

## Structural and mechanistic insights into the Artemis endonuclease and strategies for its inhibition

Yuliana Yosaatmadja<sup>1†</sup>, Hannah T Baddock<sup>2†</sup>, Joseph A Newman<sup>1</sup>, Marcin Bielinski<sup>3</sup>, Angeline E Gavard<sup>1</sup>, Shubhashish M M Mukhopadhyay<sup>1</sup>, Adam A Dannerfjord<sup>1</sup>, Christopher J Schofield<sup>3</sup>, Peter J McHugh<sup>2\*</sup>, Opher Gileadi<sup>1\*</sup>.

<sup>1</sup>Centre for Medicines Discovery, University of Oxford, ORCRB, Roosevelt Drive, Oxford, OX3 7DQ, United Kingdom;

<sup>2</sup>Department of Oncology, MRC-Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, United Kingdom;

<sup>3</sup>Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford, OX1 3TA, United Kingdom.

\* To whom correspondence should be addressed. email: [opher.gileadi@cmd.ox.ac.uk](mailto:opher.gileadi@cmd.ox.ac.uk)  
Correspondence may also be addressed to [peter.mchugh@imm.ox.ac.uk](mailto:peter.mchugh@imm.ox.ac.uk).

† These authors contributed equally

### ABSTRACT

Artemis (DCLRE1C) is an endonuclease that plays a key role in development of B- and T-lymphocytes and in DNA double-strand break repair by non-homologous end-joining (NHEJ). Artemis is phosphorylated by DNA-PKcs and acts to open DNA hairpin intermediates generated during V(D)J and class-switch recombination. Consistently, Artemis deficiency leads to radiosensitive congenital severe immune deficiency (RS-SCID). Artemis belongs to a structural superfamily of nucleases that contain conserved metallo- $\beta$ -lactamase (MBL) and  $\beta$ -CASP (CPSF-Artemis-SNM1-Pso2) domains. Here, we present crystal structures of the catalytic domain of wild type and variant forms of Artemis that cause RS-SCID Omenn syndrome. The truncated catalytic domain of the Artemis is a constitutively active enzyme that with similar activity to a phosphorylated full-length protein. Our structures help explain the basis of the predominantly endonucleolytic activity of Artemis, which contrast with the predominantly exonuclease activity of the closely related SNM1A and SNM1B nucleases. The structures also reveal a second metal binding site in its  $\beta$ -CASP domain that is unique to Artemis. By combining our structural data that from a recently reported structure we were able model the interaction of Artemis with DNA substrates. Moreover, co-crystal structures with inhibitors indicate the potential for structure-guided development of inhibitors.

### INTRODUCTION

Nucleases hydrolyse the phosphodiester bonds of nucleic acids and are grouped into two broad classes: exonucleases and endonucleases. Exonucleases are often non sequence-specific, while endonucleases can be further grouped into sequence-specific

endonucleases, such as restriction enzymes, and structure-selective endonucleases [1]. Artemis (also known as SNM1C or DCLRE1C), along with SNM1A (DCLRE1A) and Apollo (SNM1B or DCLRE1B), are human nucleases that fall into the extended structural family of metallo- $\beta$ -lactamase (MBL) fold enzymes [2,3]. The N-terminal region of Artemis is predicted to have a core MBL fold (aa 1–155, 385–361) with an inserted  $\beta$ -CASP (CPSF73, Artemis, SNM1 and PSO2) domain (aa 156–384).  $\beta$ -CASP domains are present within the larger family of eukaryotic nucleic acid processing MBLs and confer both DNA/RNA binding and nuclease activity [2]. The C-terminal region of Artemis mediates protein-protein interactions, contains post translational modification (PTM) sites, directs subcellular localisation, and may modulate catalytic activity [4–7].

Although SNM1A, SNM1B, and Artemis are predicted to have similar core structures for their catalytic domains, each have distinct cellular functions and substrate specificities. While SNM1A and SNM1B/Apollo are exclusively 5' to 3' exonucleases, the predominant activity of Artemis is endonucleolytic [5,7], although a minor 5' to 3' exonuclease activity has been reported [8]. Human SNM1A localises to sites of DNA damage, can digest past DNA damage lesions *in vitro*, and is involved in the repair of interstrand crosslinks (ICLs) [9–11]. SNM1B/Apollo is a shelterin-associated protein required for resection at newly-replicated leading-strand telomeres to generate the 3'-overhang necessary for telomere loop (t-loop) formation and telomere protection [12–14]. Both SNM1A and SNM1B/Apollo prefer ssDNA substrates *in vitro*, with an absolute requirement for a free 5'-phosphate [3,15]. By contrast, Artemis prefers hairpins and DNA junctions as substrates for its endonuclease activity, although it is able to process ssDNA substrates [16–18]

The endonuclease activity of Artemis is responsible for hairpin opening in variable (diversity) joining (V(D)J) recombination [19] and contributes to end-processing in the canonical non-homologous end joining (c-NHEJ) DNA repair [20–23]. V(D)J recombination is initiated by the recognition and binding of recombination-activating gene proteins (RAG1 and RAG2) to the recombination signal sequences (RSSs) adjacent to the V, D, and J gene segments. Upon binding, the RAG proteins induce double-strand breaks (DSBs) and create a hairpin at the coding ends [24–26]. The Ku heterodimer recognises the DNA double-strand break and recruits DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and Artemis to mediate hairpin opening [17]. Following hairpin opening, the NHEJ machinery containing the XRCC4/XLF(PAXX)/DNA-Ligase IV complex is recruited to catalyse the processing and ligation reactions of the DNA ends [20,27,28]. V(D)J recombination is an essential process in antibody maturation [16,29,30]. Mutations in the Artemis gene cause aberrant hairpin opening resulting in severe combined immune deficiency (RS-SCID), with sensitivity to ionising radiation due to impairment of the predominant DSB repair pathway in mammalian cells, NHEJ [17,19,27], and another form of SCID (Omenn syndrome) associated with hypomorphic Artemis mutations [31,32]

One of the most common mutations leading to Artemis loss-of-function are large deletions in the first four exons and a nonsense founder mutation, as found in Navajo and Apache Native Americans [33]. In addition, missense mutations and in-frame deletions in the highly conserved residues such as H35, D165 and H228 can also abolish Artemis' protein function [34].

Owing to the key roles of Artemis and related DSB repair enzymes in both programmed V(D)J recombination and non-programmed c-NHEJ DSB repair, they are attractive pharmacological targets for the radiosensitisation of tumours.

Here, we present a high-resolution crystal structure of the catalytic core of Artemis (aa1–361) containing both MBL and a  $\beta$ -CASP domains. This reveals that Artemis possesses a unique feature, that is not present in SNM1A and SNM1B/Apollo, i.e., a second metal binding site in its  $\beta$ -CASP domain that bears a resemblance to classical Cys<sub>2</sub>His<sub>2</sub> zinc finger motifs. We propose that this second metal coordination site is involved in Artemis stabilisation and substrate specificity. We also present a model for Artemis DNA binding based on our data and another recently published structure. The Artemis DNA model is compared with models of DNA binding from related nucleases to reveal distinct features that define a role for Artemis in the end-joining reaction. Following development of an assay suitable for inhibitor screens, we identified drug-like molecules that could potentially inhibit both the Artemis active site and its essential zinc finger-like motif.

## MATERIAL AND METHODS

### **Cloning and site directed mutagenesis of WT and mutant Artemis (aa 1-362)**

The Artemis MBL- $\beta$ -CASP domain (WT and mutant) encoding constructs were cloned into the baculovirus expression vector pBF-6HZB which combines an N-terminal His<sub>6</sub> sequence and the Z-basic tag (GenBank<sup>TM</sup> accession number KP233213.1) for efficient purification and to promote solubility. The Artemis gene was cloned using ligation independent cloning (LIC) [35]. Site directed mutagenesis was carried out using an inverse PCR experiment whereby an entire plasmid is amplified using complementary mutagenic primers (oligonucleotides) with minimal cloning steps [36]. Using the high-fidelity and high-processivity enzyme Herculase II Fusion DNA Polymerase (Agilent), a PCR was performed to amplify a whole plasmid. The PCR product was then added to a KLD enzyme mix (NEB) reaction and was incubated at room temperature for 1 hour, prior to transformation into *Escherichia coli* cells.

### **Expression and purification of WT and mutant Artemis with IMAC (aa 1-362)**

Baculovirus generation was performed as previously described [3]. Recombinant proteins were produced in *Sf9* cell at 2 x 10<sup>6</sup> cells/ mL infected with 1.5 mL of P2 virus for WT and 3 mL of P2 virus for mutants respectively. Infected *Sf9* cells were harvested 70 h after infection by centrifugation (900 x g, 20 min). The cell pellet was resuspended in 30 mL/ L

lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% (v/v) glycerol and 1 mM TCEP), snap frozen in liquid nitrogen, then stored at  $-80^{\circ}\text{C}$  for later use.

Thawed cell aliquots were lysed by sonication. The lysates were clarified by centrifugation (40,000 g, 30 min), then the supernatant was passed through a  $0.80\ \mu\text{m}$  filter (Millipore) and loaded onto an equilibrated (lysis buffer) immobilised metal affinity chromatography column (IMAC) (Ni-NTA Superflow Cartridge, Qiagen). The immobilised protein was washed with lysis buffer, then eluted using a linear gradient of elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 300 mM imidazole, 5% v/v glycerol, and 1 mM TCEP). The protein containing fractions were pooled and passed through an ion exchange column (HiTrap<sup>®</sup> SP FF GE Healthcare Life Sciences) pre-equilibrated in the SP buffer A (25 mM HEPES pH 7.5, 300 mM NaCl, 5% (v/v) glycerol and 1 mM TCEP). The protein was eluted using a linear gradient of SP buffer B (SP buffer A with 1 M NaCl), and fractions containing the tag-free Artemis were identified by electrophoresis.

Artemis containing fractions were pooled and dialysed overnight at  $4^{\circ}\text{C}$  in SP buffer A and supplemented with recombinant tobacco etch virus (TEV) protease for cleavage of the  $_{6}\text{His-ZB}$  tag. The protein was subsequently loaded into an ion exchange column (HiTrap<sup>®</sup> SP FF GE Healthcare Life Sciences), pre-equilibrated in the SP buffer A to remove  $_{6}\text{His-ZB}$  tag and uncleaved protein. The protein was eluted using a linear gradient of SP buffer B, and fractions containing the tag-free Artemis were identified by electrophoresis. Artemis-containing fractions from the SP column elution were combined and concentrated to 1 mL using a 30 kDa MWCO centrifugal concentrator. The protein was then loaded on to a Superdex 75 increase 10/300 GL equilibrated with SEC buffer (25 mM HEPES pH 7.5, 300 mM NaCl, 5% (v/v) glycerol, 2 mM TCEP).

Mass spectrometric analysis of the purified proteins revealed masses of 41716.5 Da, 41650.5 Da, 41672.2 Da, 41639.9 Da for WT, H35A, D37A and H35D proteins, respectively. The calculated masses are 41715.09, 41649.2, 41671.2 and 41639.2, respectively, all within 1.5 Da of the measured masses.

### **Expression and purification of WT truncated Artemis catalytic domain without IMAC (aa 1-362)**

The truncated Artemis protein was expressed and purified in a similar manner as described above except for the first purification step. We used 5 mL HiTrap<sup>®</sup> SP Fast Flow (GE Health Care) column as the first step of purification. Following an overnight TEV cleavage the protein was subjected to a second ion exchange step (5 mL HiTrap<sup>®</sup> SP Fast Flow (GE Health Care)) for the removal of the Z-Basic protein tag. The protein was further purified by size exclusion chromatography (Highload<sup>®</sup> 16/200 Superdex<sup>®</sup> 200).

### **Cloning, expression and purification of full-length WT Artemis (aa 1-692)**

The full-length Artemis encoding construct was cloned into pFB-CT10HF-LIC, a baculovirus expression vector containing a C-terminal His<sub>10</sub> and FLAG tag. pFB-CT10HF-LIC was a gift from Nicola Burgess-Brown (Addgene plasmid # 39191; <http://n2t.net/addgene:39191>; RRID: Addgene\_39191). As for the truncated protein, the full-length Artemis gene was also cloned using ligation independent cloning (LIC) [35].

The baculovirus mediated expression of the full length DCLRE1C/ Artemis gene was performed in a manner similar to that used for the truncated protein. However, instead of infection with 1.5 mL of P2 Virus, 3.0 mL of P2 virus was used to infect *Sf9* cells at 2 x 10<sup>6</sup> cells/ mL for the expression of the full-length Artemis construct.

Cell harvesting and the initial IMAC purification steps were performed as described for the catalytic domain. Following IMAC chromatographic purification, TEV cleavage overnight in dialysis buffer (50 mM HEPES pH 7.5, 0.5 M NaCl, 5% glycerol and 1 mM TCEP) gave protein which was then passed through a 5 mL Ni-sepharose column; the flowthrough fractions were collected. The Artemis protein was then concentrated using a centrifugal concentrator (Centricon, MWCO 30 kDa) before loading on a Superdex S200 HR 16/60 gel filtration column in dialysis buffer. Fractions containing purified Artemis protein were pooled and concentrated to 10 mg/mL.

### **Electrospray mass spectrometry (ESI-QTOF)**

Reversed-phase chromatography was performed in-line prior to mass spectrometry using an Agilent 1290 uHPLC system (Agilent Technologies inc. – Palo Alto, CA, USA). Concentrated protein samples were diluted to 0.02 mg/ml in 0.1% formic acid and 50 µl was injected on to a 2.1 mm x 12.5 mm *Zorbax* 5µm 300SB-C3 guard column housed in a column oven set at 40 °C. The solvent system used consisted of 0.1% formic acid in ultra-high purity water (Millipore) (solvent A) and 0.1 % formic acid in methanol (LC-MS grade, Chromasolve) (solvent B). Chromatography was performed as follows: Initial conditions were 90 % A and 10 % B and a flow rate of 1.0 ml/min. A linear gradient from 10 % B to 80 % B was applied over 35 seconds. Elution then proceeded isocratically at 95 % B for 40 seconds followed by equilibration at initial conditions for a further 15 seconds. Protein intact mass was determined using a 6530 electrospray ionisation quadrupole time-of-flight mass spectrometer (Agilent Technologies Inc. – Palo Alto, CA, USA). The instrument was configured with the standard ESI source and operated in positive ion mode. The ion source was operated with the capillary voltage at 4000 V, nebulizer pressure at 60 psig, drying gas at 350°C and drying gas flow rate at 12 L/min. The instrument ion optic voltages were as follows: fragmentor 250 V, skimmer 60 V and octopole RF 250 V.

