



Artemis (DCLRE1C, SNM1C)



A Target Enabling Package (TEP)

Gene ID / UniProt ID / EC	64421 / Q96SD1 / 3.1.-.-
Target Nominator	Internal
Authors	Yuliana Yosaatmadja ¹ , Joseph Newman ¹ , Hannah Baddock ² , Marcin Bielinski ³ , Peter McHugh ² , Christopher Schofield ³ , Opher Gileadi ¹
Target PI	Opher Gileadi
Therapeutic Area(s)	Cancer
Disease Relevance	Artemis is a human endonuclease that plays an important role in non-homologous end joining (NHEJ) and in variable (diversity) joining (V(D)J) recombination.
Date Approved by TEP Evaluation Group	27 th November 2020
Document version	1.0
Document version date	November 2020
Citation	Yuliana Yosaatmadja, Joseph Newman, Hannah Baddock, Marcin Bielinski, Peter McHugh, Christopher Schofield, & Opher Gileadi. (2021). Artemis (DCLRE1C, SNM1C): A Target Enabling Package [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4423402 .
Affiliations	¹ SGC, University of Oxford ² Department of Oncology, Weatherall Institute of Molecular Medicine, University of Oxford. ³ Department of Chemistry, University of Oxford.

USEFUL LINKS



(Please note that the inclusion of links to external sites should not be taken as an endorsement of that site by the SGC in any way)

SUMMARY OF PROJECT

Currently several radio-sensitising and chemotherapeutic agents are used in conjunction with radiotherapy to enhance the efficacy of cancer treatment. DCLRE1C/Artemis is a major player in both programmed (V(D)J) recombination and non-programmed c-NHEJ DSB repair. This makes Artemis and other DSB repair enzymes an attractive pharmacological target for the radiosensitisation of tumours. This TEP includes expression and purification methods for producing the full-length (aa 1-692), and the catalytic domain (aa 1-362) of Artemis for high-throughput activity and inhibitor assays, and a robust crystallisation method that is able to generate reproducible crystals for small molecule compound soaking. We also present an Artemis structure in complex with the β -lactam anti-bacterial compound ceftriaxone.

SCIENTIFIC BACKGROUND

DCLRE1C Artemis is a structure-specific DNA endonuclease that is essential for the development of B and T lymphocytes. Its endonucleolytic activity is responsible for hairpin opening in variable (diversity) joining (V(D)J) recombination (1) and contributes to end-processing in canonical non-homologous end joining (c-NHEJ) DNA repair (2–5). It is therefore unsurprising that Artemis deficiency causes congenital radiosensitive severe acquired immune deficiency (RS-SCID).

Artemis has a very similar fold to DCLRE1A (PDB: [5Q2A](#)) and DCLRE1B/Apollo (PDB: [5AHO](#)). Artemis has a metallo-beta-lactamase (MBL) domain and a beta-CASP (CPSF-Artemis-SNM1-Pso2) domain. While both DCLRE1A and DCLRE1B/Apollo are exclusively 5'-to-3' exonucleases, the predominant activity of Artemis is endonucleolytic (6–8), although 5' - 3' exonuclease activity has been reported (9). Artemis possesses a unique feature, that is not present in either DCLRE1A or DCLRE1B/Apollo; a second metal binding site in its β -CASP domain that is similar to the classical Cys₂His₂ zinc finger motif. We propose that the zinc finger like motif in the beta-CASP domain of Artemis plays an important role in the protein stability. This metal coordination site may also provide a second small molecule binding site.

One of the most common causes for Artemis loss-of-function mutations are large deletions in the first four exons, and a nonsense founder mutation that was found in Navajo and Apache Native Americans (10). However, missense mutations and in-frame deletion in the highly conserved residues such as H35, and D165 can also abolish Artemis protein function (11). Mutation in the zinc coordinating residues in the β -CASP domain of Artemis (H228N and H254L) trigger radiosensitive SCID (RS-SCID) in humans (11, 12). Patients with these mutations suffers from impaired V(D)J recombination which leads to underdeveloped B and T lymphocytes. The importance of histidine 254 was highlighted previously by de Villartay *et al.*, (12), who showed that the full-length H254A Artemis variant is unable to carry out V(D)J recombination *in vivo* and is incapable of maintaining its endonucleolytic activity *in vitro*. This TEP provides components that allow for further study of the protein, as well as commercially available compounds which could be used as the basis for medicinal chemistry campaigns.

RESULTS – THE TEP

Proteins purified

Both the full-length (aa 1-692) and the catalytic domain (aa 1-362) were cloned in pFastBac vector and expressed recombinantly using the baculovirus and SF9 insect cells expression system. The full-length protein is multiply phosphorylated and has the same level of nuclease activity as the isolated catalytic domain.

Structural data

Our current crystal form of DCLRE1C consists of one molecule of the catalytic domain of DCLRE1C with resolution better than 1.5 Å. The active site of the protein is accessible and (unlike the DCLRE1A and DCLRE1B structures) free of any ligands which allows soaking of small molecules (**Fig 1A**). Here, we present the structure of the DCLRE1C catalytic domain with ceftriaxone bound in the metal centre of the protein (**Fig 1B**). The structure was solved to 1.9 Å resolution. Artemis protein crystal was soaked with 5 mM Ceftriaxone in crystallisation solution for one hour. The structure was determined using molecular replacement (PDB: [7AF1](#)), in the space group P1 with one protein molecule in the asymmetric unit. A list of structures deposited in the PDB is provided below.

