The CRISPR-associated Cas4 protein from *Leptospira interrogans* demonstrate versatile nuclease activity

Bhuvan Dixit, Vineet Anand, Md. Saddam Hussain, Manish Kumar [*]

Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, 781039, Assam, India

A B S T R A C T

The Cas4 protein is one of the core CRISPR-associated (Cas) proteins implicated in the adaptation module in many variants of the CRISPR-Cas system in prokaryotes against the invading genetic elements. Cas4 is recognized as a DNA exonuclease that contains a RecB nuclease domain and a Fe-S cluster-binding module. In *Leptospira interrogans* serovar Copenhageni strain Fiocruz L1–130, the *cas4* gene is functionally transcribed as an active component of the CRISPR-Cas I-B system. Investigation of nuclease activity of Cas4 (LinCas4) of the *L. interrogans* illustrated divalent-metal cofactor (Mn^{2+} or Mg^{2+}) dependent endonuclease activity on the DNA substrate. In agreement, mutation of the selective metal interacting residues (Asp^{74} and Glu^{87}) curtails the DNA cleavage activity in LinCas4. Computational modeling shows metal-ion interacting residues (Asp^{74} and Glu^{87}) in the LinCas4 to be a part of the RecB motifs II and III, the same as other Cas4 orthologs. The mutation of a potential DNA interacting residue in the LinCas4 ($LinCas4^{Y132A}$) or one of the four cysteine residues ($LinCas4^{C18A}$) involved in coordinating the 4Fe-4S cluster did not perturb its DNase activity. Iron chelation assay of the purified LinCas4 demonstrated it in the apostate conformation. Reconstitution of the Fe-S cluster in the LinCas4 under *in vitro* condition displayed its coordination with four iron atoms per LinCas4 monomer and was confirmed by the UV and CD spectroscopy studies.

1. Introduction

CRISPR-Cas (Clustered regularly interspaced short palindromic repeat-CRISPR associated proteins) constitutes a prokaryotic acquired defense system, which fends off the invasion of foreign genetic materials such as viruses and mobile elements (Barrangou et al., 2007; Yosef et al., 2012; Heler et al., 2014; Hille et al., 2018). The CRISPR-Cas systems have been classified into six major types (Type I-VI), and each type is further divided into various subtypes based on the presence of signature *cas* genes and the locus arrangement (Makarova et al., 2015; Koonin et al., 2017). The process of acquiring the foreign genetic material into CRISPR arrays constitutes the first stage of CRISPR immunity (Westra et al., 2012). In this stage, the short foreign DNA fragments are excised from the invasive genetic element and integrated into the CRISPR array as new spacers by the conserved core proteins Cas1 and Cas2 (Makarova et al., 2015; Koonin et al., 2017). Protospacer-adjacent motifs (PAM) in the DNA of invasive genetic elements determine the acquisition efficiency and integration specificity during the spacer acquisition process (Deveau et al., 2008; Mojica et al., 2009). In CRISPR-Cas types or subtypes (I-A, I-B, I-C, I-D, I-U, and II-B), an additional conserved factor most of which belong to the Cas4 family is functionally associated with the adaptation event of prespacer DNA maturation and high-fidelity integration (Makarova et al., 2015; Koonin et al., 2017).

Evolutionarily, unlike many other Cas proteins, Cas4 can often be encoded as solo-Cas4 outside the CRISPR-Cas loci and is also found in the mobile genetic elements (MGE-Cas4) (Hudaiberdiev et al., 2017). In comparison to the archaea that encodes Cas4 in 90% of its genome, only about 20% of the bacterial genomes possess Cas4 (Hudaiberdiev et al., 2017). Recently, in *Pyrococcus furiosus*, two distinct *cas4* genes (*cas4–1* and *cas4–2*) together have been suggested to control the length, orientation, and PAM selection for the new spacers during the CRISPR expansion (Shiimori et al., 2018). In the same line, crenarchaeon *Sulfolobus islandicus* has shown that the two paralogous *cas4* genes (one denoted as *csa1*) in the *cas* operon are essential, indicating that these proteins could form a complex in which the two paralogs would execute different tasks (Liu et al., 2017). The *Sulfolobus solfataricus* genome encodes five Cas4-like proteins (Sso0001, Sso1392, Sso1449, Sso1391, Sso1451), which stake low sequence similarity (15 to 30% sequence

---

identity) (Lemak et al., 2013).

The crystal structures of the Cas4 proteins from *S. solfataricus* (Sso0001) and *Pyrobaculum calidifontis* (PcaCas4) indicate that they carry two domains: N-terminal RecB-like nuclease domain and a C-terminal domain containing a Fe-S cluster coordinated by four conserved cysteine residues (Zhang et al., 2012; Lemak et al., 2014). The Sso0001 possesses a 4Fe-4S cluster coordinated by four cysteine residues (Cys32, 188, 191, and 197) (Lemak et al., 2013). Mutation of one of the cysteine residues of Sso0001 results in the loss of its nuclease activity, suggesting that the 4Fe-4S cluster is possibly required to maintain protein stability (Lemak et al., 2013). On the other hand, PcaCas4 possesses a 2Fe-2S cluster, which is also coordinated by four cysteine residues (Cys64, 200, 203, and 209) (Lemak et al., 2014). However, mutation of the cysteine residues does not hinder the nuclease activity in PcaCas4 (Lemak et al., 2014). Furthermore, Mn^{2+} and Mg^{2+} ions are bound in the active site by aspartate, glutamate, and histidine residues of Sso0001 and PcaCas4, respectively (Lemak et al., 2013, 2014). The crystal structure of Sso0001 revealed a decameric toroid formed by five tightly bound dimers, whereas crystal structure and size-exclusion chromatography shows that PcaCas4 exists in the monomeric form (Lemak et al., 2013, 2014). The well-characterized nuclease moieties of proteins involved in DNA recombination and repair in bacteria (RecB, AddB) and eukaryotes (Dna2) are homologous to Cas4 protein (Krajewski et al., 2014; Zhou et al., 2015; Wilkinson et al., 2016). To date, various biochemical analyses have ascertained that the Cas4 can exhibit 5′-to-3′ and 3′-to-5′ DNA exonuclease activity, signifying it to be involved in yielding single-stranded DNA overhangs intermediates essential for the insertion of the new spacer in a CRISPR array (Zhang et al., 2012; Lemak et al., 2013, 2014). The Sso0001, Sso1391, and PcaCas4 cleave ssDNA substrates with the formation of products containing 5′-hydroxyls and 3′-phosphates (Lemak et al., 2013, 2014). In contrast, the characterized RecB-like nucleases and λ exonuclease have been shown to generate the products with 5′-phosphates and 3′-hydroxyls (Little, 1967; Pingoud et al., 2005).

The pathogenic and intermediate strains of *Leptospira* that cause leptospirosis disease in animals and human possess three subtypes (subtype I-B, subtype I-C, and subtype I-E) of the CRISPR-Cas system whereas, type V CRISPR-Cas system has been recently reported in saprophytic strain of *Leptospira* (*L. biflexa*) (Makarova et al., 2011; Xiao et al., 2019; Arbas et al., 2021). In *Leptospira* type I-B CRISPR-Cas system computational studies have revealed a weakly conserved PAM sequence (5′-TAC-3′) that lies upstream to the protospacers targeted by Cas protein (Xiao et al., 2019). Until recently, genetic manipulation in pathogenic species of *Leptospira* has proved to be an arduous task due to its lower efficiency (Croda et al., 2008; Pappas et al., 2015). The reason for the recalcitrant genetic manipulation of *Leptospira* is still unknown, but it is believed that the CRISPR-Cas system might be one of the possible malefactors (Xiao et al., 2019). However, recent advances in the genetic manipulation of *Leptospira* through CRISPR interference which employs the type II CRISPR/Cas9 system of *Streptococcus pyogenes,* have been used for generating mutants of saprophytic *Leptospira biflexa* and pathogenic *L. interrogans* (Fernandes et al., 2019, 2021). Therefore, understanding the CRISPR-Cas system of *Leptospira* is important since it may enable us to generate an endogenous tool for performing genetic manipulation in the future.

The CRISPR-Cas subtype I-B system of *L. interrogans* hauls a functional *cas4* gene that has previously been disclosed to be transcriptionally active (Dixit et al., 2016). As part of our ongoing investigation of the CRISPR-Cas system in *Leptospira*, in this study, we explored the biochemistry of Cas4 protein (LinCas4) present in the *L. interrogans*. We demonstrate that the recombinant LinCas4 is a magnesium or manganese-dependent endo-deoxyribonuclease (DNase) under *in vitro* conditions. LinCas4 can hold an iron-sulfur cluster, as has been documented in other Cas4 orthologs from *S. solfataricus* and *P. calidifontis* (Zhang et al., 2012; Lemak et al., 2014).

2. Results

2.1. Leptospira Cas4 demonstrates nuclease activity on the double-stranded DNA

The purified recombinant LinCas4 used in the nuclease assays (Fig. S1A and S1B) exhibits dsDNase activity in the presence of Mn^{2+} ion, which results in the cleavage of the circular dsDNA (Fig. 1A). In the presence of Ethylenediaminetetraacetic acid (EDTA) due to the chelation of divalent metal ion, the endonuclease activity of the LinCas4 was hampered. In another time-bound dsDNase assay, the nuclease activity of the LinCas4 (6 μM) indicates that it completely cleaves the circular dsDNA (12.5 nM) in 3 h (Fig. 1B). On substitution of metal-ion (Mn^{2+}) with the other divalent metals, the LinCas4 endonuclease activity was the same in the presence of the Mg^{2+} ion, whereas it was inhibited in the presence of the Ca^{2+} or Ni^{2+} ion (Fig. 1C). Functional metal ions are organized at the active site of the metalloenzymes and carry out a myriad of operations, such as substrate recognition/binding and the catalysis that jointly serves a wide variety of biological roles (Chen et al., 2018). Moreover, the LinCas4 displays nuclease activity at a wide pH range of 6.0–10.0 (Fig. 1D). The nuclease activity of the LinCas4 was inhibited below pH 6.0 and above pH 10.0. Also, at pH 3.0 and 11.0, a reduction in the DNA substrate migration is observed on the agarose gel electrophoresis, possibly due to retaining its affinity for DNA binding and losing its nuclease activity. The effect of salt on the endonuclease activity of the LinCas4 in the presence of the Mn^{2+} was analyzed using several monovalent-ion (KCl, NaCl, or NH_4Cl) individually at an increasing concentration. Noticeably, among the three monovalent salts, LinCas4 prefers KCl over NaCl and NH_4Cl for the nuclease activity (Fig. 1E). In the given set of nuclease assay, the LinCas4 may have retained some nicking ability on the circular DNA substrate resulting in the difference in the conformation of the plasmid DNA than the control. Non-specific nucleases are implicated in a comprehensive range of tasks that encompasses extra- and intracellular digestion, defense, replication, recombination, and repair (Rangarajan and Shankar, 2001; Yang, 2011).