### **Protein crystallisation and Soaking**

Artemis (PDB:6TT5) was crystallised using the sitting drop vapour diffusion method by mixing 50 nL protein with 50 nL crystallisation solution comprising 0.2 M ammonium chloride, 20% (v/v) PEG 3350. Crystals grew after 2 weeks and reached maximum size within 3 weeks. An unliganded crystal was flash frozen in liquid nitrogen, cryoprotected with the mother liquor supplemented with 20% (v/v) ethylene glycol solution.

The non-IMAC purified Artemis (PDB:7AF1) was crystallised in a similar manner, with the addition of 20 nL of crystal seed solution obtained from previous crystallisation experiment. The crystals were grown in a solution comprising 0.25 M ammonium chloride and 30% (v/v) PEG 3350 at 4°C. Crystals grew after one day and reached a maximum size within one week.

Artemis variants (mutants H33A and H35D) were crystallised using the sitting drop vapour diffusion method by mixing 50 nL protein with 50 nL crystallisation solution comprising 0.1 M sodium citrate pH 5.5, 20% PEG 3350, while the D37A was crystallised in 0.2 M ammonium acetate, 0.1 M bis-TRIS pH 5.5, 25% PEG 3350. All Artemis variants were crystallised in the presence of 20 nL of crystal seed solution obtained from previous crystallisation experiment. Crystals grew after one day at 4°C. and reached maximum size within one week

#### **Data collection and refinements**

Data were collected at Diamond Light Source I04, I03, or I24 beamlines. Diffraction data were processed using DIALS [37] and structures were solved by molecular replacement using PHASER [38] and the PDB coordinates 5Q2A. Model building and the addition of water molecules were performed in COOT [39] and structures refined using REFMAC [40]. Data collection and refinement statistics are given in Table I. The X-ray fluorescence data was collected at Diamond Light Source I03 (6TT5) using 100% transmission and 12.7 eV, and I24 (7AF1) using 1% transmission and 12.8 eV (Suppl. Figure 1).

#### **Generation of 3'-radiolabelled substrates**

10 pmol of single-stranded DNA (Eurofins MWG Operon, Germany) were labelled with 3.3 pmol of  $\alpha$ -<sup>32</sup>P-dATP (Perkin Elmer) by incubation with terminal deoxynucleotidyl transferase (TdT, 20 U; ThermoFisher Scientific), at 37°C for 1 hour. This solution was then passed through a P6 Micro Bio-Spin chromatography column (BioRad), and the radiolabeled DNA was annealed with the appropriate unlabeled oligonucleotides (1:1.5 molar ratio of labelled to unlabeled oligonucleotide) (Supplementary Table 1 for sequences) by heating to 95°C for 5 min, and cooling to below 30°C in annealing buffer (10 mM Tris-HCl; pH 7.5, 100 mM NaCl, 0.1 mM EDTA).

#### **Gel-based nuclease assays**

Standard nuclease assays were carried out in reactions containing 20 mM HEPES-KOH, pH 7.5, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 0.05% (v/v) Triton X-100, 5% (v/v) glycerol (final volume:

10  $\mu$ L), and the indicated concentrations of Artemis. Reactions were started by the addition of DNA substrate (10 nM), incubated at 37°C for the indicated time, then quenched by addition of 10  $\mu$ L stop solution (95% formamide, 10 mM EDTA, 0.25% (v/v) xylene cyanole, 0.25% (v/v) bromophenol blue) with incubation at 95 °C for 3 min.

Reaction products were analysed by 20% denaturing polyacrylamide gel electrophoresis (made from 40% solution of 19:1 acrylamide:bis-acrylamide, BioRad) and 7 M urea (Sigma Aldrich)) in 1 x TBE (Tris-borate EDTA) buffer. Electrophoresis was carried out at 700 V for 75 minutes; gels were subsequently fixed for 40 minutes in a 50% methanol, 10% acetic acid solution, and dried at 80°C for two hours under a vacuum. Dried gels were exposed to a Kodak phosphor imager screen and scanned using a Typhoon 9500 instrument (GE).

### **Fluorescence-based nuclease assay.**

The protocol of Lee *et al* [41] was adapted for structure-specific endonuclease activity. A ssDNA substrate was utilised containing a 5' FITC-conjugated T and a 3' BHQ-1 (black hole quencher)-conjugated T (Suppl. Table 1). As the FITC and BHQ-1 are located proximal to one another, prior to endonucleolytic incision, the intact substrate does not fluoresce. Following endonucleolytic incision by DCLRE1C/Artemis, there is uncoupling of the FITC from the BHQ-1 and an increase in fluorescence. Inhibitors (at increasing concentrations) were incubated with Artemis for 10 minutes at room temperature, before the reaction was started with the addition of DNA substrate. Assays were carried out in a 384-well format, in a 25  $\mu$ L reaction volume. The buffer was the same as for the gel-based nuclease assays, Artemis concentration was 50 nM, and the DNA substrate was at 25 nM. Fluorescence spectra were measured using a PHERAstar FSX (excitation: 495 nm; emission: 525 nm) with readings taken every 150 sec, for 35 min, at 37 °C.

## **RESULTS**

### **Human Artemis (SNM1C or DCLRE1C) has a core catalytic fold similar to SNM1A and Apollo/SNM1B**

The core catalytic domain of Artemis (aa 3–361) was produced in baculovirus-infected *Sf9* cells fused to a highly basic His<sub>6</sub>-Zb tag, which confers tight binding to cation exchange columns. The protein was purified using immobilised metal affinity chromatography (IMAC) on a Nickel-Sepharose column as the initial step. Subsequent preparations were performed without the use of IMAC, to avoid the introduction of Ni<sup>2+</sup> ions during purification.

Artemis protein was purified as detailed in the Methods & Materials, and crystals were subsequently grown and diffracted to 1.6 Å resolution (Table 1); the structure was solved using a structure of SNM1A (PDB coordinates 5Q2A) as the molecular replacement model. The resultant Artemis structure (PDB coordinates 7AF1) contains a single molecule in the

asymmetric unit, with two zinc ions coordinated at the active site. The metal ions were identified using X-ray fluorescence (XRF) analysis during data collection at the Diamond Light source. When using the protein purified using IMAC, the first zinc ion in the active site can be replaced by a nickel ion (PDB: 6TT5) (Figure 2E). The presence of the nickel ion was also confirmed in the crystal by XRF. The X-ray fluorescence analysis of the metal ions present in the structures are shown in Suppl. Figure 1. This metal ion coordination pattern has been observed with other member of the family, such as SNM1A and SNM1B/Apollo [3,9].

The overall fold of human Artemis protein catalytic core is very similar to that of human SNM1A and SNM1B/Apollo (2 Å RMSD). It has the key structural characteristics of human MBL fold nucleases, with the di-metal containing active site interfaced between the MBL and  $\beta$ -CASP domains (Figure 1A). As anticipated, the MBL domain (Figure 1A and 1B, in pink) of Artemis has the typical  $\alpha/\beta$ - $\beta/\alpha$  sandwich MBL fold [42] and contains all of the highly conserved motifs 1–4 (Figure 1C and D; and sequence alignment, Suppl. Figure 2) which are typical for the whole MBL superfamily, and motif A–C which are typical of the  $\beta$ -CASP fold containing family [2,27,43,44].

Motifs 1–4 (Figures 1D and 2C) are responsible for metal ion coordination in both DNA and RNA processing MBL enzymes [44]. As previously observed in crystal structures of human SNM1A and Apollo/SNM1B, Artemis can coordinate one or two metal ions in its active site. One zinc ion (Zn1) in the active site is coordinated by four residues (His33, His35, His115, and Asp116) and two water molecules (H<sub>2</sub>O 506 and 611) in an octahedral manner (Figure 2C). The second zinc ion (Zn2) was refined with 30% occupancy and is coordinated by three residues (Asp37, His38, and Asp136) and two water molecules (H<sub>2</sub>O 506 and 529). The low occupancy of the second zinc ion, together with the two conformations (0.5 occupancy for each conformation) observed for Asp37 (Figure 2E) suggest that this site binds a metal ion less tightly than the Zn1 site, consistent with studies on other human MBL fold nucleases [3, Baddock *et. al.*, 2020]

The structure of human SNM1A (PDB: 5AHR) [3] was solved with a single zinc ion coordinated in the active site (Figure 2A). By contrast, SNM1B structures solved with a bound AMP (Baddock *et. al.*, accompanying paper) showed that both metal ions are positioned to coordinate the phosphate group of the AMP in an octahedral manner (Figure 2B). In summary, the octahedral coordination sites for the first zinc ion are contributed to by three histidines, one aspartate residue, and either water molecules or a phosphate oxygen of the substrate; the second metal ion is more weakly coordinated in the SNM1 protein family, with one histidine and two aspartates, with the remaining three positions occupied by water or a phosphate oxygen of the DNA substrate. This can explain the partial occupancy of the second zinc position in SNM1 enzyme structures, where the full occupancy may be achieved only in presence of substrate. We propose that Artemis would



coordinate a phosphate group of its substrate in a similar manner. The structure of human CPSF-73 (PDB: 2I7T), an RNA processing nuclease, with two active site bound zinc ions and a phosphate molecule [45], shows that the two zinc ions are coordinated in a very similar geometry with the human MBL DNA processing enzymes. [46,47]. A striking difference between the MBL RNA and DNA nucleases is that the second metal ion (M2) in the RNA processing nucleases is coordinated by an additional histidine residue (His418 for CPSF-73) [45] that is absent in the DNA processing enzymes (Figure 2D).

### **The structure of Artemis reveals a novel zinc-finger like motif in the $\beta$ -CASP domain.**

Proteins with a  $\beta$ -CASP fold form a distinct sub-group within the MBL-fold superfamily that specifically act on nucleic acids [2]. Artemis'  $\beta$ -CASP domain is comprised of residues 156–384 and it is the second globular domain (Figure 1A, shown in white) in the catalytic region; inserted within the Artemis MBL fold sequence between the small  $\alpha$ -helices 6 and 7 (figure 1 B). The  $\beta$ -CASP domain has been proposed to facilitate substrate recognition and binding in the nucleic acid processing MBL fold-containing family of enzymes [2,48].

Another metal ion coordination site, unique to Artemis, is present in the  $\beta$ -CASP domain, with similarity to the canonical Cys<sub>2</sub> His<sub>2</sub> zinc-finger motif [49,50]. Many DNA binding proteins, including transcription factors and a substantial number of DNA repair factors (including those involved in NHEJ), possess the classical Cys<sub>2</sub> His<sub>2</sub> zinc finger motif, that serves as a structural feature stabilising the DNA binding domain [49–52]. A typical Cys<sub>2</sub> His<sub>2</sub> zinc coordinating finger (Figure 3A) has a  $\beta\beta\alpha$  motif, wherein the zinc ion is coordinated between an  $\alpha$ -helix and two antiparallel  $\beta$ -sheets. The zinc ion confers structural stability and hydrophobic residues located at the sides of the zinc coordination site enable specific binding of the zinc finger in the major groove of the DNA [49,50,53,54]. Similar to the canonical Cys<sub>2</sub> His<sub>2</sub> zinc finger motif, the zinc ion coordination in Artemis'  $\beta$ -CASP domain adopts a tetrahedral geometry, with coordination by two cysteine (Cys256 and Cys272) and two histidine (His228 and His254) residues (Figure 3 B). However, in the case of Artemis the metal ion coordination site is sandwiched between two  $\beta$ -sheets instead of an  $\alpha$ -helix and two antiparallel  $\beta$ -sheets.

Almost all the residues in the zinc-finger like motif (His228, Cys256, and Cys272) are unique to Artemis (sequence alignment Suppl. Figure 2), with only His254 being well conserved within the SNM1-family. However, these four residues that forms the zinc-finger like motif are highly conserved in Artemis across different species (from human to marine sponge), implying functional importance (sequence alignment Suppl. Figure 3). Consistently, substitution of His228 and His254 (H228N and H254L), two of the zinc coordinating residues in the  $\beta$ -CASP domain of Artemis, cause RS-SCID in humans [34,44,55]. Patients with these inherited mutations suffer from impaired V(D)J recombination, leading to underdeveloped B and T lymphocytes. The importance of histidine 254, has been highlighted by de Villartay *et al.* [44], who showed that the full-

length H254A Artemis variant is unable to carry out V(D)J recombination *in vivo* and has no discernible endonucleolytic activity *in vitro*.

### **Comparison of Artemis structure with 6W00 and 6WNL**

During the preparation of this manuscript a structural study on the catalytic core of Artemis was published [56]. This study described reported two Artemis structures (PDB: 6W00 and 6WNL) that are similar to our Artemis structure (PDB code 7AF1) (backbone RMSDs of 0.48 Å and 0.54 Å respectively), with identical relative positioning of the MBL and  $\beta$ -CASP domains (Figure 4A and Suppl. Figure 4A). The only significant difference was that whilst we refined our structure with two zinc ions in the active site, both of the crystal forms reported by Karim *et al.* were modelled with a single active site zinc ion (Zn1), reinforcing the proposal of weaker metal ion binding at the Zn2 site.

### **Re-analysis of the 6W00 and 6WNL structures**

An unusual aspect of the Karim *et al.* structures is that both crystal forms were obtained in the presence of DNA and were reported to require DNA for their growth; the crystals showed a fluorescence signal supporting the presence of DNA (the oligonucleotides used contained a cyanine dye fluorophore), yet neither of the models presented contain DNA. The authors referred to some broken stacking electron density in 6WNL in a solvent channel and a patch of unsolved density approaching the active site in 6W00, but state that the DNA “did not bind to the protein in a physiological way, and likely bound promiscuously to promote crystallization” [56].

We performed a careful re-examination of these structures looking closely at the residual electron density. For the 6WNL structure we were able to locate a distorted duplex DNA of around 13 base pairs which we propose may be the product of duplex annealing of the oligonucleotide used for crystallization (a semi-palindromic 13-mer that was designed to form a hairpin with phospho-thioate linkages in the single-stranded region) (Suppl. Figure 4B). For this structure, we are in general agreement with Karim *et al* that the DNA does not appear to make meaningful interactions with the protein that inform on the mechanism of nuclease activity, although this mode of association with DNA may possibly be relevant to alternative binding modes relating to higher order complexes containing Artemis. By contrast, for the 6W00 structure we were able to confidently build a DNA molecule that contacts the Artemis active site in a manner that we believe to be relevant to the Artemis nuclease activity.