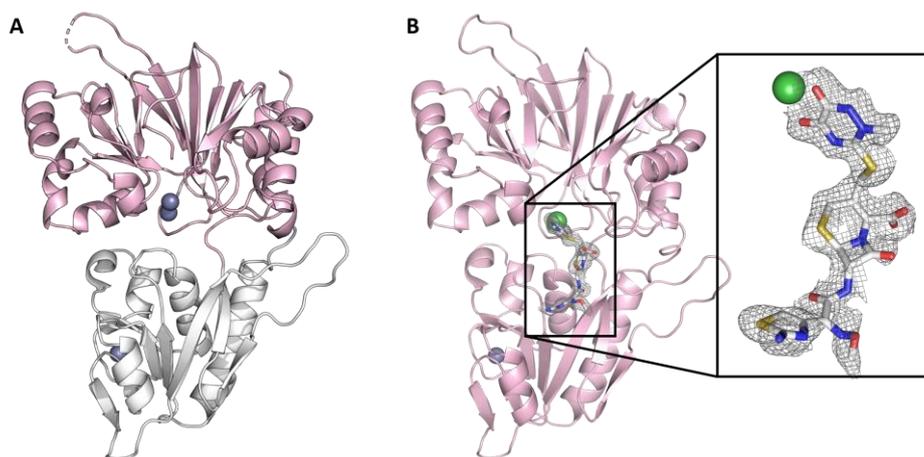


Figure 1. Representation of the structures of DCLRE1C/Artemis. **(A)** Apo structure of DCLRE1C/ Artemis catalytic domain. The MBL domain shown in pink and the β -CASP domain (white) contains a novel zinc-finger like motif that is not present in other MBL/ β -CASP nucleic acid processing enzymes. The three zinc ions are represented by grey spheres. **(B)** The structure of DCLRE1C/ Artemis catalytic domain in complex with ceftriaxone. The electron density is shown in grey mesh around the ceftriaxone molecule. A nickel ion is shown as a green sphere.

Assays

In collaboration with the McHugh lab (Department of Oncology, Oxford) and the Schofield lab (Department of Chemistry, Oxford) a high throughput *in vitro* assay to detect the cleavage of single stranded DNA substrate containing both a fluorescein-conjugated T and a BHQ-1-conjugated (black hole quencher) T has been developed (**Fig 2A**). As the fluorescein and BHQ-1 are located proximally to one another, when the ssDNA substrate is intact there is no fluorescence, however, following endonucleolytic cleavage by DCLRE1C/Artemis, there is uncoupling of the fluorescein from the BHQ and an increase in fluorescence is observed. This assay is carried out in 384-well format, with readings taken every 150 sec for 35 min. This real-time fluorescence assay protocol enables the calculation of enzyme kinetics and can also be used for inhibitor screening and testing. In addition to this, a second assay protocol has been developed which utilises 3'-radiolabelled (α - ^{32}P -dATP) substrates in a gel-based nuclease assay (**Fig 2D**). In this protocol, ssDNA is radiolabelled on the 3' end, and this can be annealed with complementary sequences to generate DNA substrates with secondary structure. These relevant DNA substrates are then incubated with DCLRE1C/Artemis at 37°C at increasing enzyme concentrations, and this allows us to probe biochemical features of the nuclease, including (but not limited to) substrate preference, site(s) of endonucleolytic incision, metal ion dependency, and activity over time. This protocol can also be utilised as a validation of inhibitors generated via the high-throughput assay approach.