2.2. Leptospira Cas4 demonstrates nuclease activity on the single-stranded DNA

To comprehend the nuclease activity diversity of the LinCas4, a viral single-stranded DNA was used as a substrate. LinCas4 (6 μM) could cleave the linear ssDNA (M13mp18; 21.4 nM) and the circular ssDNA (φX174; 29 nM) in 60 min, but specifically in the presence of a divalent-metal cofactor (Mn^{2+}) (Fig. 2A). In agreement, PcaCas4 (0.4 μM), Sso0001 (8 μM), and the Sso1391 (0.5 μM) also cleave the viral ssDNA (5 nM) in 20 to 45 min in the presence of cofactor Mn^{2+} (Lemak et al., 2013). This ascertains that the Cas4 proteins in varying concentrations display Mn^{2+} dependent endonucleolytic activity on the ssDNA. Furthermore, the PcaCas4 and Sso0001 have also been illustrated to exhibit a metal-dependent 5′-to-3′ exonuclease activity against the ssDNA substrates (Zhang et al., 2012; Lemak et al., 2013, 2014), whereas the Sso1391 has been reported to cleave the ssDNA in both the 5′-to-3′ and the 3′-to-5′ directions (Lemak et al., 2013, 2014). Thus, deviations in the exonuclease activity have been recorded among different Cas4 isoforms within the same organism. Influenced by the observation of the directional inconsistency in exonuclease activity, a similar time-dependent nuclease assay was carried out using the LinCas4 (1 μM) against 5′-FAM-labeled short oligonucleotide (36-mer; 250 nM) (Fig. 2B) and 3′-FAM-labeled short oligonucleotide (50-mer; 250 nM) (Fig. 2C). LinCas4 demonstrated a complete degradation of each of the labeled short oligonucleotides in 60 min and no difference in the cleavage of the short oligonucleotides labeled either at the 5′ or 3′ was noticed. The labeled oligo cleavage is in contrast to the reported PcaCas4 (2.5 μM) and the SsoCas4 activity (Sso0001; 8 nM and Sso1391; 2.2 μM), which can cleave directionally the 5′ and/or 3′-labeled short oligonucleotides (100 nM) into a series of products of the varying chain lengths
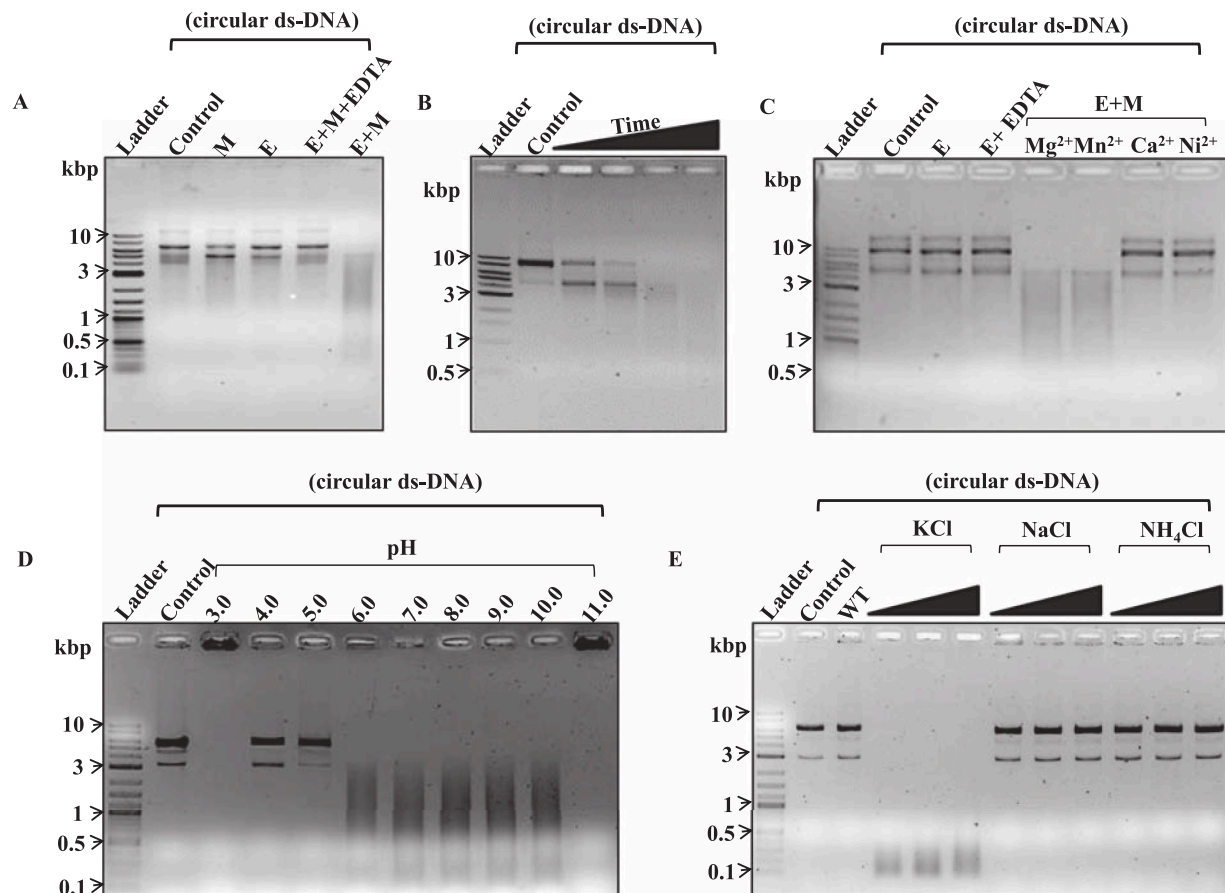
Figure 1

Fig. 1. LinCas4 nuclease activity on the double-stranded DNA. (A) Metal-ion dependent dsDNase activity. LinCas4 activity on the plasmid DNA (pTZ57R/T; 2.8 kbp; 12.5 nM) was performed in the presence or absence of Mn^{2+} with or without the EDTA and analyzed by the agarose gel electrophoresis. **(B)** Time-dependent dsDNase activity. Nuclease activity on the pTZ57R/T plasmid DNA (12.5 nM) for 3 h (0.5 h, 1 h, 1.5 h, 2 h, 2.5 h, and 3 h). **(C)** Effect of different divalent-metal on the dsDNA substrate cleavage. **(D)** Nuclease activity at different pH on the dsDNA. DNase activity in the LinCas4 was exhibited at pH 6.0–10.0. **(E)** Effect of the different monovalent-ions on the dsDNase activity. The DNA cleavage reaction of the LinCas4 shows optimal activity in KCl (50–150 mM). WT indicates nuclease activity without any monovalent-ions. In all panels, 'E' represents LinCas4 (6 μM), and 'M' denotes Mn^{2+}. All the reactions were carried out for 1 h at 37 °C and resolved on the 0.8% agarose gel unless stated. The results presented were confirmed by two independent experiments.
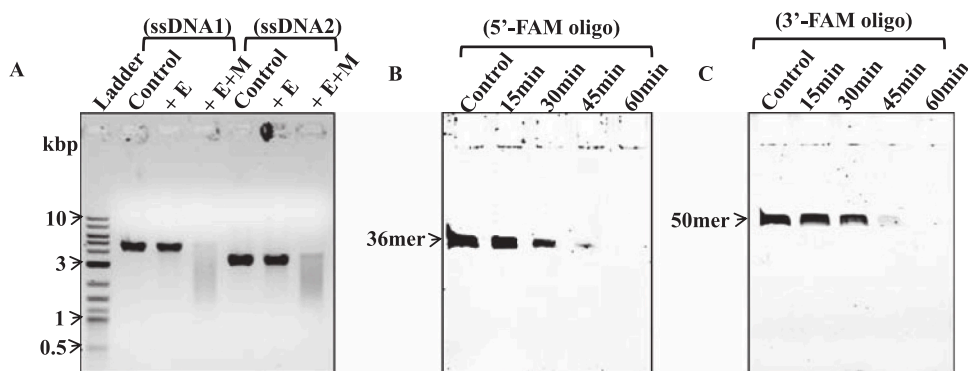

Fig. 2. LinCas4 nuclease activity on the single-stranded DNA. (A) Nuclease activity in the LinCas4 (6 μM) on the linear ssDNA (M13mp18 vector; 7.2 kb; 21.4 nM, ssDNA1) and the circular ssDNA (φX174 genome; 5.3 kb; 29 nM, ssDNA2) shows that Mn^{2+} ion is required for cleaving both the types of ssDNA. The reaction is resolved on the 0.8% agarose gel. **(B)** LinCas4 (1 μM) activity on the 5′-FAM-labeled short ssDNA oligonucleotide substrate (36-mer; 250 nM) show cleavage in the presence of a metal-ion and in a time-dependent manner (0, 15, 30, 45, and 60 min). **(C)** Lin-Cas4 (1 μM) activity on the 3′-FAM-labeled short ssDNA oligonucleotide substrate (50-mer; 250 nM), the) show cleavage in metal-ion presence and in a time-dependent manner (0, 15, 30, 45, and 60 min). The nuclease reaction in B and C are resolved on the 20% Urea PAGE. In all panels, 'E' represents LinCas4, and 'M' denotes Mn^{2+}. All the reactions were executed for 1 h at 37 °C. The results illustrated were confirmed by two independent experiments.

within 15–20 min of reaction commencement (Lemak et al., 2013).

2.3. Computational characterization of the LinCas4

A homology search for LinCas4 revealed its closest ortholog to be the SsoCas4 of *S. solfataricus* (sequence identity: 25.6% and query coverage: 74%). Low sequence similarity was observed with the other Cas4 orthologs present in the thermophilic organisms, such as SisCas4 of *S. islandicus* (sequence identity: 29.63% and query coverage: 38%) and one (PfuCas4–2) of the two Cas4 proteins of *P. furiosus* (18.46% and 31%). Interestingly, the *Pfucas4–2* is remotely located and not associated with the CRISPR (Shiimori et al., 2018). The least sequence similarity was observed with the PfuCas4–1 (26.09% and 11%), whereas no notable similarity was detected to the available crystallized structure of PcaCas4 of *P. calidifontis*.

