3946–3957 *Nucleic Acids Research, 2016, Vol. 44, No. 8* *Published online 11 March 2016*
doi: 10.1093/nar/gkw143

Structural characterization of the virulence factor Sda1 nuclease from *Streptococcus pyogenes*

Andrea F. Moon[1,*], Juno M. Krahn[1], Xun Lu[2], Matthew J. Cuneo[2] and Lars C. Pedersen[1]

[1]Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA and [2]Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Received December 21, 2015; Revised February 24, 2016; Accepted February 25, 2016

ABSTRACT

Infection by Group A *Streptococcus pyogenes* (GAS) is a leading cause of severe invasive disease in humans, including streptococcal toxic shock syndrome and necrotizing fasciitis. GAS infections lead to nearly 163,000 annual deaths worldwide. Hypervirulent strains of *S. pyogenes* have evolved a plethora of virulence factors that aid in disease—by promoting bacterial adhesion to host cells, subsequent invasion of deeper tissues and blocking the immune system's attempts to eradicate the infection. Expression and secretion of the extracellular nuclease Sda1 is advantageous for promoting bacterial dissemination throughout the host organism, and evasion of the host's innate immune response. Here we present two crystal structures of Sda1, as well as biochemical studies to address key structural features and surface residues involved in DNA binding and catalysis. In the active site, Asn211 is observed to directly chelate a hydrated divalent metal ion and Arg124, on the putative substrate binding loop, likely stabilizes the transition state during phosphodiester bond cleavage. These structures provide a foundation for rational drug design of small molecule inhibitors to be used in prevention of invasive streptococcal disease.

INTRODUCTION

The Group A *Streptococcus pyogenes* (GAS), a strictly human pathogen, is a leading cause of over 700 million cases per year of superficial infections (pharyngitis and impetigo), as well as greater than 650,000 cases of invasive infections (1). Examples of invasive infection include cellitis, bacteraemia, streptococcal toxic shock syndrome (STSS) and necrotizing fasciitis, and GAS infections have been estimated to be the primary cause of nearly 163,000 deaths annually. The 1980s witnessed an explosive resurgence of invasive streptococcal infections worldwide, which has been attributed largely to the emergence of a highly virulent GAS subclone with serotype M1T1 (2,3). This clone differs from other, less pathogenic GAS strains in that it contains three prophages—ΦM1T1.X, ΦM1T1.Y and ΦM1T1.Z. ΦM1T1.Y is similar to the phage 307.3 of the M1-SF370 strain, but ΦM1T1.X and ΦM1T1.Z, encoding virulence factors SpeA2 and Sda1, respectively, are unique to the M1T1 subclone (4). The acquisition of multiple prophages allows bacteria to exchange virulence factors, toxins and gain antibiotic resistance, and is an evolutionarily beneficial process by which bacteria adapt to a changing environment (3,5,6).

GAS has developed a vast arsenal of virulence factors that work synergistically to promote efficient bacterial infection of the host organism. These factors play a variety of different roles during infection, including adhesion to host cells, invasion and spread to deeper tissues, toxins to break down those tissues and evasion of the host immune response (7,8). One of the two virulence factors distinguishing the M1T1 strain, SpeA2 (Streptococcal pyrogenic exotoxin A2), is a superantigen whose function involves simultaneous engagement of major histocompatibility complex-class II molecules and T-cell receptor (9). This interaction may cause a catastrophic overproduction of cytokines and is directly correlated with severe forms of STSS (10,11).

Sda1, identified as a homologue of SdaD, displays potent sequence-nonspecific nuclease activity on DNA substrates in a divalent cation-dependent manner (12,13). Overexpression of secreted Sda1 occurs during phase-shifts to hypervirulent invasive streptococcal infection (14,15) and is thought to play a role in evasion of the host's innate immune response in a variety of different ways. Sda1 is thought to be involved primarily in degradation of the DNA component of chromatin-rich neutrophil extracellular traps (NETs), and the similar DNA-based extracellular traps extruded by macrophages (13). The secreted nuclease activity of Sda1 provides a protective effect against bacterial killing by activated neutrophils and macrophages. Additionally, it has been recently reported that Sda1 may also provide a protective effect by avoidance of TLR9-mediated recognition

*To whom correspondence should be addressed. Tel: +1 919 316 4677; Fax: +1 919 541 7880; Email: moon@niehs.nih.gov

of unmethylated CpG-rich bacterial DNA, and subsequent cytokine overproduction (13). Therefore, the presence of Sda1 may represent an evolutionarily advantageous adaptation that aids in severe streptococcal invasion and disease.

In order to gain a better understanding of Sda1 function as a virulence factor in *S. pyogenes* infection, catalytically inactive forms (general base His188 replaced with glycine) of the recombinant nuclease were expressed and purified for structural characterization. Two X-ray crystal structures of Sda1(H188G) were determined—the catalytic core with a bound zinc ion in the active site (1.95 Å) and the other included a partial structure of the C-terminal region unique to Sda1 (1.90 Å). Detailed structural comparison to other structurally related bacterial nucleases led to identification of surface residues contributing to catalytic activity, which were tested using an imidazole chemical rescue strategy. Asn211 was discovered to directly chelate the divalent metal cation in the active site, and His188 serves as the general base. On the putative DNA substrate binding loop, Arg124 likely stabilizes the transition state intermediate during cleavage of the phosphodiester bond. Analysis of the Sda1 active site reveals that the canonical hydrogen bonding network that exists in related nucleases is not conserved and have been functionally replaced by a different web of interactions in this nuclease. This study provides a deeper understanding of the relationship between enzyme function and invasive streptococcal disease.

MATERIALS AND METHODS

Cloning, expression and purification of Sda1 for biochemical assays

Sequences encoding Sda1 (Figure 1A and B) were cloned into the NotI and BamHI restriction sites of the pGEXM (16) expression vector, fusing Sda1 to the C-terminus of glutathione-S-transferase (GST). Due to historical difficulties with solubly expressing wild-type nucleases in *Escherichia coli*, all constructs contained a putatively deactivating mutation of the general base (His188) to glycine. The resulting constructs were transformed into Rosetta2 (DE3) cells (Novagen) and expressed in LB medium. Cultures were grown at 37°C, with shaking at 275 rpm, to an OD_{600nm} of 0.6–0.8, at which point the temperature was decreased to 18°C. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.4 mM and continued overnight. Cells were pelleted by centrifugation and lysed by sonication in 25 mM Tris pH 8, 500 mM NaCl. The resulting lysate was clarified by further centrifugation and the soluble Sda1 was bound in-batch to glutathione sepharose 4B resin (GE Healthcare) at 4°C. The N-terminal GST fusion tag was removed by on-resin Tobacco Etch Virus protease cleavage. Soluble, cleaved Sda1 proteins were concentrated and further purified by size exclusion chromatography. For detailed information on purification and crystallization of proteins, see the Supplementary Materials.

Generation of Sda1 mutations and deletion variants

Several loop deletion variants and point mutations on the surface of Sda1 (residues Glu38-Glu390) were generated on the background of the H188G mutant, using QuikChange mutagenesis (Agilent). The resulting variants were expressed and purified in small scale as GST-fusion proteins, as described above for the H188G mutant alone. The mutants were concentrated to 1–4 mg/ml in a storage buffer containing 25 mM Tris pH 8, 100 mM NaCl. All mutants behaved indistinguishably from Sda1 (H188G) in size exclusion chromatography experiments.

Plasmid conversion nuclease activity assays

The supercoiled population of the pBluescript SK(+) plasmid was purified by gel extraction used as the substrate for the plasmid conversion assay. Sda1 proteins lacking His188 were diluted to 20 nM in 20 mM Tris pH 7, 0.1 mM $MgCl_2$ and incubated with 15 ng/μl of supercoiled pBluescript SK(+) plasmid for 40 min, with or without 30 mM imidazole (pH 7). The reactions were quenched by addition of loading dye containing EDTA. Samples were run on an 0.8% (w/v) agarose gel, dissolved in 1X Tris-Acetate-EDTA buffer. DNA species in the gel were visualized by ethidum bromide staining, scanned with a Typhoon fluorescence imager and analysed using ImageQuant TL. Nuclease activity assays were performed in triplicate.

RESULTS

Sequence analysis of Sda1 domain structure

Sequence analysis of Sda1 from the M1T1 strain of *S. pyogenes* shows a high degree of similarity to the SdaD nuclease from the M49 strain (12). BLAST sequence alignment (17) also reveals that this nuclease is related to DNA/RNA non-specific endonucleases (streptodornases) from other *Streptococcus* species, including Nuclease A (GBS_NucA) from *S. agalactiae* (18), DNA-entry nuclease EndA from *S. pneumoniae* (19) and another phage-encoded nuclease from *S. pyogenes* known as Spd1 (20). However, strong sequence similarity exists only for the catalytic core of the enzyme (approximately Tyr55-Ser294) (Figure 1B). The C-terminal region (Thr295-Glu390) of the protein displays no similarity to any other proteins but streptodornases similar to SdaD from strain M49. The function of this domain remains undetermined.