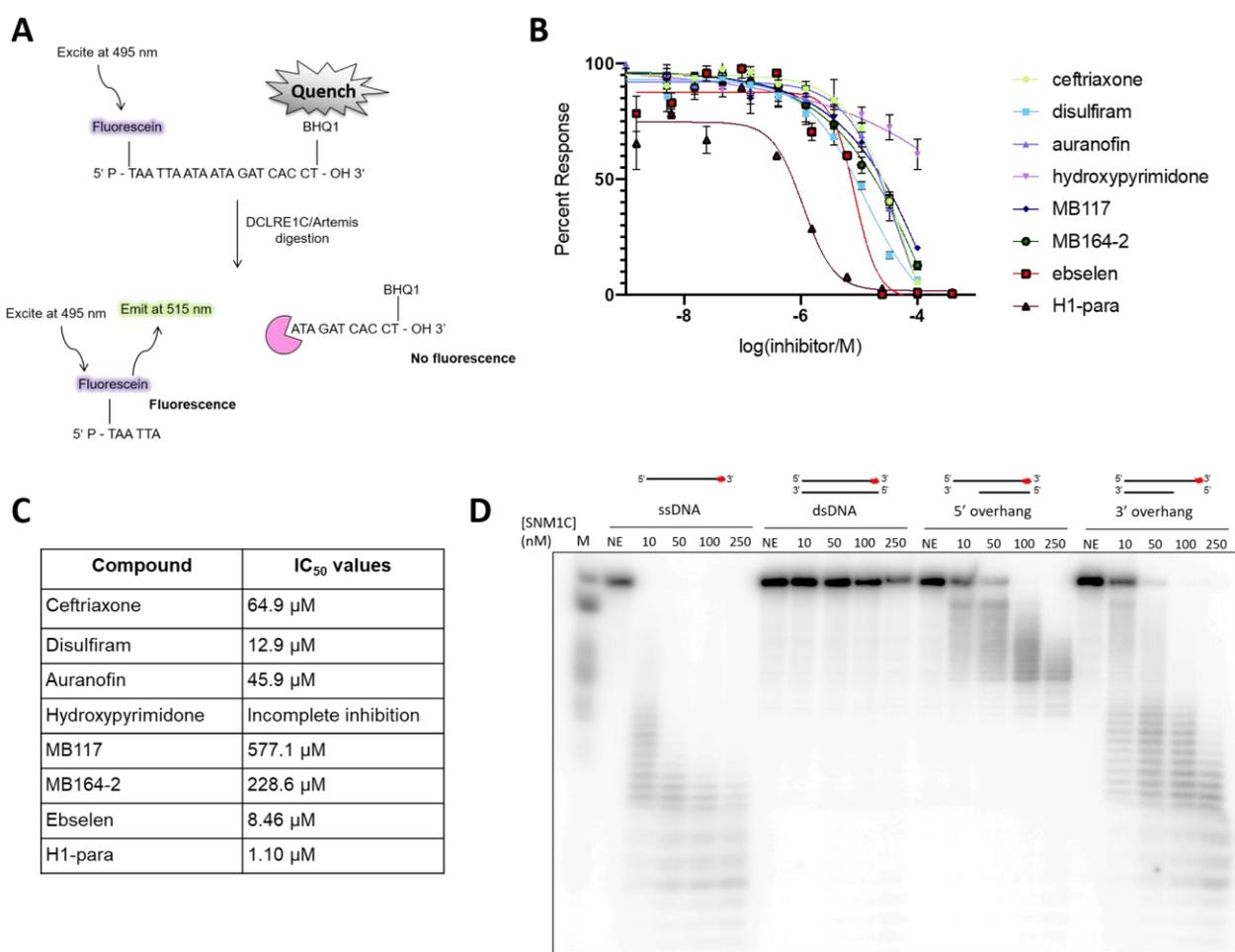


Figure 2. *In vitro* activity assay for DCLRE1C. **(A)** High throughput fluorescent assay, observing the release of the fluorescein-conjugated T after cleavage. **(B)** Inhibition curves of DCLRE1C upon addition of small molecules. **(C)** Table showing the IC₅₀ values for DCLRE1C inhibition. **(D)** Gel-based activity assay for DCLRE1C showing the digestion of various DNA substrates.

Chemical Matter

We intend to target DCLRE1C at 2 sites: the active site and the second metal binding site (zinc finger like motif) in the β -CASP domain. We identified two compounds that chelates the metal in the active site (**Fig 3**). We have fortuitously identified ceftriaxone (a β -lactam anti-bacterial compound) as a weak inhibitor (IC₅₀ = 64.9 μ M). The DCLRE1C ceftriaxone bound structure shows that most of the interaction happens in the metal centre of the active site between the cyclic 1,2 diamide functional group of ceftriaxone and the nickel ion (**Fig 1B**).

The second sets of compounds could potentially target the zinc finger motif. We found both ebselen and disulfiram (IC₅₀ 8.46 μ M and 12.9 μ M respectively) are better at inhibiting DCLRE1C compared to Auranofin with IC₅₀ 45.9 μ M.

Ceftriaxone	IC ₅₀ =64.9 μ M	
-------------	--------------------------------	--

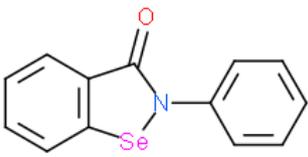
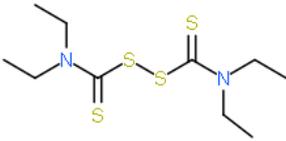
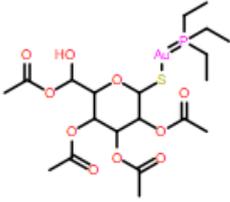
Ebselen	IC ₅₀ =8.46μM	
Disulfiram	IC ₅₀ =12.9μM	
Auranofin	IC ₅₀ =45.9μM	

Figure 3. A table of initial compound hits for DCLRE1C with IC₅₀ derived from the high throughput fluorescence assay.

IMPORTANT: Please note that the existence of small molecules within this TEP indicates only that chemical matter might bind to the protein in potentially functionally relevant locations. The small molecule ligands are intended to be used as the basis for future chemistry optimisation to increase potency and selectivity and yield a chemical probe or lead series. As such, the molecules within this TEP should not be used as tools for functional studies of the protein, unless otherwise stated, as they are not sufficiently potent or well-characterised to be used in cellular studies.

