1 and the TCC PAM placed at the 5′-end, in the middle, or 3′-end of the bottom strand, which was 3′-[^{32}P]-labeled (Figures 3e and S3c). SSO0001 exhibited different levels of unwinding and cleavage activities against these substrates with the absence of the noticeable correlation with the presence or location of PAM (Figures 3e and S3c). Thus, our results are consistent with a lack of significant sequence preference in SSO0001 toward the

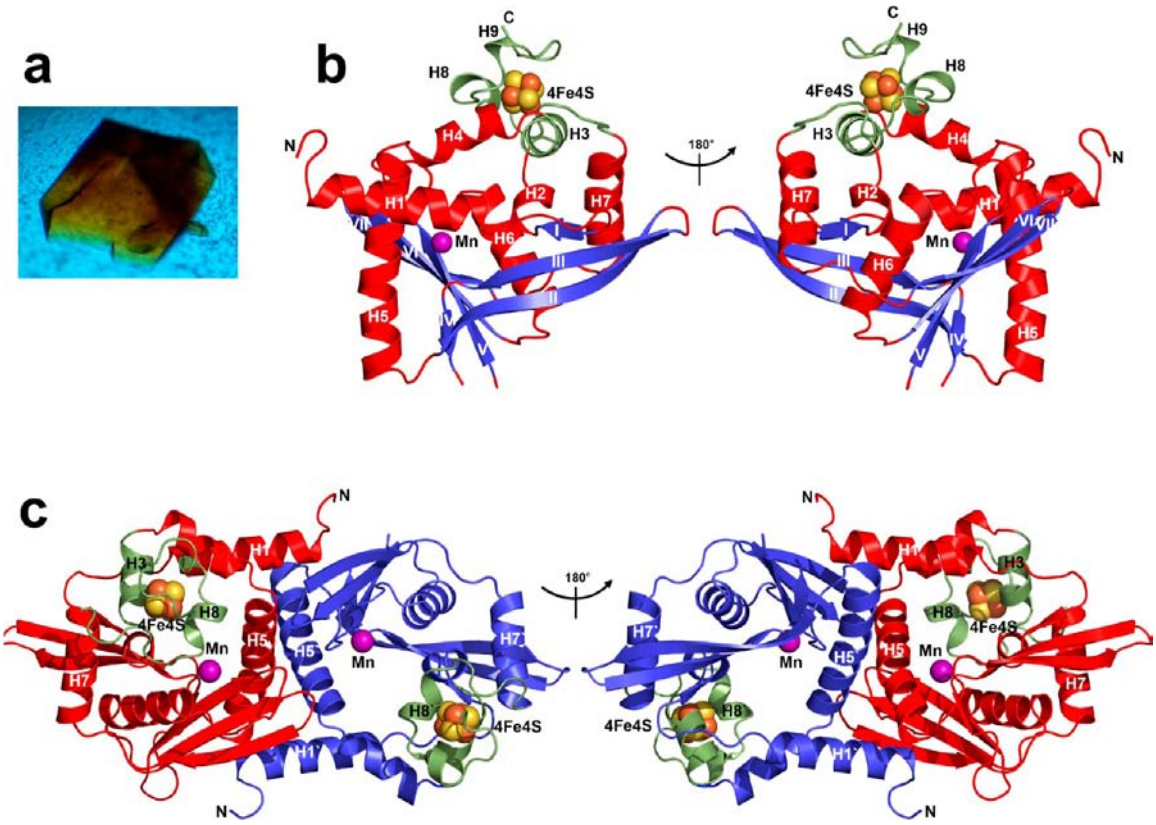


Figure 4. Crystal structure of SSO0001. (a) Photograph of the SSO0001 crystal (~300 × 300 × 50 μm). (b) Overall structure of the SSO0001 protomer: two views related by a 180° rotation. The protein core domain is colored in red (helices) and blue (strands), whereas the [Fe-S] cluster-binding subdomain is colored in green. The [4Fe-4S] clusters are shown as orange and yellow balls, whereas the position of the active site is indicated by the Mn^{2+} ion (magenta balls). (c) Two views of the SSO0001 "dimer" related by a 90° rotation. The two protomers are colored in red and blue, whereas their [Fe-S] cluster-coordinating subdomains are colored in green.

substrates containing the *S. solfataricus* PAMs in DNA binding and cleavage reactions *in vitro*.

The Crystal Structure of SSO0001 Reveals Bound Mn^{2+} and a [4Fe-4S] Cluster. Purified SSO001 was crystallized and produced brown-colored crystals, suggesting the presence of intact [Fe-S] clusters (Figure 4a). The structure of the seleno-methionine-substituted protein was solved at 2.35 Å resolution (Table S3) and revealed the core $\alpha\beta\alpha$ protein domain capped by a small, mostly α-helical subdomain (Figure 4b). The core domain contains the V-shaped β-sheet, which is surrounded by α-helices α1, α2, α3, α4, and α5, creating a large elongated opening (12−24 Å) on one side and a smaller round hole (9 Å) on the other side. The small subdomain is formed by three helices (α3, α8, and α9) and the long loop 177−186. This domain coordinates a cubane-like [4Fe-4S] cluster with the typical Fe−Fe (2.9 Å) and Fe−S (2.1 Å) distances (Figures 4b and 5d). The [4Fe-4S] cluster is coordinated by the side chains of four Cys residues (Cys32, Cys188, Cys191, and Cys197), which are conserved in all Cas4 proteins and AddB-like nucleases (Figure S1). Thus, the SSO0001 structure confirms that Cas4 proteins are members of a class of "iron-staple" nucleases, which also include AddB and Dna2.[17,26]