A multiple sequence alignment (MSA) of the Cas4 proteins was performed to spot the conserved residues in the LinCas4 from its known

orthologs (Fig. 3). Cas4 orthologs reflected conservation of the metal-ion interacting residues and the RecB motifs through MSA (Fig. 3). The conserved residues in the protein sequences of LinCas4, Sso0001, and PcaCas4 have been represented in the form of cartoons (Fig. 4). The MSA analysis revealed three conserved metal-ion interacting residues (His[44], Asp[74], and Glu[87]) at the N-terminal domain of the LinCas4 (Fig. 3 and 4A). In the LinCas4, residues of RecB motif-I (Glu[58]), -II (Asp[74]), -III (Glu[87] and Lys[89]), and -QxxxY (Gln[109]-Tyr[113]) were also found to be conserved (Fig. 3 and 4A) with its ortholog Sso0001 and PcaCas4 (Fig. 3, 4B, and 4C) (Lemak et al., 2013, 2014). The Asp[74] and Glu[87] metal-ion interacting residues in the LinCas4 are part of the RecB motifs-II and -III in all the Cas4 orthologs (Fig. 3 and 4). The reported studies of the PcaCas4 and the Sso1391 suggest that the residues of the RecB motif -II, -III, and -QxxxY are critical for the nuclease activity (Lemak et al., 2014). In addition, a conserved iron-sulfur cluster is known to be present in the PcaCas4 and Sso0001, which is coordinated by four cysteine residues (Fig. 4B and 4C) (Bernick et al., 2012; Zhang et al., 2012; Lemak et al.,
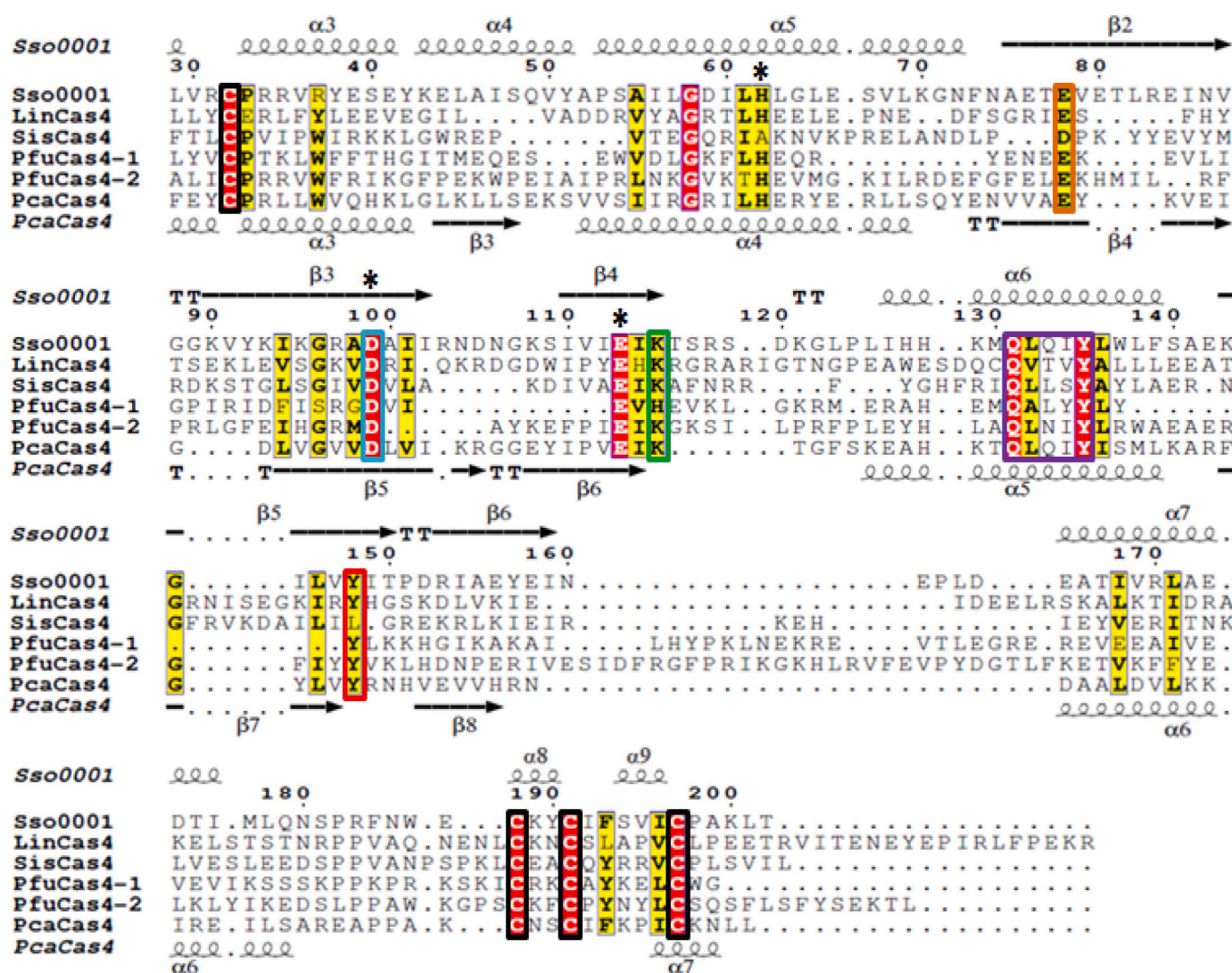

Fig. 3. Multiple sequence alignment of the Cas4 proteins along with the LinCas4. The LinCas4 (LinCas4; Q72TS4) sequence of the CRISPR-Cas I-B is compared with the available Cas4 sequence of the *Pyrobaculum calidifontis* (PcaCas4; A3MTK6), *Pyrococcus furiosus* (PfuCas4–1; Q8U1T6 and PfuCas4–2; Q8U027), *Sulfolobus islandicus* (SisCas4; F0NH45), and *Sulfolobus solfataricus* (Sso0001; Q97TX9). The secondary structure elements represented at the top are inferred based on the known crystal structures of the Sso0001 and at the bottom one of PcaCas4. Four conserved cysteine residues (Csy[18], Cys[177], Cys[180], and Cys[186]) of the LinCas4 are highlighted in a black box. Three metal-ion interacting residues (His[44], Asp[74], and Glu[87]) of the LinCas4 are highlighted with a black asterisk. RecB motifs I (Glu[58]), II (Asp[74]), III (Glu[87] and Lys[89]) are highlighted in brown, light blue, and green boxes, respectively. QxxxY motif (Gln[109]-Tyr[113]) is highlighted in the violet box. The DNA coordinating residue (Tyr[132]) is highlighted in the red box. The semi-conserved residues are highlighted in yellow boxes. Only the partial MSA has been shown for clarity.
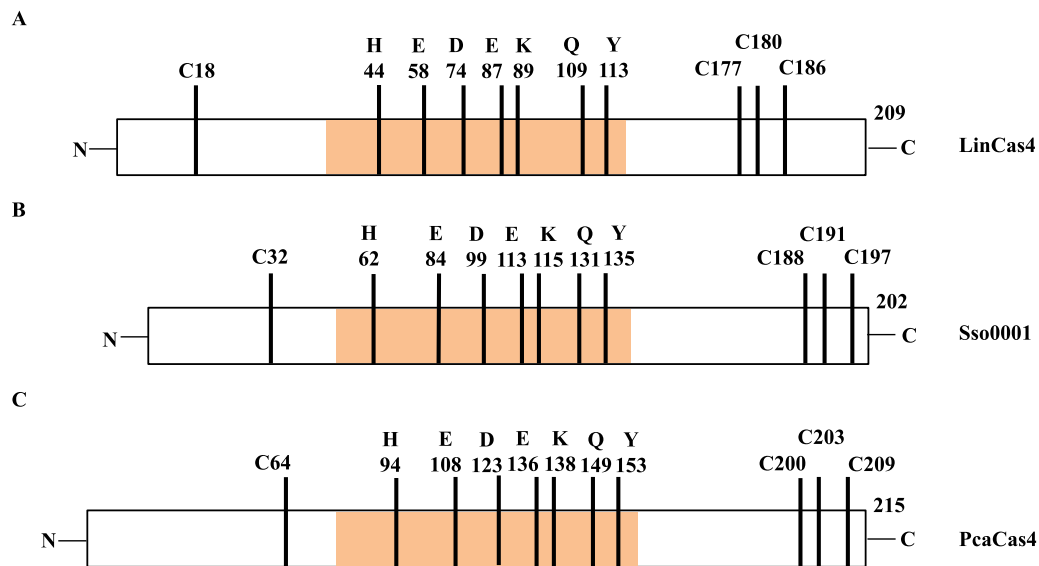
Fig. 4. Representation of protein sequences of Cas4 orthologs. The protein sequence of **(A)** LinCas4, **(B)** Sso0001, and **(C)** PcaCas4 depicts the conserved metal-ion coordinating residues and RecB motifs -I, -II, -III, and QxxxY forming the RecB nuclease active site, denoted by the orange shading and are bordered by the four cysteines serving as iron-sulfur ligands.

2014). Similarly, in LinCas4, three cysteines were found to be located at the C-terminal (Cys[177, 180, 186]) and one towards the N-terminal end (Cys[18]) as depicted in the cartoon representation of LinCas4, Sso0001, and PcaCas4 protein sequence (Fig. 4). Reportedly (Lemak et al., 2014), a proposed DNA coordinating tyrosine residue in the PcaCas4 (Tyr[166]) and Sso0001 (Tyr[148]) was also found to be conserved in the LinCas4 (Tyr[132]) (Fig. 3).