Future Plans

We are currently in the process of pilot screening fragments for Artemis in collaboration with the EU-OPENSREEN-Drive.

We are currently testing and developing Artemis inhibitors in collaboration with the McHugh lab (Department of Oncology, Oxford) and the Schofield lab (Department of Chemistry, Oxford).

CONCLUSION

Radiotherapy remains a mainstay of cancer therapy. The effectiveness of radiotherapy relies on inducing DNA double-strand breaks (DSBs) that contain complex, chemically modified ends that must be processed prior to repair (13, 14). Around 80% of double strand break DNA repair are processed through the NHEJ pathway in proliferating cells (15). DCLRE1C/Artemis is one of the key players in both NHEJ and V(D)J recombination pathways. Therefore, combining radiation therapy in conjunction with c-NHEJ inhibitors could give a synergistic effect for a more efficient cancer therapy.

We solved the first structure of human DCLRE1C catalytic domain that shows a novel zinc finger like motif in the β-CASP domain which could provide a second small molecule target site. We provide a robust expression, purification, and crystallisation method for DCLRE1C catalytic domain. This TEP provides a platform for high-throughput DCLRE1C inhibitor screen through *in vitro* biochemical assays and crystallography.

TEP IMPACT

Publications arising from this work:

- To be published: Yuliana Yosaatmadja¹, Hannah T Baddock², Joseph A Newman¹, Marcin Bielinski³, Angeline E Gavard¹, Solene Goubin¹, Shubhashish M M Mukhopadhyay¹, Adam A Dannerfjord¹, Chris J Schofield³, Peter J McHugh², Opher Gileadi¹. *The structure and function of the endonuclease Artemis. In Prep.*

FUNDING INFORMATION

The work was funded by Wellcome [106169/ZZ14/Z] and CRUK grant A24759.

ADDITIONAL INFORMATION

Structure Files

PDB ID	Structure Details
6TT5	The structure of DCLRE1C catalytic domain with Ni and Zn
7AF1	The structure of DCLRE1C catalytic domain with two Zn
7APV	The structure of DCLRE1C catalytic domain with Ceftriaxone

Materials and Methods

Protein Expression and Purification

Construct: DCLRE1CA-c401

DCLRE1C catalytic domain

Boundaries: residues 1-362

Vector: pFB-6HZB (Genbank: [JF682518.1](#))

Tag and additions: TEV-cleavable His6 tag at the N-terminus with the addition of Z-basic tag

Expression cell: Sf9

Protein sequence (Tag sequence is underlined, and the TEV protease cleavage site is indicated with *):

MGHHHHHSSGVDNKFNKERRRARREIRHLPNLNREQRRAFIRSLRDDPSQSANLLAEAKKLNDAPKGTENLYFQ*SMSS
FEGQMAEYPTISIDRFDRENLRARAYFLSHCHKDHMKGLRAPTLKRRLECSLKVYLYCSPVTKELLTSPKYRFWKKRIISIEIETP
TQISLVDEASGEKEEIVVTLTPAGHCPGSVMFLFQGNNGTVLYTGDFRLAQGEAARMELLHSGGRVKDIQSVYLDITFC DPR
FYQIPSREECLSGVLELVRSWITRSPYHVWLNCKAAYGYEYLFNLSEELGVQVHVNKLDMFRNMPEILHLLTDRNTQIHA
CRHPKAEYFQWSKLP CGITSRNRIPLHIISIKPSTMWFGERSRKTNVIVRTGESSYRACFSFHSSYSEIKDFLSYLCPVNAYPNVI
PVGTTMDKVVEILKPLCRS

Predicted mass: 50730.3

Protein after tag cleavage:

SMSSFEGQMAEYPTISIDRFDRENLRARAYFLSHCHKDHMKGLRAPTLKRRLECSLKVYLYCSPVTKELLTSPKYRFWKKRIISI
EIEPTQISLVDEASGEKEEIVVTLTPAGHCPGSVMFLFQGNNGTVLYTGDFRLAQGEAARMELLHSGGRVKDIQSVYLDITFC
DPRFYQIPSREECLSGVLELVRSWITRSPYHVWLNCKAAYGYEYLFNLSEELGVQVHVNKLDMFRNMPEILHLLTDRNT
QIHACRHPKAEYFQWSKLP CGITSRNRIPLHIISIKPSTMWFGERSRKTNVIVRTGESSYRACFSFHSSYSEIKDFLSYLCPVNA
YPNVIPVGTTMDKVVEILKPLCRS

Predicted mass: 41715.2

DNA sequence (ORF):