A Dali search[27] for the SSO0001 structural homologues identified several exonucleases including the alkaline exonuclease LHK-Exo from *Laribacter honkongensis* (PDB code 3SZ4; Z-score 11.9, root-mean-square deviation (rmsd) 3.2 Å), λ exonuclease (PDB 3SLP; Z-score 11.0, rmsd 3.2 Å), and the C-terminal RecB-like nuclease domain of the *B. subtilis* AddAB helicase−nuclease complex (PDB ; Z-score 10.6, rmsd 3.3 Å). The nuclease domain of the *E. coli* RecE was also highly similar to SSO0001 (PDB 3H4R; Z-score 10.5, rmsd 3.7 Å), whereas the *E. coli* RecB nuclease domain showed lower similarity (PDB 1W36; Z-score 6.9, rmsd 3.9 Å). All these enzymes are metal-dependent 5′→3′ exonucleases that bind to dsDNA ends and processively digest the 5′-strand to produce long 3′-overhangs.[12,16,17,28,29]

SSO0001 Protomers Form Decameric Toroids. The final model of SSO0001 revealed 10 protomers assembled as a pentamer of "dimers" creating a toroidal structure (105 Å × 67 Å) with a large central channel (55 Å diameter) (Figure 5). The tightly sealed "dimers" are formed through the interactions between the two long helices α5 with the contribution from the residues located on the α1 helix and the β-strands β6 and β7 (Figure 4c). Analysis of the crystal contacts in the SSO0001 structure using quaternary prediction server PISA also suggested a decameric state arranged as a pentamer of "dimers". The decameric state of SSO0001 in solution was experimentally confirmed using analytical ultracentrifugation (243 138 Da or 9.5 protomers; predicted molecular mass for the His-tagged protomer 25.6 kDa) and size-exclusion chromatography (216−268 kDa or 9−10.7 protomers). The SSO0001 toroid is formed by five vertically positioned and slightly twisted "dimers" creating a complex with two pentameric rings with D5 symmetry (Figure 5). The C-terminal part of each protomer coordinates a [4Fe-4S] cluster and also interacts with two protomers from a neighbor "dimer", suggesting that the intact state of the [Fe-S] cluster might be important for the maintenance of the SSO0001 toroid. This is

dx.doi.org/10.1021/ja408729b | J. Am. Chem. Soc. 2013, 135, 17476−17487

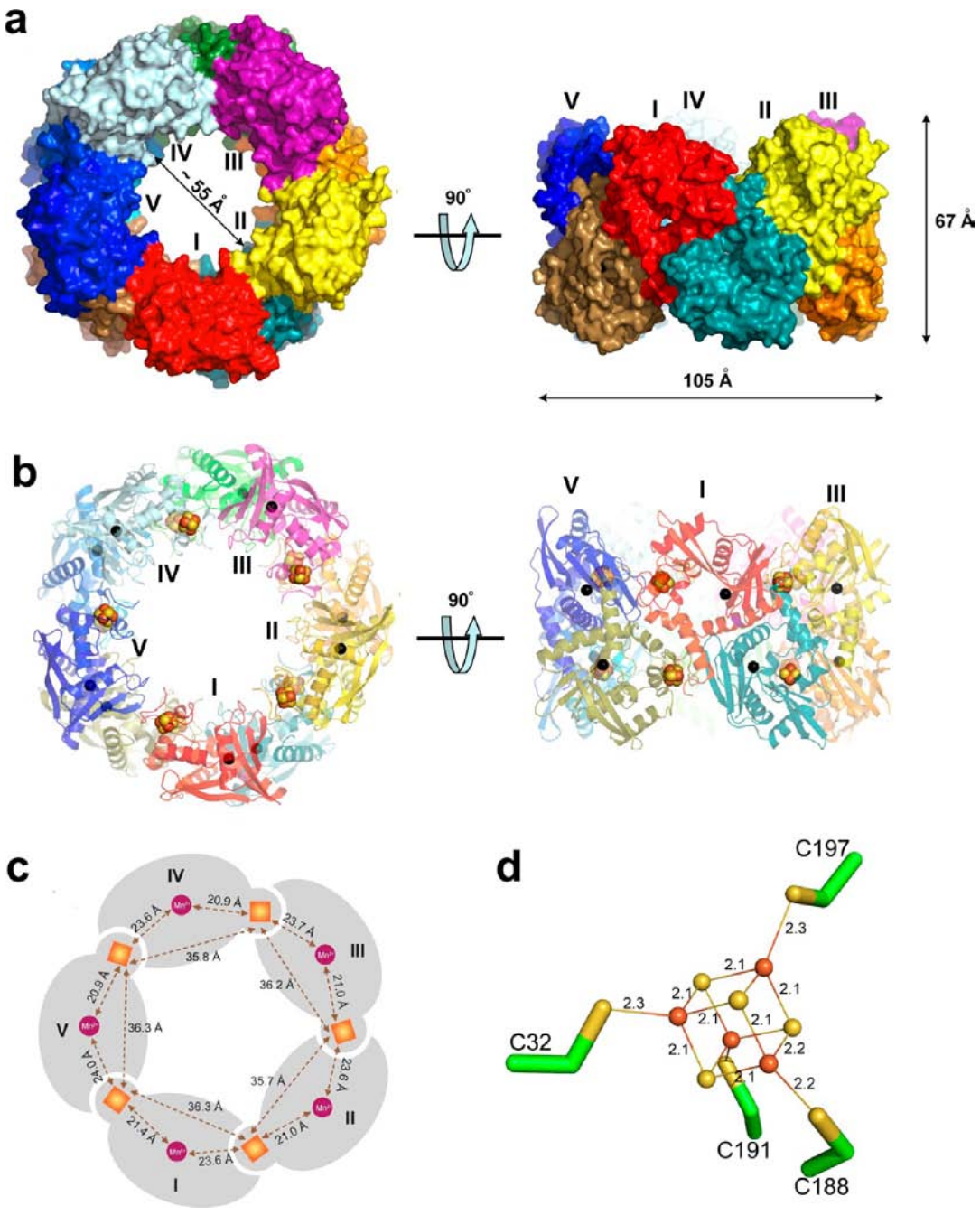


Figure 5. Quaternary structure of SSO0001. (a) Two views of the decameric toroid related by 90° rotation: a surface presentation. The protomers are shown in different colors. (b) Two views of the decameric toroid, with protomers shown as semitransparent ribbons to demonstrate the position of [4Fe-4S] clusters (orange and yellow balls) and Mn^{2+} (black balls). (c) Schematic diagram showing the arrangement of protomers in the pentameric ring and distances between [Fe-S] clusters and Mn^{2+} ions. (d) Coordination of [4Fe-4S] cluster in SSO0001 and interatomic distances (in Å). The Cys side chains are shown as sticks, the Fe atoms as orange balls, and S^{2-} as yellow balls.

supported by a gel-filtration analysis of the oligomeric state of the SSO0001 Cys mutants (Cys32A, Cys188A, Cys191A, and Cys197A), which showed no brown color after purification and existed as a mixture of high-molecular-weight aggregates (385−400 kDa; 58−73%), pentamers (117−123 kDa; 20−30%), and monomers (27−29 kDa; 7−14%) in solution (Figure S3f).