Hydropathy plots of the Sda1 protein sequence suggest that this nuclease contains a putative transmembrane α-helix including residues Leu8-Phe21 (21) (Figure 1A). Since previously published studies have shown secreted Sda1 in the supernatents of cultured M1T1 *S. pyogenes* clinical isolates (12), it seems likely that this nuclease may contain a signal peptide for extracellular extrusion and peptide cleavage. Analysis of the N-terminal region by the SignalP server (22) reveals a possible cleavage site between Ala37 and Glu38, which would yield a mature protein containing residues Glu38-Glu390. Secondary structure prediction suggests that the boundaries of the ordered catalytic domain may extend from Lys60 through Asp285 (23,24) (Supplementary Figure S1). Though the C-terminal region appears to be largely unstructured, there is one predicted α-helix (Glu373-Thr378) and three small putative β-strands (Thr340-Ala344, Gln348-Trp353 and Val369-Met371).

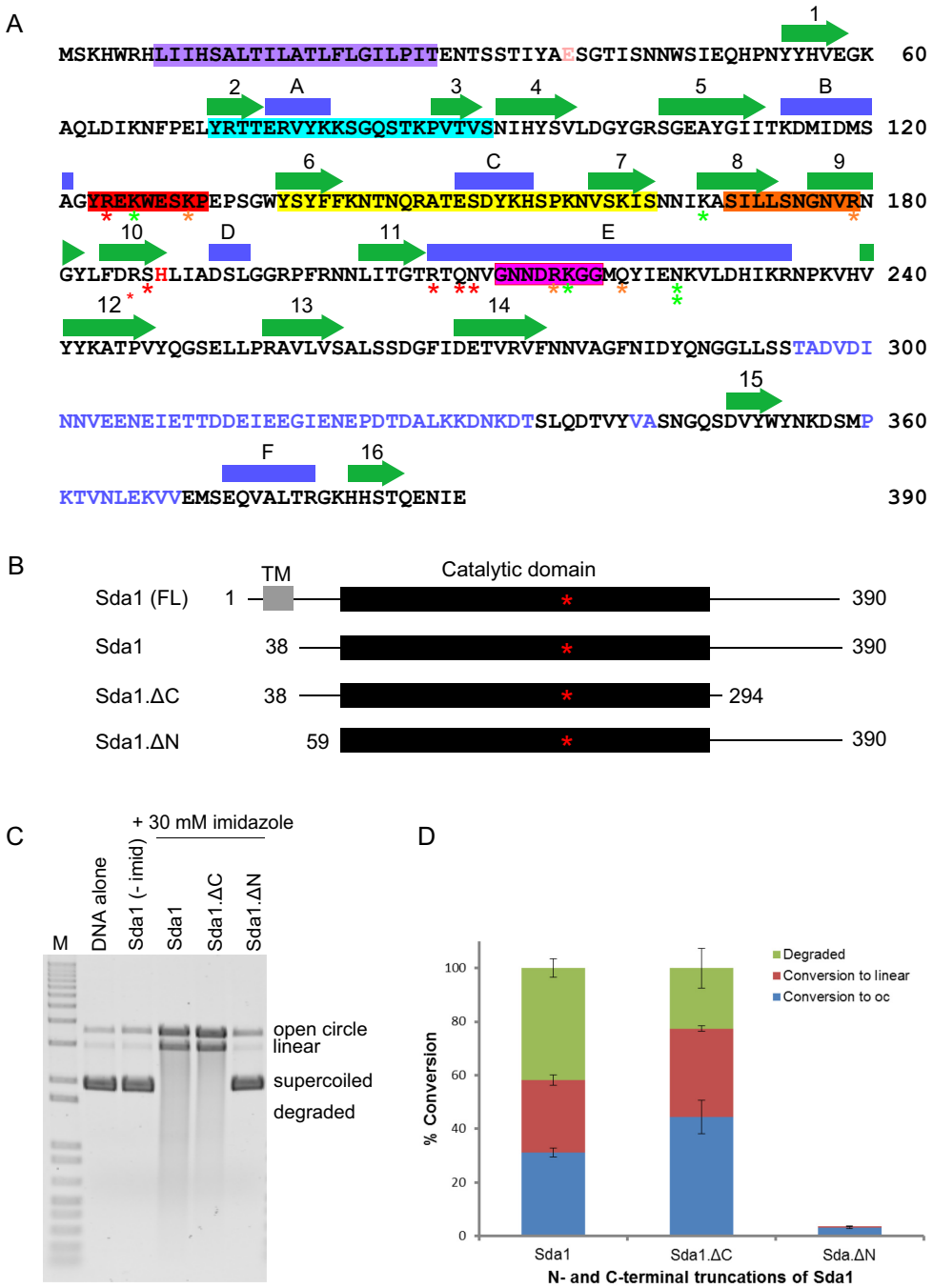

Figure 1. Sequence, domain structure and activity of Sda1. (**A**) Protein sequence of Sda1. The putative transmembrane α-helix is boxed in purple, with the likely cleavage site immediately upstream of Glu38 (pink). The 'Finger Loop' and Insertions (Ins.) I, II and III are boxed in magenta, cyan, yellow and orange, respectively. The putative DNA substrate binding loop is shown in a red box. Secondary structural elements are indicated by green arrows (β-strands, labelled alphabetically) or blue rectangles (α-helices, labelled numerically). The location of the general base (His188) is shown in red. Surface residues mutated in this study are marked by asterisks: green, orange or red based on having minimal, moderate or severe effects on activity, respectively. Disordered residues of the C-terminal domain are marked in light blue. (**B**) Sda1 construct designs used in this study, with the beginning and ending residues of Sda1 indicated. (**C–D**) Nuclease activity assays for the different construct lengths of Sda1(H188G), as measured using the imidazole chemical rescue strategy. The DNA substrate is a supercoiled plasmid, which forms an open circle when nicked (blue bar), and is linearized upon accumulation of double-strand breaks (red bar). The linear form is further degraded to a variety of smaller fragments, as visualized by a lower molecular weight smear (green bar). All reactions were performed in triplicate, and the resulting relative product populations were calculated using ImageQuant TL. The indicated error bars represent the standard deviation for each calculation.

To explore the roles of the N- and C-terminal regions and the nuclease activity of Sda1, three constructs of varying lengths were generated for bacterial expression: the mature secreted protein (Glu38-Glu390, henceforth referred to as Sda1), a C-terminal truncation (Glu38-Ser294, Sda1.ΔC) and an N-terminal truncation (Gly59-Glu390, Sda1.ΔN) (Figure 1B). Historically, wild-type endonucleases are difficult to express in *E. coli*, due to toxicity resulting in degradation of the bacterial genome. Expression of recombinant Sda1 in *E. coli* yielded only very small quantities of active protein—sufficient for limited biochemical assays, but not for structural characterization (12). Therefore, inactive variants of Sda1, containing a mutation of the putative general base His188 to glycine, were created for each construct length. The resulting constructs expressed solubly and were observed to be well-behaved monodisperse populations, as assayed by size-exclusion chromatography (Supplementary Figure S2). Interestingly, the Sda1 construct representing the mature protein (40.1 kDa) exhibited a decreased retention time, relative to the chicken ovalbumin standard protein (44 kDa), suggesting the possibility of dimerization. In contrast, the Sda1.ΔC construct lacking the C-terminal region (29.2 kDa) exhibited an increased retention time, as compared to the horse myoglobin (17 kDa). The drastic change in retention time cannot be accounted for by simply reducing the protein size by 10.9 kDa. Taken together, these observations suggest that the mature protein could be forming a dimer in solution, possibly mediated by the C-terminal region.

Evaluation of Sda1 oligomeric state by small-angle neutron scattering (SANS)

To determine whether the C-terminal domain of Sda1 might play a role in dimerization, the solution scattering behaviours of Sda1 and Sda1.ΔC constructs were analysed using SANS. Comparison of the Guinier plots for Sda1 and Sda1.ΔC suggests that the proteins are monodisperse and display no aggregation at the protein concentrations used in the SANS studies (Supplementary Figure S3). The intensity at zero scattering angle (I_o) was used to determine molecular mass. I_o analysis yielded molecular weights of 39,883 Da and 32,384 Da for Sda1 (actual MW = 40,169 D) and Sda1.ΔC (actual MW = 29,297), respectively (Supplementary Table S1 and Supplementary Figure S3). Therefore, the SANS data indicates that both constructs of Sda1 exist in a monomeric state in solution.