Our model contains an 8-nucleotide 5'-single-stranded extension with a short 2-base pair region of duplex DNA that reaches into the Artemis active site making close contacts with the metal ion centre in a manner consistent with the proposed catalytic mechanism (Figure 4B). The sequence of the longest strand corresponds to the 10-nucleotide cy-5 labelled strand (cy5-GCGATCAGCT) with some residual density at the 5'-end that may be attributed to the cyanine fluorophore which we did not include in our model. The complementary

strand used for crystallization was 13-nucleotides long and was intended to produce a 5'-overhang, but only two bases and three phosphates could be located in the density. The abrupt manner in which the electron density apparently disappears from either end of this strand suggests that this is the product of a cleavage reaction, although it is possible that remaining nucleotides are not located due to disorder.

The analysis of electron density at this site is complicated by the proximity to a crystallographic 2-fold symmetry axis, which brings a symmetry copy of the DNA molecule into a position where atoms partially overlap and the extended 5' strands form a pseudo duplex (Suppl. Figure 5A). The occupancy of the entire DNA molecule is thus limited to 0.5, and the lower occupancy is reflected in the electron density map which requires a lower contour level than would usually be applied (Suppl. Figure 5B). After carefully building and refining the afore-described DNA bound model, significant positive electron density was revealed for the second metal ion (Zn<sup>2+</sup> site) which we also included in the model with the same occupancy (0.5) as the DNA. Our model was refined to similar crystallographic R-factors as 6W00 and has been deposited with PDB accession number 7ABS (refinement statistics are given in Table I).

### **Model for Artemis DNA binding**

Using the crystallographically observed DNA as a template we were able to make a model for Artemis binding to a longer section of double-stranded DNA by complementing unpaired bases on the single-stranded DNA overhang with canonical base pairs, whilst maintaining acceptable geometry of the sugar phosphate backbone (Figure 5A). The duplex section of this model deviates slightly from the ideal B-form geometry [57], in a manner that is reminiscent of certain transcription factor DNA complexes [58,59]. We have also extended the metal ion contacting strand by three nucleotides to form a 5'-overhang; the positioning of the overhang nucleotides is more speculative, nevertheless it was possible to avoid clashes with protein residues whilst maintaining relaxed geometry (Figure 6C).

In the extended DNA complex model, Artemis contacts both strands of the DNA model in several areas; notably a single phosphate lies above the di-metal ion bearing active site and ligates to both metal ions in the same manner as observed in structures of related enzymes with phosphate or phosphate-containing compounds (Baddock *et.al.* 2020) [60]. The two downstream nucleotides on this strand pass close to the protein surface, forming possible interactions with both the main chain of Asp37 and sidechain of Lys36, whilst subsequent nucleotides are not close to the protein (Figure 6A). The overhang portion of this strand continues with a slightly altered trajectory, potentially contacting Artemis in the vicinity of the cleft separating the MBL and  $\beta$ -CASP domains, with the potential to form favourable interactions with both positively charged (Arg 172) and aromatic residues (Phe 173 and Trp 293) (Figure 6C).

The complementary strand forms interactions with the protein *via* backbone contacts that span a 4-nucleotide stretch between 5- and 8-bases from the 3'-terminus that contact positively charged sidechains in the MBL domain (Lys36, Lys40, Arg43, and Lys74) (Figure 6A). The 3'-end of this strand apparently terminates directly above a cluster of polar or positively charged residues in the  $\beta$ -CASP domain (Lys207, Lys288, Asn205) (Figure 6B). Whilst the experimental (as used in co-crystallisation) DNA substrate and our model both contain a 3'-hydroxyl group, the model implies that addition of a 3'-phosphate could be accommodated and may be expected to make favourable interactions with the basic cluster of residues. Thus, our model illustrates a preferred binding mode for Artemis for DNA with a 5'-overhang binding at the junction between double- and single-stranded regions, and the expected product of this reaction would be a blunt ended DNA with a 5'-phosphate. In the case of hairpin DNA substrates our model indicates the possibility for Artemis to accommodate a loop connecting the two strands possibly of around 4-nucleotides or more, with the cleavage product being DNA with a 3'-overhang cleaved from the last paired base of the duplex.

### **Comparison of the Artemis DNA binding mode with that of other nucleases**

We have recently determined the structure of SNM1B/Apollo in complex with two deoxyadenosine monophosphate nucleotides and through a similar process of extrapolation to that outlined above we have independently built a model for SNM1B binding to DNA containing a 3'-overhang (one of its preferred substrates) (Baddock *et al.* accompanying paper). The overall mode of DNA binding is similar in the two models (Figure 5), with the two DNA duplexes being roughly parallel and forming contacts to similar regions on the MBL domain. The most important differences lie in the nature of the contacts formed to the active site and the paths of the various overhangs. In the SNM1B model extensive contacts are made to the 5'-phosphate in a well-defined phosphate binding pocket. Both human SNM1A and SNM1B are exclusively 5'-phosphate exonucleases, with most of these phosphate binding residues being highly conserved (sequence alignment Suppl. Figure 1 in yellow) (Baddock *et al.* accompanying paper). Interestingly, Artemis lacks these key phosphate binding residues and the 5'-phosphate binding pocket of SNM1A and SNM1B. Instead, this pocket in Artemis is partially filled by the side chain of Phe318. These contacts appear to define a high-affinity binding pocket exclusively for the 5'-phosphate of the DNA terminus, thus explaining the major differences in nuclease activities within the family, i.e., SNM1A and SNM1B being exonucleases and Artemis being an endonuclease. Further differences between Artemis and SNM1B/SNM1A are found in the loop connecting  $\beta$ -strands G and H (using Artemis numbering), which in Artemis is significantly longer and occupies a different position contacting residues in the MBL domain (Figure 6B), compared to the loops in SNM1B and SNM1A that form part of the phosphate binding pocket and make potential contacts with the 3'-overhang. This loop displacement in Artemis may contribute to its ability in accommodating DNA substrates

with either 5'-overhangs or hairpins, thus facilitating its function as a structure specific endonuclease.

Interestingly, the surface of the Artemis interface between the MBL and  $\beta$ -CASP domains contains a belt of positively charged residues (Figure 5A). These positive surface charges are proposed to facilitate productive DNA binding to the active site, consistent with our DNA binding model. One of the striking differences, at least in the available structures, is that the active site of Artemis is more open compared to those of SNM1A or SNM1B. This openness may reflect an ability to accommodate different substrate conformation including hairpins, 3'- and 5'-overhangs, as well as DNA flaps and gaps. Both human SNM1A and SNM1B appear to have a more sequestered active site that would only fit a single strand of DNA, which is consistent with previous findings on their preferred substrate selectivity [10].

### **Biochemical characterisation of truncated Artemis catalytic domain (aa 1-361)**

To investigate the activity of our different versions of recombinant Artemis, we performed nuclease assays using radiolabelled DNA substrates. We compared the catalytic domain purified using IMAC (which contained  $\text{Ni}^{2+}$  in the active site) with protein purified using ion exchange (and avoiding IMAC), which contained predominantly  $\text{Zn}^{2+}$ . We also tested the activity of full-length phosphorylated Artemis. The results show that both truncated enzymes have identical activities, which is also very similar to that of the full-length enzyme (Suppl. Figure 6).

One notable difference between our full-length protein and that reported by Ma *et al.* [18], is that our full-length protein is active in the absence of DNA-PKcs. Intact protein mass spectrometric analysis of our full-length protein shows that the protein has undergone up to five phosphorylation events (Suppl. Figure 7). Poinsignon *et al.* have shown that Artemis is constitutively phosphorylated in cultured mammalian cells and is the target of additional phosphorylation in response to induced DNA damage [61]; it is interesting that the capacity to phosphorylate Artemis to produce an active form is also conserved in insect cells. We observed exonuclease activity with full-length Artemis at 10 nM (Suppl. Figure 8), though this was weak compared to its endonuclease activity at the same concentration. We observed no exonuclease activity for the truncated Artemis construct, it is possible that phosphorylation alters the balance between endonuclease and exonuclease activity, though the biological relevance of this, if any, remains to be validated.

As mentioned above, both human SNM1A and SNM1B require a 5'-phosphate for their activity [3,9,15]. To investigate whether there is a similar requirement for Artemis, we tested the activity of truncated Artemis against single-stranded and overhang DNA substrates with different 5'-end groups, including a phosphate, hydroxyl group, and biotin groups (Figure 7A). The results imply that, at least under the tested conditions, Artemis is

agnostic to the different end modifications, exhibiting comparable digestion of all substrates.

Extensive evidence demonstrates that full-length Artemis in complex with DNA-PKcs has structure specific endonuclease activity [17,18,27]. These studies reported that Artemis can digest substrates including overhangs, hairpins, stem-loops, and splayed arms (pseudo-Y). To investigate the activity of truncated Artemis catalytic domain (aa 1–361) we performed nuclease assays using a variety of radio-labelled DNA substrates (Figure 7B). The results show that truncated Artemis has substrate specific endonuclease activity, with a preference for single-stranded DNA substrates, and those that contain single stranded character (e.g. 5'- and 3'-overhangs, splayed arms, and a lagging flap structure), compared with double stranded DNA structures (e.g. ds DNA and a replication fork). This is in accordance with previous research, where Artemis has been reported to cleave around ss- to dsDNA junctions in DNA substrates (perhaps cite Chang et al, 2015 for this). The truncated Artemis catalytic domain also exhibits hairpin opening activity, in accordance with what has previously been reported (Suppl. Fig. 9). On a duplex substrate (YM117 from Ma *et al*) [18] with a 20 nt hairpin region, Artemis cleaves adjacent to the hairpin, consistent with previous data. It is clear that truncated form of Artemis exhibits nuclease activity closely comparable to the phosphorylated full-length Artemis protein [18], indicating that the structural studies presented here reveal mechanistic insights of direct relevance to the DNA-PKcs-associated form of Artemis that engages in end-processing reactions *in vivo*.

### **Structural and biochemical characterisation of Artemis point mutations**

Previous site-directed mutagenesis studies by Pannicke *et al.* targeting the metal ion coordinating residues in the active site (D17N, H33A, H35A, D37N) of full-length Artemis (aa 1–692) established the importance of active site motifs 1–4 for activity [27]. Each of these substitutions markedly reduced or abolished Artemis' ability to carry out its role in V(D)J recombination *in vitro*.

We mutated, expressed, purified, and crystallised three forms of truncated Artemis (aa 1–361) with substitutions in several of these metal ion co-ordinating residues, i.e. D37A, H33A, and the Omenn Syndrome patient mutation H35D [31,55]. We found that the overall architecture of the three variants is almost identical to the WT (Figure 8A). The D37A structure retains one Zn ion (Zn1) whilst losing the second, (Zn2) (Figure 8B) in the active site. Both the H33A and H35D variants additionally exhibited loss of the Zn1 ion. All three variants retained the Zn ion in the zinc finger-like motif of the  $\beta$ -CASP domain.

The position of the active site residues and the surrounding residues in the D37A variant, superposed perfectly with WT Artemis. As previously mentioned, Asp37 can adopt two conformations as seen in 6TT5 structure, noting that the coordinated zinc ion is generally present at about 30% occupancy in both of the WT Artemis structures (PDB 6TT5 and 7AF1).

Therefore, it is unsurprising that mutation of Asp37 to alanine results in the loss of Zn<sup>2+</sup>. Differential scanning fluorimetry (DSF) experiments were carried out to investigate the stability of the three variants. DSF analysis showed that the D37A variant has similar thermal stability as the WT, suggesting that the protein is stable and folded in the presence of a single metal ion in the active site, whilst both H33A and H35D variants were substantially destabilized with  $\Delta T_m$  around -13 °C compared to the WT Artemis (Suppl. Figure 10).

Histidines 33 and 35 are the first two histidine residues in the HxHxDH motif (motif 2) in the MBL domain. Their role is to coordinate the first metal ion (Zn<sup>1</sup> site) in the catalytic site. In the absence of metal ions in the catalytic site, the loop comprising residues 113–119 moves away from the active site (Figure 8C and D). Another small rearrangement occurs in helix  $\alpha 8$  (residues 348–358) of the MBL domain. In both the H33A and H35D variants, helix  $\alpha 8$  moves slightly closer toward strand  $\beta 14$ , compared to the WT and D37A variant. Surprisingly, the biggest rearrangement occurs in  $\beta$ -strand E (residues 268–270) and  $\alpha$ -helix E (residues 261–267); both located near the zinc finger motif in the  $\beta$ -CASP domain (Figure 8A). In H33A and H35D variants, both  $\beta$ -strand E and  $\alpha$ -helix E shifted upward and away from the zinc finger like motif. These conformational changes may suggest some allosteric regulation in terms of substrate binding and catalytic activity of the enzyme.

We also tested the activity of the D37A, H33A, and H35D Artemis variants *in vitro* using single-stranded 3' end radiolabelled DNA as a substrate in a gel-based assay. All three variants lost their ability to digest the DNA substrate *in vitro* (Figure 9). These observations are in agreement with the results obtained with full-length variants by Ege *et al.* and Pannicke *et al.* [27,31]. Their studies show that the full-length Artemis variants H33A, H35D, and D37N are able to interact with and be phosphorylated by DNA-PKcs, however, have lost the ability to digest DNA substrates *in vitro*. The combined results reveal the importance of the HxHxDH motif and highlight the importance of the di-metal catalytic core in the SNM1 family, not only in directly catalysing hydrolysis, but also likely in conformational changes involved in catalysis [43].