Furthermore, the tertiary structure of the LinCas4 monomer was modeled using the I-TASSER server and found to be analogous to that of

the Sso0001 (Fig. 5A and 5B) (Yang et al., 2015). A Dali search for the LinCas4 orthologs confirmed Sso0001 as the closest ortholog (PDB 4IC1, Z-score 13.2, rmsd 2.5 Å) compared to the PcaCas4 (PDB 4R5Q, Z-score 11.4, rmsd 2.5 Å) (Holm, 2020). Similar to the Sso0001 and PcaCas4 (Lemak et al., 2013, 2014), the LinCas4 structure reflects the presence of the core αβα protein domain and a small α-helical sub-domain (Fig. 5A-C). The core domain contains the five stranded (β1-β5) V-shaped β-sheet, which is surrounded by the six α-helices (H1, H2, H3, H4, H5, and H6) constituting a large elongated opening (14–24 Å) on one
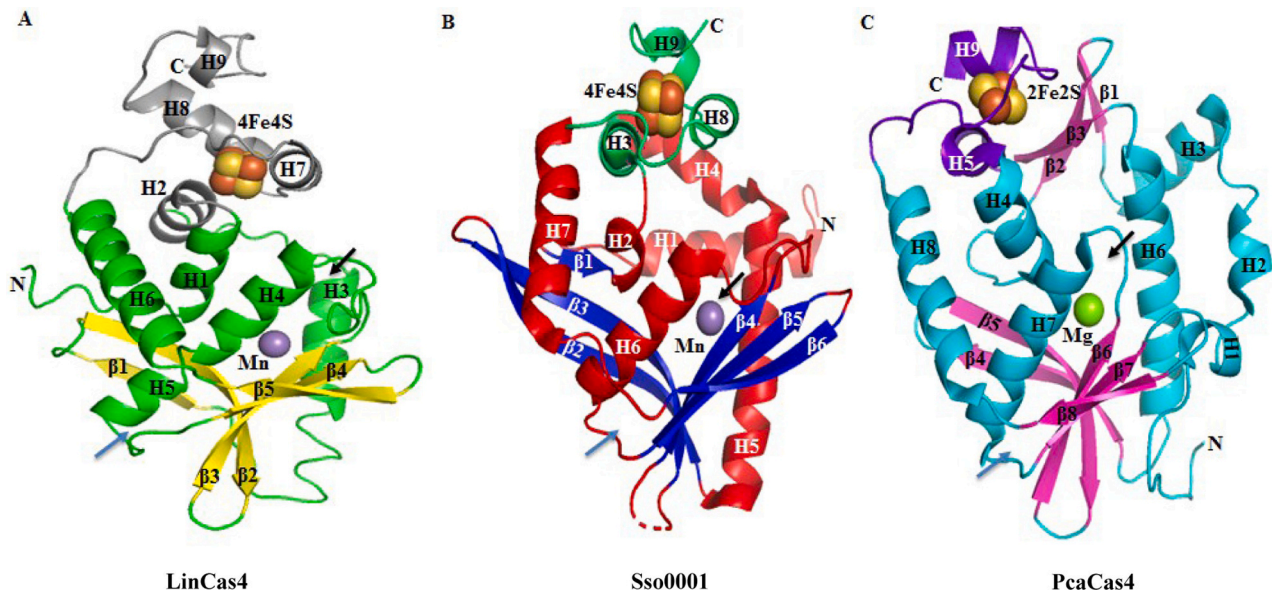

Fig. 5. Structural comparison of the modeled LinCas4. (A) The modeled tertiary structure of the LinCas4 is displayed with its N-terminal domain embodied in green (α-helix; H) and yellow (β-strands) while the C-terminal Fe-S cluster (brown iron atoms and yellow sulfur atoms) binding domain is embodied in gray helices and loops. The metal-ion (Mn^{2+}) in the active site is represented with a purple sphere. **(B)** The tertiary structure of the SsoCas4 monomer of the *S. solfataricus* (Sso0001; PDB 4IC1) for comparison with the LinCas4. The N-terminal domain is illustrated in red (α-helix; H) and dark blue (β-strand). The secondary structure of the C-terminal 4Fe-4S cluster (brown iron atoms and yellow sulfur atoms) binding domain is embodied in green. The metal-ion (Mn^{2+}) in the active site is represented with a purple sphere. **(C)** The tertiary structure of the Cas4 monomer of *P. calidifontis* (PcaCas4; PDB 4R5Q) for comparison with the LinCas4. The N-terminal domain is illustrated in cyan (α-helix; H) and magenta (β-strand). The secondary structure of the C-terminal, 4Fe-4S cluster (brown iron atoms and yellow sulfur atoms) binding domain is embodied in purple. The metal-ion (Mg^{2+}) in the active site is depicted with a green sphere. In all images, the black and blue arrow denotes a possible substrate entrance site and product exit channel in the LinCas4, Sso0001, and PcaCas4.

side and a smaller round hole (9 Å) on the other side as substrate entrance and product release channels, respectively, as described for the Sso0001 and PcaCas4 (Fig. 5A-C) (Lemak et al., 2013, 2014). The active site in the LinCas4 is predicted to be located in a cavity formed by the H3, H4, H5, β2, β3, and β4 in the core domain (Fig. 5A). The LinCas4 exhibits the DNase activity using either Mn^{2+} or Mg^{2+} ions coordinated by the Asp^{74} (3.1 Å) and Glu^{87} (2.7 Å) of RecB-nuclease motifs -II and -III, and His^{44} (6.1 Å) in the active site, which is juxtaposed to the divalent metal-ion coordination observed in the Sso0001 (His^{62}; 2.3 Å, Asp^{99}; 2.1 Å, and Glu^{113}; 2.1 Å) and PcaCas4 (His^{94}; 4.2 Å, Asp^{123}; 2.4 Å, and Glu^{136}; 2.1 Å) (Fig. 6A and 6B) (Lemak et al., 2013, 2014). In the predicted 3D model of LinCas4, four helices (H2, H7, H8, and H9) together with a ten residue (164–173) long loop shape a small sub-domain- very similar to what reported in Sso0001 (Fig. 5A and 5B). (Lemak et al., 2013). This sub-domain is predicted to bind the Fe-S cluster (possibly 4Fe-4S) based on the relative closeness in the similarity of the sub-domain with the Sso0001 *versus* PcaCas4 (Fig. 5) (Lemak et al., 2013). The Fe-S cluster in the LinCas4 is coordinated by Cys^{18} (3.1 Å), Cys^{177} (2.5 Å), Cys^{180} (2.3 Å), and Cys^{186} (2.7 Å) of H2, H7, and H8 helices (Fig. 5A and 6C), which coincide with the Cys^{32} (2.3 Å), Cys^{188} (2.2 Å), Cys^{191} (2.1 Å), and Cys^{197} (2.3 Å) of the Sso0001 (Fig. 6C). The LinCas4 differs from the PcaCas4 since it holds a three-stranded anti-parallel β-sheet (β1, β2, and β3), which may contribute to the

stabilization of the 2Fe-2S cluster domain of this protein besides two helices (H5 and H9) and a long loop (191–207 residues) (Fig. 5C) (Lemak et al., 2014). In addition, the active site bottom loop in LinCas4 (90–99) connecting β3-strand and H4 α-helix, and in Sso0001 (116–126) connecting β4-strand and H6 α-helix has shifted away, thereby opening the product exit channel, whereas the PcaCas4 has a shorter loop (138–142) connecting β6-strand and the H7 α-helix oriented toward the active site (Fig. 5A, 5B, and 5C). In Cas4 orthologs (Sso0001 and PcaCas4), the nuclease domain encompassing metal-ion coordinating residues and RecB motifs (I, II, III, and QxxxY) is end-to-end stapled by the N-terminal cysteine and three C-terminal cysteines, which comes together to coordinate the iron-sulfur cluster. A similar iron staple nuclease domain is also present in the bacterial helicase-nuclease complex AddAB and eukaryotic DNA replication/repair factor Dna2, where an iron-sulfur cluster has been demonstrated to be important for structural stability and nuclease activity of the protein (Yeeles et al., 2009; Mariotti et al., 2020).

2.4. The site-directed mutagenesis of LinCas4

Based on the multiple sequence alignments of the LinCas4, site-directed mutagenesis of the predicted metal-ion interacting residues (Asp^{74} and Glu^{87}), the DNA interacting residue (Tyr^{132}), and the Fe-S
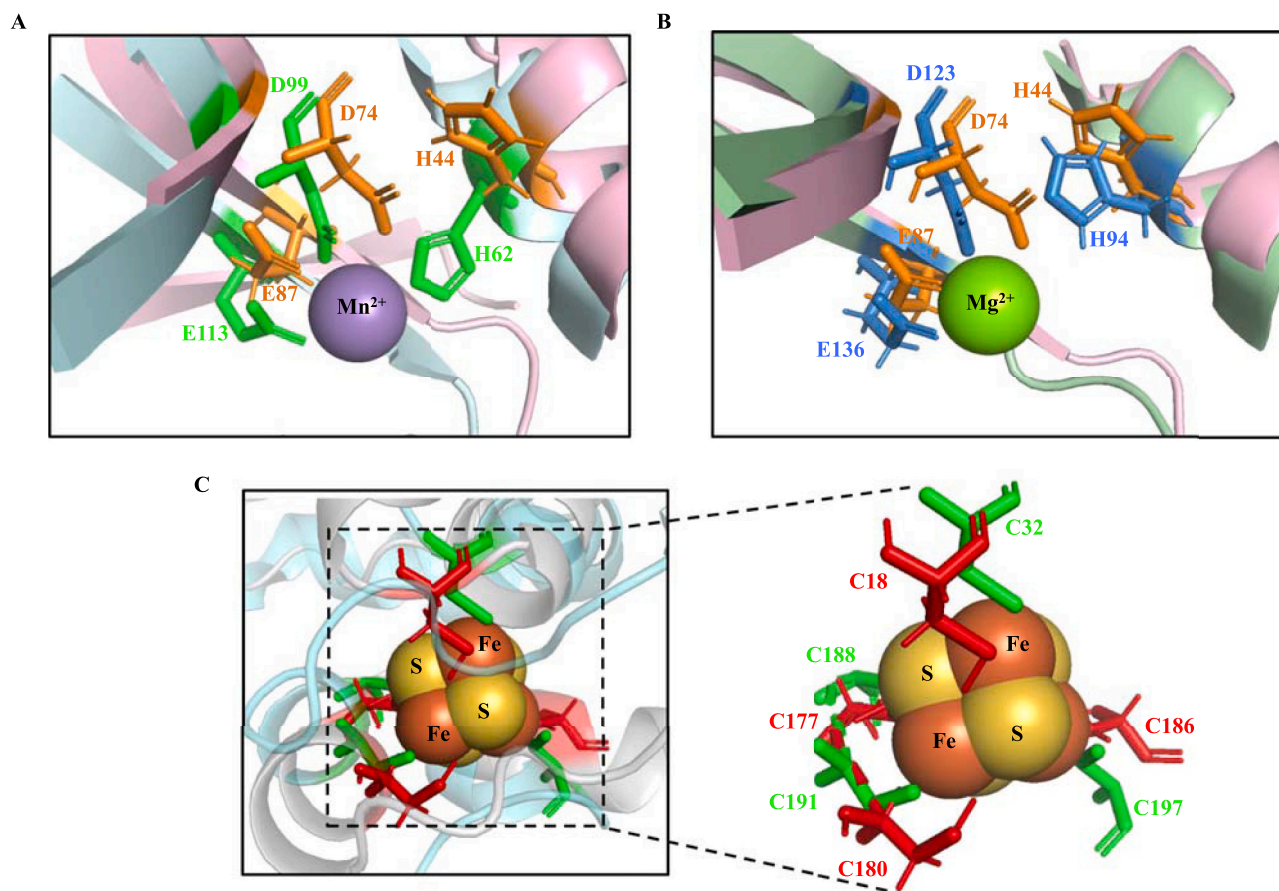