Assessment of the contributions of the Sda1 N- and C-termini to nuclease activity

Constructs of Sda1(H188G) were tested for nuclease activity in a plasmid conversion assay, using an imidazole chemical rescue strategy (25,26). This strategy is based on the principle that exogenously added imidazole may effectively substitute for the histidine (His188) serving as the general base in the active site of Sda1 (27). Imidazole could extract a proton from a nearby water molecule in a similar fashion to a histidine side chain, activating the water for a nucleophilic attack on the scissile phosphate (28). In this assay, Sda1 constructs are presented with a supercoiled plasmid substrate, and the subsequent conversion from this state to either nicked open circle or linearized forms, or even further degradation to smaller fragments, is measured. The Sda1(H188G) mutant has nearly undetectable activity in the absence of imidazole, and its nuclease activity is readily rescued by imidazole addition (Figure 1C and D). Comparison of the nuclease activity levels of the mature Sda1 versus Sda1.ΔC shows that removing the C-terminal domain has only a minimal effect on substrate conversion. Both constructs completely transformed the supercoiled substrate to a different form, but the C-terminal truncation variant displayed decreased conversion to the more degraded fragments, possibly indicating a slight decrease in the overall catalytic rate (Figure 1C and D). Conversely, removing the first 20 residues of the Sda1.ΔN construct severely abrogated its nuclease activity to barely detectable levels.

Structural characterization of Sda1 using X-ray crystallography

In order to gain an understanding of the structure/function relationship for Sda1, different constructs of the nuclease were screened for crystallization potential. The crystal structure of Sda1.ΔC(H188G) was determined at 1.95 Å resolution (Table 1) and contained four molecules in the asymmetric unit. This structure reveals a globular fold for the catalytic domain, wherein a central antiparallel β-sheet, largely open to solvent on the 'back' face of the sheet, is flanked on the 'front' face by the ββα-metal finger motif (ββα-Me) that forms the active centre (Figure 2A–C). The secondary structural elements comprising the ββα motif are β-strands 10 and 11, and α-helix E. Like in other ββα-Me nucleases (25,28,29), the catalytic site is very compact, with its breadth and depth determined by residues in the ββα motif and by β-strands 12–14 in the central β-sheet, respectively. The Sda1.ΔC(H188G) active site contains a bound divalent, partially hydrated zinc ion in the active site (Figure 2C). The zinc ion is buried within the compact active site and is directly coordinated by Asn211 alone. Other nearby residues within the ββα motif coordinate three water molecules hydrating the metal (Glu225, Ser187 and the backbone carbonyls of Asn211 and Gly188). The position of the zinc ion is consistent with the pentahydrated catalytic metal ions observed in other ββα-Me nuclease apoprotein structures, as is the observation that only a single asparagine side chain from the protein coordinates the metal (25,28,29). Canonical octahedral geometry for the metal is disrupted in this structure, due to bifurcated coordination of the ion by the C-terminal carboxyl group of a neighbouring molecule in the crystal lattice (Supplementary Figure S4A and B). This interaction is likely a crystallographic artefact, since truncation of the nuclease at this location is an artificial modification used to enhance crystallization.

Though residues Glu38-Asn45 of the N-terminus are disordered, residues Trp46-Glu58 are visible in the electron density and provide some insight into the decrease in nuclease activity in their absence (Figure 2D). Trp46 partially stacks over Tyr54, which begins a β-strand (β-strand 1, Tyr54-Lys60) that extends the central β-sheet. Tyr55 on this strand forms a stacking interaction with Tyr223 on α-

Table 1. Data collection and refinement statistics

	EGFPX-Sda1.ΔN(H188G)	Sda1.ΔC (H188G)
PDB ID code	5FGU	5FGW
Data collection		
Space group	$C222_1$	P1
Cell dimensions		
a, *b*, *c* (Å)	81.41, 139.37, 120.97	51.10, 66.54, 83.54
α, β, γ (°)	90, 90, 90	108.72, 90.01, 92.36
Resolution (Å)	50–1.90 (1.93–1.90)[a]	50–1.95 (1.98–1.95)
Unique reflections	54586	73933
R_{sym}	9.9 (65.9)	9.2 (32.5)
$I / \sigma I$	25.6 (3.54)	16.44 (2.91)
Completeness (%)	100.0 (99.6)	97.1 (88.4)
Redundancy	7.3 (6.4)	4.5 (2.9)
Model Refinement		
Resolution (Å)	50–1.90	50–1.95
No. reflections	54532	73913
R_{work} / R_{free} (%)[b]	17.68/20.12	21.50/26.01
No. atoms		
Protein	3779	7478
Ions	46	55
Water	327	427
B-factors		
Protein (Mol A/Mol B/Mol C/Mol D)	34.86	29.51/30.91/28.69/29.61
Ions	57.28	46.95
Water	38.33	30.45
R.m.s. deviations		
Bond lengths (Å)	0.011	0.007
Bond angles (°)	1.217	0.978
Ramachandran plot		
Most favourable (%)	98.0	96.9
Allowed (%)	100	100

[a] Values in parentheses refer to the highest resolution shell.
[b] $R_{work} = \Sigma \, | |F_{obs}| - |F_{calc}| | \, / \, \Sigma \, |F_{obs}|$, where R_{free} is calculated for a randomly chosen 5% of reflections, which were not used for structure refinement and R_{work} is calculated for the remaining reflections.

helix E. Additionally, there is a putative hydrogen bond between the backbone carbonyl of Asn53 and the backbone amide of Gly220 on the 'finger loop'. The additional N-terminal residues cover a small hydrophobic patch comprised of residues Gly219, Gly220, Val273 and Val275. It is possible that the presence of the N-terminal residues in this region may stabilize the central β-sheet, the 'finger loop' and the position of α-helix E, aiding in assembly of correct active site architecture for catalysis.

Comparison of each of the four Sda1.ΔC(H188G) molecules in the asymmetric unit reveals only slight structural variation (RMS deviation of <0.2 Å between each molecule in the asymmetric unit, Supplementary Figure S4C). Insertions (Ins.) I and II are disordered to some extent in two of the Sda1.ΔC(H188G) molecules (molecules A and C). Ins.II displays a small amount of 'hinge' flexibility at the base of its 'pseudoknot' structure, which shifts the position of that loop slightly, relative to the rest of the catalytic domain. Intriguingly, the putative DNA substrate binding loop (Figure 2A, red) is ordered in all four molecules of Sda1.ΔC(H188G) in the asymmetric unit, though the N-terminal portion of the loop shows four slightly different conformations in each molecule (Supplementary Figure S4C). These conformational variations may be influenced by packing interactions with neighbouring molecules in the crystal.

Constructs containing the C-terminal domain did not yield usable crystals without the aid of a 'fixed-arm' fusion protein carrier (30). Sda1.ΔN(H188G), when used in conjunction as a fixed-arm fusion downstream of EGFP (31), generated diffraction-quality crystals that provided a 1.90 Å resolution structure (Table 1 and Figure 3A and B). The EGFP fusion carrier protein is located proximal to the back face of the nuclease, with the three alanine 'fixed-arm' linker residues doubling back on the Sda1 N-terminus (Gly59) to accommodate the fusion (Figure 3A). Superposition of the EGFPX-Sda1.ΔN(H188G) and Sda1.ΔC(H188G) structures shows a high degree of structural similarity between the two crystal forms (Figure 3C and D). Unsurprisingly, the configuration of the N-terminus of Sda1.ΔC(H188G) differs from that observed in the EGFP fusion and is likely influenced by the limited flexibility of the linker.

The C-terminal domain of EGFPX-Sda1(H188G), comprised residues Thr295-Glu390, is largely disordered, which is consistent with the lack of predicted secondary structure. However, small portions of the region are visible in the electron density, wedged between the EGFP and the Sda1 catalytic domain (Figure 3A and B). Residues Asp350-Trp353 and His382-Thr385 create a small, two-stranded, antiparallel β-sheet (β-strands 15 and 16) and Glu373-Arg379 form a short α-helix (α-helix F). Though the degree of thermal motion is higher for this C-terminal domain, the amino acid sequence could be determined by the electron density.

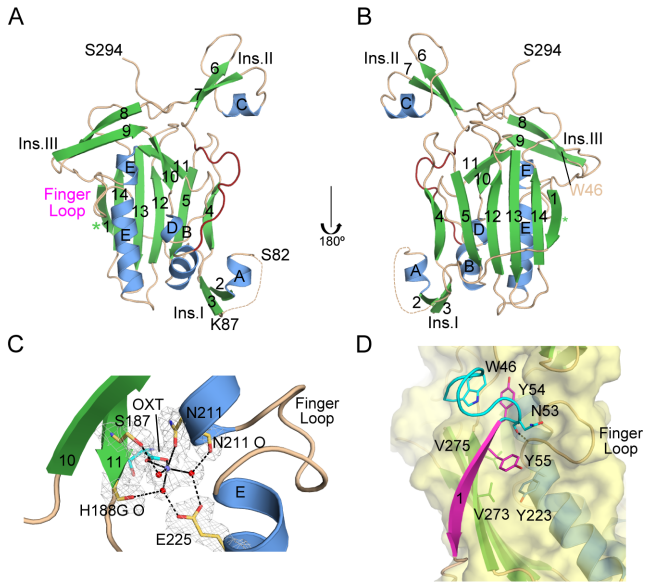

Figure 2. Structural characterization of Sda1.ΔC. (**A**–**B**) Cartoon diagram of Sda1.ΔC(H188G), viewing the 'front' (A) and 'back' (B) faces of the Sda1 catalytic domain (180° y-axis rotation, as indicated by black arrows), with key structural features coloured and labelled as in Figure 1A. (**C**) Diagram of interactions involving ββα motif (β-strands in green, α-helix in blue) and the divalent zinc ion (purple). Water molecules (red spheres) filling out the hydration sphere are shown. Direct chelation of the metal ion is depicted as solid black lines, and putative hydrogen bonding interactions as dashed black lines. Amino acid side chains are shown in yellow sticks, with interactions from the C-terminus of the neighbouring molecule in cyan. 2*Fo-Fc* electron density is shown in grey, contoured at 1σ. (**D**) Interactions involving the N-terminus of Sda1.ΔC(H188G). N-terminal resides Trp46-Asn53 are shown in cyan, and β-strand 1 (Asn54-Lys60) in cartoon in magenta. Putative hydrogen bonding between the backbone carbonyl of Asn53 and the backbone amide of Gly220 is marked with a dashed black line. All structural figures were created using PyMOL (http://www.pymol.org).