### Identification of small molecule inhibitors of Artemis

Radiotherapy is a mainstay of cancer therapy; its effectiveness relies on inducing DNA double-strand breaks (DSBs) that contain complex, chemically modified ends that must be processed prior to repair [62,63]. The canonical non-homologous end-joining (c-NHEJ) pathway repairs ~80% of DNA double-strand breaks in mammalian cells [20,23]. Therefore, combining radiation therapy in conjunction with c-NHEJ inhibitors could selectively radiosensitise tumours. Weterings *et al.* have reported a compound that interferes with the binding of Ku70/80 to DNA, thereby increasing sensitivity to ionising radiation in human cell lines [64]; ATM inhibitors are also in advanced clinical development and represent the most developed strategy to inhibit DSB repair to increase the efficacy of radiotherapy [65]

Artemis, along with SNM1A and SNM1B, possess a conserved MBL-fold domain that is similar to the true bacterial MBLs. Previous studies on human SNM1A and SNM1B/Apollo, showed that ceftriaxone (Rocephin), a widely used  $\beta$ -lactam antibacterial (third generation cephalosporin) inhibits the nuclease activity of both SNM1A and SNM1B [66]. To investigate if this class of  $\beta$ -lactam anti-bacterial compounds could inhibit Artemis we performed fluorescence-based nuclease assays with three cephalosporins, i.e. ceftriaxone, cefotaxime and 7-aminocephalosporanic acid (Figure 10A). The results show that neither cefotaxime nor the parent compound, 7-aminocephalosporanic acid, potently inhibit Artemis' activity, whilst ceftriaxone inhibits Artemis with a modest  $IC_{50}$  of 65  $\mu$ M (Figure 10B).

We solved the structure of ceftriaxone bound to the catalytic domain of Artemis (purified by IMAC) at 1.9 Å resolution (Figure 10C) by soaking an Artemis crystal with ceftriaxone. This structure was solved by molecular replacement (using PDB: 6TT5 as a model), in the space group P1 with one protein molecule in the asymmetric unit. As before, in this structure Artemis possesses the canonical bilobar MBL and  $\beta$ -CASP fold with an active site containing one nickel ion, possibly related to the purification method.

Ceftriaxone binds to the protein surface in an extended manner making interactions with the active site, towards the  $\beta$ -CASP domain (Figure 10C). There is no evidence for cleavage of the  $\beta$ -lactam ring nor of loss of the C-3' cephalosporin side chain, reactions that can occur during 'true' MBL catalysed cephalosporin hydrolysis. The electron density at the active site clearly reveals the presence of the ceftriaxone side chain in a position to coordinate the nickel ion (at the Zn1 site) replacing water molecules (waters 72 and 106) compared with the apo structure (Figure 10D). Despite the conservation of key elements of the active site of the MBL fold nucleases and the 'true'  $\beta$ -lactam hydrolysing MBLs [67], ceftriaxone, does not interact with the nickel ion *via* its  $\beta$ -lactam ring (as occurs for the true MBLs), but *via* both carbonyl oxygens of the cyclic 1, 2 diamide in its sidechain (Ni-O distances: 2 Å and 2.2 Å), i.e. it is not positioned for productive  $\beta$ -lactam hydrolysis. The amino-thiazole group (N7) of ceftriaxone forms hydrogen bonds with the side chain of Asn205, while the S1 of the 7-aminocephalosporanic acid core of the compound interacts with the hydroxyl of Tyr212 through an ethylene glycol molecule. The rest of the molecule appears to be flexible. The binding mode of ceftriaxone to Artemis shown in Figure 10C is near identical to that observed for ceftriaxone with SNM1A (PDB: 5NZW) structure (Suppl. Figure 11).

One notable difference between the Ceftriaxone-bound Artemis structure and the apo structure, is the loss of a second metal ion at the active site (Figure 10D). In the apo structure (Figure 2C), this zinc ion is coordinated by residues Asp37, His38, and Asp130. With a single metal coordination in the ceftriaxone bound structure, the Asp37 side chain is positioned away from the active site (Figure 10D), as seen in the nickel bound (PDB: 6TT5) structure (Figure 2E).



To investigate the possibility of inhibiting Artemis through binding to the zinc finger motif in the  $\beta$ -CASP domain, we used the fluorescence-based nuclease assay to test three compounds known to react with thiol groups present in zinc fingers and which result in zinc ion displacement, i.e. ebselen [68], auranofin [69] and disulfiram [70]. We found that both ebselen and disulfiram inhibit Artemis with  $IC_{50}$  values around 8.5  $\mu$ M and 10.8  $\mu$ M respectively, whilst Auranofin inhibits less potently ( $IC_{50}$  46  $\mu$ M) (Suppl. Figure 12), indicating additional possible inhibitory strategies.

## DISCUSSION

The DCLRE1C/Artemis gene was first discovered in 2000, following work with children with severe combined immunodeficiency disease (SCID) [71]. Subsequent studies have shown that Artemis is a key enzyme in V(D)J recombination [16,17,72] and the c-NHEJ DNA repair pathway [21,44,73]; and that it is structure specific endonuclease, and member of MBL fold structural superfamily [2]. Our structures of wild-type and catalytic site mutants of SNM1C/DCLRE1C or Artemis protein show that, like SNM1A and SNM1B/Apollo and the RNA processing enzyme CPSF73, Artemis has a typical  $\alpha/\beta$ - $\beta/\alpha$  sandwich fold in its MBL domain and has a  $\beta$ -CASP domain, the latter a characteristic feature of MBL fold nucleases. However, both our Artemis structures and those recently reported by Karim *et. al* [56] reveal a unique structural feature of Artemis in its  $\beta$ -CASP domain that is not reported in other human MBL enzymes, i.e. a classical zinc-finger like motif. Moreover, collectively, these structures allow us to assign a likely mode of DNA substrate interaction for Artemis.

The role of the newly-described zinc-finger like motif remains unknown. However, zinc-finger motifs are common structural features in DNA binding proteins such as transcription factors [50,54], but are also observed in a substantial number of required and accessory NHEJ proteins [52]. These zinc fingers provide structural stability and enhance substrate selectivity rather than being involved in catalytic reactions, and we propose that this is likely to be the case for Artemis. The fact that the residues (His 228, His 254, Cys 256, and Cys 272) that are involved in the zinc-finger like motif are highly conserved across different Artemis species suggests the importance of this structural feature. Furthermore, point mutations in His 228 and His 254 (H228N and H254L) have been reported in patients with a SCID phenotype [55].

The presence of one or two metal ions coordinated by the HxHxDH motif at the active site of Artemis reflects a hallmark of the SNM1 enzyme family [9,74]; the available evidence implies that metal ion binding at one site (Zn1 site in standard MBL nomenclature) is stronger than at the other (Zn2 site). By analogy with studies on the true MBLs, these metal ions are proposed to activate a water molecule that act as the nucleophile for the phosphodiester cleavage. Our structure (PDB: 7AF1) suggests that the native metal ion(s) residing in the active site of Artemis is zinc, although a nickel ion can also occupy the same

site depending on the how the protein was purified (PDB: 6TT5). Neither the presence of Ni ion in the active site, nor the truncation of the C-terminal tail appear to inhibit, at least substantially, the activity of Artemis. Thus, using radio-labelled gel-based nuclease assays, we showed that the truncated Artemis catalytic domain (aa 1–361) with either Zn or Ni ions in the active site (as observed crystallographically in the same preparations) have similar activity with the full-length Artemis construct (aa 1- 693). Therefore, it seems likely that nickel ions are able to replace zinc ions in solution, but catalysis of MBL fold enzymes, including hydrolytic reactions, with metal ions other than zinc is well-precedented [67,75]

We also solved structures of three Artemis catalytic mutants; D37A, H33A, and an Omenn syndrome patient mutation, H35D. Using gel-based nuclease assays, we showed that these variants are biochemically inactive. Overall, the three variant structures are similar to the WT structure, even though H33A and H35D entirely lack any metal ions in the active site, although zinc was present in the zinc finger. Mutation of Asp37 to alanine results in the loss of the second metal in the catalytic site, likely explaining the loss of activity, although the first metal ion is still present. Note that some MBL fold hydrolases uses two metal ions (e.g., B1 and B3 subfamilies of the true MBLs and RNase J1 from *Bacillus subtilis*) (Suppl. Figure 13A and C) [67,76] whereas others, sometimes with apparently very similar active sites, only use one metal ion (e.g. the B2 subfamily of the true MBLs and RNase J from *Staphylococcus epidermis*) [67,77](Suppl. Figure 13B and D). Thus, whilst our results support the importance of having both metals for the nuclease activity by Artemis, subtle features can influence MBL fold enzyme activity [67,74].

Following re-analysis of the Karim *et.al* structure (PDB code 6W00), we were able to generate a model of a DNA overhang I. complex with Artemis that informs on the substrate binding mode. Our model shows that Artemis interacts with the DNA substrate in the interface between the MBL and the  $\beta$ -CASP domains. This interaction is mediated through the combination of polar or positive residues and aromatic residues of Artemis and the DNA substrate (Figure 6).

Artemis is the only identified MBL/ $\beta$ -CASP DNA processing enzyme that possesses substantive endonuclease activity. By contrast both SNM1A and SNM1B/Apollo are strictly 5'-phosphate exonucleases [3,9,15]. The recent structure of SNM1B/Apollo in complex with two deoxyadenosine monophosphate nucleotides (PDB code: 7A1F) reported by Baddock *et. al.* 2020, reveals a cluster of residues that form a 5'-phosphate binding pocket, adjacent to the metal centre. Structural sequence alignments of the three proteins shows that these residues are highly conserved in SNM1A and SNM1B (Suppl. Figure 1). Apart from Ser 317, none of these conserved phosphate binding pocket residues are present in Artemis. Instead, the pocket is partially occupied by the Phe318 side chain, which is absent in both SNM1A and SNM1B. Artemis also possesses a longer and more flexible loop connecting  $\beta$ -strands 27 and 28 (Figure 6B), compared to the same loop in SNM1A and SNM1B that make

up part of the 5'-phosphate binding pocket. The flexibility of this loop could enable accommodation with different types of DNA structures, such as hairpins, and 5'-overhangs. These differences plausibly explain Artemis' substrate preferences and its primary activity as a structure-selective endonuclease.

Of the three of the  $\beta$ -lactam anti bacterial compounds previously shown to inhibit SNM1A [66], we only observed inhibition of Artemis with ceftriaxone. Although the potency of inhibition is moderate ( $IC_{50}$  65  $\mu$ M), we were able to solve the structure of ceftriaxone in complex with Artemis. Notably, ceftriaxone does not bind with its  $\beta$ -lactam carbonyl located at the active site where it ligates to one zinc (or other metal) ion, but instead binds the single nickel ion in bidentate manner *via* the carbonyls of its cyclic 1,2 diamide on its C-3' sidechain. [74,78]. Studies with the true MBLs have shown that appropriate derivatisation of weakly binding molecules can lead to highly potent and selective inhibitors.

In proof of principle attempts to inhibit Artemis though its novel structural feature compared to other MBL fold nucleases, i.e. *via* its zinc-finger like motif, we tested three covalent inhibitors with thiol-reactive groups. Ebselen, disulfiram and auranofin have the potential to interact with zinc fingers, including *via* zinc ejection with consequent protein destabilization [69,70,79,80]. Both ebselen and auranofin are reported have some antimicrobial properties [81], ebselen is in clinical trials for a variety of conditions, ranging from stroke to bipolar disorder [82], and auranofin is used for treatment of rheumatoid arthritis [83]. Recent studies have also shown that ebselen inhibits enzymes from SARS-CoV-2, i.e. the main protease (Mpro) and the exonuclease ExoN (nsp14<sup>ExoN</sup>-nsp10) complex [84,85]. Disulfiram is a known acetaldehyde dehydrogenase inhibitor used in treatment for alcohol abuse disorder [86]. Our results show that both ebselen and disulfiram inhibit Artemis ( $IC_{50}$ s 8.5  $\mu$ M and 10.8  $\mu$ M, respectively), whilst auranofin is less potent ( $IC_{50}$ 46  $\mu$ M).

Studies focussed on inhibiting the MBL fold nucleases are at an early stage compared with work on the true MBLs. The structures and assays results presented here provide starting points with established drugs, from which it might be possible to generate selective Artemis inhibitors, either binding at the active site or elsewhere (including the apparently unique zinc finger of Artemis), in order to radiosensitise cells.

### **ACCESSION NUMBERS**

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 6TT5, 7AF1, 7AFS, 7AFU, 7AGI, 7APV and 7ABS.

### **SUPPLEMENTARY DATA**

Supplementary Data are available at NAR online.

## ACKNOWLEDGEMENT

We are very grateful to Dr Rod Chalk and Tiago Moreira for mass spectrometry, and Dr Neil Patterson for the helpful discussions and data collections at Diamond Light Source. We acknowledge Diamond Light Source for time on Beamlines I03, I04 and I24 under Proposal MX19301.

## FUNDING

This work was supported by a Cancer Research UK Programme Award [A24759 to PJM, OG and CJS] and Wellcome trust grant [106169/ZZ14/Z to OG and 106244/Z/14/Z to CJS].