Fig. 6. Comparison of the metal-ion and the Fe-S cluster binding pocket of the modeled LinCas4. (A) Conserved residues at the metal-ion binding pocket of the Cas4 protein from *S. solfataricus* (Sso0001; cyan; PDB 4IC1) and the model LinCas4 (pink) depicted as a stick model with green and orange color for the Sso0001 and LinCas4, respectively, along with the Mn^{2+} divalent-metal embodied as a purple sphere in the center. The number signifies the metal-ion binding amino acid residue of the Sso0001 (green) and the LinCas4 (orange). **(B)** Conserved residues at the metal-ion binding pocket of the Cas4 protein from *P. calidifontis* (PcaCas4; pale-green; PDB 4R5Q) and the model LinCas4 (pink) depicted as a stick model with sky blue and orange for the PcaCas4 and LinCas4, respectively along with the Mg^{2+} divalent-metal embodied as a green sphere in the center. The number signifies metal-ion binding amino acid residue of the PcaCas4 (sky blue) and the LinCas4 (orange). **(C)** 4Fe-4S cluster coordinating with the conserved residues of the Sso0001 (cyan) and the model LinCas4 (gray). Next to it is a close-up view of the conserved cysteine residues depicted as a stick model with green and red color for the Sso0001 and the LinCas4, respectively, along with the cubane-like 4Fe-4S cluster embodied as brown iron atoms and yellow sulfur atoms in the center. The number signifies cysteine residues of the Sso0001 (green) and the LinCas4 (red).

cluster coordinating cysteine residue (Cys[18]) of the LinCas4 was performed. The LinCas4 mutant proteins (LinCas4[D74A], LinCas4[E87A], LinCas4[Y132A], and LinCas4[C18A]) were overexpressed and purified with a similar yield like the LinCas4 (Fig. S1C, S1D, and S1E). The mutants LinCas4[D74A] and LinCas4[E87A] exhibited a significant reduction in the DNA cleavage (Fig. 7A), suggesting that the metal-ion interaction is indispensable for the DNase activity.

The mutation of the potential DNA interacting residue in the LinCas4 (LinCas4[Y132A]) did not interfere with the DNase activity (Fig. 7A). This advocates that the residue Tyr[132] may be dispensable for the DNase activity of the LinCas4, unlike Tyr[148] of the Sso0001 and the Tyr[166] residue of the PcaCas4 where mutation provokes loss in the DNA binding and cleavage activity (Lemak et al., 2013, 2014). Furthermore, mutation of one of the cysteines (LinCas4[C18A]) involved in interacting with the presumed 4Fe-4S cluster did not affect the DNase activity of LinCas4 (Fig. 7B). The nuclease activity of LinCas4[C18A] suggests that the Fe-S cluster is not necessary for LinCas4 nuclease activity, unlike in Sso0001 (Lemak et al., 2013).

2.5. Presence of the 4Fe-4S cluster in the LinCas4

The iron chelation assay revealed that the purified LinCas4 was in an apostate conformation lacking the Fe-S cluster. Therefore, the holo form of LinCas4 was generated through the reconstitution of the Fe-S cluster in LinCas4. The reconstituted LinCas4 showed the existence of a 4Fe-4S cluster (Table 2). In vitro reconstitution of LinCas4 with Fe-S cluster was further substantiated by UV–visible spectroscopy where holo-LinCas4 exhibited a shoulder peak between 380 and 420 nm, characteristic of the 4Fe-4S cluster, while no such peak was observed in the apo-LinCas4 (Fig. 8A) (Lemak et al., 2013). The near UV-visible circular dichroism (CD) spectroscopy analysis of apo and holo forms of LinCas4 revealed two characteristic maxima at 330 and 425 nm in holo-LinCas4, characteristic of the 4Fe-4S cluster containing proteins (Fig. 8B) (Mapolelo et al., 2012; Freibert et al., 2018). This confirms the reconstitution of the 4Fe-4S cluster in the LinCas4, which was lost during the process of purification. Furthermore, analysis of the far UV-CD spectra of the apo-LinCas4 and holo-LinCas4 revealed similar secondary structure characteristics, where 35% α-helix and 17% β-strand is present in the apo-LinCas4 while 36% α-helix and 19% β- strand is present in the holo-LinCas4 (Fig. 9). This suggests that the purified LinCas4, in the absence of the 4Fe-4S cluster, can retain its structural stability. The consistency of the secondary structure characteristics of the apo and holo forms of the LinCas4 is reminiscent to their biochemical characteristics since, both the forms of the LinCas4 cleaved the circular dsDNA substrate without any significant difference in the cleavage product (Fig. S2). This is further validated by the nuclease activity demonstrated by the mutant LinCas4[C18A], lacking the cysteine residue essential for forming the 4Fe-4S cluster. It suggests that unlike Sso0001, LinCas4 does

not require a Fe-S cluster for maintaining the structural integrity of the RecB domain. The retaining of structural integrity of RecB domain in the absence of the Fe-S cluster in LinCas4 is in line with the reported PcaCas4 where mutation of the cysteine residues does not hinder the nuclease activity of the protein (Lemak et al., 2013, 2014).

3. Discussion

Among all the known Cas4 nucleases, LinCas4 appears to be distinct since it cleaves circular dsDNA and does not show a preference to cleave the 5′ or 3′ free ends of the short DNA oligos. This stands in contrast to the reported nuclease activity of Sso0001, PcaCas4, and Sso1391 which do not cleave circular dsDNA but demonstrate cleavage of the short oligos either at 5′ or 3′ ends (Zhang et al., 2012; Lemak et al., 2013, 2014). Nevertheless, the cleavage activity of LinCas4 on viral ssDNA substrates is in agreement to Cas4 orthologs including PcaCas4, Sso0001, and the Sso1391 (Lemak et al., 2013). The endonuclease activity of LinCas4 might have resulted from an ATP-independent DNA unwinding activity towards dsDNA that has recorded for Sso0001 and PcaCas4 (Zhang et al., 2012; Lemak et al., 2013, 2014). Furthermore, LinCas4 may cleave ssDNA via a nucleophilic attack on the 3′ side of the scissile phosphate resulting in the formation of products containing 5′-hydroxyls and 3′-phosphates as suggested for Sso0001, Sso1391, and PcaCas4 (Lemak et al., 2013, 2014).

Recently, using dCas9 and sgRNA expressing plasmid, stable silencing of the genes in the pathogenic L. interrogans has been reported and established to be maintained for several passages (Fernandes et al., 2021). This contradicts the nuclease activity described for Cas proteins of Leptospira under in vitro condition. We speculate that the possible discrepency may be due to the low expression level of Cas proteins (LinCas4 and LinCas2) in Leptospira, and divalent ions inside the Leptospira may additionally modulate this. It has been documented elsewhere (Liu et al., 2015) that in the S. islandicus type I-A system, Csa3a, a positive transcriptional regulator binds to the cas1 promoter and facilitate the expression of downstream cas genes involved in adaptation. Similarly, in the Pseudomonas aeruginosa type I-F system, CRP, a cAMP receptor protein, promotes the expression of the adaptation cas genes and facilitates CRISPR mediated immunity in response to the glucose level (Patterson et al., 2015). Thus, regulating the expression of the Cas protein in Leptospira by unknown regulatory genes cannot be overruled. The maintenance of the plasmids expressing dCas9 and sgRNA (almost 10 kb in size) for several passages in pathogenic leptospires may also be possible due to the occurrence of only primed adaptation in the Leptospira CRISPR I-B system. Primed adaptation has been established in Haloarcula hispanica CRISPR I-B system (Li et al., 2014). In primed adaptation, a pre-existing spacer is a prerequisite to initiate CRISPR adaptation (Li et al., 2014). Lastly, the cas genes in L. interrogans Copenhageni CRISPR-Cas I-B are not constitutively expressed. In
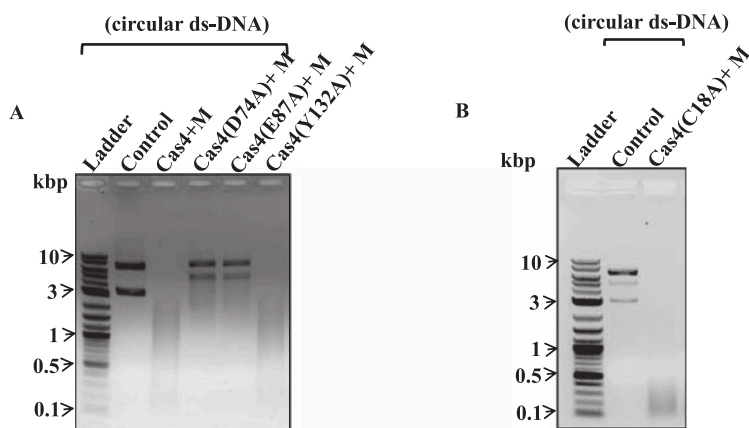