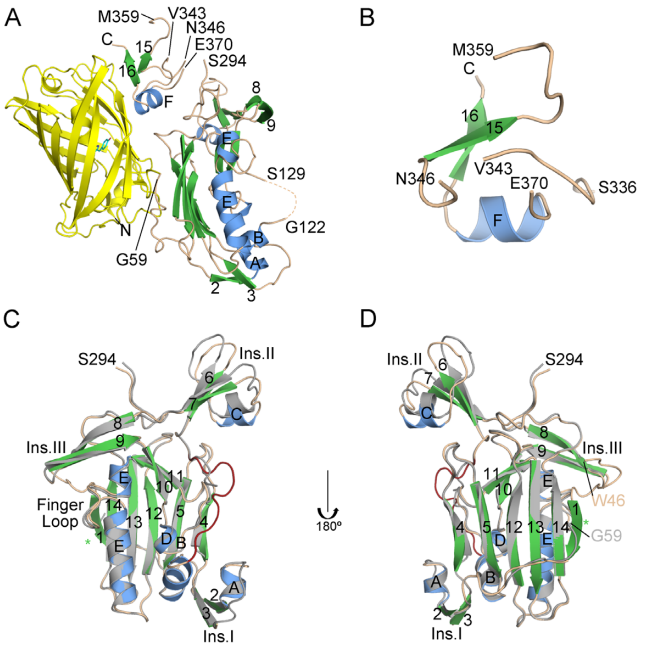

Figure 3. Structure of EGFPX-Sda1.ΔN(H188G) and comparison to Sda1.ΔC(H188G). (**A**) Cartoon diagram of EGFPX-Sda1.ΔN(H188G). The EGFP fusion carrier protein (yellow), with its fluorphore (cyan). Sda1.ΔN(H188G) is shown in a 'side' view, looking down the plane of the central β-sheet, with the 'front' and 'back' faces to the right and left, respectively. The C-terminal domain is wedged between the EGFP and the Sda1 catalytic domain (top, centre). (**B**) Cartoon diagram of the ordered regions of the C-terminal domain, with the ends of each fragment labelled. (**C**–**D**) Superposition of EGFPX-Sda1.ΔN(H188G) (grey) with molecule B of Sda1.ΔC(H188G), coloured as indicated in Figure 1A. 'Front' (C) and 'back' (D) faces of the superimposed catalytic domains are shown, with the location of β-strand 1 marked (green asterisk). Relative locations of the N-termini of the different constructs are shown, with EGFPX-Sda1.ΔN(H188G) in grey and Sda1.ΔC in wheat. The C-terminal domain has been excluded from panels C and D for simplicity.

It should be noted that the connectivity for this domain could not be determined, since more than 50 amino acid residues are disordered between the C-terminal end of the catalytic domain (Ser294) and the first ordered residues of the C-terminal domain (Ser349), in addition to the nearby location of Ser294 from a neighbouring symmetry-related molecule (Supplementary Figure S5).

One intriguing aspect of ββα-Me nucleases is that their active sites have evolved to bind a hydrated divalent metal, yet the active site in this structure of EGFPX-Sda1.ΔN(H188G) contains only solvent molecules, despite the presence of divalent magnesium ions in the crystallization condition. The crystallization condition also contained a high concentration of sulfate ions, and a few such ions are visible near the active centre in this crystal structure. Since sulfate ions could mimic the negative charge presented by the phosphate backbone of the DNA, the location of these sulfate ions could suggest the likely binding position of the substrate. The region of Sda1 that is commensurate with the DNA substrate binding loop (Tyr123-Pro131) in other ββα-Me nucleases (25,29) is disordered in this structure, which suggests a certain amount of flexibility in the absence of a bound substrate.

Structural comparison of Sda1 with other ββα-Me nucleases

Analysis of Sda1.ΔC(H188G) and the nuclease moiety of the EGFPX-Sda1.ΔN(H188G) fusion protein using the Dali server (32) reveals that Sda1 shows the highest degree of structural similarity to GBS_NucA from *S. agalactiae*, EndA from *S. pneumoniae* and Spd1 encoded by prophage SF370.1 (Supplementary Table S2). Values for structure-based sequence identity ranged from 23 to 26%. Z-scores were higher for the Sda1.ΔC(H188G), as compared to the EGFPX-Sda1.ΔN(H188G). Sda1 also shows similarity to SM nuclease from *Serratia marcescens*, NucA from *Anabaena* sp., CSP-6 from *C. elegans* and human mitochondrial EndoG. However, the extent of similarity is lower, as indicated by markedly decreased Z-scores and sequence identity, and increased RMS deviation (Supplementary Table S2).

Structural superposition of Sda1.ΔC(H188G) with the catalytic domain of Spd1 (PDB ID code 2XGR (20), Supplementary Table S3), another prophage-encoded ββα-Me nuclease, reveals that the two nucleases share a similar core architecture, yet display some striking differences (Figure 4A–C and Supplementary Figure S6A). The disposition of the N-terminus differs between Sda1 and Spd1, with that of