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

1. Yang W (2011) *Nucleases: Diversity of structure, function and mechanism*.
2. Callebaut I, Moshous D, Mornon J-P & De Villartay JP (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res.* **30**, 3592–3601.
3. Allerston CK, Lee SY, Newman JA, Schofield CJ, Mchugh PJ & Gileadi O (2015) The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res.* **43**, 11047–11060.
4. Goodarzi AA, Yu Y, Riballo E, Douglas P, Walker SA, Ye R, Härer C, Marchetti C, Morrice N, Jeggo PA & Lees-Miller SP (2006) DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO J.* **25**, 3880–3889.
5. Malu S, De Ioannes P, Kozlov M, Greene M, Francis D, Hanna M, Pena J, Escalante CR, Kurosawa A, Erdjument-Bromage H, Tempst P, Adachi N, Vezzone P, Villa A, Aggarwal AK & Cortes P (2012) Artemis C-terminal region facilitates V(D)J recombination through its interactions with DNA Ligase IV and DNA-pkcs. *J. Exp. Med.* **209**, 955–963.
6. Niewolik D, Pannicke U, Lu H, Ma Y, Wang LCV, Kulesza P, Zandi E, Lieber MR & Schwarz K (2006) DNA-PKcs dependence of artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs. *J. Biol. Chem.* **281**, 33900–33909.
7. Niewolik D, Peter I, Butscher C & Schwarz K (2017) Autoinhibition of the nuclease ARTEMIS is mediated by a physical interaction between its catalytic and C-terminal domains. *J. Biol. Chem.* **292**, 3351–3365.
8. Li S, Chang HH, Niewolik D, Hedrick MP, Pinkerton AB, Hassig CA, Schwarz K & Lieber MR (2014) Evidence that the DNA endonuclease ARTEMIS also has intrinsic 5'-exonuclease

- activity. *J. Biol. Chem.* **289**, 7825–7834.
9. Baddock HT, Yosaatmadja Y, Newman JA, Schofield CJ, Gileadi O & McHugh PJ (2020) The SNM1A DNA Repair Nuclease. *DNA Repair (Amst)*. **594**, 102941.
  10. Yan Y, Akhter S, Zhang X & Legerski R (2010) The multifunctional SNM1 gene family: not just nucleases. *Future Oncol.* **6**, 1015–1029.
  11. Wang AT, Sengerova B, Cattell E, Inagawa T, Hartley JM, Kiakos K, Burgess-Brown NA, P SL, H EJ, Schofield CJ, Gileadi O, Hartley JA & Mchugh PJ (20AD) Human SNM1A and XPF–ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev.* **25**, 1859–1870.
  12. Lenain C, Bauwens S, Amiard S, Brunori M, Giraud-Panis MJ & Gilson E (2006) The Apollo 5' Exonuclease Functions Together with TRF2 to Protect Telomeres from DNA Repair. *Curr. Biol.* **16**, 1303–1310.
  13. van Overbeek M & de Lange T (2006) Apollo, an Artemis-Related Nuclease, Interacts with TRF2 and Protects Human Telomeres in S Phase. *Curr. Biol.* **16**, 1295–1302.
  14. Demuth I, Digweed M & Concannon P (2004) Human SNM1B is required for normal cellular response to both DNA interstrand crosslink-inducing agents and ionizing radiation. *Oncogene* **23**, 8611–8618.
  15. Sengerová B, Allerston CK, Abu M, Lee SY, Hartley J, Kiakos K, Schofield CJ, Hartley JA, Gileadi O & McHugh PJ (2012) Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J. Biol. Chem.* **287**, 26254–26267.
  16. Mansilla-Soto J & Cortes P (2003) VDJ recombination: Artemis and its in vivo role in hairpin opening. *J. Exp. Med.* **197**, 543–547.
  17. Ma Y, Pannicke U, Schwarz K & Lieber MR (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794.
  18. Ma Y, Schwarz K & Lieber MR (2005) The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair (Amst)*. **4**, 845–851.
  19. Moshous D, Callebaut I, De Chasseval R, Corneo B, Cavazzana-Calvo M, Le Deist F, Tezcan I, Sanal O, Bertrand Y, Philippe N, Fischer A & De Villartay JP (2001) Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105**, 177–186.
  20. Lieber MR (2011) The mechanism of DSB repair by the NHEJ. *Annu. Rev. Biochem.* **79**, 181–211.
  21. Gu J, Li S, Zhang X, Wang LC, Niewolik D, Schwarz K, Legerski RJ, Zandi E & Lieber MR (2010) DNA-PKcs regulates a single-stranded DNA endonuclease activity of Artemis. *DNA Repair (Amst)*. **9**, 429–437.
  22. Srivastava M & Raghavan SC (2015) DNA double-strand break repair inhibitors as cancer therapeutics. *Chem. Biol.* **22**, 17–29.
  23. Pannunzio NR, Watanabe G & Lieber MR (2018) Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *J. Biol. Chem.* **293**, 10512–10523.
  24. Shockett PE & Schatz DG (1999) DNA Hairpin Opening Mediated by the RAG1 and RAG2

- Proteins. *Mol. Cell. Biol.* **19**, 4159–4166.
25. De P, Peak MM & Rodgers KK (2004) DNA Cleavage Activity of the V(D)J Recombination Protein RAG1 Is Autoregulated. *Mol. Cell. Biol.* **24**, 6850–6860.
  26. Kim MS, Lapkouski M, Yang W & Gellert M (2015) Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature* **518**, 507–511.
  27. Pannicke U, Ma Y, Hopfner KP, Niewolik D, Lieber MR & Schwarz K (2004) Functional and biochemical dissection of the structure-specific nuclease ARTEMIS. *EMBO J.* **23**, 1987–1997.
  28. Barnes DE, Stamp G, Rosewell I, Denzel A & Lindahl T (1998) Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. *Curr. Biol.* **8**, 1395–1398.
  29. Roth DB, Menetski JP, Nakajima P, Bosma MJ & Gellert M (1992) V ( D ) J Recombination : Broken DNA Molecules with Covalently Sealed ( Hairpin ) Coding Ends in scid Mouse Thymocytes. **70**, 983–991.
  30. Bassing CH, Swat W & Alt FW (2002) The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109**, 45–55.
  31. Ege M, Ma Y, Manfras B, Kalwak K, Lu H, Lieber MR, Schwarz K & Pannicke U (2005) Plenary paper Omenn syndrome due to ARTEMIS mutations. *Blood* **105**, 4179–4186.
  32. Volk T, Pannicke U, Reisli I, Bulashevskaya A, Ritter J, Björkman A, Schäffer AA, Fliegau M, Sayar EH, Salzer U, Fisch P, Pfeifer D, Virgilio M Di, Cao H, Yang F, Zimmermann K, Keles S, Schindler D, Hammarström L, Caliskaner Z, Rizzi M, Hummel M, Pan-hammarström Q, Schwarz K & Grimbacher B (2015) DCLRE1C ( ARTEMIS ) mutations causing phenotypes ranging from atypical severe combined immunodeficiency to mere antibody deficiency. **24**, 7361–7372.
  33. Li L, Moshous D, Zhou Y, Wang J, Xie G, Salido E, Hu D & Cowan MJ (2002) A Founder Mutation in Artemis, an SNM1-Like Protein, Causes SCID in Athabascan-Speaking Native Americans. *J. Immunol.* **168**, 6323–6329.
  34. Felgentreff K, Lee YN, Frugoni F, Du L, Van Der Burg M, Giliani S, Tezcan I, Reisli I, Mejstrikova E, De Villartay JP, Sleckman BP, Manis J & Notarangelo LD (2015) Functional analysis of naturally occurring DCLRE1C mutations and correlation with the clinical phenotype of ARTEMIS deficiency. *J. Allergy Clin. Immunol.* **136**, 140-150.e7.
  35. Savitsky P, Bray J, Cooper CDO, Marsden BD, Mahajan P, Burgess-Brown NA & Gileadi O (2010) High-throughput production of human proteins for crystallization: The SGC experience. *J. Struct. Biol.* **172**, 3–13.
  36. Dominy CN & Andrews DW (2003) Site-directed mutagenesis by inverse PCR. *Methods Mol. Biol.* **235**, 209–223.
  37. Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, Fuentes-Montero L, Vollmar M, Michels-Clark T, Young ID, Sauter NK & Evans G (2018) DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 85–97.
  38. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007)

- Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
39. Emsley P & Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132.
  40. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F & Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367.
  41. Lee SY, Brem J, Pettinati I, Claridge TDW, Gileadi O, Schofield CJ & McHugh PJ (2016) Cephalosporins inhibit human metallo  $\beta$ -lactamase fold DNA repair nucleases SNM1A and SNM1B/apollo. *Chem. Commun.* **52**, 6727–6730.
  42. Carfi A, Pares S, Duée E, Galleni M, Duez C, Frère JM & Dideberg O (1995) The 3-D structure of a zinc metallo- $\beta$ -lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J.* **14**, 4914–4921.
  43. Li X & Moses RE (2003) The  $\beta$ -lactamase motif in Snm1 is required for repair of DNA double-strand breaks caused by interstrand crosslinks in *S. cerevisiae*. *DNA Repair (Amst.)* **2**, 121–129.
  44. de Villartay JP, Shimazaki N, Charbonnier JB, Fischer A, Mornon JP, Lieber MR & Callebaut I (2009) A histidine in the  $\beta$ -CASP domain of Artemis is critical for its full in vitro and in vivo functions. *DNA Repair (Amst.)* **8**, 202–208.
  45. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL & Tong L (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953–956.
  46. Alberts IL, Nadassy K & Wodak SJ (1998) Analysis of zinc binding sites in protein crystal structures. *Protein Sci.* **7**, 1700–1716.
  47. Ataie NJ, Hoang QQ, Zahniser MPD, Tu Y, Milne A, Petsko GA & Ringe D (2008) Zinc Coordination Geometry and Ligand Binding Affinity: The Structural and Kinetic Analysis of the Second-Shell Serine 228 Residue and the Methionine 180 Residue of the Aminopeptidase from *Vibrio proteolyticus* †. *Biochemistry* **47**, 7673–7683.
  48. Ishikawa H, Nakagawa N, Kuramitsu S & Masui R (2006) Crystal Structure of TTHA0252 from *Thermus thermophilus* HB8, a RNA Degradation Protein of the Metallo- $\beta$ -lactamase Superfamily. **542**, 535–542.
  49. Matthews JM & Sunde M (2002) Zinc fingers - Folds for many occasions. *IUBMB Life* **54**, 351–355.
  50. Wolfe SA, Nekludova L & Pabo CO (2000) DNA Recognition by Cys 2 His 2 Zinc Finger Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212.
  51. Krishna SS, Majumdar I & Grishin N V. (2003) Structural classification of zinc fingers. *Nucleic Acids Res.* **31**, 532–550.
  52. Singh JK & van Attikum H (2020) DNA double-strand break repair: Putting zinc fingers on the sore spot. *Semin. Cell Dev. Biol.*
  53. Wu X, Bishopric NH, Discher DJ, Murphy BJ & Webster KA (1996) Physical and functional sensitivity of zinc finger transcription factors to redox change. *Mol. Cell. Biol.* **16**, 1035–1046.

54. Laity JH, Lee BM & Wright PE (2001) Zinc finger proteins: New insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**, 39–46.
55. Pannicke U, Hönig M, Schulze I, Rohr J, Heinz GA, Braun S, Janz I, Rump EM, Seidel MG, Matthes-Martin S, Soerensen J, Greil J, Stachel DK, Belohradsky BH, Albert MH, Schulz A, Ehl S, Friedrich W & Schwarz K (2010) The most frequent DCLRE1C (ARTEMIS) mutations are based on homologous recombination events. *Hum. Mutat.* **31**, 197–207.
56. Karim MF, Liu S, Laciak AR, Volk L, Rosenblum M, Lieber MR, Wu M, Curtis R, Huang N, Carr G & Zhu G (2020) Structural analysis of the catalytic domain of Artemis endonuclease/SNM1C reveals distinct structural features. *J. Biol. Chem.* **444**, jbc.RA120.014136.
57. Ussery DW (2002) DNA Structure: A-, B- and Z-DNA Helix Families. *Encycl. Life Sci.*
58. Elrod-Erickson M, Rould MA, Nekludova L & Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 Å: A model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171–1180.
59. Locasale JW, Napoli AA, Chen S, Berman HM & Lawson CL (2009) Signatures of Protein-DNA Recognition in Free DNA Binding Sites. *J. Mol. Biol.* **386**, 1054–1065.
60. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL & Tong L (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953–956.
61. Poinsignon C, Moshous D, Callebaut I, De Chasseval R, Villey I & De Villartay JP (2004) The Metallo-β-Lactamase/β-CASP Domain of Artemis Constitutes the Catalytic Core for V(D)J Recombination. *J. Exp. Med.* **199**, 315–321.
62. Jekimovs C, Bolderson E, Suraweera A, Adams M, O'Byrne KJ & Richard DJ (2014) Chemotherapeutic compounds targeting the DNA double-strand break repair pathways: The good, the bad, and the promising. *Front. Oncol.* **4 APR**, 1–18.
63. Shibata A & Jeggo P (2019) A historical reflection on our understanding of radiation-induced DNA double strand break repair in somatic mammalian cells; interfacing the past with the present. *Int. J. Radiat. Biol.* **95**, 945–956.
64. Weterings E, Gallegos AC, Dominick LN, Cooke LS, Bartels TN, Vagner J, Matsunaga TO & Mahadevan D (2016) A novel small molecule inhibitor of the DNA repair protein Ku70/80. *DNA Repair (Amst)*. **43**, 98–106.
65. Jin MH & Oh DY (2019) ATM in DNA repair in cancer. *Pharmacol. Ther.* **203**, 107391.
66. Lee SY, Brem J, Pettinati I, Claridge TDW, Gileadi O, Schofield CJ & McHugh PJ (2016) Cephalosporins inhibit human metallo β-lactamase fold DNA repair nucleases SNM1A and SNM1B/apollo. *Chem. Commun.* **52**, 6727–6730.
67. Palzkill T (2013) Metallo-β-lactamase structure and function. *Ann. N. Y. Acad. Sci.* **1277**, 91–104.
68. Spraggon G, Koesema E, Scarselli M, Malito E, Biagini M, Norais N, Emolo C, Barocchi MA, Giusti F, Hilleringmann M, Rappuoli R, Lesley S, Covacci A, Massignani V & Ferlenghi I (2010) Supramolecular organization of the repetitive backbone unit of the *Streptococcus pneumoniae pilus*. *PLoS One* **5**.