Fig. 7. DNase activity in the mutant LinCas4. (A) LinCas4 metal coordinating residue mutants (LinCas4[D74A] and LinCas4[E87A]; 6 μM each) exhibits reduced dsDNase activity on circular dsDNA (plasmid pTZ57R/T; 2.8 kbp; 12.5 nM) when compared to the LinCas4 and the LinCas4[Y132A] (DNA coordinating residue mutant; 6 μM), which displays similar DNase activity. **(B)** LinCas4[C18A] (6 μM) exhibits dsDNase activity on circular dsDNA (plasmid pTZ57R/T; 2.8 kbp; 12.5 nM) comparable to the LinCas4. In all panels, 'M' denotes Mn[2+]. All the reactions were carried out for 1 h at 37 °C and resolved on the 0.8% agarose gel. The results presented were confirmed by two independent experiments.
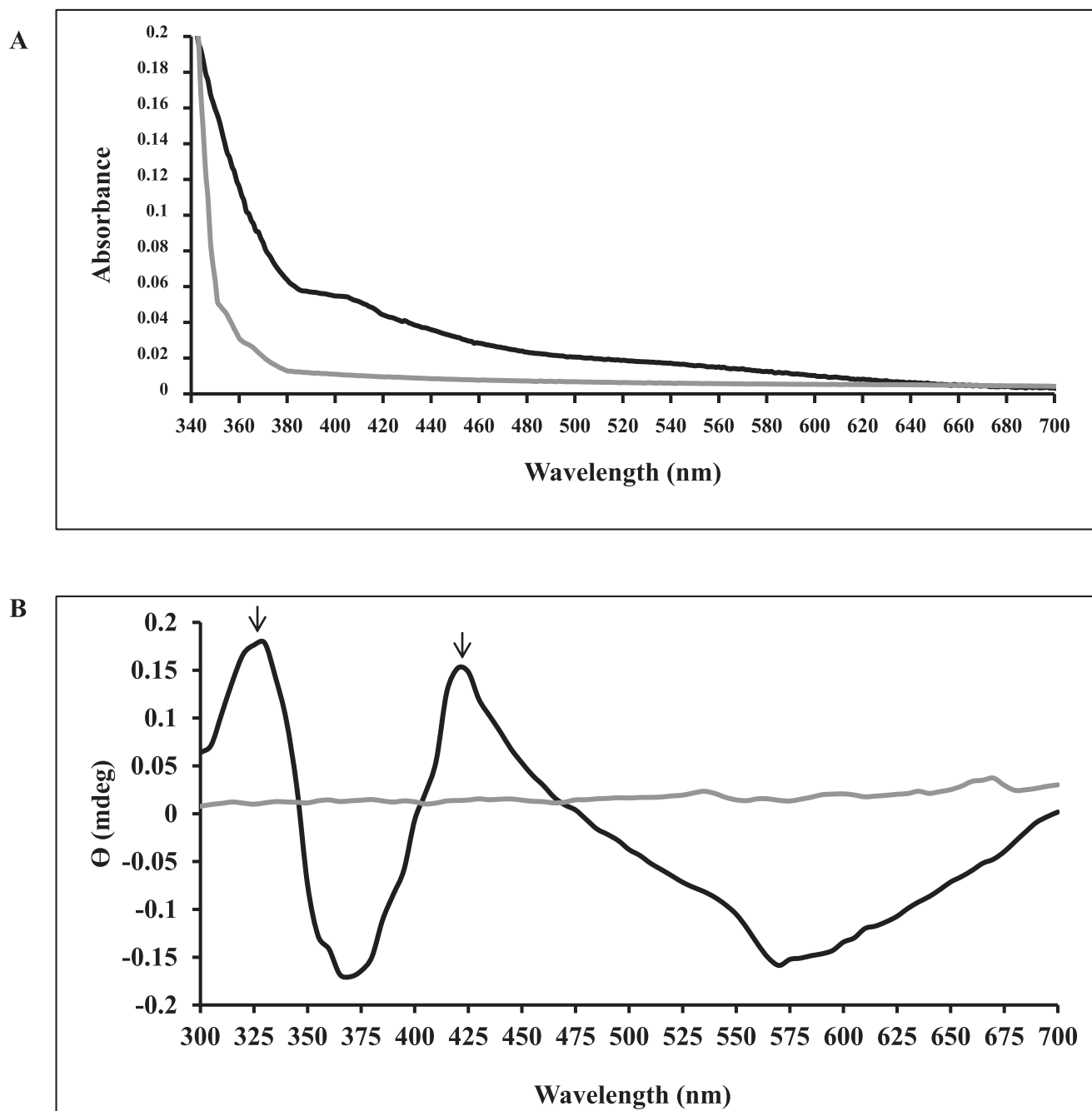
Fig. 8. Spectroscopic characterization of LinCas4. (A) UV-visible spectra of holo-LinCas4 (black line) and apo-LinCas4 (gray line) at 300–700 nm demonstrate a shoulder peak between 380 and 420 nm in holo-LinCas4, suggesting the presence of the 4Fe-4S cluster. **(B)** Near UV-visible CD spectra of holo-LinCas4 (black line) and apo-LinCas4 (gray line) at 300–700 nm shows two maxima at 330 nm and 425 nm in holo-LinCas4 (arrow marked).

Leptospira CRISPR-Cas I-B, the *cas* genes (*cas1, cas2,* and *cas4*) have been reported (Dixit et al., 2016) to be actively transcribed under *in vitro* growth condition. Nevertheless, the translation of these *cas* genes has not been validated to date. The expression level of the *cas* genes in *Leptospira* may be so minimal that it fails to be detected using the conventional immuno blot technique.

In silico characterization of LinCas4 revealed the presence of three conserved metal-ion interacting residues (His[44], Asp[74], and Glu[87]) at the N-terminal domain of the LinCas4. Upon substituting the Asp[74] and Glu[87] residues with Ala, the LinCas4 showed a drastic reduction in its DNase activity, similar to the loss of the DNase activity reported in Sso0001, Sso1391, and PcaCas4 (Zhang et al., 2012; Lemak et al., 2013, 2014). The metal-ion interacting residues Asp[74] and Glu[87] in the LinCas4

correspond to the identical residues in the PfuCas4–1, PfuCas4–2, and SisCas4, which regulates the foreign DNA acquisition efficiency, define the spacer length, its oriented integration, and the 5′ PAM or 3′-protospacer motif recognition under *in vivo* conditions (Shiimori et al., 2018; Zhang et al., 2019). This leads us to speculate that the interacting metal residues (Asp[74] and Glu[87]) of the LinCas4 may conceivably be crucial for the CRISPR acquisition in the *L. interrogans*. In LinCas4, the mutation of a probable DNA interacting residue, Tyr[132], did not hinder the DNase activity, which in contrast is essential for PcaCas4 and Sso0001 DNase activity as previously reported (Lemak et al., 2013, 2014). We speculate that the endonuclease activity of pure LinCas4 under *in vitro* condition may alter when allowed to associate with LinCas1 and LinCas2 mandated for CRISPR adaptation. Pure Cas2 protein is an example of
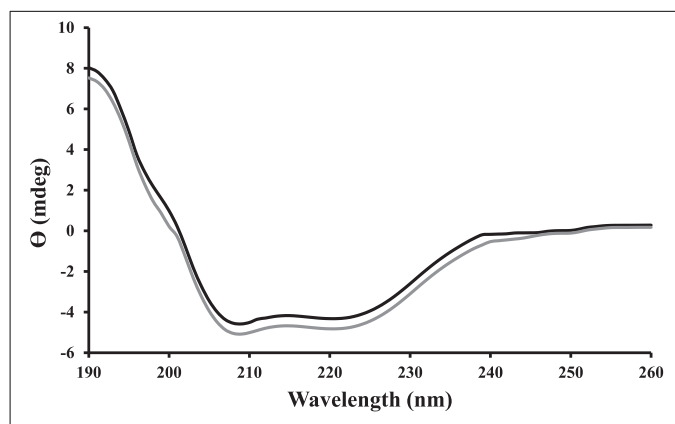
Fig. 9. Comparison of secondary structure composition in apo-LinCas4 and holo-LinCas4. Far-UV circular dichroism (CD) spectra for apo-LinCas4 (gray line) and holo-LinCas4 (black line) depicted over a wavelength range of 190–260 nm and analyzed using the K2D3 program.

such a nuclease, which alone has been reported to possess endonuclease activity under *in vitro* condition (Samai et al., 2010; Nam et al., 2012; Ka et al., 2014; Gunderson et al., 2015; Dixit et al., 2016). However, Cas2 catalytic activity is lost or is not required when it forms a complex with Cas1 during CRISPR adaptation under *in vivo* and *in vitro* conditions (Nuñez et al., 2014; Nunez et al., 2015; Rollie et al., 2015; Wang et al., 2015).

Furthermore, mutation of one of the Fe-S cluster coordinating cysteine (Cys[18]) in LinCas4 did not had any adverse effect on its DNase activity which suggests that Fe-S cluster is not a prerequisite for LinCas4 stability and nuclease activity. Moreover, the iron chelation assay of the purified LinCas4, which has been earlier shown to exhibit DNase activity, revealed absence of any form of Fe-S cluster. Fe-S cluster was reconstituted in the purified LinCas4 under *in vitro* condition to determine the role of the Fe-S cluster in LinCas4. The binding of 4Fe-4S cluster to the reconstituted LinCas4 (holo-LinCas4) was confirmed via iron chelation assay, UV-visible spectroscopy, and near UV-CD spectroscopy. The presence of 4Fe-4S cluster in LinCas4 shows that it is structurally similar to Sso0001, which too possesses 4Fe-4S cluster, compared to PcaCas4, which possesses 2Fe-2S cluster (Zhang et al., 2012; Lemak et al., 2013, 2014). However, rescconstituting the 4Fe-4S cluster in the LinCas4 did not result in any significant change in the secondary structure compared to apo-LinCas4 when analyzed via far UV-CD spectroscopy. Furthermore, the two forms of the LinCas4 could not be differentiated owing to their similar DNase activity on circular dsDNA. We speculate that the 4Fe-4S cluster is an expendable entity in LinCas4. The dispensability of the DNA interacting residue (Tyr[132]) and Fe-S cluster for its nuclease activity further adds to the versatile nature of LinCas4. Also, based on the *in silico* studies, the LinCas4 demonstrate resemblance with RecB and AddB nucleases that is known to play a role in DNA repair mechanism. The endonuclease activity in LinCas4 may provide a possible advantage with extra role in addition to the conventional CRISPR adaptation.

The lower yield (0.27 mg/L) of the LinCas4 purification using the Ni-NTA affinity chromatography was the limiting factor in our investigation. Despite multiple attempts of using different hosts, vectors, and improvising the induction conditions, the yield of rLinCas4 could not be improved. The amount and concentration of pure LinCas4 achieved was adequate only for carrying out nuclease assays, spectroscopy analysis, and iron chelation assay, discussed in this study. The concentration of LinCas4 was inadequate for the generation of LinCas4-Cas1-Cas2 complex required for other experiments such as, protein-protein interaction of LinCas4 with LinCas1 and LinCas2 and *in vitro* spacer acquisition.

4. Conclusion

The LinCas4 possesses DNase activity, conditionally relying on the divalent-metal, salt, and pH. The site-directed mutagenesis of the metal interacting residues in LinCas4 (LinCas4[D74A] and LinCas4[E87A]), which are also part of RecB motif -II and -III, suggests that the RecB domain is vital for the DNase activity. However, the plausible DNA interacting residue (Tyr[132]) in the LinCas4 is dispensable for the DNase activity. The confirmation of the existence of the 4Fe-4S cluster in holo-LinCas4 depicts that it belongs to a family of "iron staple" nucleases, which includes AddB, Exonuclease V, and Dna2 proteins (Yeeles et al., 2009; Mariotti et al., 2020). The purified LinCas4 in this study has been shown to lack the Fe-S cluster. Nevertheless, the DNase activity of the purified LinCas4 remained unperturbed. This was further validated by mutating one of the four Fe-S cluster coordinating cysteines (LinCas4[C18A]) which retained similar DNase activity levels as the LinCas4. This suggests that the Fe-S cluster is not essential for the nuclease activity of LinCas4. Thus, the nuclease activity of the LinCas4 appears to support the hypothesis that Fe-S clusters in many proteins are purely structural features and are perhaps evolutionary relics of an anaerobic past when Fe-S clusters were much more common in proteins (White, 2009). Further, *in vitro* and *in vivo* studies mandated in association with the Cas1 and Cas2 for deciphering the role of LinCas4 in the *Leptospira* CRISPR adaptation process are underway in our laboratory.