ATGGGCCACCATCATCATCATTCTTCTGGTGTGGATAACAAGTTCAACAAGGAGCGTCGAAGAGCTCGCCGTGAAA
TTCGCCATCTGCCAACCTGAACCGCAACAGCGTCGCGCATTATTCGCAGCCTGCGCGATGATCCGAGCCAGAGCGC
GAACCTGCTGGCGGAAGCGAAGAAGCTGAACGATGCGCAGCCGAAGGGTACAGAGAACCTGTACTTCCAATCCATGA
GTTCTTTGAGGGGCGAGATGGCCGAGTATCCAACATATCTCCATAGACCGCTTCGATAGGGGAGAACCTGAGGGCCCGCG
CCTACTTCTGTCCACTGCCACAAAGATCACATGAAAGGATTAAGAGCCCTACCTTGAAAAGAAGGTTGGAGTGCAG
CTTGAAGGTTTATCTATACTGTTACCTGTGACTAAGGAGTTGTTGTTAACGAGCCCGAAATACAGATTTTGAAGAAA
CGAATTATATCTATTGAAATCGAGACTCCTACCCAGATATCTTTAGTGGATGAAGCATCAGGAGAGAAGGAAGAGATT
GTTGTGACTCTTACCAGCTGGTCACTGTCCGGGATCAGTTATGTTTTATTTTCAGGGCAATAATGGAAGTGCCTGTA
CACAGGAGACTTCAGATTGGCGCAAGGAGAAGCTGCTAGAATGGAGCTTCTGCACTCCGGGGGCGAGAGTCAAAGACA
TCCAAAGTGTATTTGGATACTACGTTCTGTGATCCAAGATTTTACCAAATCCAAGTCGGGAGGAGTGTTAAGTGG
AGTCTTAGAGCTGGTCCGAAGCTGGATCACTCGGAGCCCGTACCATGTTGTGTGGCTGAACTGCAAAGCGGCTTATGG
CTATGAATATTTGTTACCAACCTTAGTGAAGAATTAGGAGTCCAGGTTTCATGTGAATAAGCTAGACATGTTTAGGAAC
ATGCCTGAGATCCTTCATCATCTCACAACAGACCGCAACACTCAGATCCATGCATGCCGGCATCCCAAGGCAGAGGAAT
ATTTTCAGTGGAGCAAATTACCCTGTGGAATTACTCCAGAAATAGAATTCCTCCACATAATCAGCATTAAAGCCATCC
ACCATGTGGTTTGGAGAAAGGAGCAGAAAAACAAATGTAATTGTGAGGACTGGAGAGAGTTCATACAGAGCTTGTTTT
TCTTTTCACTCCTACAGTGAGATTAAGATTTCTTGAGCTACCTCTGCTGTGAACGCATATCCAAATGTCATTCCA
GTTGGCACAACACTATGGATAAAGTTGTCGAAATCTTAAAGCCTTTATGCCGGTCTTGA

Expression and purification of DCLRE1C catalytic domain (aa 1-362)

Baculovirus generation was performed as previously described in Allerston *et.al* (16). The proteins were expressed in Sf9 cell at 2×10^6 cells/ mL infected with 1.5 mL of P2 virus. Infected Sf9 cells were harvested 70 h at 27 °C after infection by centrifugation at 900g for 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, and stored at -80 °C for later use.

Thawed cell aliquots were lysed by sonication. The lysate was clarified by centrifugation at 40,000g for 30 min, and the supernatant was passed through a 0.80 µm filter (Millipore) before being 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 and 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® 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).

The protein-containing fractions were pooled and dialysed overnight at 4 °C in SP buffer A and supplemented with recombinant tobacco etch virus (TEV) protease for cleavage of the 6His-ZB tag. The protein was subsequently loaded into an ion exchange column (HiTrap® SP FF GE Healthcare Life Sciences) pre-equilibrated in the SP buffer A to remove 6His-ZB tag and un-cleaved protein. The protein was eluted using a linear gradient of SP buffer B.

Artemis-containing fractions from the SP column elution were combined and concentrated down to 1 mL using a 30 kDa MWCO Amicon® 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).

DCLRE1C Full length

Boundaries: residues 1-692

Vector: pFB-CT10HF (Addgene <http://www.addgene.org/39191/>)

Tag and additions: TEV-cleavable His10 tag at the C-terminus with the addition of FLAG tag

Expression cell: Sf9

Protein sequence (Tag sequence is underlined, *:TEV protease cleavage site):

MSSFEGQMAEYPTISIDRFDRENLRARAYFLSHCHKDHMKGLRAPTLKRRLECSLKVYLYCSPVTKELLTSPKYRFWKKRIISIE
IETPTQISLVDEASGEKEEIVVTLTPAGHCPGSVMFLFQGNNGTVLYTGDFRLAQGEAARMELLHSGGRVKDIQSVYLDTTFC
DPRFYQIPSRREECLSGVLELVRSWITRSPYHVWLNCKAAYGYEYLFNLSEELGVQVHVNKLDMFRNMPEILHHLTDRNT
QIHACRHPKAEYFQWSKLP CGITSRNRIPLHIISIKPSTMWFGERSRKTNVIVVRTGESSYRACFSFHSSYSEIKDFLSYLCPVNA
YPNVIPVGTMDKVV EILKPLCRSSQSTEPKYKPLGKLRARTVHRDSEEDDYLFDDPLPIPLRHKVPYPETFFHPEVFSMTAVS
EKQPEKLRQTPGCCRAECMQSSRFTNFVDCEESNSESEEEVGIPASLQGDGLGSLVHLQKADGDVPQWEVFFKRNDITDES
ENFPSSTVAGGSQSPKLFSDSDGESTHISSQNSSQSTHITEQGSQGWDSQSDTVLLSSQERNSGDITSLDKADYRPTIKENIPA
SLMEQNVICPKDYS DLKSRDKDVTIVPSTGEPTT L SETHIPEEKSLNLTNADSQSSDFEVPSTPEAELPKREHLQYLYEKL
ATGESIAVKKRKC LLDTAENLYFQ*SHHHHHHHHHHDYKDDDDK