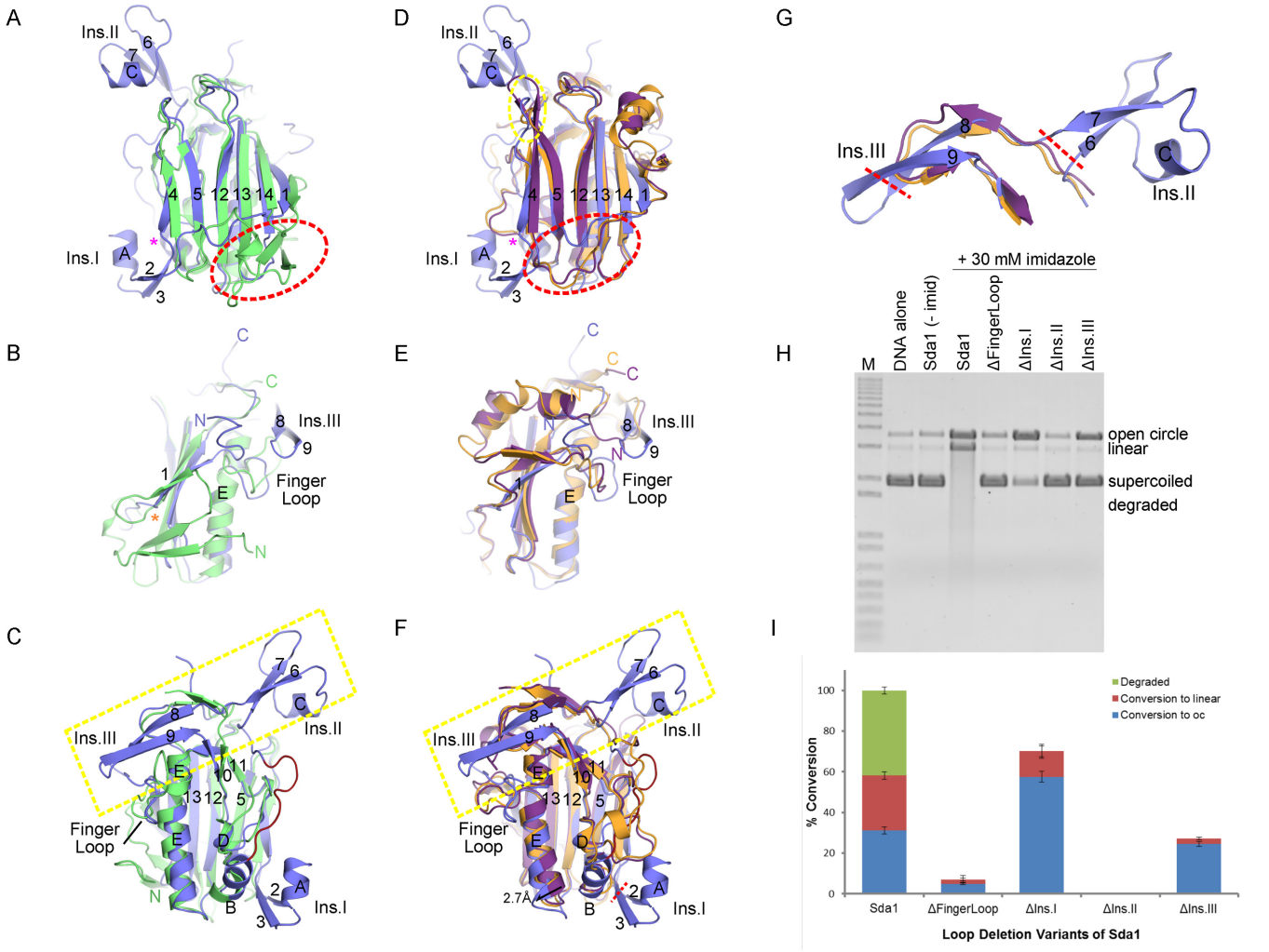

Figure 4. Contribution of Sda1 loop regions to nuclease activity. (**A–C**) Superposition of the Sda1 catalytic domain (blue) with that of SF370.1 prophage-encoded Spd1 (PDB ID code 2XGR (20), green), showing 'back' (A), 'side' (B) and 'front' (C) views, relative to the central β-sheet. (**D–F**) Superposition of Sda1 catalytic domain with that of EndA from *S. pnuemoniae* (PDB ID code 3OWV (25), purple) and GBS_NucA from *S. agalactiae* (PDB ID code 4QH0 (29), orange), showing 'back' (D), 'side' (E) and 'front' (F) views, relative to the central β-sheet. Magenta asterisk and yellow or red dashed circles mark regions of structural differences between the three nucleases and Sda1. Orange asterisk marks the Spd1 β-strand that is structurally but not sequentially homologous to β-strand 1 in Sda1. Secondary structural elements are labelled using Sda1 nomenclature, as described in Figure 1A. (**G**) Insertions II and III in Sda1 compared to similar regions in EndA and GBS_NucA. These loop regions were truncated and assayed for nuclease activity. (**H**) Plasmid conversion nuclease activity assay for Sda1 and Sda1 loop truncation variants. (**I**) Conversion of the supercoiled plasmid substrate from H to either nicked open circle (blue bar), linearized (red bar) or degraded forms (green bar). Each assay was performed in triplicate, and the error bars for each population represent the standard deviation.

Sda1 beginning on the 'back' face of the enzyme, running β-strand 1 alongside the central β-sheet, then traversing the entire length of that face to connect to Ins.I (Figure 4A). In contrast, the N-terminus of Spd1 begins on the 'front' face of the enzyme, then skirts the side of the central β-sheet in a β-strand that forms a small, two-stranded antiparallel β-sheet with another strand from a C-terminal region that is not conserved in Sda1 (Figure 4B and C), before it too traverses the 'back' face to the other side of the sheet. Interestingly, β-strand 1 that proved to be critical for Sda1 nuclease activity is structurally homologous with a short β-strand (residues Lys219-Ile222) from Spd1, though the connectivity is entirely dissimilar and the sequence is not conserved (Figure 4B, orange asterisk). Conservation of a β-strand at this position underscores the critical importance of stabilizing the central β-sheet. On the 'front' face of these enzymes, the architecture of the ββα motif is very similar, though the 'finger loop' extrusion from the long α-helix E varies in length and structure between the two nucleases (Figure 4C). Additionally, the return of the 'finger loop' extrusion to the α-helix appears to cause a mild distortion at that position in Spd1, which is not observed for Sda1. One critical difference between Sda1 and Spd1 is the length and structure of the putative substrate binding loop. This loop is ordered in the structure of Sda1.ΔC(H188G), and disordered in the structure of EGFPX-Sda1.ΔN(H188G), which implies flexibility in this region in the absence of a DNA substrate. This loop is considerably longer in Spd1 than in Sda1 (Supplementary Figure S6A) and is disordered over a larger area. The substrate binding loop is also disordered in EndA (25),

and certain structures of GBS_NucA (29). Given its length in all known ββα-Me nucleases, this loop could play a substantial role in substrate binding for these enzymes, by determining the length, breadth and charge distribution of the DNA binding cleft.

Structural superposition of Sda1.ΔC(H188G) with the structures of GBS_NucA (9) and EndA (25) reveals some intriguing structural differences (Figure 4D–G and Supplementary Table S3). First, the random coil traversing the 'back' face of the enzyme (residues Lys60-Tyr72 in Sda1) displays different conformations in each enzyme (Figure 4D, red circle). Slight variations in the conformation of this coil could alter the solvent-exposed landscape on this 'face' of the nuclease. Second, the length of the loop between β-strands 4 and 5 in Sda1varies between the three nucleases. This loop is longest in EndA, shortest in GBS-NucA and is an intermediate length in Sda1 closer to that of GBS_NucA (Figure 4D). The N-terminal end of β-strand 4 in Sda1 is slightly splayed outward from that of β-strand 5 in the plane of the central β-sheet, as compared to EndA and GBS_NucA, which wedges open the sheet at that position. There is also a significant variation in the conformation of the N-termini of these nucleases (Figure 4E). GBS_NucA and EndA display a high degree of similarity in their N-termini, with an α-helix traversing the 'shoulder' of the molecule from the 'front' face toward the 'back' face. GBS_NucA and EndA also have a small two-stranded antiparallel β-sheet on that 'shoulder', which places the second β-strand of this sheet (Ile69-Leu72 of GBS_NucA and Ala77-Val80 of EndA) in a position roughly similar to that of β-strand 1 in Sda1. As seen for Spd1, there are differences in the sequences and structures of the 'finger loop' extrusion from the long α-helix (α-helix E for Sda1) in the ββα motif, though its length appears to be conserved between Sda1, GBS_NucA and EndA (Figure 4F and Supplementary Figure S6B). Additionally, there is a slight variation in the C-terminal end of this helix. In GBS_NucA and Enda, this helix is slightly bent over the length of the helix, downstream of the 'finger loop', which ultimately shifts the end of the helix by 2.7 Å, as compared to the same location in Sda1. The structure of the DNA substrate binding loop differs between Sda1 and GBS_NucA. For GBS_NucA, this loop (referred to as the S-loop) contains a short α-helix. Given the similarity of loop length and sequence in this region, EndA has been hypothesized to have a homologous α-helix at this position (25,29). In Sda1, however, this loop is slightly shorter, has little sequence conservation beyond a positively charged residue near the N-terminal end of the loop (Arg124). Subsequently, there is no S-loop α-helix observed for Sda1. The C-termini of the GBS_NucA and EndA catalytic domains occupy very similar positions, but which differ from that of the Sda1 catalytic domain (Figure 4E).

Sda1 contains three insertion regions on the 'front' face of the enzyme, which are unique structural features as compared to Spd1, GBS_NucA or EndA (Figure 4C, F, and Supplementary Figure S6). Ins.I (Tyr72-Ser92) is an insertion between β-strands 1 and 4 in Sda1, and consists of a two small β-strands forming a short, antiparallel β-sheet, with a short α-helix in the loop between them. The position of this loop could likely affect the splaying of the end of β-strand 4, along the bottom of the central β-sheet. Ins.II (Tyr134-Ser164), an extrusion between the C-terminal end of the substrate binding loop and the beginning of β-strand 8, forms an intriguing 'pseudoknot' type of structure. β-strands 6 and 7 form a two-stranded antiparallel β-sheet, the loop between the two strands contains the short α-helix C and several random coil residues that flatten back over the short β-sheet. This region displays no structural similarity to any known protein, as surveyed by the Dali Server (32). In GBS_NucA and EndA, the substrate binding loop flows smoothly to the following β-strand (Figure 4F and G). Finally, Ins.III (Ser170-Arg179), which connects β-strands 8 and 9 in Sda1, is up to seven residues longer than in GBS_NucA or EndA. The confluence of Insertions II and III in Sda1 drastically alters the landscape of that particular region, as compared to Spd1, GBS_NucA or EndA.

Contributions of Sda1 loop regions to nuclease activity

In order to determine if any of the insertions in Sda1 contributes to its nuclease activity, each insertion was deleted (ΔIns.I:ΔTyr72-Val91, connected with glycine; ΔIns.II:ΔTyr134-Ser164; ΔIns.III:ΔIle171-Arg179, connected with glycine) and subsequently tested in the plasmid conversion assay, using the imidazole chemical-rescue strategy. The 'finger loop' extrusion from α-helix E was also deleted (ΔLoop: ΔAsn214-Gly219), similarly to what had previously been done in EndA (25), such that the helical structure was not disrupted. None of the deletions had any apparent effect on protein folding, since each deletion variant expressed solubly and behaved as a uniform population, well-resolved from the void volume peak during size exclusion chromatography (Supplementary Figure S7). In the plasmid conversion assay, deletion of Ins.I had a limited effect on nuclease activity, while the Ins.III had a more pronounced effect (Figure 4H and I). Surprisingly, the 'finger loop' deletion variant had nearly undetectable levels of nuclease activity, and deletion of Ins.II yielded catalytically inactive protein. These results were unexpected, given that Ins.II lies a considerable distance (∼18 Å) from the catalytic centre and that a similar 'finger loop' deletion in EndA had no effect on either DNA substrate binding or catalysis (25).