SSO0001 Active Site Is Located in a Tunnel. The position of the SSO0001 active site is indicated by the location of a metal ion (identified as Mn^{2+} using an X-ray fluorescence analysis), which is bound at the bottom of the elongated opening (between α5 and β5) close to the residues of the RecB-like sequence motif (Asp99, Glu113, Lys115, and Tyr148) (Figures 6 and S6). In the SSO0001 toroid, all 10 active sites are exposed to the interior side of the central channel, where the [Fe-S] clusters are located and whose surface is dominated by positively charged residues (Figures 6 and S7). The opposite side of the SSO0001 protomers (exposed to the external side of the ring) has a predominantly negatively charged surface with small portals (Figure S7). Inside the SSO0001 protomer, the two openings are connected by the internal tunnel with ∼10−12 Å diameter accommodating the Mn^{2+} ion (Figure 6a). Compared to the λ exonuclease trimeric

dx.doi.org/10.1021/ja408729b | J. Am. Chem. Soc. 2013, 135, 17476−17487

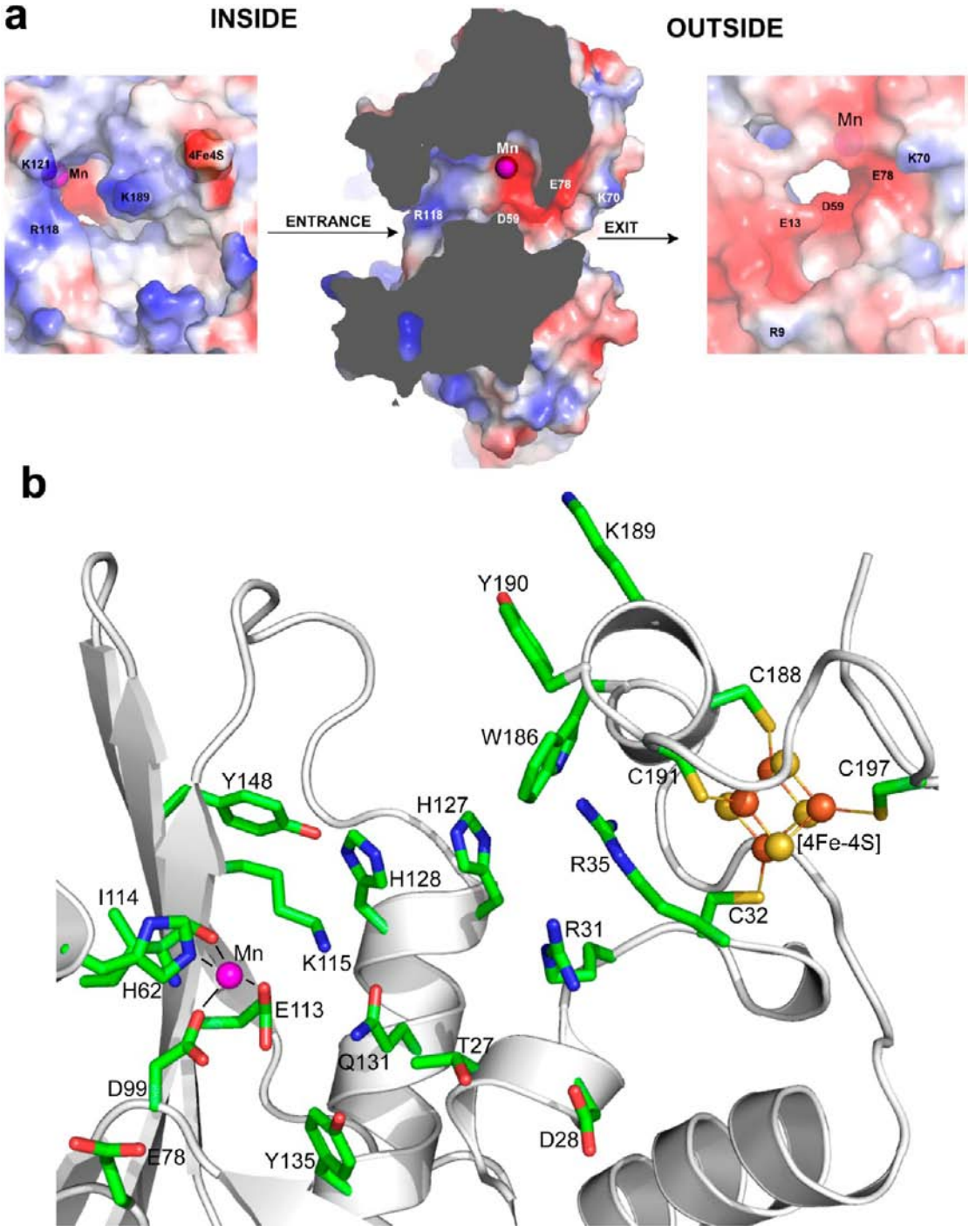


Figure 6. Internal tunnel and active site of SSO0001. (a) Three views of the internal tunnel related by 90° rotation: from the toroid interior, an axial slice through the "dimer", and from the toroid exterior. The protein is shown in surface charge presentation with a blue (positively charged) to red (negatively charged) gradient and low degree transparency to reveal the position of Mn^{2+} (shown as magenta colored balls). For clarity, only one tunnel is shown, and several charged residues are labeled to indicate their positions in different images. Electrostatic potential is indicated: −25kTe-1 in red and 25kTe-1 in blue. (b) Close-up view of the SSO0001 active site showing the position of the Mn^{2+} ion (magenta-colored ball), the [4Fe-4S] cluster (orange and yellow balls), and the active site residues (shown as sticks) along a SSO0001 ribbon (gray).

and RecE tetrameric rings, the SSO0001 toroid has a much larger central channel and is composed of two pentameric rings.