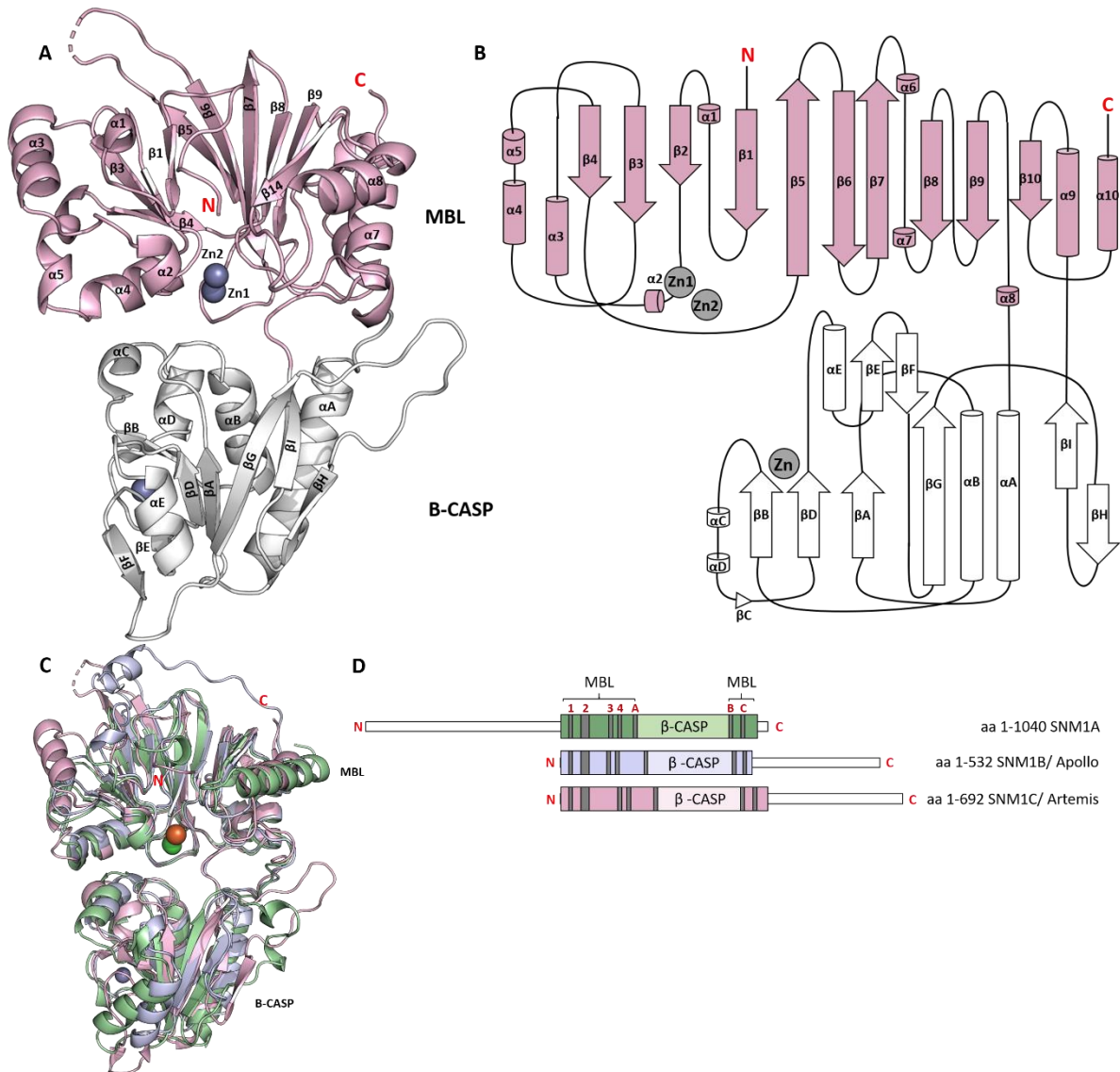


69. Abbehausen C (2019) Zinc finger domains as therapeutic targets for metal-based compounds - an update. **11**.
70. Chen S, Jeng K & Lai MMC (2017) Zinc Finger-Containing Cellular Transcription Corepressor ZBTB25 Promotes Influenza Virus RNA Transcription and Is a Target for Zinc Ejector Drugs. **91**, 1–20.
71. Brandt VL & Roth DB (2003) Artemis: Guarding small children and, now, the genome. *J. Clin. Invest.* **111**, 315–316.
72. Vázquez-Torres A (2012) Redox active thiol sensors of oxidative and nitrosative stress. *Antioxidants Redox Signal.* **17**, 1201–1214.
73. Mao Z, Bozzella M, Seluanov A & Gorbunova V (2008) Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)*. **7**, 1765–1771.
74. Pettinati I, Brem J, Lee SY, Mchugh PJ & Scho CJ (2016) The Chemical Biology of Human Metallo-  $\beta$ -Lactamase Fold Proteins. *Trends Biochem. Sci.* **41**, 338–355.
75. Cahill ST, Tarhonskaya H, Ryzdik AM, Flashman E, McDonough MA, Schofield CJ & Brem J (2016) Use of ferrous iron by metallo- $\beta$ -lactamases. *J. Inorg. Biochem.* **163**, 185–193.
76. Newman JA, Hewitt L, Rodrigues C, Solovyova A, Harwood CR & Lewis RJ (2011) Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure* **19**, 1241–1251.
77. Raj R, Nadig S, Patel T & Gopal B (2020) Structural and biochemical characteristics of two Staphylococcus epidermidis RNase J paralogues RNase J1 and RNase J2 . *J. Biol. Chem.*, jbc.RA120.014876.
78. Hamed RB, Gomez-Castellanos R, Henry L, Ducho C, McDonough MA & Schofield CJ (2013) The enzymes of B-lactam biosynthesis. *Nat. Prod. Rep.* **30**, 1–204.
79. Antony S & Bayse CA (2013) Density Functional Theory Study of the Attack of Ebselen on a Zinc- Finger Model. *Inorg. Chem.* **52**, 13803–13805.
80. Lee Y, Wang Y, Duh Y, Yuan HS & Lim C (2013) Identification of Labile Zn Sites in Drug-Target Proteins. *J. Am. Chem. Soc.* **135**, 14028–14031.
81. May HC, Yu JJ, N. Guentzel M, Chambers JP, Cap AP & Arulanandam BP (2018) Repurposing auranofin, ebselen, and PX-12 as antimicrobial agents targeting the thioredoxin system. *Front. Microbiol.* **9**, 1–10.
82. Noguchi N (2016) Ebselen, a useful tool for understanding cellular redox biology and a promising drug candidate for use in human diseases. *Arch. Biochem. Biophys.* **595**, 109–112.
83. Roder C & Thomson MJ (2015) Auranofin: Repurposing an Old Drug for a Golden New Age. *Drugs R D* **15**, 13–20.
84. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L, Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z & Yang H (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293.
85. Baddock HT, Brolih S, Yosaatmadja Y, Ratnaweera M, Bielinski M, Swift L, Cruz-Migoni A,

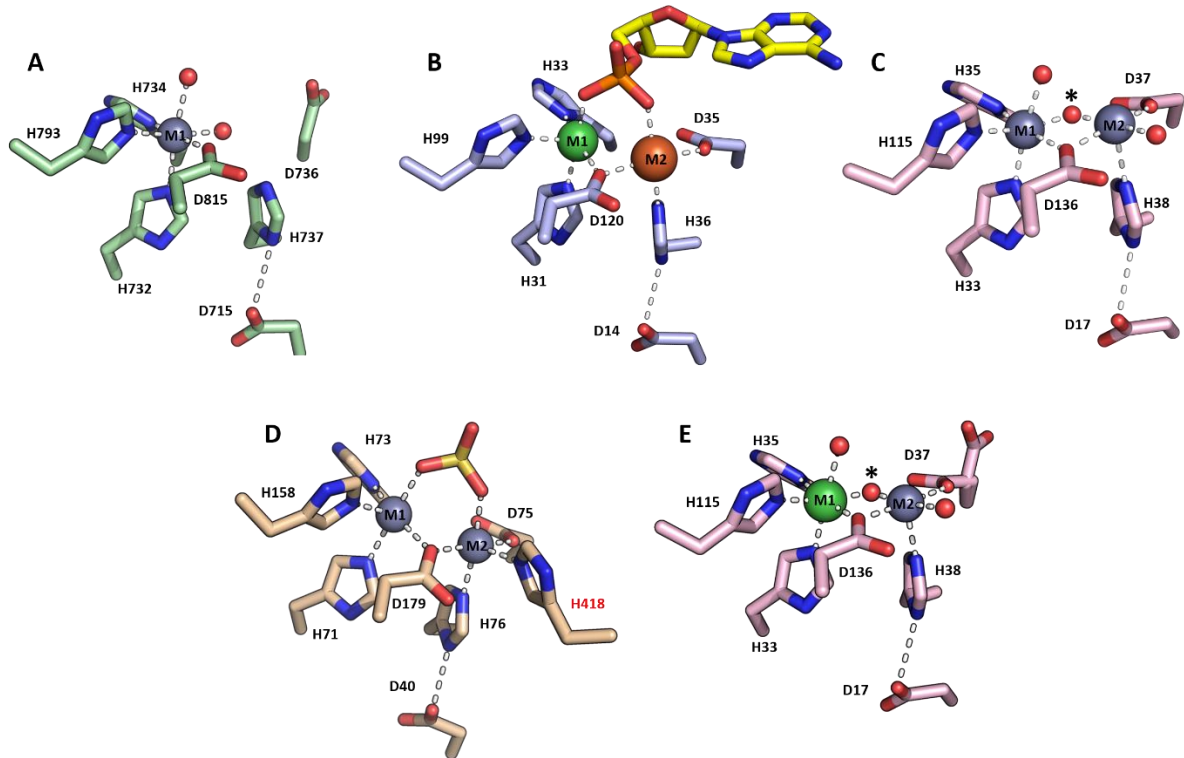
Morris GM, Schofield CJ, Gileadi O & McHugh PJ (2020) Characterisation of the SARS-CoV-2 ExoN (nsp14<sup>&gt;ExoN</sup>-nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *bioRxiv*, 2020.08.13.248211.

86. Skinner MD, Lahmek P, Pham H & Aubin HJ (2014) Disulfiram efficacy in the treatment of alcohol dependence: A meta-analysis. *PLoS One* **9**.

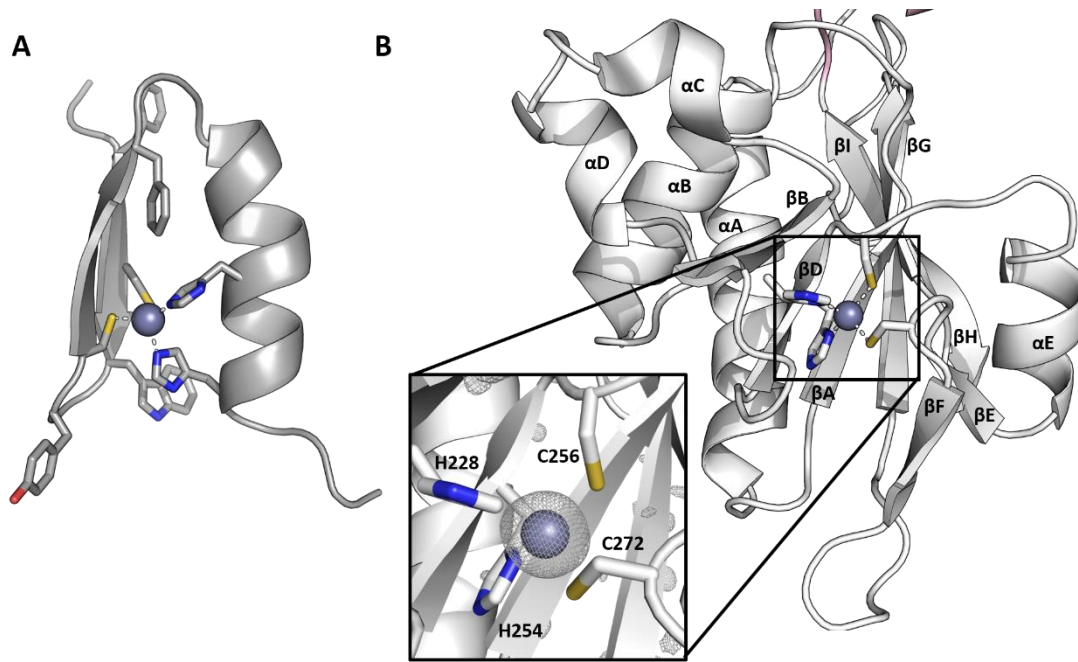
## FIGURES



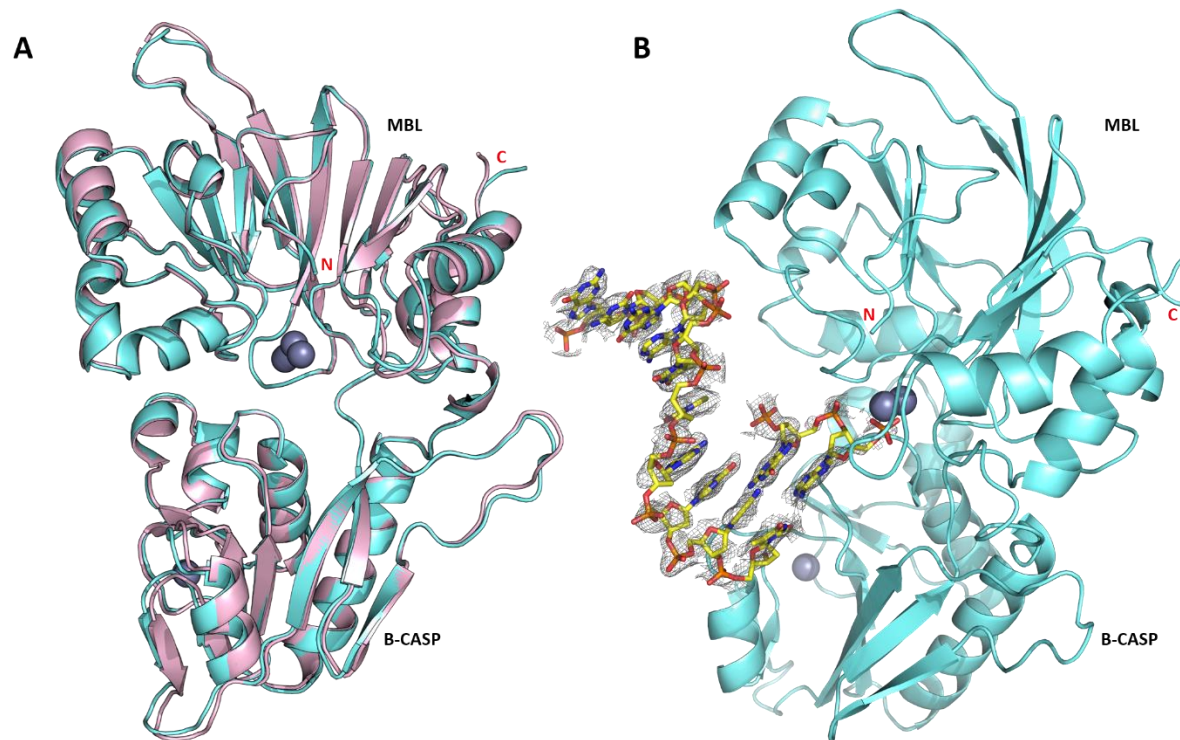
**Figure 1: Overall architecture of Artemis.** **A.** A cartoon representation of the structure of human SNM1C/ Artemis. The active site containing MBL domain is in pink; the  $\beta$ -CASP domain (white) contains a novel zinc-finger like motif, that has not been identified in other MBL/  $\beta$ -CASP nucleic acid processing enzymes. The three zinc ions are represented by grey spheres. **B.** Topology diagram of Artemis protein. The  $\beta$ -strands are represented arrows and  $\alpha$ -helices by cylinders. The MBL domain (pink) has the typical  $\alpha/\beta$ - $\beta/\alpha$  sandwich of the MBL superfamily, with an insert of the  $\beta$ -CASP domain (white) between the small helix  $\alpha_6$  and helix  $\alpha_7$ . **C.** Overlay of structures of the human SNM1 Family members: SNM1A, SNM1B and SNM1C. **D.** Cartoon representation of amino acid sequence alignment for human SNM1A, SNM1B and SNM1C, showing the conserved MBL and  $\beta$ -CASP domains.