Surveying Sda1 surface residues for involvement in catalysis

Analysis of DNA substrate binding by Sda1 was attempted, by co-crystallization, soaking of apoprotein crystals, electrophoretic mobility shift assays and fluorescence polarization anisotropy, all of which were unsuccessful. It is likely that the lack of sequence specificity and low binding affinity both contribute to difficulties in obtaining detailed information pertaining to DNA substrate interactions. Therefore, the crystal structure of the nuclease VVN from *Vibrio vulnificus* was used as a model for DNA binding to Sda1. VVN is currently the only ββα motif-containing nuclease with a DNA substrate bound in a crystal structure (33,34). Though the amino acid sequence and core protein architecture display no similarity to that of Sda1, it is possible to superimpose the two structures, based solely on the ββα motif (Figure 5A). Using this method, the asparagine residues (Asn211 in Sda1 and Asn127 in VVN) that chelate the hydrated divalent metal, and the metals themselves, align well.

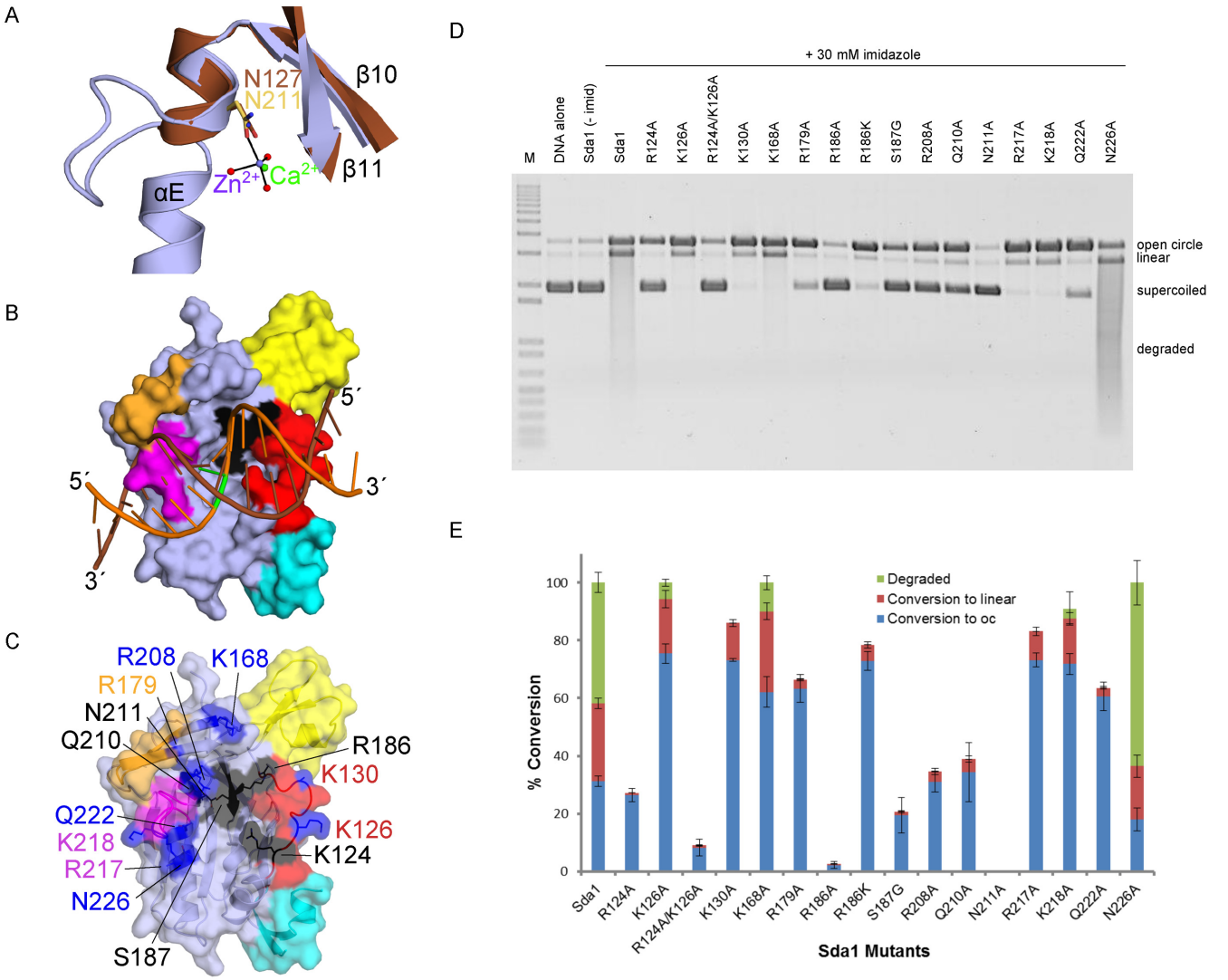

Figure 5. Substitution mutagenesis survey of surface residue involvement in Sda1 catalysis. (**A**) Structural superposition of Sda1.ΔC(H188G) with VVN from *Vibrio vulnificus* (PDB ID codes 2IVK (34) and 1OUP (33)), based on the ββα motif (Asp184-Gly188 and Leu203-Val212 from Sda1, blue; Ile76-Ala80 and Leu119-Gly128 from VVN, brown). The position of the asparagine (Asn211 for Sda1 and Asn127 for VVN) chelating the zinc (Sda1, purple) or calcium (VVN, green) is shown. (**B**) Putative position of DNA substrate (16 base pair duplex bound to VVN, PDB ID code 2IVK (34)) binding to Sda1 catalytic domain (light blue surface). The cleavable strand (orange) and its complement (brown) are drawn as a cartoon, with the scissile phosphate in green. The 'finger loop' (magenta), substrate binding loop (red), Insertions I, II and III (cyan, yellow and light orange, respectively) and the catalytic residues on β-strand 10 (black) of the ββα motif are shown, relative to the DNA substrate. (**C**) Positions of Sda1 surface residues mutated in this study are shown in stick. Residues hypothesized to play a role in catalysis are coloured black, while those with a putative role in DNA binding are coloured blue. Orientation of the molecule is as shown in panel B, but with the DNA removed. (**D**) Plasmid conversion nuclease activity assay for Sda1 and Sda1 mutants. (**E**) Conversion of the supercoiled plasmid substrate from D to either nicked open circle (blue bar), linearized (red bar) or degraded forms (green bar). Each assay was performed in triplicate, and the error bars for each population represent the standard deviation.

Therefore, it is likely that a duplex DNA substrate would bind the ββα motif of Sda1 in a similar orientation (Figure 5B).

Using this model, it was possible to perform a structural survey for residues that may contribute to DNA binding and/or catalysis (Figure 5C). These residues were mutated using site-directed mutagenesis, on the background of the H188G mutant. Therefore, the effects of the surface alterations could be assayed using the imidazole chemical rescue strategy (Figure 5D and E). The K126A, K130A, K168A, R217A and K218A mutants were nearly indistinguishable from the H188G mutant alone, while R179A, R186K and Q222A displayed slight, but statistically significant decreases in activity. The S187G, R208A, Q210A and R124A mutants showed a more severe effect on catalysis, and the R124A/K126A double mutant exhibited significantly lower activity than either of the two individual mutants combined. The N211A mutant, which would negate binding of the divalent metal in the active site, produced a catalytically inactive enzyme. Though not entirely inactive, R186A had barely detectable activity levels.

DISCUSSION

Previously published reports have established that Sda1 serves as a virulence factor for invasive strains of *S. pyrogenes*, degrading NETs, and allowing bacterial dissemination throughout the host organism (35). In this study, we have attempted to elucidate the mechanism of nuclease activity by Sda1, through structural and biochemical methods. We present two X-ray crystal structures of an inactivated Sda1 variant (general base, His188, replaced with glycine), which yields insight into the putative catalytic mechanism. These structures also provide a limited structural portrait of the C-terminal domain (Figure 3A and B) and underscore the importance of an N-terminal β-strand for stabilization of the central β-sheet. Detailed analyses of these structures, compared with those of related nucleases, were the foundation of structure-driven mutagenesis experiments that broaden our understanding of nuclease function.

Sda1 is a ββα-metal finger nuclease that contains a DRGH-like catalytic motif similar to those of EndA and GBS_NucA, and prophage-encoded Spd1 from other *Streptococcus* species (16,20,25). However, in Sda1, this motif has evolved as DRSH, with Ser187 replacing the glycine residue. Despite this alteration in the ββα motif, Sda1 is expected to utilize a similar catalytic mechanism as has been described for other structurally related DRGH nucleases (25,28). Substitution of the general base, His188, allows for soluble overexpression of the nuclease in an *E. coli* expression system and provides an effective means of rescuing nuclease activity by addition of exogenous imidazole (25,26). This imidazole chemical rescue strategy identified a few residues that contribute to catalysis by Sda1. As expected, alanine substitution of Asn211, the only residue directly chelating the active site metal, produced a catalytically inactive enzyme. Similarly, mutations along β-strand 10 (S187G and R186A) had severely deleterious effects on catalysis. A conservative mutation of Arg186 to lysine maintained the positive charge in this area and had only a minimal effect on nuclease activity.