5. Materials and methods

5.1. Bacterial strain and culture media

L. interrogans serovar Copenhageni strain Fiocruz L1–130 was procured from the Indian Council of Medical Research (ICMR), Regional Medical Research centre (RMRC), Port Blair, Andaman, and Nicobar Island, India. Spirochetes were maintained in EMJH (Ellinghausen-McCullough-Johnson-Harris) medium (Difco, Becton Dickinson India Pvt. Ltd., New Delhi, India) at 29 °C by regular subculturing at an interval of 6–7 days. *E. coli* strains DH5α and BL21 (DE3) (Novagen) was used to maintain plasmid constructs and over-expression of recombinant proteins, respectively. Both *E. coli* strains were grown in Luria-Bertani (LB)-broth (Himedia) or LB-agar (Himedia) medium supplemented with antibiotics at 37 °C.

5.2. Recombinant DNA techniques and isolation of nucleic acids

Himedia bacterial genomic purification kit (cat#MB505) was used to extract *the Leptospira* genome from the 7-day old culture of *Leptospira* in the EMJH media. Plasmids were isolated from *E. coli* culture, and standard procedures were used to generate the recombinant plasmid. HiPurA plasmid DNA miniprep purification spin kit (MB508) was used to extract the bacterial plasmid. The concentration of plasmid was measured using a nanodrop before storing it at −20 °C. The QIAquick gel extraction kit (Qiagen) was used to isolate the DNA fragments from the agarose gels. DNA sequencing was conducted after cloning by outsourcing the constructs to the available local companies. The single-stranded DNA substrates (M13mp18 and φX174), along with all the enzymes used for the DNA manipulations, were obtained from the New England Biolabs (NEB) or the Fermentas.

5.3. Nuclease activity assays

Nuclease activity of the recombinant LinCas4 protein was investigated on various substrates *viz.* circular double-stranded (ds) DNA plasmid (pTZ57R/T; 2.8 kbp; 12.5 nM), linear ssDNA (M13mp18; 7.2 kb; 21.4 nM), circular ssDNA (φX174 genome; 5.3 kb; 29 nM), 5′-FAM (6-Fluorescein phosphoramidite)-labeled short oligonucleotides (36-mer; 250 nM) and the 3′-FAM-labeled short oligonucleotides (50-mer; 250 nM). These substrates were incubated with the LinCas4 or the Fe-S

cluster reconstituted LinCas4 in a 25 µl reaction buffer (25 mM HEPES pH 7.0, 100 mM KCl, and 10 mM $MnSO_4$ or $MgCl_2$) at 37 °C for 1 hour (h) unless stated. Divalent metal-ion dependence for the nuclease activity was determined with the various other metals (Mn^{+2}, Ca^{+2}, and Ni^{+2} ions). The optimal pH was determined by substituting the buffer with the either 25 mM sodium citrate (pH 3.0 to 5.0), MES (pH 6.0), HEPES (pH 7.0), Tris–HCl (pH 8.0), or the CAPS (pH 9.0 to 11.0). The effect of salt on the LinCas4 nuclease activity was studied using the NaCl, KCl, and NH_4Cl, at varying concentrations (50 - 150 mM) independently. The time-bound nuclease assay was performed to determine the duration required to degrade the substrate. All the above resulting reaction products were resolved on the 0.8% (w/v) agarose gel electrophoresis except the FAM-labeled ssDNA short oligo reaction products, which were resolved on a 20% (w/v) 8 M denaturing Urea PAGE and imaged in the gel documentation instrument (Biorad).

5.4. Multiple sequence alignments and the site-directed mutagenesis

Towards an exploration of the conserved motifs like the metal-ion coordinating residues in the LinCas4, the amino acid sequences of the Cas4 orthologs from different organisms were procured from the UniProtKB database (Boutet et al., 2007). Multiple sequence alignment (MSA) was performed using the program Clustal Omega with a default set of parameters (Sievers and Higgins, 2014). The aligned sequences were enriched using the webserver ESPript (Easy Sequencing in PostScript) (Gouet et al., 2003).

A single amino acid mutation was inducted into the potential metal-binding, DNA interacting, and the possible Fe-S cluster binding site of the LinCas4 using the Q5 site-directed mutagenesis kit (NEB). Primers were designed (Table 1) using the NEBaseChanger tool program to mutate the residues Asp^{74} (codon GAT) to Ala (GCT), Glu^{87} (GAG) to Ala (GCG), Tyr^{132} (TAC) to Ala (GCC), and Cys^{18} (TGT) to Ala (GCT) in the LinCas4. All the derived mutants were validated by sequencing the constructs before the protein's purification using the Ni-NTA affinity chromatography.

5.5. Structure prediction of the Leptospira Cas4

The three-dimensional structure of the LinCas4 of CRISPR-Cas I-B system was modeled using the program I-TASSER (Iterative Threading ASSEmbly Refinement) server (Yang et al., 2015). Further improvement and energy minimization of the modeled structures was performed using the webserver 3D^{refine} (Bhattacharya et al., 2016) and YASARA (Yet Another Scientific Artificial Reality Application) (Krieger et al., 2009). Structural homologs of the LinCas4 were obtained using the Dali server. Structures were visualized and superimposed using PYMOL (The PyMOL Molecular Graphics System, Schr€odinger, LLC). The three-dimensional atomic coordinates of the protein structures used in the study were downloaded from the Protein Data Bank (Berman et al., 2000).

Table 2
Quantification of iron bound to the LinCas4 (holo and apo form) using the ferrozine reagent.

Protein Sample	Concentration (µM)	Absorbance at 562 nm	Iron Conc. (µM)	Fe^{2+} molar ratio per protein monomer
Holo-LinCas4	10	0.258	36.27 ± 0.242	3.627 ± 0.024
Apo-LinCas4/ purified LinCas4	10	0.004	0.562 ± 0.015	0.056 ± 0.001

5.6. Reconstitution of the iron-sulfur cluster in LinCas4

To generate an iron-sulfur cluster containing holo-LinCas4 protein, the collected elutes of LinCas4 protein in denaturing elution buffer were dialyzed in dialysis buffer (50 mM Tris–HCl and 100 mM NaCl). Subsequently, the obtained elutes were lyophilized and was resuspended in 1 ml of nuclease-free water, resulting in 50 µM or 1.35 mg/ml of LinCas4 protein. The reconstitution of the iron-sulfur cluster in LinCas4 was performed in an inert gas glove box (containing 99.99% inert argon and ≤ 1 ppm O_2) (Jacomex), where LinCas4 (50 µM) was reduced anaerobically with 5 mM DTT (Dithiothreitol) (containing inert argon and ≤ 1 ppm O_2) for 1 h. After this pretreatment, $FeCl_3$ containing 5 mM DTT was added in a stoichiometry of 5:1 (iron to protein) and incubated for approximately 10 min until the observed color change to red. An equal amount of Na_2S containing 5 mM DTT was added slowly, and the sample was incubated for 15 min before unbound iron and sulfide were removed via buffer exchange with Tris–HCl (50 mM; pH 8.0) buffer containing NaCl (100 mM) and DTT (5 mM) using concentrators (Corning centricon spin-x-UF). All the buffers and salt solutions were prepared inside the anaerobic chamber using Milli-Q water purged with argon gas. The UV/Visible (UV/Vis) spectra were recorded on a Cary 60 UV–Vis spectrometer (Agilent Technologies).

5.7. Circular dichroism (CD) spectroscopy

Circular dichroism (CD) spectroscopy measurements were performed for apo-LinCas4 and holo-LinCas4 at room temperature using a Jasco 1500 spectrometer (Japan Spectroscopic, Tokyo) at a scanning speed of 100 nm min^{-1}. The near and far-ultraviolet CD spectra were measured using a 1 mm path-length cell at 1.0 nm intervals. The near and far-UV CD spectra were presented as an average of three scans recorded from 300 to 700 nm and 190 to 260 nm respectively. The molar ellipticity (Φ) is expressed in degrees cm^2 dmol^{-1}. The far-UVspectra data were submitted to the K2D3 web server (Louis-Jeune et al., 2012), which calculated the secondary structure content from the ellipticity experimental data. The theoretical secondary structure was calculated using PSIPRED (Jones, 1999).

Table 1
List of primers used in this study.

Sequence	5′—3′	Length
Cas4_NheI F	CTAGCTAGCATGGACCAATACGGAAACA	28
Cas4_XhoI R	CCGGCTCGAGTTACCTTTTTTCCGGAAAG	28
Cas4 (D74A)_F	GGGCAAAGTAGcTCGTATTCAAAAAC	26
Cas4 (D74A)_R	GAAACTTCCAATTTTTCACTAG	23
Cas4 (E87A)_F	GATTCCTTACGcGCATAAAAGAG	23
Cas4 (E87A)_R	CAATCCCCATCTCGTTTTTG	21
Cas4 (Y132A)_F	AAAGATCCGAgcCCACGGAAGTAAAG	26
Cas4 (Y132A)_R	CCTTCTGATATATTTCTGCC	20
Cas4 (C18A)_F	CCTACTTTATgcTGAAAGGCTTTTCTATTTAGAAG	35
Cas4 (C18A)_R	GAATGAATTCCCATCGCG	18
5′-FAM oligo	TTCTAAACCGCCTATCGGCATCAAAGTTATATTCAG	36
3′-FAM oligo	TTTTTTTTTTTTTTTTATTATCTGAGGGTTTAATCTTATTAATCTCTTACTA	50

5.8. Iron chelation assay

The iron bound to the LinCas4 was quantified using the ferrozine method (Hadley and Nolan, 2019) after purification from the heterologous host *E. coli*. Briefly, a given volume (100 μl) of the apo-LinCas4 (10 μM) and holo-LinCas4 (10 μM) was individually mixed with the 35% concentrated HCl (5 μl) and 0.5 M NH$_2$OH.HCl (11.5 μl) and incubated for 30 min at room temperature. The mixtures were then centrifuged at 13,000 $\times$ *g* for 1 min. The supernatant from each mixture was transferred into 2 ml tubes to which NaOH (25 μl of 1 M) was added to bring the pH in the range of 6.0–7.0. After pH adjustment, freshly prepared 100 mM of ferrozine (1.5 μl) was added and mixed well. The reactions were incubated at room temperature for 1 h before measuring the absorbance (562 nm) to determine the iron concentration using the molar extinction coefficient of the complex, $\mathcal{E} = 25.4$ mM^{-1}cm^{-1}. A standard graph consisting of known amounts of FeSO$_4$ (0–100 μM) was prepared using the same procedure, as mentioned for LinCas4.

Author contributions statement

MK conceived and supervised the study; MK designed experiments; BD, VA, and SH performed experiments. BD performed modeling, and structural studies; MK and BD analyzed data; MK and BD wrote the manuscript.