The Mn^{2+} ion is located 20 Å away from the [4Fe-4S] cluster and is coordinated by the side chains of conserved or semiconserved residues His62 (2.3 Å), Asp99 (2.1 Å), and Glu113 (2.1 Å), as well as the main-chain carbonyl oxygen of Ile114 (2.1 Å) (Figures 6b and S6). Similar coordination of the metal ion was found in the λ exonuclease (Asp119, Glu129, and the backbone carbonyl of Leu130), *E. coli* RecB (His956, Asp1067, Asp1080, and the backbone carbonyl of Tyr1081), and *E. coli* RecE (His652, Asp748, Asp759, and the backbone carbonyl of Val760)[16,30,31] (Figure S8). In SSO0001, Asp99 and Glu113 are part of the RecB motifs II and III, which also include the closely located Lys115 (motif III), Gln131, and Tyr135 (QxxxY motif) (Figures 6b and S1). Based on the SSO0001 structure, its RecB motif I residue is Glu78 (7.1 Å

dx.doi.org/10.1021/ja408729b | J. Am. Chem. Soc. 2013, 135, 17476−17487

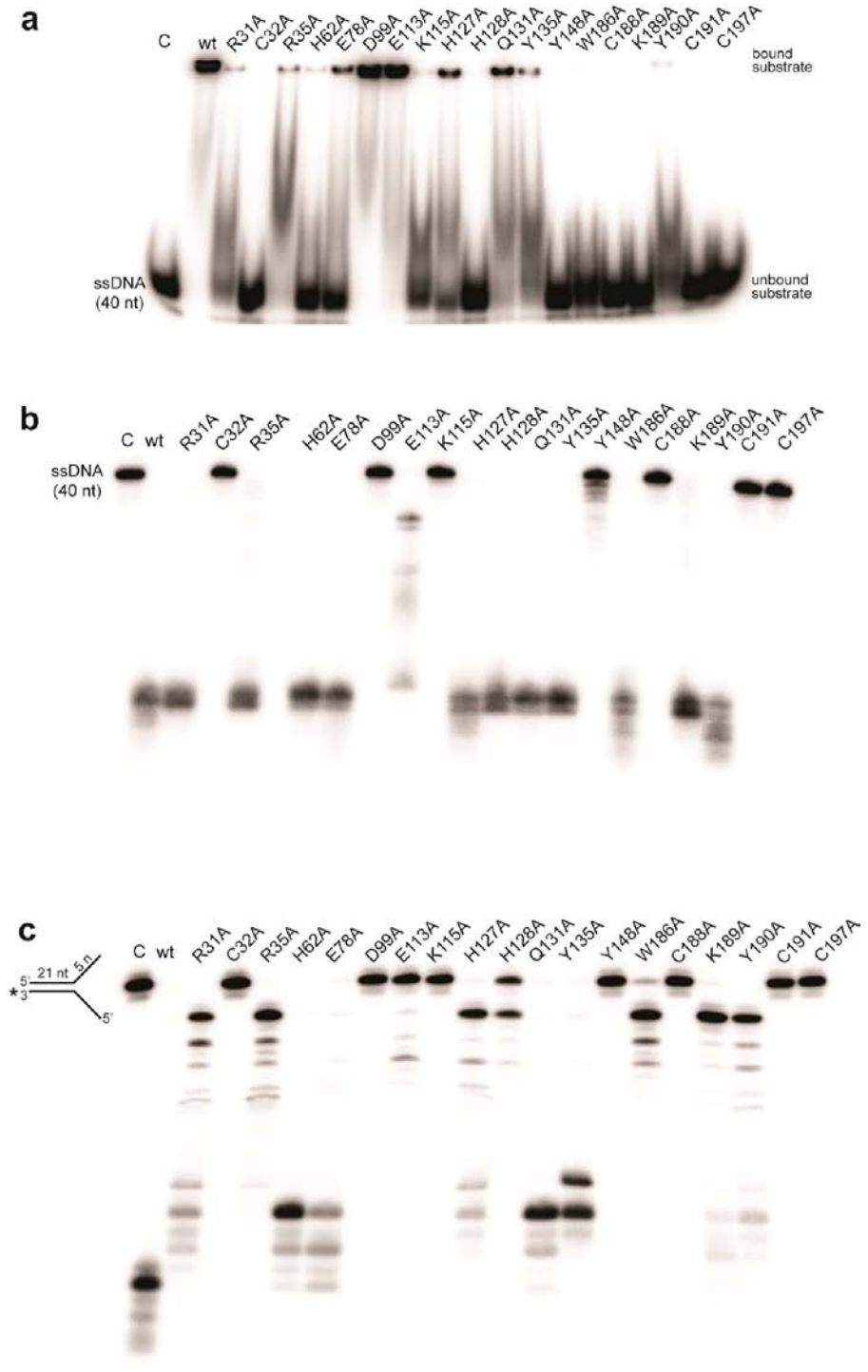


Figure 7. Site-directed mutagenesis of SSO0001. Binding (a) and cleavage (b,c) of the 3′-[^{32}P]-labeled ssDNA (390 nM SSO0001) or splayed arm substrate (65 nM SSO0001) by the purified wild-type (wt) or mutant SSO0001 proteins. Cleavage of ssDNA was carried out for 12 min at 50 °C, while cleavage of splayed arm substrates was for 20 min at 45 °C. Lanes C show incubation without SSO0001 addition.

from the motif II Asp99), which is located near the exit from the internal channel and might contribute to product release (Figure 6). The bottom part of the active site (between Mn^{2+} and [Fe-S] cluster) is occupied by the side chains of semiconserved Arg31, Arg35, His127, His128, and Trp186 representing the potential binding site for a nucleic acid substrate, whereas the closely located Lys189 and Tyr190 might form a wedge structure involved in DNA strand separation (Figure 6a).