On the putative substrate binding loop, mutation of two positively charged residues, Arg124 and Lys126, yielded interesting results. When mutated individually, K126A had only a minimal effect, while R124A severely diminished activity. Mutating both residues simultaneously produced a synergistic decrease in activity that surpassed R124A alone. Since deletion of the 'finger loop' drastically decreased catalysis by Sda1, residues Arg217 and Lys218 were also mutated to determine the extent of their individual contribution to the loss of activity. However, K218A was nearly indistinguishable from the H188G mutant alone, and R217A displayed only a slight decrease in catalysis. Arg179 on Ins.III was mutated in a similar fashion and did produce a decrease in plasmid conversion. However, mutation of this residue alone cannot account for the loss of activity exhibited by the ΔIns.III mutant (66.4% versus 27.2% supercoiled substrate conversion for R179A and ΔIns.III, respectively).

Gln210 lies deeper within the active site and may serve as a structural substitute for Gln174 in GBS_NucA (Figure 6A) and Gln186 in EndA, which is not conserved in Sda1 (Gly206). The Q210A mutation has a moderate effect

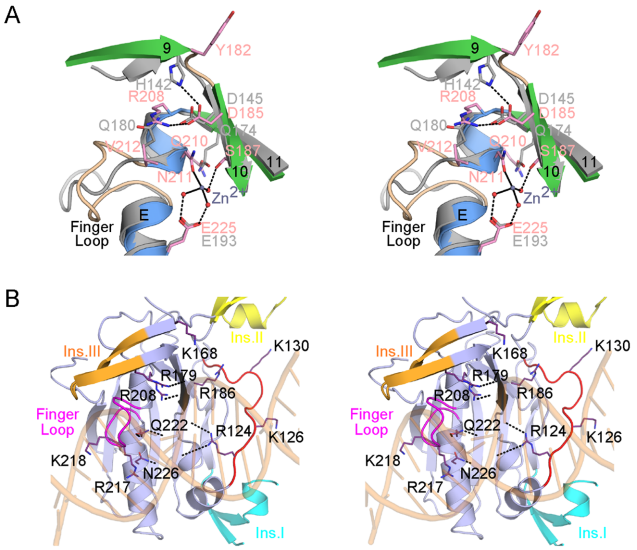

Figure 6. Functionally important residues for Sda1. (**A**) Stereo diagram of the hydrogen bonding network surrounding the hydrated metal cluster (zinc, purple sphere; water molecules, red spheres) in Sda1.ΔC(H188G). Secondary structural elements of the ββα motifs for Sda1 (β-strands, green; α-helix, blue; coils, wheat, labelled as in Figure 1A) and GBS_NucA (PDB ID code 4QH0 (29), grey) are superimposed. Amino acid side chains for each protein are drawn in stick (Sda1, pink; GBS_NucA, grey). Metal coordination and putative hydrogen bonding interactions are marked as solid and dashed black lines, respectively. (**B**) Stereo diagram of putative interactions that could exist between Sda1 surface residues (amino acids in stick, purple) and a hypothetical bound DNA molecule (modeled from VVN crystal structure PDB ID code 2IVK(34)). The DNA (brown) is drawn as a transparent cartoon. Insertions I–III are coloured as previously described (Ins.I in cyan, Ins.II in yellow, Ins.III in orange, 'finger loop' in magenta). Location of β-strand 10 in the ββα motif is marked in black.

on Sda1 catalysis. Gln222 and Asn226 lie on the solvent-accessible side of α-helix E and may be capable of interacting with the DNA. Interestingly, Q222A displayed a decrease in catalytic activity, while N226A was considerably more active than the H188G alone. This mutant converted more of the supercoiled substrate to degraded product forms (63.6% versus 41.8% for H188G alone). Attempts to influence DNA binding and catalysis by mutating positively charged residues along the DNA binding cleft met with only limited success. The K168A mutant was nearly indistinguishable from the H188G alone, and K130A displayed only a slight decrease in activity. R208A was moderately inhibited, to a similar extent of the Q210A mutant (34.7% substrate conversion for R208A and 38.9% substrate conversion for Q210A).

Taken together, the results from the mutagenesis experiments are intriguing, and highlight essential differences between Sda1 and other related streptococcal nucleases. In each of the enzymes, only a single amino acid residue directly chelates the divalent metal ion and several neighboring active site residues putatively hydrogen bond with the water molecules comprising the hydration sphere around the metal (25,29). The same is true for Sda1, where Asn211 chelates the metal, while Ser187, Glu225 and some backbone carbonyls coordinate the water molecules (Figure 6A). These residues comprise the 'first shell' interactions around the hydrated metal cluster and represent the catalytic cen-

tre. In EndA and GBS_NucA, a 'second shell' of residues creates a complex hydrogen bonding network that appears to stabilize the first shell. In Sda1, however, many of the residues involved in this network are not conserved, and different hydrogen bonding interactions appear to stabilize the first shell residues. For example, in GBS_NucA, His142 (Figure 6A and Supplementary Figure S6B) putatively hydrogen bonds with the side chain of Asp145, which then hydrogen bonds with Asn179, the chelating amino acid. However, in Sda1, the structural equivalent of the histidine is Tyr182, which is rotated away from the active site. Asp185 is conserved, but now mediates stabilizing interactions with Arg208 and Asn211. Gln180 on the 'finger loop' of GBS_NucA was found to play a role in second shell interactions, however, this residue is not conserved in Sda1, and is replaced with valine (Val212). One of the most interesting differences between these enzymes is that the structural equivalent of Gln174 in GBS_NucA is Gly206 in Sda1. Gln174 in GBS_NucA and Gln186 in EndA coordinate one of the water molecules hydrating the metal ion, in the first shell. In Sda1, the side chain of Ser187 coordinates the equivalent water, and Gln210 may serve as a structural surrogate for Gln174, indirectly stabilizing the position of Ser187 (Figure 6A). As might be expected, mutation of first shell residues (N211A and S187G) appears to have a more deleterious effect than alterations of second shell residues (Q210A and R208A) (Figure 5D and E).

Though DNA substrate binding to Sda1 could not be directly observed, it is possible to glean some information from the mutagenesis studies. For example, modeling the DNA substrate from VVN onto the Sda1 structure reveals a potential clash between protein and DNA, at the confluence of the 'finger loop' and Ins.III (Figures 5B and 6B). It seems logical that the 'finger loop' might alter its conformation to permit binding of a DNA substrate. Given their proximity to the DNA, it was surprising that mutation of Arg217 and Lys218 on the 'finger loop' had negligible effect on nuclease activity. However, in a different loop conformation, these residues may lie farther from the DNA substrate. Lys130 and Lys168 were also more distal from the DNA binding cleft, so lack of involvement in DNA substrate binding is perhaps unsurprising.

Interpreting the roles of Gln222 and Asn226 is more difficult. Both are positioned to make contact with the DNA. Given their position, it was hypothesized that altering them to smaller side chains might create a more open cleft in which the substrate might bind. Since the N226A mutation generated a more active enzyme, such a hypothesis might be accurate. However, Q222A exhibited decreased activity. It is possible that this side chain may adopt a different rotamer in the presence of a DNA substrate, bringing it closer to the first shell residues and allowing it to stabilize the conformation of Glu225.

Arg124 on the putative substrate binding loop currently lies just outside of hydrogen bonding range of the DNA substrate and the catalytic centre. However, with a slightly different rotamer conformation, it could be precisely placed to hydrogen bond with the scissile phosphate. Such placement would be consistent with a role in transition state stabilization, as hypothesized for the structurally homologous residues in other ββα-Me nucleases (Arg127 in EndA (25), Arg116 in GBS_NucA (29) and Arg93 from *Anabaena* sp. NucA (28)).

The structural and biochemical data presented in this study provide a foundation for better understanding the activities of secreted sequence/structure-nonspecific nucleases as virulence factors during invasive *Streptococcus* sp. infection. These nucleases represent novel therapeutic targets that attack the virulence factor independently of the invading bacteria. Inhibiting these enzymes would maximize bacterial destruction by the innate immune response and minimize the effect of either the infection or the resulting treatment on the host. This work reinforces the catalytic roles of many conserved residues (Arg186, Asn211 and Arg124), while revealing different effects on nuclease activity by other nonconserved residues (Arg179, Ser187, Gln222, Asn226) and loop regions, such as the 'finger loop' and Ins. regions II and III. Ongoing and future studies will focus on delineating the roles of specific surface residues in substrate specificity differences—endo- versus exonucleolytic activity on deoxyribo- versus ribonucleotide containing substrates. As such, obtaining structures of these enzymes in complex with a variety of substrates would aid in this endeavor. Such structures would significantly contribute to the growing database of information on these and related virulence factors, and could facilitate rational drug design of small molecule inhibitors—targeting DNA binding and possible conformational changes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank G. Mueller and D. Singh for critical reading of the manuscript, and C. Stanley of the EQ-SANS beamline at the Spallation Neutron Source for help with data collection.