**Figure 2: Active site views of human MBL/  $\beta$ -CASP nuclease family enzymes.** Each of these catalytic sites contains 4 highly conserved motifs (1-4, in red). Motif 1 = Asp, motif 2 = 3 His and 1 Asp (HxHxDH), motif 3 = His and motif 4 = Asp. **A.** The human SNM1A active site with a single octahedral zinc ion (grey) coordination (PDB: 5AHR). **B.** The active site of human SNM1B/ Apollo (PDB: 7A1F) with a nickel ion (green) and an iron ion (orange) with a coordinating AMP molecule. **C.** Human SNM1C/ Artemis (PDB: 7AF1) purified in the absence of IMAC with two zinc ions (in grey) in its active site. A water molecule shared (asterisk\*) between the two metals is the proposed nucleophile for the hydrolytic reaction. **D.** The active site of the human RNA processing enzyme CPSF73 (PDB: 2I7T). The second zinc ion (M2) is coordinated by an additional histidine residue (His 418) which has no counterpart in the SNM1 proteins. **E.** The active site of human SNM1C/ Artemis (PDB: 6TT5) purified with IMAC. A nickel ion is present in the first metal coordination site.

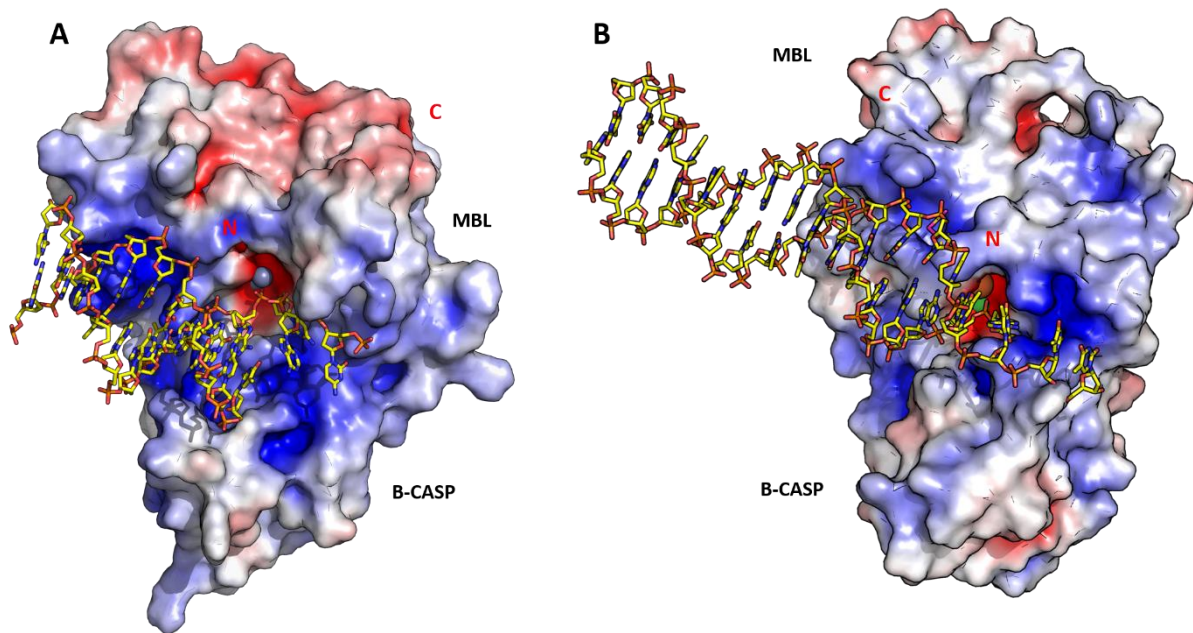


**Figure 3: Comparison of a novel zinc-finger like motif in the  $\beta$ -CASP domain of Artemis with a canonical zinc-finger motif.** **A.** Cartoon representation of the classical Cys<sub>2</sub>His<sub>2</sub> zinc-finger motif, from transcription factor SP1F2 (PDB code: 1SP2). This has a  $\beta\beta\alpha$  fold, where two Cys- and two His-residues are involved in zinc ion coordination and the sidechains of three conserved hydrophobic residues are shown. **B.** The  $\beta$ -CASP region of Artemis has a novel zinc-finger like motif. The inset shows the four residues (two His and two Cys) coordinating the zinc ion (grey). The  $F_o - F_c$  electron- density map (scaled to  $2.5\sigma$  in PyMOL) surrounding the zinc ion before it was included in refinement.

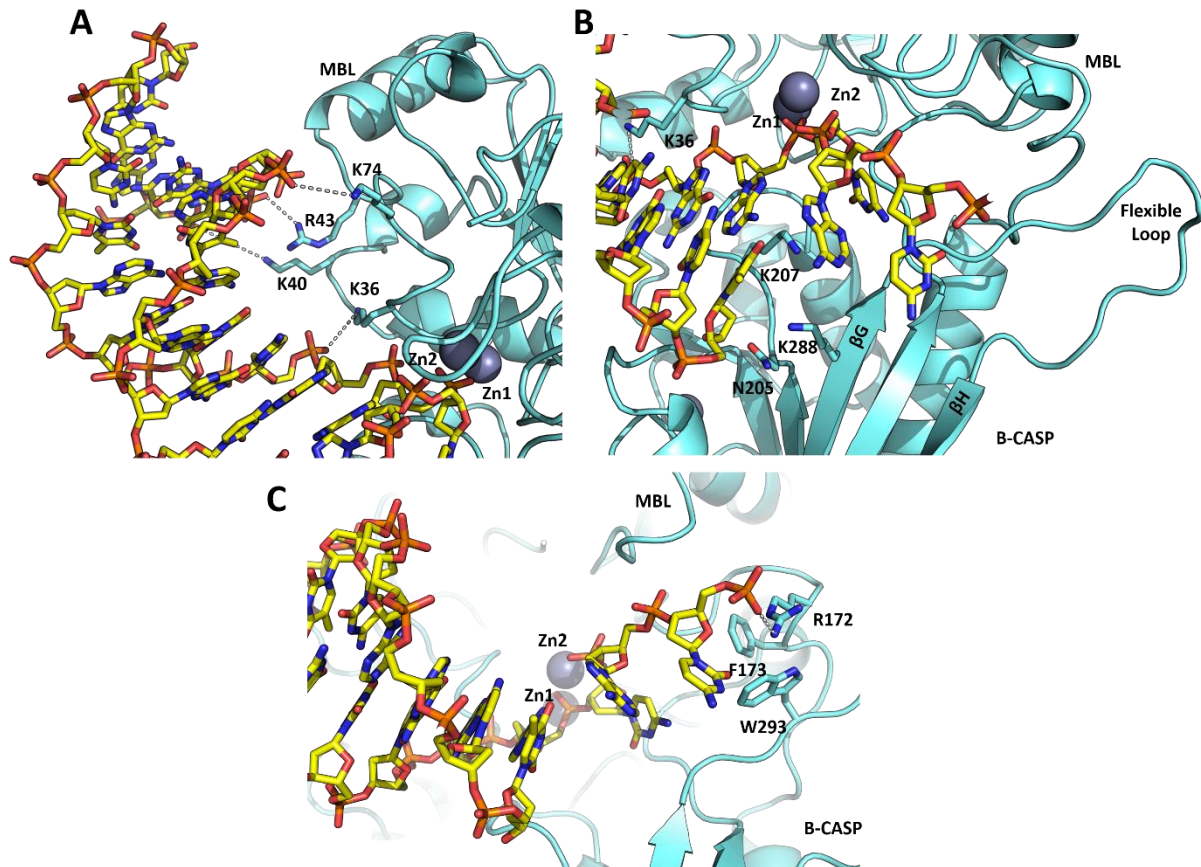


**Figure 4: Overall structure representation of the Artemis /SNM1C fold.** **A.** Overlay of our wt Artemis structure (PDB code: 7AF1) (pink) with that of Karim *et. al.* PDB code: 6W00 (aquamarine) (backbone RMSD 0.48 Å). **B.**

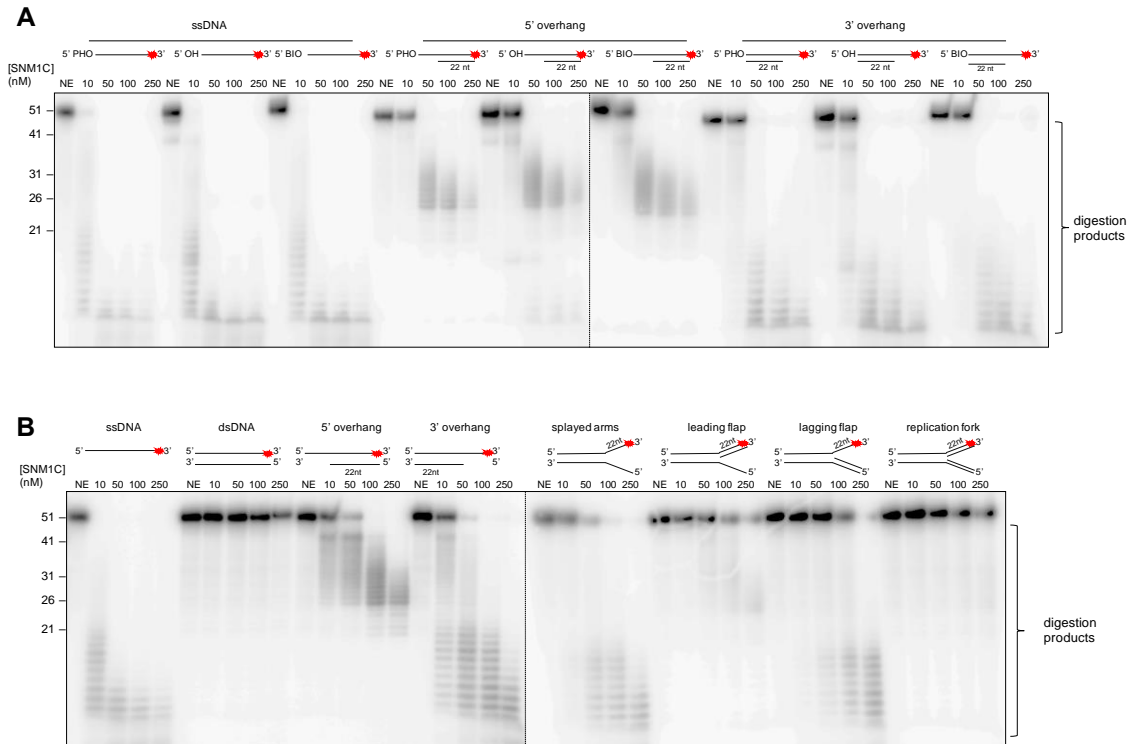
Re-analysis of the latter (6W00) structure re-refined with a DNA molecule present (PDB code; 7ABS). The 2Fo-Fc map (contoured at  $0.6\sigma$  in pymol) is represented by grey mesh surrounding the DNA (yellow).



**Figure 5: Electrostatic surface potentials of DNA bound model for Artemis/SNM1C (A) and Apollo/SNM1B (B).** The blue colour represents a more electropositive surface potential and the red show a more electronegative cluster. The active site contains the two metal ions represented in grey sphere for Zinc, orange sphere for Iron and green sphere for Nickel ion. N- and C- terminal of the protein are indicated in red. The electrostatic surface potentials were generated using PyMOL (electrostatic range at +/- 5).

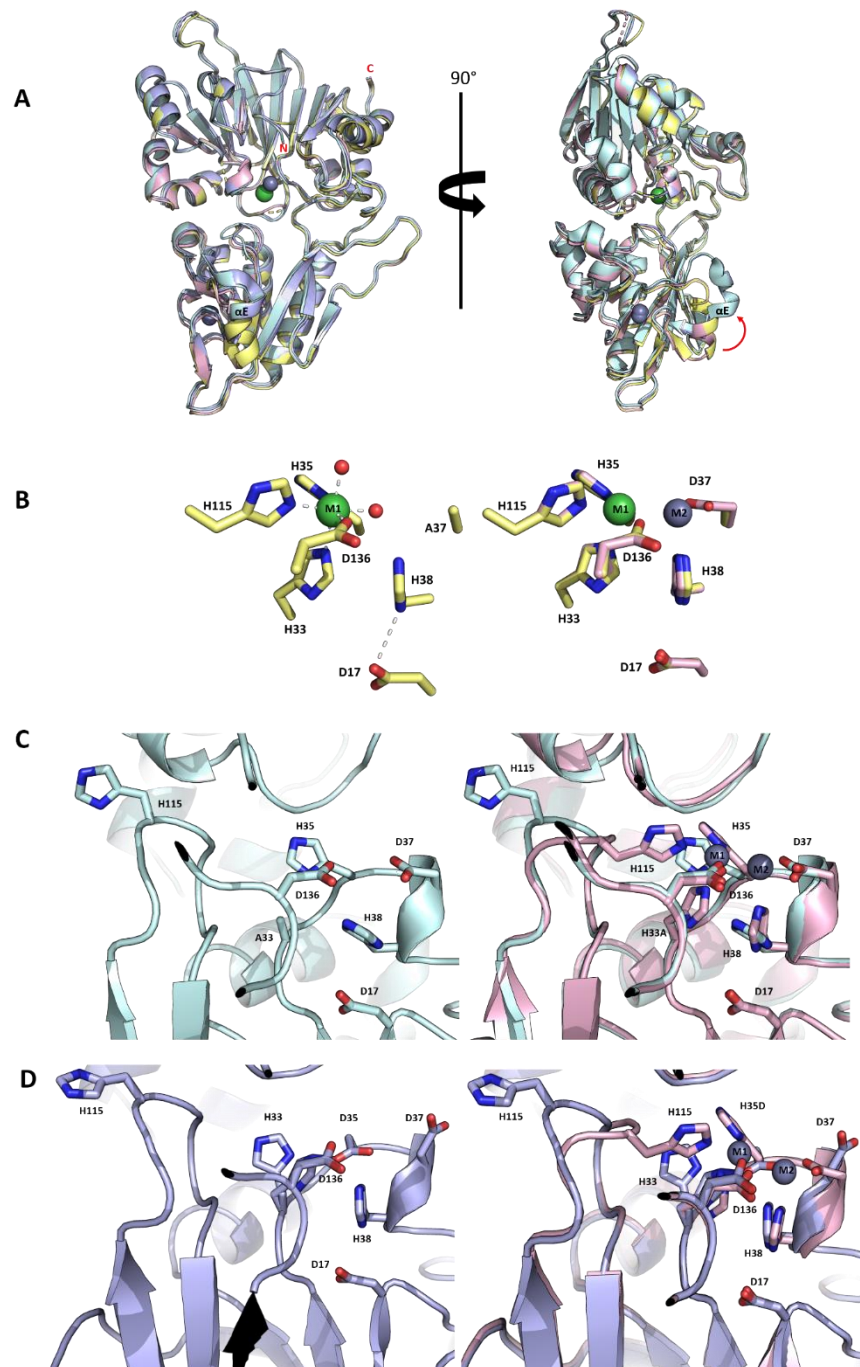


**Figure 6: Proposed interactions of DNA with Artemis (PDB: 7ABS).** Model for DNA binding to Artemis showing the residues contacting DNA. The two zinc ions at the active site are represented by grey spheres. **A.** A row of positively charged residue is on the surface of the MBL domain interact with the phosphate backbone of the DNA. **B.** A DNA overhang is located at the active site. A cluster of polar residues (N205, K207 and K288) is located in the  $\beta$ -CASP domain. The extended flexible loop, which is unique to Artemis compared to SNM1A and SNM1B, that connects b27 and b28 is indicated on the right. **C.** The DNA overhang forms a hydrogen bond with Arginine 172 and interacts with a cluster of hydrophobic residues at the interface between MBL and  $\beta$ -CASP domains.