Expected mass: 81756

Protein sequence (after tag cleavage):

MSSFEGQMAEYPTISIDRFDRENLRARAYFLSHCHKDHMKGLRAPTLKRRLECSLKVYLYCSPVTKELLTSPKYRFWKKRIISIE
IETPTQISLVDEASGEKEEIVVTLTPAGHCPGSVMFLFQGNNGTVLYTGDFRLAQGEAARMELLHSGGRVKDIQSVYLDTTFC
DPRFYQIPSRREECLSGVLELVRSWITRSPYHVWLNCKAAYGYEYLFNLSEELGVQVHVNKLDMFRNMPEILHHLTDRNT
QIHACRHPKAEYFQWSKLP CGITSRNRIPLHIISIKPSTMWFGERSRKTNVIVVRTGESSYRACFSFHSSYSEIKDFLSYLCPVNA

YPNVIPVGTMDKVVVEILKPLCRSSQSTEPKYKPLGKLRARTVHRDSEEDDYLFDDPLIPLRHKVPYPETFHPEVFSMTAVS
EKQPEKLRQTPGCCRAECMQSSRFNFDCEESNSESEEEVGIPLASLQGLDLSVLHLQKADGDVPQWEVFFKRNDITDESL
ENFPSSTVAGGSQSPKLFSDSDGESTHISSQSSQSTHITEQGSQGWDSQSDTVLLSSQERNSSGDTSLDKADYRPTIKENIPA
SLMEQNVICPKDYSDLKSRDKDVTIVPSTGEPTLLSSETHIPEEKSLNLSNADSSQSSDFEVPSTPEAELPKREHLQLYEKL
ATGESIAVKRKRCSLLDTAENLYFQ

Predicted mass: 79302.5

DNA sequence (ORF):

ATGAGTTCTTTGAGGGGCGAGATGGCCGAGTATCCAACATCTCCATAGACCGCTTCGATAGGGAGAACCTGAGGGCC
CGCGCCTACTTCTGTCCCCTGCCACAAAGATCACATGAAAGGATTAAGAGCCCCTACCTTAAAAGAAGGTTGGAGT
GCAGCTTGAAGGTTTATCTATACTGTTACCTGTGACTAAGGAGTTGTTGTTAACGAGCCGAAATACAGATTTTGGAA
GAAACGAATTATCTATTGAAATCGAGACTCCTACCCAGATATCTTTAGTGGATGAAGCATCAGGAGAGAAGGAAGA
GATTGTTGTGACTCTTACCAGCTGGTCACTGTCCGGGATCAGTTATGTTTTTATTTACAGGGCAATAATGAACTGTCC
TGTACACAGGAGACTTCAGATTGGCGCAAGGAGAAGCTGCTAGAATGGAGCTTCTGCACTCCGGGGCAGAGTCAA
GACATCCAAAGTGTATTTGGATACTACGTTCTGTGATCCAAGATTTTACCAAATCCAAGTCGGGAGGAGTGTAA
GTGGAGTCTTAGAGCTGGTCCGAAGCTGGATCACTCGGAGCCCGTACCATGTTGTGTGGCTGAACTGCAAAGCGGCTT
ATGGCTATGAATATCTGTTACCAACCTTAGTGAAGAATTAGGAGTCCAGGTTTCATGTGAATAAGCTAGACATGTTTAG
GAACATGCCTGAGATCCTTCATCATCTCACAAACAGACCGCAACACTCAGATCCATGCATGCCGGCATCCCAAGGCAGAG
GAATATTTTTCAGTGGAGCAAATTACCCTGTGGAATTACTTCCAGAAATAGAATTCCTCCACATAATCAGCATTAAAGCC
ATCCACCATGTGGTTTGGAGAAAGGAGCAGAAAAACAAATGTAATTGTGAGGACTGGAGAGAGTTCATACAGAGCTTG
TTTTCTTTTCACTCCTCCTACAGTGAGATTAAGATTTCTTGAGCTACCTCTGTCTGTGAACGCATATCCAAATGTCATT
CCAGTTGGCACAACCTATGGATAAAGTTGTGCAAACTTAAAGCCTTTATGCCGGTCTTCCCAAAGTACGGAGCCAAAGT
ATAAACCACTGGGAAAACCTGAAGAGAGCTAGAACAGTTACCGAGACTCAGAGGAGGAAGATGACTATCTCTTTGATG
ATCCTCTGCCAATACCTTTAAGGCACAAAGTTCATACCCGAAACTTTTACCCTGAGGTATTTCAATGACTGCAGTAT
CAGAAAAGCAGCCTGAAAACTGAGACAAACCCAGGATGCTGCAGAGCAGAGTGTATGCAGAGCTCTCGTTTACAA
ACTTTGTAGATTGTGAAGAATCCAACAGTGAAGTGAAGAAGAAGTAGGAATCCAGCTTCACTGCAAGGAGATCTGG
GCTCTGTACTTCACCTGCAAAAGGCTGATGGGGATGTACCCAGTGGGAAGTATTCTTTAAAGAAATGATGAAATCAC
AGATGAGAGTTTGGAAAACCTTCCCTTCCACAGTGGCAGGGGGATCTCAGTACCAAAGCTTTTTCAGTACTCTGAT
GGAGAATCAACTCACATCTCCTCCAGAAATTCTCCAGTCAACACACATAACAGAACAAGGAAGTCAAGGCTGGGACA
GCCAATCTGATACTGTTTTGTTATCTTCCCAAGAGAGAAACAGTGGGGATATTACTTCTTGACAAAGCTGACTACAG
ACCAACAATCAAAGAGAATATTCCTGCCTCTCTCATGGAACAAAATGTAATTTGCCCAAAGGATACTTACTCTGATTTGA
AAAGCAGAGATAAAGATGTGACAATAGTTCCTAGTACTGGAGAACCAACTACTCTAAGCAGTGAACACATATACCCG
AGGAAAAAGTTTGCTAAATCTTAGCACAAATGCAGATTTCCAGAGCTCTTCTGATTTTGAAGTTCCCTCAACTCCAGAA
GCTGAGTTACCTAAACGAGAGCATTTACAATATTTATATGAGAAGCTGGCAACTGGTGAAGATATAGCAGTCAAAAA
AGAAAATGCTCACTTAGATACCGCAGAGAACCTTACTTCCAATCGCACCATCATCACCATCACCATCACCACCATGA
TTACAAGGATGACGACGATAAGTGA