FUNDING

Division of Intramural Research of the National Institute of Environmental Health Sciences; National Institutes of Health Grant 1ZIA ES102645–03 (to L.C.P.); US Department of Energy, Office of Science, Office of Basic Energy Sciences Contract W-31–109-Eng-38. The research at the Spallation Neutron Source at Oak Ridge National Laboratory was sponsored by the Scientific User Facilities Division, Office of Basic Energy Sciences, US Department of Energy; Office of Biological and Environmental Research supported the research at Oak Ridge National Laboratory's Center for Structural Molecular Biology (CSMB) using facilities supported by the Scientific User Facilities Division, Office of Basic Energy Sciences, US Department of Energy.
Conflict of interest statement. None declared.

REFERENCES

1. Carapetis,J.R., Steer,A.C., Mulholland,E.K. and Weber,M. (2005) The global burden of group A streptococcal diseases. *Lancet Infect. Dis.*, **5**, 685–694.
2. Cleary,P.P., Kaplan,E.L., Handley,J.P., Wlazlo,A., Kim,M.H., Hauser,A.R. and Schlievert,P.M. (1992) Clonal basis for resurgence

of serious Streptococcus pyogenes disease in the 1980s. *Lancet*, **339**, 518–521.
3. Aziz,R.K. and Kotb,M. (2008) Rise and persistence of global M1T1 clone of Streptococcus pyogenes. *Emerg. Infect. Dis.*, **14**, 1511–1517.
4. Aziz,R.K., Edwards,R.A., Taylor,W.W., Low,D.E., McGeer,A. and Kotb,M. (2005) Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated M1T1 clone of Streptococcus pyogenes. *J. Bacteriol.*, **187**, 3311–3318.
5. Ben Zakour,N.L., Davies,M.R., You,Y., Chen,J.H., Forde,B.M., Stanton-Cook,M., Yang,R., Cui,Y., Barnett,T.C., Venturini,C. *et al.* (2015) Transfer of scarlet fever-associated elements into the group A Streptococcus M1T1 clone. *Sci. Rep.*, **5**, 15877–15883.
6. Davies,M.R., Holden,M.T., Coupland,P., Chen,J.H., Venturini,C., Barnett,T.C., Zakour,N.L., Tse,H., Dougan,G., Yuen,K.Y. *et al.* (2015) Emergence of scarlet fever Streptococcus pyogenes emm12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. *Nat. Genet.*, **47**, 84–87.
7. Cunningham,M.W. (2000) Pathogenesis of group A streptococcal infections. *Clin. Microbiol. Rev.*, **13**, 470–511.
8. Cole,J.N., Barnett,T.C., Nizet,V. and Walker,M.J. (2011) Molecular insight into invasive group A streptococcal disease. *Nat. Rev. Microbiol.*, **9**, 724–736.
9. Sundberg,E.J., Deng,L. and Mariuzza,R.A. (2007) TCR recognition of peptide/MHC class II complexes and superantigens. *Semin. Immunol.*, **19**, 262–271.
10. Norrby-Teglund,A., Chatellier,S., Low,D.E., McGeer,A., Green,K. and Kotb,M. (2000) Host variation in cytokine responses to superantigens determine the severity of invasive group A streptococcal infection. *Eur. J. Immunol.*, **30**, 3247–3255.
11. Kotb,M., Norrby-Teglund,A., McGeer,A., El-Sherbini,H., Dorak,M.T., Khurshid,A., Green,K., Peeples,J., Wade,J., Thomson,G. *et al.* (2002) An immunogenetic and molecular basis for differences in outcomes of invasive group A streptococcal infections. *Nat. Med.*, **8**, 1398–1404.
12. Aziz,R.K., Ismail,S.A., Park,H.W. and Kotb,M. (2004) Post-proteomic identification of a novel phage-encoded streptodornase, Sda1, in invasive M1T1 Streptococcus pyogenes. *Mol. Microbiol.*, **54**, 184–197.
13. Uchiyama,S., Andreoni,F., Schuepbach,R.A., Nizet,V. and Zinkernagel,A.S. (2012) DNase Sda1 allows invasive M1T1 Group A Streptococcus to prevent TLR9-dependent recognition. *PLoS Pathog.*, **8**, e1002736.
14. Aziz,R.K., Pabst,M.J., Jeng,A., Kansal,R., Low,D.E., Nizet,V. and Kotb,M. (2004) Invasive M1T1 group A Streptococcus undergoes a phase-shift in vivo to prevent proteolytic degradation of multiple virulence factors by SpeB. *Mol. Microbiol.*, **51**, 123–134.
15. Walker,M.J., Hollands,A., Sanderson-Smith,M.L., Cole,J.N., Kirk,J.K., Henningham,A., McArthur,J.D., Dinkla,K., Aziz,R.K., Kansal,R.G. *et al.* (2007) DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection. *Nat. Med.*, **13**, 981–985.
16. Moon,A.F., Pryor,J.M., Ramsden,D.A., Kunkel,T.A., Bebenek,K. and Pedersen,L.C. (2014) Sustained active site rigidity during synthesis by human DNA polymerase mu. *Nat. Struct. Mol. Biol.*, **21**, 253–260.
17. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
18. Derre-Bobillot,A., Cortes-Perez,N.G., Yamamoto,Y., Kharrat,P., Couve,E., Da Cunha,V., Decker,P., Boissier,M.C., Escartin,F., Cesselin,B. *et al.* (2013) Nuclease A (Gbs0661), an extracellular nuclease of Streptococcus agalactiae, attacks the neutrophil extracellular traps and is needed for full virulence. *Mol. Microbiol.*, **89**, 518–531.
19. Puyet,A., Greenberg,B. and Lacks,S.A. (1990) Genetic and structural characterization of endA. A membrane-bound nuclease required for transformation of Streptococcus pneumoniae. *J. Mol. Biol.*, **213**, 727–738.
20. Korczynska,J.E., Turkenburg,J.P. and Taylor,E.J. (2012) The structural characterization of a prophage-encoded extracellular DNase from Streptococcus pyogenes. *Nucleic Acids Res.*, **40**, 928–938.
21. Hofmann,K.S.W. (1993) TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, http://www.ch.embnet.org/software/tmbase/TMBASE_doc.html.
22. Petersen,T.N., Brunak,S., von Heijne,G. and Nielsen,H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.
23. Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
24. Buchan,D.W., Minneci,F., Nugent,T.C., Bryson,K. and Jones,D.T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.*, **41**, W349–W357.
25. Moon,A.F., Midon,M., Meiss,G., Pingoud,A., London,R.E. and Pedersen,L.C. (2011) Structural insights into catalytic and substrate binding mechanisms of the strategic EndA nuclease from Streptococcus pneumoniae. *Nucleic Acids Res.*, **39**, 2943–2953.
26. Midon,M., Gimadutdinow,O., Meiss,G., Friedhoff,P. and Pingoud,A. (2012) Chemical rescue of active site mutants of S. pneumoniae surface endonuclease EndA and other nucleases of the HNH family by imidazole. *Chembiochem*, **13**, 713–721.
27. Lehoux,I.E. and Mitra,B. (1999) (S)-Mandelate dehydrogenase from Pseudomonas putida: mutations of the catalytic base histidine-274 and chemical rescue of activity. *Biochemistry*, **38**, 9948–9955.
28. Ghosh,M., Meiss,G., Pingoud,A., London,R.E. and Pedersen,L.C. (2005) Structural insights into the mechanism of nuclease A, a betabeta alpha metal nuclease from Anabaena. *J. Biol. Chem.*, **280**, 27990–27997.
29. Moon,A.F., Gaudu,P. and Pedersen,L.C. (2014) Structural characterization of the virulence factor nuclease A from Streptococcus agalactiae. *Acta Crystallogr. D Biol. Crystallogr.*, **70**, 2937–2949.
30. Moon,A.F., Mueller,G.A., Zhong,X. and Pedersen,L.C. (2010) A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci.*, **19**, 901–913.
31. Mueller,G.A., Pedersen,L.C., Lih,F.B., Glesner,J., Moon,A.F., Chapman,M.D., Tomer,K.B., London,R.E. and Pomes,A. (2013) The novel structure of the cockroach allergen Bla g 1 has implications for allergenicity and exposure assessment. *J. Allergy Clin. Immunol.*, **132**, 1420–1426.
32. Holm,L. and Rosenstrom,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**(Suppl), W545–W549.
33. Li,C.L., Hor,L.I., Chang,Z.F., Tsai,L.C., Yang,W.Z. and Yuan,H.S. (2003) DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site. *EMBO J.*, **22**, 4014–4025.
34. Wang,Y.T., Yang,W.J., Li,C.L., Doudeva,L.G. and Yuan,H.S. (2007) Structural basis for sequence-dependent DNA cleavage by nonspecific endonucleases. *Nucleic Acids Res.*, **35**, 584–594.
35. Buchanan,J.T., Simpson,A.J., Aziz,R.K., Liu,G.Y., Kristian,S.A., Kotb,M., Feramisco,J. and Nizet,V. (2006) DNase expression allows the pathogen group A Streptococcus to escape killing in neutrophil extracellular traps. *Curr. Biol.*, **16**, 396–400.