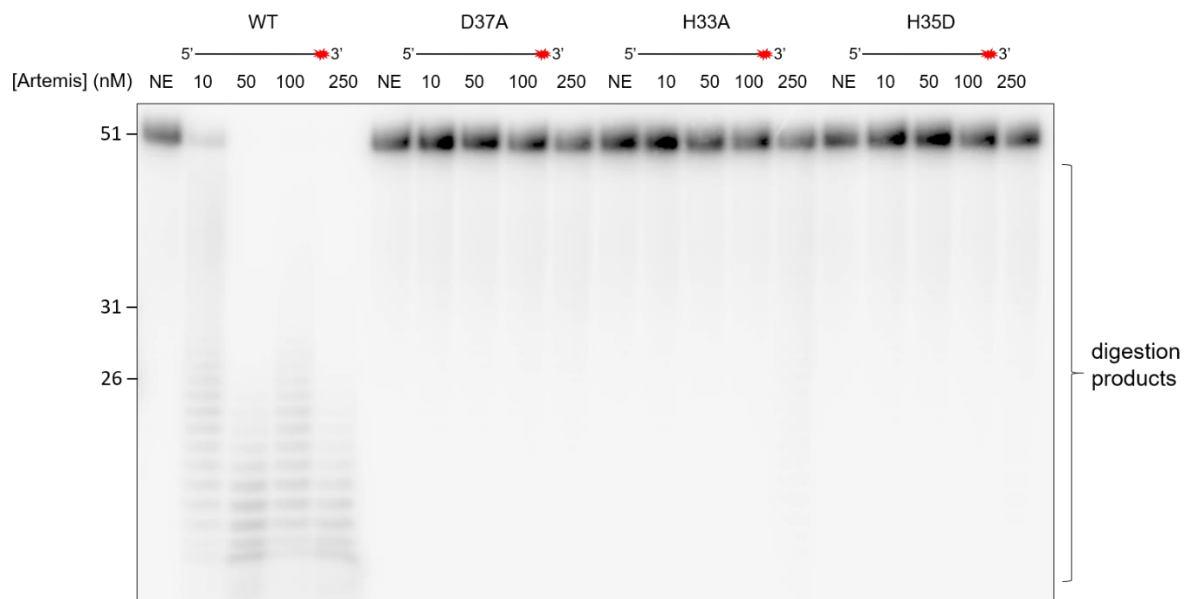


**Figure 7: Nuclease assay utilizing truncated Artemis (aa 3–361) with various DNA substrates. A.** The nuclease activity of Artemis is indifferent to the 5' end group, indicative of true endonuclease activity. Increasing concentrations of Artemis from 0 (NE; no enzyme) to 250 nM incubated with 10 nM ssDNA with either a 5' phosphate, 5' hydroxyl, or 5' biotin moiety for 45 min at 37°C. **B.** Artemis is able to cleave DNA substrates containing single-stranded regions. Increasing amounts of Artemis incubated with structurally diverse DNA substrates (10 nM) for 45 min at 37°C. Products for A and B were analysed by 20% denaturing PAGE. The DNA substrates utilised are represented at the top of the lanes and a red asterisk indicates the position of the 3' radiolabel. The positions of DNA size markers run as a reference are indicated on the left, with sizes in nt.

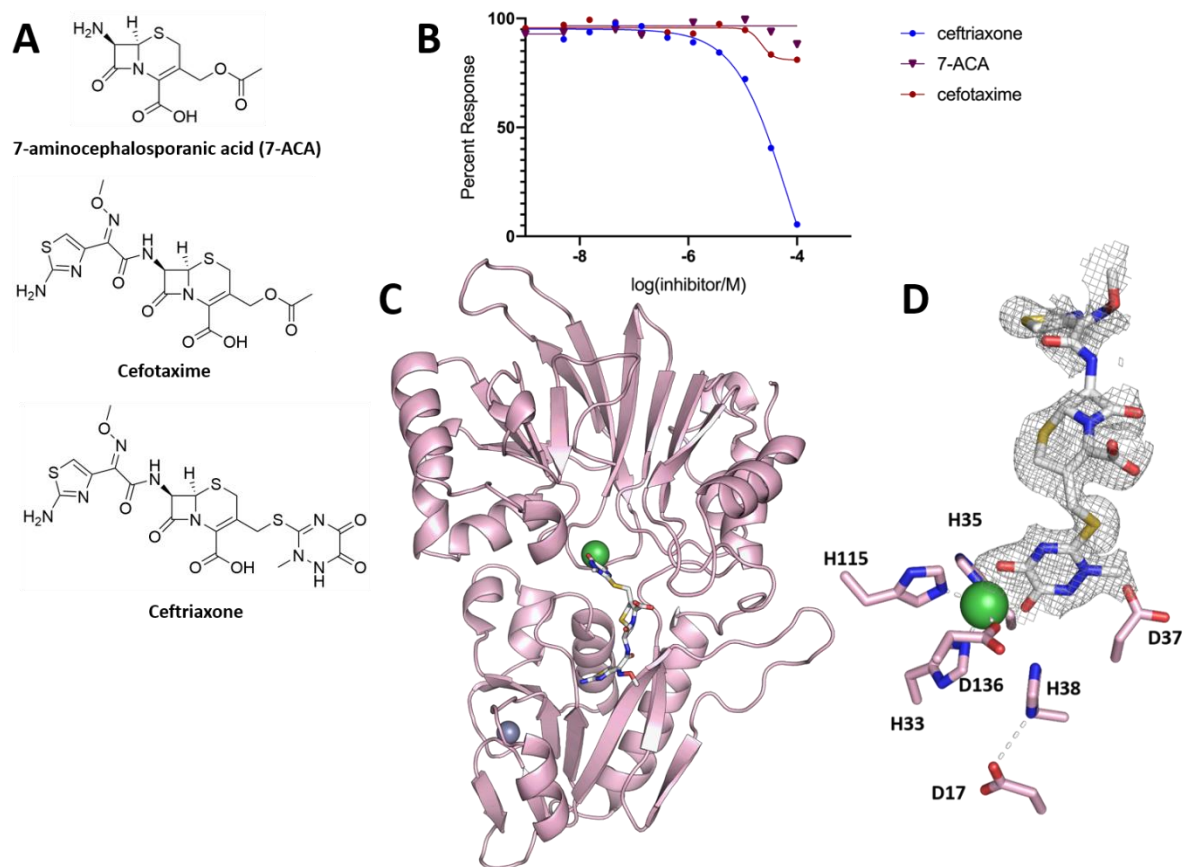




**Figure 8: Views from structure of Artemis D37A, H33A and H35D variants.** **A.** An overlay of the four Artemis structures: WT (PDB:7AF1) in pink, D37A variant (PDB:7AFS) in yellow, H33A variant (PDB:7AFU) in cyan and H35D patient mutation (PDB:7AGI) in blue, showing the general architecture of the three variants are the same as the W structure. The nickel ion is represented as green sphere, and the zinc ions as grey spheres. The movement of helix  $\alpha E$  in the  $\beta$ -CASP domain is indicated by a red arrow. **B.** Left: The active site of D37A variant has a single nickel ion with two complexing water molecules (red spheres); Right: the active site residues of WT Artemis (pink), superimposed with those of the D37A variant (yellow). Aside from loss of the second metal in the D37A mutant, there is little movement at the active site. **C.** Active site residues of the H33A variant (cyan; left) and an overlay (right) with WT Artemis (pink). The two distinguishing features of H33A variant are a lack of metal ions and movement of the loop containing His115. **D.** The active site of the H35D variant (blue; left) and an overlay (right) with WT Artemis (pink). The H35D point substitution is present in patients with Omen syndrome. This variant lacks both metal ions, similarly to the H33A variant.



**Figure 9: Comparing the activity of Artemis variants vs WT protein.** Increasing amounts (from 0 to 250 nM) of WT and mutant Artemis proteins (as indicated) were incubated with 10 nM of 51 nucleotide ssDNA substrate for 30 min at 37 °C. Reaction products were subsequently analysed by 20% denaturing PAGE. The size (in nucleotides) of the marker oligonucleotides are indicated on the left-hand side of the corresponding bands.



**Figure 10: Artemis inhibition by  $\beta$ -lactam antibiotics.** **A.** Structures of selected  $\beta$ -lactam antibacterial. **B.** Inhibitor profiles of  $\beta$ -lactams on the nuclease activity of Artemis was assessed *via* a real-time fluorescence-based nuclease assay. **C.** Cartoon representation of the structure of truncated Artemis (aa1-361) with a ceftriaxone molecule (in white) bound at the active site. **D.** Active site residues with the electron density ( $F_o - F_c$ ) contoured map around the modelled Ceftriaxone. The map is contoured at the  $1.0 \sigma$  level and was calculated before the Ceftriaxone molecule was included in the refinement.

## TABLES

**Table 1: Data collections and refinement statistics.** \* Data in parentheses is for the high-resolution shell.

PDB ID	6TT5		7AFS	7AFU	7AGI	7APV	7ABS
	(Ni and Zn)	7AF1 (2 Zn)	(D37A)	(H33A)	(H35D)	(Ceftriaxone)	(DNA bound)
<b>Data Collection and processing</b>							
Diffraction Source	DLS (I04)	DLS (I24)	DLS (I03)	DLS (I03)	DLS (I03)	DLS (I03)	APS 17-IDD
Wavelength (Å)	0.979	0.976	0.976	0.976	0.976	0.976	1.00
Space group	P1	P1	P1	P1	P1	P1	P21212
Cell Dimensions							
$a, b, c$ (Å)	35.87, 47.99, 48.25	35.75, 47.97, 48.15	35.91, 48.06, 48.21	35.97, 47.90, 48.37	35.97, 48.05, 48.44	35.88, 48.10, 48.25	72.81, 111.00, 55.17
$\alpha, \beta, \gamma$ (°)	82.61, 76.37, 85.98	82.68, 76.35, 85.81	82.89, 76.43, 86.38	82.51, 75.94, 87.73	82.43, 76.01, 87.33	82.76, 76.29, 86.30	90.00, 90.00, 90.00
Resolution (Å) *	47.55 - 1.50 (1.53 - 1.50)	47.53 - 1.70 (1.73 - 1.70)	35.35 - 1.70 (1.73 - 1.70)	35.53 - 1.56 (1.59 - 1.56)	47.62 - 1.70 (1.73 - 1.70)	47.69 - 1.95 (2.00 - 1.95)	40 - 1.97 (2.02 - 1.97)
$R_{\text{merge}}$ (%)*	6.0 (64.4)	13.3 (79.5)	5.3 (53.2)	4.9 (31.5)	4.8 (23.2)	11.6 (65.0)	5.0 (82.8)
$I/\sigma(I)$	13.4 (3.2)	5.8 (2.0)	11.7 (2.2)	11.3 (2.6)	13.4 (3.8)	8.6 (2.8)	16 (2.4)
Completeness (%)	94.7 (66.4)	97.4 (95.6)	97.3 (96.0)	94.2 (63.7)	97.4 (95.4)	98.0 (97.2)	99.6 (99.6)
Multiplicity	3.6 (3.2)	3.5 (3.5)	3.6 (3.7)	3.6 (3.3)	3.7 (3.8)	3.5 (3.6)	6.4 (6.6)
<b>Refinements</b>							
Resolution (Å)	47.55 - 1.50	47.53 - 1.70	47.66 - 1.70	35.53 - 1.56	47.62 - 1.70	47.96 - 1.95	40 - 1.97
No. of reflections	44746	31282	34470	39478	31745	21086	32168
$R_{\text{work}}$	0.17	0.19	0.18	0.19	0.19	0.18	0.22
$R_{\text{free}}$	0.19	0.21	0.22	0.21	0.21	0.23	0.28
No. of Atom							
Protein	2927	2989	2957	2985	3122	2975	2923
Water	224	144	163	255	164	107	105
Zinc/ Nickle	3	3	2	1	1	2	3
Ethylene glycol	24	28	12	44	32	16	-
DNA	-	-	-	-	-	-	253
Ceftriaxone	-	-	-	-	-	1	-
B-factors	16.3	19.3	25.9	17.5	18.4	22.6	56
r.m.s. deviations							
Bond length (Å)	0.003	0.008	0.008	0.006	0.007	0.01	0.007
bond angles (°)	1.23	1.33	1.29	1.28	1.32	1.5	1.47