Expression and purification of full-length DCLRE1C (aa 1-692)

The baculovirus generation and expression of the full length DCLRE1C was performed in a similar manner as the catalytic domain. Instead of infection with 1.5 mL of P2 Virus, 3.0 mL of P2 virus was used to infect SF9 cell at 2×10^6 cells/ mL for the expression of the full-length construct

Cell harvest and the initial IMAC purification step was performed as described above. Following IMAC purification TEV cleavage overnight in dialysis buffer (50 mM HEPES pH 7.5, 0.5 M NaCl, 5% glycerol and 1 mM TCEP). The protein was then passed through a 5 mL Ni-sepharose column, and the flowthrough fractions were collected. DCLRE1C protein was then concentrated on a centrifugal concentrator (Amicon®, MWCO 30 kDa) before loading on a Superdex S200 HR 16/60 gel filtration column in dialysis buffer. Fractions containing purified DCLRE1C protein were pooled and concentrated to 10 mg/mL.

Protein crystallisation and compound soaking

The catalytic domain of DCLRE1C was crystallised using a sitting drop vapour diffusion method by mixing 50 nL protein with 50 nL crystallisation solution comprising 0.2 M ammonium chloride, 20% PEG 3350 and 20 nL of crystal seed solution obtained from previous crystallisation experiments. The crystals were grown at 4 °C. Crystals grew after 1 day and reached maximum size within 1 week.

Artemis protein crystal was soaked with 5 mM ceftriaxone in crystallisation solution supplemented with the addition of 20% ethylene glycol solution for one hour prior to flash freezing in liquid nitrogen. Data was collected at Diamond Light Source beamlines I03, I04 and I24, unless otherwise stated.

Structure determination

PDB Code 6TT5

Data Collection: Data were collected to 1.50 Å resolution at Diamond light source beamline I04 and was processed using DIALS.

Data Processing: The structure was solved by molecular replacement using the program PHASER and the structure of DCLRE1A (**5Q2A**) as a search model. Refinement was performed using REFMAC to a final Rfactor = 17 %, Rfree = 19 %.

PDB Code 7AF1

Data Collection: Data were collected to 1.70 Å resolution at Diamond light source beamline I24 and was processed using DIALS.

Data Processing: The structure was solved by molecular replacement using the program PHASER and the structure of DCLRE1C (**6TT5**) as a search model. Refinement was performed using REFMAC to a final Rfactor = 19 %, Rfree = 21 %.

PDB Code 7APV

Data Collection: Data were collected to 1.95 Å resolution at Diamond light source beamline I03 and was processed using DIALS.

Data Processing: The structure was solved by molecular replacement using the program PHASER and the structure of DCLRE1C (**6TT5**) as a search model. Refinement was performed using REFMAC to a final Rfactor = 18 %, Rfree = 23 %.

Assays

Gel-based nuclease assays

Standard nuclease assays were carried out in 10 µL reactions containing 20 mM HEPES-KOH, pH 7.5, 50 mM KCl, 10 mM MgCl₂, 0.05% Triton X-100, 5% glycerol, 0.5 mM TCEP and the indicated amount of Artemis enzyme. Reactions were started by the addition of 3' radiolabelled (α-³²P-dATP) DNA substrate (10 nM) incubated at 37 °C for the indicated time period and quenched by the addition of 10 µL stop solution (95% formamide, 10 mM EDTA, 0.25% xylene cyanol, 0.25% bromophenol blue) and incubating at 95 °C for 3 minutes.