Funding

Declaration of Competing Interest

There are no conflicts to declare

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.crmicr.2021.100040.

References

Arbas, S.M., Narayanasamy, S., Herold, M., Lebrun, L.A., Hoopmann, M.R., Li, S., Lam, T. J., Kunath, B.J., Hicks, N.D., Liu, C.M., 2021. Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. Nature Microbiol. 6 (1), 123–135.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315 (5819), 1709–1712.

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., Bourne, P., 2000. Th e protein data bank. Nucleic Acids Res. 28, 235–242.

Bernick, D.L., Cox, C.L., Dennis, P.P., Lowe, T.M., 2012. Comparative genomic and transcriptional analyses of CRISPR systems across the genus Pyrobaculum. Front Microbiol 3, 251.

Bhattacharya, D., Nowotny, J., Cao, R., Cheng, J., 2016. 3Drefine: an interactive web server for efficient protein structure refinement. Nucleic Acids Res. 44 (W1), W406–W409.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bairoch, A., 2007. Uniprotkb/ swiss-prot. Plant Bioinformatics. Springer, pp. 89–112.

Chen, A.Y., Adamek, R.N., Dick, B.L., Credille, C.V., Morrison, C.N., Cohen, S.M., 2018. Targeting metalloenzymes for therapeutic intervention. Chem. Rev. 119 (2), 1323–1455.

Croda, J., Figueira, C.P., Wunder, E.A., Santos, C.S., Reis, M.G., Ko, A.I., Picardeau, M., 2008. Targeted mutagenesis in pathogenic Leptospira species: disruption of the LigB gene does not affect virulence in animal models of leptospirosis. Infect. Immun. 76 (12), 5826–5833.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S., 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. 190 (4), 1390–1400.

Dixit, B., Ghosh, K.K., Fernandes, G., Kumar, P., Gogoi, P., Kumar, M., 2016. Dual nuclease activity of a Cas2 protein in CRISPR–Cas subtype I-B of Leptospira interrogans. FEBS Lett. 590 (7), 1002–1016.

Fernandes, L., Guaman, L., Vasconcellos, S., Heinemann, M.B., Picardeau, M., Nascimento, A., 2019. Gene silencing based on RNA-guided catalytically inactive Cas9 (dCas9): a new tool for genetic engineering in Leptospira. Sci Rep 9 (1), 1–14.

Fernandes, L.G.V., Hornsby, R., ALTOd, Nascimento, Nally, J., 2021. Genetic manipulation of pathogenic Leptospira: CRISPR interference (CRISPRi)-mediated gene silencing and rapid mutant recovery at 37°C. Sci. Rep. 11 (1), 1–12.

Freibert, S.-A., Weiler, B.D., Bill, E., Pierik, A.J., Mühlenhoff, U., Lill, R., 2018. Biochemical reconstitution and spectroscopic analysis of iron–sulfur proteins. Methods Enzymol. Elsevier 599, 197–226.

Gouet, P., Robert, X., Courcelle, E., 2003. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. 31 (13), 3320–3323.

Gunderson, F.F., Mallama, C.A., Fairbairn, S.G., Cianciotto, N.P., 2015. Nuclease activity of Legionella pneumophila Cas2 promotes intracellular infection of amoebal host cells. Infect. Immun. 83 (3), 1008–1018.

Hadley RCNolan, E.M., 2019. Preparation and iron redox speciation study of the Fe (II)-binding antimicrobial protein calprotectin. Calcium-binding proteins of the EF-hand superfamily. Springer 397–415.

Heler, R., Marraffini, L.A., Bikard, D., 2014. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. Mol. Microbiol. 93 (1), 1–9.

Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S., Charpentier, E., 2018. The biology of CRISPR-Cas: backward and forward. Cell 172 (6), 1239–1259.

Holm, L., 2020. DALI and the persistence of protein shape. Protein Sci. 29 (1), 128–140.

Hudaiberdiev, S., Shmakov, S., Wolf, Y.I., Terns, M.P., Makarova, K.S., Koonin, E.V., 2017. Phylogenomics of Cas4 family nucleases. BMC Evol. Biol. 17 (1), 232.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292 (2), 195–202.

Ka, D., Kim, D., Baek, G., Bae, E., 2014. Structural and functional characterization of Streptococcus pyogenes Cas2 protein under different pH conditions. Biochem. Biophys. Res. Commun. 451 (1), 152–157.

Koonin, E.V., Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 37, 67–78.

Krajewski, W.W., Fu, X., Wilkinson, M., Cronin, N.B., Dillingham, M.S., Wigley, D.B., 2014. Structural basis for translocation by AddAB helicase–nuclease and its arrest at χ sites. Nature 508 (7496), 416–419.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., Karplus, K., 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. Proteins 77 (S9), 114–122.

Lemak, S., Nocek, B., Beloglazova, N., Skarina, T., Flick, R., Brown, G., Joachimiak, A., Savchenko, A., Yakunin, A.F., 2014. The CRISPR-associated Cas4 protein Pcal_0546 from Pyrobaculum calidifontis contains a [2Fe-2S] cluster: crystal structure and nuclease activity. Nucleic Acids Res. 42 (17), 11144–11155.

Lemak, S., Beloglazova, N., Nocek, B., Skarina, T., Flick, R., Brown, G., Popovic, A., Joachimiak, A., Savchenko, A., Yakunin, A.F., 2013. Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from Sulfolobus solfataricus. J. Am. Chem. Soc. 135 (46), 17476–17487.

Li, M., Wang, R., Zhao, D., Xiang, H., 2014. Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res. 42 (4), 2483–2492.

Little, J.W., 1967. An exonuclease induced by bacteriophage λ II. Nature of the enzymatic reaction. J. Biol. Chem. 242 (4), 679–686.

Liu, T., Li, Y., Wang, X., Ye, Q., Li, H., Liang, Y., She, Q., Peng, N., 2015. Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR de novo spacer acquisition. Nucleic Acids Res. 43 (2), 1044–1055.

Liu, T., Liu, Z., Ye, Q., Pan, S., Wang, X., Li, Y., Peng, W., Liang, Y., She, Q., Peng, N., 2017. Coupling transcriptional activation of CRISPR–Cas system and DNA repair genes by Csa3a in Sulfolobus islandicus. Nucleic Acids Res. 45 (15), 8978–8992.

Louis-Jeune, C., Andrade-Navarro, M.A., Perez-Iratxeta, C., 2012. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. Proteins 80 (2), 374–381.

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., 2011. Evolution and classification of the CRISPR–Cas systems. Nature Rev. Microbiol. 9 (6), 467–477.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., 2015. An updated evolutionary classification of CRISPR–Cas systems. Nature Rev. Microbiol. 13 (11), 722.

Mapolelo, D.T., Zhang, B., Naik, S.G., Huynh, B.H., Johnson, M.K., 2012. Spectroscopic and functional characterization of iron–sulfur cluster-bound forms of Azotobacter vinelandii NifIscA. Biochemistry 51 (41), 8071–8084.

Mariotti, L., Wild, S., Brunoldi, G., Piceni, A., Ceppi, I., Kummer, S., Lutz, R.E., Cejka, P., Gari, K., 2020. The iron–sulphur cluster in human DNA2 is required for all biochemical activities of DNA2. Commun. Biol. 3 (1), 1–11.

Mojica, F., Diez-Villasenor, C., Garcia-Martinez, J., Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155 (3), 733–740.

Nam, K.H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M.P., Ke, A., 2012. Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. J. Biol. Chem. 287 (43), 35943–35952.

Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., Doudna, J.A., 2014. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nat. Struct. Mol. Biol. 21 (6), 528–534.

Nunez, J.K., Lee, A.S., Engelman, A., Doudna, J.A., 2015. Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. Nature 519, 193–198.

Pappas, C.J., Benaroudj, N., Picardeau, M., 2015. A replicative plasmid vector allows efficient complementation of pathogenic Leptospira strains. Appl. Environ. Microbiol. 81 (9), 3176–3181.

Patterson, A.G., Chang, J.T., Taylor, C., Fineran, P.C., 2015. Regulation of the Type IF CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. Nucleic Acids Res. 43 (12), 6038–6048.

Pingoud, A., Fuxreiter, M., Pingoud, V., Wende, W., 2005. Type II restriction endonucleases: structure and mechanism. Cellular Mol. Life Sci. 62 (6), 685.

Rangarajan ESShankar, V, 2001. Sugar non-specific endonucleases. FEMS Microbiol. Rev. 25 (5), 583–613.

Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L., White, M.F., 2015. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. Elife 4, e08716.

Samai, P., Smith, P., Shuman, S., 2010. Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. Acta Crystallogr. Sec. F 66 (12), 1552–1556.

Shiimori, M., Garrett, S.C., Graveley, B.R., Terns, M.P., 2018. Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. Mol. Cell 70 (5), 814–824 e6.

Sievers FHiggins, D.G, 2014. Clustal Omega, Accurate Alignment of Very Large Numbers of sequences. Multiple sequence Alignment Methods. Springer, pp. 105–116.

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., Wang, Y., 2015. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. Cell 163 (4), 840–853.

Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J., van der Oost, J., 2012. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annu. Rev. Genet. 46, 311–339.

White, M.F., 2009. Structure, function and evolution of the XPD family of iron–sulfur-containing 5′→ 3′ DNA helicases. Biochem. Soc. Trans. 37 (3), 547–551.

Wilkinson, M., Chaban, Y., Wigley, D.B., 2016. Mechanism for nuclease regulation in RecBCD. Elife 5, e18227.

Xiao, G., Yi, Y., Che, R., Zhang, Q., Imran, M., Khan, A., Yan, J., Xa, Lin, 2019. Characterization of CRISPR-Cas systems in Leptospira reveals potential application of CRISPR in genotyping of Leptospira interrogans. Apmis 127 (4), 202–216.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12 (1), 7–8.

Yang, W., 2011. Nucleases: diversity of structure, function and mechanism. Q. Rev. Biophys. 44 (1), 1.

Yeeles, J.T., Cammack, R., Dillingham, M.S., 2009. An iron-sulfur cluster is essential for the binding of broken DNA by AddAB-type helicase-nucleases. J. Biol. Chem. 284 (12), 7746–7755.

Yosef, I., Goren, M.G., Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. 40 (12), 5569–5576.

Zhang, J., Kasciukovic, T., White, M.F., 2012. The CRISPR associated protein Cas4 Is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. PLoS ONE 7 (10).

Zhang, Z., Pan, S., Liu, T., Li, Y., Peng, N., 2019. Cas4 nucleases can effect specific integration of CRISPR spacers. J. Bacteriol. 201 (12) e00747-18.

Zhou, C., Pourmal, S., Pavletich, N.P., 2015. Dna2 nuclease-helicase structure, mechanism and regulation by Rpa. Elife 4, e09832.