Site-Directed Mutagenesis and Potential Catalytic Mechanism of SSO0001. The four SSO0001 mutant proteins with conserved cysteines mutated to Ala (C32A, C188A, C191A, and C197A) produced colorless preparations, whose gel-filtration profiles showed the formation of various protein aggregates (Figure S3f). These mutant proteins were inactive in both binding and cleavage of ssDNA and splayed arm substrate, indicating that the [Fe-S] cluster domain contributes to the stabilization of the SSO0001 active site (Figure 7). In addition,

dx.doi.org/10.1021/ja408729b | J. Am. Chem. Soc. 2013, 135, 17476−17487

the [Fe-S] cluster domains are oriented toward the central channel of the SSO0001 toroid, providing a potential "wedge" structure for DNA strand separation (Figure 5b). Similarly, alanine replacement of the [Fe-S] cluster-coordinating cysteines of AddB eliminated the ability of this protein to bind duplex DNA.[26] In SSO0001, a strong negative effect of mutation on both substrate binding and cleavage was also observed in the K115A and Y148A proteins, whereas D99A and E113A showed significant binding but very low or no nuclease activity, confirming the importance of Mn^{2+} coordination and RecB motif residues for nuclease activity of SSO0001 (Figure 7). The absence of nuclease activity in the SSO0001 D99A protein was also reported by Zhang et al.[20] The H62A and H128A proteins retained high nuclease activity against ssDNA but showed reduced ssDNA binding, suggesting that these residues contribute to substrate coordination. The mutant proteins H127A, Q131A, and Y135A exhibited both significant DNA binding and nuclease activity against ssDNA (Figure 7). Interestingly, the mutant proteins E113A and Y148A appear to be affected in both activity and ability to cleave short ssDNA, as they produced only a limited number of long products and no short products (Figure 7). In the SSO0001 structure (Figure 6b), the highly conserved Tyr148 is positioned close to the catalytic Lys115 (motif-III) and likely contributes to the coordination of the 5′-DNA strand in the active site. In addition, the R31A, R35A, H127A, H128A, W186A, K189A, and Y190A proteins retained high nuclease activity against ssDNA with the formation of short products, but they were unable to produce short products from the splayed arm substrate, confirming our hypothesis that these residues are involved in the unwinding activity of SSO0001 and supporting our assumption that the closely located large opening represents the entry gate into the SSO0001 active site (Figure 7).

Most characterized nucleases, including the RecB-like enzymes, have been proposed to use a catalytic mechanism based on the two metal ion centers.[22] Although the apo (nonliganded) structure of the λ exonuclease showed the presence of one metal ion,[30] the structure of this enzyme in complex with dsDNA revealed two Mg^{2+} ions (Mg^A and Mg^B) bridged by the side chain of Asp119 on one side and the scissile phosphate of the DNA strand on the other.[29] Mg^A (which is present in the apo-structure) is bound to Glu129, the main-chain carbonyl of Leu130, and two water molecules, one of which is also coordinated by Lys131 and is positioned for in-line attack on the scissile phosphate[29] (Figure S8). Mg^B is bound only in the presence of DNA and is proposed to stabilize the transition state through the bidentate coordination to the scissile phosphate. A similar catalytic mechanism can also be proposed for SSO0001. In this model, the catalytic water molecule is coordinated by the Mn^{2+} ion (Figure 6a), Lys115, and phosphate, whereas the second metal ion (bound to Asp99), Gln131, Tyr135, and Tyr148 might interact with the DNA phosphates.

DISCUSSION

Our work, together with the previous report on SSO0001,[20] demonstrates that Cas4 proteins represent a diverse group of metal-dependent exonucleases with 5′→3′ and 3′→5′ polarity, which also exhibit endonuclease activity against ssDNA (Figure 1). SSO0001 also demonstrated the presence of ATP-independent DNA unwinding activity against splayed arm substrates containing both 3′- and 5′-ssDNA overhangs. Potentially, SSO0001 can use a mechanism of dsDNA processing proposed for the λ and RecE exonucleases, in which the substrate 5′-end enters the active site, whereas the 3′-end moves uncleaved through the central channel.[29,31] However, the structure of SSO0001 also suggests an alternative mechanism, in which DNA strands are separated at the top edge of the ring and the uncleaved 3′-ended strand is redirected out of the central channel, whereas the 5′-ended strand of dsDNA approaches the active site through the central channel. Thus, SSO0001 and other CRISPR-associated RecB-like proteins comprise a new group of RecB-like nucleases with a distinctive and complex mode of dsDNA binding, unwinding, and processing.

In CRISPR interference, the Cas4 proteins have been predicted to be involved in the adaptation step together with Cas1 and Cas2 proteins, which is supported by the experimental observation of *in vitro* complex formation between the Cas4, Cas1, and Cas2 proteins from the crenarchaeon *T. tenax*.[13] Multiple Cas4 proteins present in *S. solfataricus* and other genomes might have different substrate or PAM preferences or can perform different functions. A potential functional role of SSO0001 and other Cas4 proteins in CRISPR immunity is based on the presence of 5′→3′ exonuclease and DNA unwinding activities, which produce 3′-ssDNA overhangs as potential intermediates in the process of new spacer addition (Figures 1 and 2).[20] Furthermore, because Cas4 proteins are likely to function in complex with Cas1 and Cas2 proteins *in vivo*,[13] the toroidal structures of SSO0001 and other Cas4 can serve as sliding clamps for other Cas proteins. Interestingly, the *E. coli* CRISPR system lacks a Cas4 protein, but its Cas1 protein (YgbT) has been shown to interact physically and genetically with RecB,[32] suggesting that the house-keeping RecB nuclease might compensate for the absence of Cas4 and contribute to CRISPR interference.