Reactions were analysed by 20% denaturing polyacrylamide gel electrophoresis (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 phosphorimager screen and scanned using a Typhoon 9500 instrument (GE).

High-throughput fluorescence-based nuclease assay.

The protocol of Lee *et al* [39] was adapted for structure-specific endonuclease activity. Briefly, a ssDNA substrate was utilised, containing a 5' fluorescein-conjugated T and a 3' BHQ-1-conjugated T (5'- [FITC] TAA TTA ATA ATA GAT CAC CT [BHQ1] - 3'). Prior to endonucleolytic incision the intact substrate does not fluoresce

due to the proximity of the BHQ-1 to the fluorescein. However, following endonucleolytic incision by Artemis/DCLRE1C there is uncoupling of the fluorescein-T and the BHQ-1 and a concomitant increase in fluorescence.

The fluorescence assays were performed at 37 °C in 384-well format (black plate, clear bottom) using PheraStar plate reader by BMG Labtech. The fluorescence spectra were measured using a PHERAstar FSX (excitation: 495 nm; emission: 525 nm) with readings taken every 150 seconds. Reactions were performed in a final volume of 25 µL in buffer containing 20 mM HEPES-KOH, pH 7.5; 50 mM KCl, 10 mM MgCl₂; 0.05% Triton-X100; 0.5 mM TCEP, 5% glycerol). To determine IC₅₀, inhibitors (at increasing concentrations) were incubated with the protein for 10 minutes at room temperature, before the reaction was started with the addition of DNA substrate.

References

1. 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., and De Villartay, J. P. (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
2. Lieber, M. R. (2011) The mechanism of DSB repair by the NHEJ. *Annu. Rev. Biochem.* **79**, 181–211
3. Gu, J., Li, S., Zhang, X., Wang, L. C., Niewolik, D., Schwarz, K., Legerski, R. J., Zandi, E., and Lieber, M. R. (2010) DNA-PKcs regulates a single-stranded DNA endonuclease activity of Artemis. *DNA Repair (Amst)*. **9**, 429–437
4. Srivastava, M., and Raghavan, S. C. (2015) DNA double-strand break repair inhibitors as cancer therapeutics. *Chem. Biol.* **22**, 17–29
5. Pannunzio, N. R., Watanabe, G., and Lieber, M. R. (2018) Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *J. Biol. Chem.* **293**, 10512–10523
6. Malu, S., De Ioannes, P., Kozlov, M., Greene, M., Francis, D., Hanna, M., Pena, J., Escalante, C. R., Kurosawa, A., Erdjument-Bromage, H., Tempst, P., Adachi, N., Vezzone, P., Villa, A., Aggarwal, A. K., and 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
7. Niewolik, D., Pannicke, U., Lu, H., Ma, Y., Wang, L. C. V., Kulesza, P., Zandi, E., Lieber, M. R., and Schwarz, K. (2006) DNA-PKcs dependence of artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs. *J. Biol. Chem.* **281**, 33900–33909
8. Niewolik, D., Peter, I., Butscher, C., and 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
9. Li, S., Chang, H. H., Niewolik, D., Hedrick, M. P., Pinkerton, A. B., Hassig, C. A., Schwarz, K., and Lieber, M. R. (2014) Evidence that the DNA endonuclease ARTEMIS also has intrinsic 5'-exonuclease activity. *J. Biol. Chem.* **289**, 7825–7834
10. Li, L., Moshous, D., Zhou, Y., Wang, J., Xie, G., Salido, E., Hu, D., and Cowan, M. J. (2002) A Founder Mutation in Artemis, an SNM1-Like Protein, Causes SCID in Athabascan-Speaking Native Americans. *J. Immunol.* **168**, 6323–6329
11. Felgentreff, K., Lee, Y. N., Frugoni, F., Du, L., Van Der Burg, M., Giliani, S., Tezcan, I., Reisli, I., Mejstrikova, E., De Villartay, J. P., Sleckman, B. P., Manis, J., and Notarangelo, L. D. (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
12. de Villartay, J. P., Shimazaki, N., Charbonnier, J. B., Fischer, A., Mornon, J. P., Lieber, M. R., and Callebaut, I. (2009) A histidine in the β-CASP domain of Artemis is critical for its full in vitro and in vivo functions. *DNA Repair (Amst)*. **8**, 202–208
13. Jekimovs, C., Bolderson, E., Suraweera, A., Adams, M., O'Byrne, K. J., and Richard, D. J. (2014) Chemotherapeutic compounds targeting the DNA double-strand break repair pathways: The good,

- the bad, and the promising. *Front. Oncol.* **4 APR**, 1–18
14. Shibata, A., and 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
 15. Mao, Z., Bozzella, M., Seluanov, A., and Gorbunova, V. (2008) Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)*. **7**, 1765–1771
 16. Allerston, C. K., Lee, S. Y., Newman, J. A., Schofield, C. J., Mchugh, P. J., and 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

We respectfully request that this document is cited using the DOI value as given above if the content is used in your work.