In addition to recombinational DNA repair, the *E. coli* RecBCD complex is involved in the protection of host cells from viral invasions through direct degradation of foreign linear DNAs,[15] a role which is similar to that of the CRISPR system. If the multiple active sites of the SSO0001 decameric toroid can function simultaneously, then it might represent a powerful DNA-degrading machine. We hypothesize that, in CRISPR interference, SSO0001 and other Cas4 proteins quickly degrade viral DNA molecules using the combination of DNA unwinding and nuclease activities, whereas the host chromosomes might be protected from the degradation by other host proteins or possibly through the recognition of specific "Chi-like" sequences.

In addition, recent data suggest that the cellular role of Cas4 proteins is not limited by CRISPR interference, but they might also contribute to stress response and DNA repair. A genome-wide transcriptional study of the *S. solfataricus* response to UV-induced DNA damage has revealed that the expression of the SSO0001 gene is induced together with other genes involved in DNA repair.[33] In addition, a recent work on the *T. tenax* CRISPR system revealed an increased expression of its *cas4* gene in UV-treated cells.[13] The potential role of Cas4 proteins in DNA repair might be based on their RecB-like exonuclease activity and formation of 3′-ssDNA overhangs. Together with previous works on other Cas proteins (*E. coli* Cas1, *P. furiosus* Cascade-like complex, *M. jannaschii* Cas7, *S. solfataricus* Cas8),[32,34−37] these findings suggest that some components of the CRISPR system can also contribute to stress response

dx.doi.org/10.1021/ja408729b | J. Am. Chem. Soc. 2013, 135, 17476−17487

and DNA repair, a role originally proposed for the thermophilic Cas proteins 10 years ago.[38]

CONCLUSIONS

This study shows that the Cas4 proteins, one of the core protein families of CRISPR interference, represent a diverse group of Mg^{2+}- or Mn^{2+}-dependent exonucleases. The Cas4 proteins SSO0001 and Pcal_0546 belong to the DUF83 family of RecB-like proteins and cleave ssDNA with 5′→3′ polarity, whereas SSO1391 (DUF911) can degrade ssDNA in both directions. All three proteins also exhibit Mn^{2+}-dependent endonuclease activity, and the presence of ATP-independent DNA unwinding activity toward dsDNA was demonstrated in SSO0001. The crystal structure of SSO0001 revealed a decameric toroid with a large 55 Å diameter central channel formed by five tightly packed "dimers". In the SSO0001 protomer, the active site is located inside the narrow tunnel and accommodates the four RecB motif residues and one Mn^{2+} ion. The 10 active sites of the SSO0001 toroid are open toward the central channel, which has a predominantly positively charged surface, suggesting that it might serve as an entry point for the DNA substrate. Near the active site, each SSO0001 protomer contains one [4Fe-4S] cluster, which is important for both maintaining the decameric state and nuclease activity. Site-directed mutagenesis of SSO0001 revealed that the RecB motif residues (Asp99, Glu113, Lys115, and Tyr148) and four conserved Cys coordinating the [Fe-S] cluster (Cys32, Cys188, Cys191, and Cys197) are important for nuclease activity. In combination with the structure, this suggests a catalytic mechanism for Cas4 nucleases. We propose that the Cas4 proteins can have at least two functions in the CRISPR mechanism, including the formation of 3′-overhangs for generation of new spacers and direct degradation of foreign DNAs using a combination of DNA unwinding and nuclease activities. Future research will provide new insights into the role and molecular mechanisms of Cas4 proteins in the CRISPR/Cas system.

EXPERIMENTAL SECTION

Protein Expression, Purification, and Mutagenesis. Cloning and purification of the 6His-tagged SSO0001, SSO1391, and Pcal_0546, as well as site-directed mutagenesis of SSO0001, were performed as described previously.[39] The oligomeric state of purified proteins was analyzed using size-exclusion chromatography on a Superdex 200 HiLoad 16/60 column (Amersham Biosciences) and analytical ultracentrifugation as described previously.[40]

Preparation of Nucleic Acid Substrates. The ssDNA and RNA oligonucleotides used in this work (Table S1) were purchased from IDT (USA). The oligonucleotides were [^{32}P]-labeled at the 3′-end using the calf thymus Terminal Transferase TdT or at the 5′-end using T4 polynucleotide kinase (PNK, BioLabs) and purified as previously described.[41] The synthetic dsDNA or splayed arm substrates were prepared by annealing the oligonucleotides shown in Tables S1 and S2. The 5′-phosphorylated and 5′-biotinylated substrates were purchased from IDT (USA).

Enzymatic and DNA Binding Assays. The reaction mixture for multiple-turnover ssDNase assays with SSO0001 contained 50 mM CAPS (pH 9.5), 100 mM KCl, 10 mM $MgCl_2$ (or 0.5 mM for other divalent cations), 1 mM DTT (for exonuclease assay), and 0.1 μM 3′- or 5′-[^{32}P]-labeled substrate, 5 nM ssDNA of the M13mp18 phage, or 20 nM pUC19 dsDNA. The reaction mixture for ssRNase assays contained 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.1 μM 3′-[^{32}P]-labeled RNA, and SSO0001 (as indicated). The reaction mixtures for ssDNase assays with Pcal_0546 and SSO1391 contained 50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$ (or 2 mM of other divalent metal ion), 50 mM KCl (for SSO1391), 1 mM DTT (for exonuclease assays), and 0.1 μM 3′- or 5′-[^{32}P]-labeled substrate (or 5 nM M13mp18 ssDNA or 20 nM pUC19 dsDNA). The solutions were incubated at 42−50 °C and processed as previously described.[41] Analysis of the ssDNA product ends was carried out using T4 PNK as previously described.[41] The reaction for the LC-MS-based analysis of SSO0001 products was performed using optimal reaction conditions with 6 μM SSO0001 and poly-dC_{10} substrate (10 μg). LC-MS analysis of the products of poly-dC_{10} cleavage by SSO0001 was performed using an Exactive orbitrap mass spectrometer with an electrospray ionization source and an Accela 1250 U-HPLC system (all from Thermo Scientific) essentially as previously described.[42] The reaction mixture for DNA-binding assays contained 50 mM Tris-HCl (pH 8.0), 5 mM KCl, 0.5 mM $CaCl_2$, 1 mM DTT, and SSO0001 (as indicated). The solutions were incubated at 25 °C for 10 min and analyzed by native gel electrophoresis (10% acrylamide). These reaction conditions are optimal for DNA binding, whereas DNA cleavage was prevented by low incubation temperature, short incubation time, and Ca^{2+} ions.

Crystallization and Structure Determination of SSO0001. SSO0001 was crystallized at room temperature using the sitting drop vapor diffusion protocol by mixing 1 μL of the selenomethionine-substituted protein (10 mg/mL) with 1 μL of the crystallization solution containing 25% (w/v) PEG 2000 MME, 0.2 M NaCl, 100 mM sodiumacetate (pH 5.6), and 0.3 M NDSB 201. The crystals were stabilized with the crystallization buffer containing 3 M NaCl followed by cryoprotection in Paratone-N prior to flash-freezing in liquid nitrogen. Data collections were carried out at the beamlines 19-ID and 19-BM of the Structural Biology Center, Advanced Photon Source, Argonne National Laboratory.[43] Prior to data collection, an X-ray fluorescence spectrum was recorded, which identified the presence of Se, Fe, and Mn in protein crystals. Data were collected from a single crystal to 2.35 Å at the wavelength of 0.9794 Å and processed using the program HKL3000[44] (Table S3). The structure of SSO0001 was determined by the Se-methionine SAD phasing, density modification, and initial model building as implemented in the Auto-Rickshaw package.[45] The initial models (~60% complete) were further built manually using the program COOT[46] and refined with the PHENIX[47] and REFMAC[48] programs. The presence and identity of metals were verified by collecting data sets (from the same crystal) at the iron X-ray K absorption edge and just above the Mn absorption edge and calculating anomalous Fourier difference maps (Figure S6). Analysis and validation of structures were performed using MOLPROBITY[49] and COOT[46] validation tools. The final model was refined to R_{work}/R_{free} = 0.205/0.249, and it shows good geometry with no outliers in the Ramachandran plot. Data collection and refinement statistics are summarized in Table S3. The atomic coordinates have been deposited in the Protein Data Bank, with accession code 4IC1.

ASSOCIATED CONTENT

Supporting Information

Experimental and structural data presented in eight supplementary figures and three supplementary tables. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

a.iakounine@utoronto.ca

Author Contributions

§S.L. and N.B. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank all members of the Structural Proteomics in Toronto (SPiT) Centre and the Structural Biology Center at Argonne National Laboratory (especially Frank Rotella) for help in

conducting the experiments. Dr. David Bernick (University of California, Santa Cruz) is thanked for sending the *Pyrobaculum calidifontis* genomic DNA. This work was supported by the Government of Canada through Genome Canada and Ontario Genomics Institute (2009-OGI-ABC-1405), Ontario Research Fund (ORF-GL2-01-004), Natural Science and Engineering Research Council of Canada, and in part by a grant from the U.S. National Institutes of Health Grant GM094585 and by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02-06CH11357.

REFERENCES

(1) Makarova, K. S.; Grishin, N. V.; Shabalina, S. A.; Wolf, Y. I.; Koonin, E. V. *Biol. Direct* **2006**, *1*, 7.
(2) Makarova, K. S.; Haft, D. H.; Barrangou, R.; Brouns, S. J.; Charpentier, E.; Horvath, P.; Moineau, S.; Mojica, F. J.; Wolf, Y. I.; Yakunin, A. F.; van der Oost, J.; Koonin, E. V. *Nat. Rev. Microbiol.* **2011**, *9*, 467.
(3) Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A. *Nature* **2012**, *482*, 331.
(4) Hale, C. R.; Zhao, P.; Olson, S.; Duff, M. O.; Graveley, B. R.; Wells, L.; Terns, R. M.; Terns, M. P. *Cell* **2009**, *139*, 945.
(5) Westra, E. R.; Swarts, D. C.; Staals, R. H.; Jore, M. M.; Brouns, S. J.; van der Oost, J. *Annu. Rev. Genet.* **2012**, *46*, 311.
(6) Brouns, S. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J.; Snijders, A. P.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der Oost, J. *Science* **2008**, *321*, 960.
(7) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. *Science* **2012**, *337*, 816.
(8) Garneau, J. E.; Dupuis, M. E.; Villion, M.; Romero, D. A.; Barrangou, R.; Boyaval, P.; Fremaux, C.; Horvath, P.; Magadan, A. H.; Moineau, S. *Nature* **2010**, *468*, 67.
(9) Haft, D. H.; Selengut, J.; Mongodin, E. F.; Nelson, K. E. *PLoS Comput. Biol.* **2005**, *1*, e60.
(10) Sorek, R.; Lawrence, C. M.; Wiedenheft, B. *Annu. Rev. Biochem.* **2013**, in press.
(11) Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P. *Science* **2007**, *315*, 1709.
(12) Fineran, P. C.; Charpentier, E. *Virology* **2012**, *434*, 202.
(13) Plagens, A.; Tjaden, B.; Hagemann, A.; Randau, L.; Hensel, R. *J. Bacteriol.* **2012**, *194*, 2491.
(14) Kinch, L. N.; Ginalski, K.; Rychlewski, L.; Grishin, N. V. *Nucleic Acids Res.* **2005**, *33*, 3598.
(15) Kowalczykowski, S. C.; Dixon, D. A.; Eggleston, A. K.; Lauder, S. D.; Rehrauer, W. M. *Microbiol. Rev.* **1994**, *58*, 401.
(16) Singleton, M. R.; Dillingham, M. S.; Gaudier, M.; Kowalczykowski, S. C.; Wigley, D. B. *Nature* **2004**, *432*, 187.
(17) Saikrishnan, K.; Yeeles, J. T.; Gilhooly, N. S.; Krajewski, W. W.; Dillingham, M. S.; Wigley, D. B. *EMBO J.* **2012**, *31*, 1568.
(18) Garrett, R. A.; Shah, S. A.; Vestergaard, G.; Deng, L.; Gudbergsdottir, S.; Kenchappa, C. S.; Erdmann, S.; She, Q. *Biochem. Soc. Trans.* **2011**, *39*, 51.
(19) Gudbergsdottir, S.; Deng, L.; Chen, Z.; Jensen, J. V.; Jensen, L. R.; She, Q.; Garrett, R. A. *Mol. Microbiol.* **2011**, *79*, 35.
(20) Zhang, J.; Kasciukovic, T.; White, M. F. *PLoS One* **2012**, *7*, e47232.
(21) Wang, J.; Chen, R.; Julin, D. A. *J. Biol. Chem.* **2000**, *275*, 507.
(22) Yang, W. *Q. Rev. Biophys.* **2011**, *44*, 1.
(23) Little, J. W. *J. Biol. Chem.* **1967**, *242*, 679.
(24) Pingoud, A.; Fuxreiter, M.; Pingoud, V.; Wende, W. *Cell. Mol. Life Sci.* **2005**, *62*, 685.
(25) Mojica, F. J.; Diez-Villasenor, C.; Garcia-Martinez, J.; Almendros, C. *Microbiology* **2009**, *155*, 733.
(26) Yeeles, J. T.; Cammack, R.; Dillingham, M. S. *J. Biol. Chem.* **2009**, *284*, 7746.
(27) Holm, L.; Rosenstrom, P. *Nucleic Acids Res.* **2010**, *38*, W545.
(28) Yang, W.; Chen, W. Y.; Wang, H.; Ho, J. W.; Huang, J. D.; Woo, P. C.; Lau, S. K.; Yuen, K. Y.; Zhang, Q.; Zhou, W.; Bartlam, M.; Watt, R. M.; Rao, Z. *Nucleic Acids Res.* **2011**, *39*, 9803.
(29) Zhang, J.; McCabe, K. A.; Bell, C. E. *Proc. Natl. Acad. Sci. U.S.A* **2011**, *108*, 11872.
(30) Kovall, R.; Matthews, B. W. *Science* **1997**, *277*, 1824.
(31) Zhang, J.; Xing, X.; Herr, A. B.; Bell, C. E. *Structure* **2009**, *17*, 690.
(32) Babu, M.; Beloglazova, N.; Flick, R.; Graham, C.; Skarina, T.; Nocek, B.; Gagarinova, A.; Pogoutse, O.; Brown, G.; Binkowski, A.; Phanse, S.; Joachimiak, A.; Koonin, E. V.; Savchenko, A.; Emili, A.; Greenblatt, J.; Edwards, A. M.; Yakunin, A. F. *Mol. Microbiol.* **2011**, *79*, 484.
(33) Frols, S.; Gordon, P. M.; Panlilio, M. A.; Duggin, I. G.; Bell, S. D.; Sensen, C. W.; Schleper, C. *J. Bacteriol.* **2007**, *189*, 8708.
(34) Boonyaratanakornkit, B. B.; Miao, L. Y.; Clark, D. S. *Extremophiles* **2007**, *11*, 495.
(35) Gotz, D.; Paytubi, S.; Munro, S.; Lundgren, M.; Bernander, R.; White, M. F. *Genome Biol.* **2007**, *8*, R220.
(36) Strand, K. R.; Sun, C.; Li, T.; Jenney, F. E., Jr.; Schut, G. J.; Adams, M. W. *Arch. Microbiol.* **2010**, *192*, 447.
(37) Williams, E.; Lowe, T. M.; Savas, J.; DiRuggiero, J. *Extremophiles* **2007**, *11*, 19.
(38) Makarova, K. S.; Aravind, L.; Grishin, N. V.; Rogozin, I. B.; Koonin, E. V. *Nucleic Acids Res.* **2002**, *30*, 482.
(39) Beloglazova, N.; Brown, G.; Zimmerman, M. D.; Proudfoot, M.; Makarova, K. S.; Kudritska, M.; Kochinyan, S.; Wang, S.; Chruszcz, M.; Minor, W.; Koonin, E. V.; Edwards, A. M.; Savchenko, A.; Yakunin, A. F. *J. Biol. Chem.* **2008**, *283*, 20361.
(40) Snider, J.; Gutsche, I.; Lin, M.; Baby, S.; Cox, B.; Butland, G.; Greenblatt, J.; Emili, A.; Houry, W. A. *J. Biol. Chem.* **2006**, *281*, 1532.
(41) Beloglazova, N.; Petit, P.; Flick, R.; Brown, G.; Savchenko, A.; Yakunin, A. F. *EMBO J.* **2011**, *30*, 4616.
(42) Lu, W.; Clasquin, M. F.; Melamud, E.; Amador-Noguez, D.; Caudy, A. A.; Rabinowitz, J. D. *Anal. Chem.* **2010**, *82*, 3212.
(43) Rosenbaum, G.; Alkire, R. W.; Evans, G.; Rotella, F. J.; Lazarski, K.; Zhang, R. G.; Ginell, S. L.; Duke, N.; Naday, I.; Lazarz, J.; Molitsky, M. J.; Keefe, L.; Gonczy, J.; Rock, L.; Sanishvili, R.; Walsh, M. A.; Westbrook, E.; Joachimiak, A. *J. Synchrotron Radiat.* **2006**, *13*, 30.
(44) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. *Acta Crystallogr. D: Biol. Crystallogr.* **2006**, *62*, 859.
(45) Panjikar, S.; Parthasarathy, V.; Lamzin, V. S.; Weiss, M. S.; Tucker, P. A. *Acta Crystallogr. D: Biol. Crystallogr.* **2005**, *61*, 449.
(46) Emsley, P.; Cowtan, K. *Acta Crystallogr. D: Biol. Crystallogr.* **2004**, *60*, 2126.
(47) 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.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 213.
(48) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr. D: Biol. Crystallogr.* **1997**, *53*, 240.
(49) Davis, I. W.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res.* **2004**, *32*, W615.