Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition

Levon Halabelian, Mani Ravichandran, Yanjun Li, Hong Zeng, Anjana Rao, L. Aravind and Cheryl H Arrowsmith

Embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein (HMCES) can covalently cross-link to abasic sites in single-stranded DNA at stalled replication forks to prevent genome instability. Here, we report crystal structures of the human HMCES SOS response-associated peptide domain (SRAP) domain in complex with DNA-damage substrates, including HMCES cross-linked with an abasic site within a 3′ overhang DNA. HMCES interacts with both single-strand and duplex segments of DNA, with two independent duplex DNA interaction sites identified in the SRAP domain. The HMCES DNA-protein cross-link structure provides structural insights into a novel thiazolidine covalent interaction between the DNA abasic site and conserved Cys 2 of HMCES. Collectively, our structures demonstrate the capacity for the SRAP domain to interact with a variety of single-strand- and double-strand-containing DNA structures found in DNA-damage sites, including 5′ and 3′ overhang DNAs and gapped DNAs with short single-strand segments.

DNA bases are constantly damaged by factors such as reactive oxygen species, chemotoxic agents, ionizing radiation and UV radiation, and are subject to physiological modification by enzymes such as activation-induced cytidine deaminase, DNA methylases and the ten-eleven translocation enzymes. These types of DNA alterations are primarily repaired by the base excision repair pathway, which is initiated by DNA glycosylases that recognize and cleave damaged or modified bases, creating apurinic or apyrimidinic sites (AP sites), also known as abasic sites. HMCES was recently reported to recognize and covalently cross-link to abasic sites in single-stranded DNA (ssDNA), generated by uracil-DNA glycosylase (UDG) at stalled replication forks. The authors suggested that these DNA-protein cross-link (DPC) intermediates prevented ssDNA breaks that may consequently occur on cleavage by AP endonucleases, which could subsequently be repaired through error-prone pathways.

Human HMCES has a highly conserved amino (N)-terminal SOS response-associated peptide domain (SRAPd) that is widely found in bacteria and eukaryotes, with a sporadic presence in certain bacteriophages and archaea. Animal SRAP proteins have an additional carboxy (C)-terminal disordered extension with multiple copies of the PCNA-interacting motif (PIP)46. Gene-neighborhood analysis in prokaryotes identified SRAPd as a novel component of predicted catalytic residues, namely Cys 2, Glu 127 and His 210, which form a covalent linkage to the deoxyribose at the abasic site of the DNA4. Here, we present the crystal structures of the human HMCES SRAP domain in DNA-free form and in complex with several 3′ overhang DNAs, including a DPC structure with an abasic site at the 3′ overhang. Through structure analysis and mutagenesis experiments we delineate the key residues of HMCES that are involved in DNA binding, revealing considerable flexibility in substrate recognition.

Results

Crystal structures of the human HMCES SRAP domain in complex with 3′ overhang DNA. To better understand the mechanism of HMCES association with DNA, we crystallized the human HMCES SRAPd (residues 2–270) in its DNA-free form (Apo_SRAPd) and in complex with several DNA-damage substrates containing 3′ overhangs of different lengths. The crystal structure of SRAPd in complex with duplex DNA containing a three-nucleotide overhang at the 3′ end (referred to here as SRAPd_3nt) revealed SRAPd binding to two DNA molecules: DNA-A interacts via the 3′ overhang and another molecule (DNA-B) via the blunt-end (Fig. 1a). Both DNA interaction surfaces are highly conserved.

SRAPd interacts with the 3′ overhang of DNA-A through a hydrophobic shelf created by Trp81 and Phe92, which form π-stacking interactions with the duplex segment of DNA at the ssDNA-dsDNA junction, which is referred to here as dsDNA-interaction site A (Fig. 1b,c). At the same time, Arg106 inserts into the minor groove of dsDNA, further stabilizing the complex. In Apo_SRAPd, Arg106 is stacked onto Trp81 instead of the nucleobase (Fig. 2a). The ssDNA segment of the 3′ overhang is sharply bent by approximately 90 degrees and lies in a narrow, positively charged cleft. The ssDNA-binding cleft includes conserved Arg98 and Arg212, which form salt-bridges with the phosphate backbone of ssDNA (Fig. 1b,d). Alanine substitutions

*Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada. †Division of Signaling and Gene Expression, La Jolla Institute for Immunology, La Jolla, CA, USA. ‡Department of Pharmacology, University of San Diego, La Jolla, CA, USA. §Moores Cancer Center, University of San Diego, La Jolla, CA, USA. ††Sanford Consortium for Regenerative Medicine, La Jolla, CA, USA. ||National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA. ¶National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. #Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. ©Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. *e-mail: cheryl.arrowsmith@uhnresearch.ca
of either of these arginine residues severely hinder ssDNA binding (Supplementary Fig. 1), and are consistent with gel-shift assays reported by Mohni et al.4.

The pocket housing the catalytic triad accommodates the 3’-hydroxyl of the ssDNA overhang (Fig. 1d). Interestingly, mutation of individual catalytic triad residues to alanine yielded SRAPd variants with higher affinity for ssDNA compared to wild-type protein, suggesting a role other than simply DNA binding (Supplementary Fig. 1).

The SRAPd_3nt structure also revealed that the blunt-end of DNA-B interacts with SRAPd via dsDNA-interaction site B, composed of residues Gly 3, Arg 4, Pro 46, Asp 47, Trp 128 (Fig. 1e). This interaction surface represents a potential binding site for 5’ overhang DNA, as SRAPd was shown to bind both 5’ and 3’ overhangs with similar affinities4. This dsDNA-interaction site B accounts for the remaining residues, which are highly conserved in SRAP domains across all superkingdoms of life and phages5, suggesting that it is a universal functional feature of this domain. It is immediately adjacent to the catalytic triad and forms a contiguous, similarly charged surface with the ssDNA-binding site (Fig. 1b). These features suggested that dsDNA-interaction site B may also be able to accommodate ssDNA extending from a longer 3’ overhang substrate bound to the dsDNA-interaction site A.

To address this question, we determined the crystal structure of SRAPd with DNA containing a six-nucleotide overhang at the 3’ end ssDNA-binding cleft and catalytic triad site.

![Fig. 1](image_url)
Articles

Nature Structural & Molecular Biology

Although SRAPd has nearly tenfold higher affinity for ssDNA compared to dsDNA (Fig. 2d), the longer 3′ overhang did not displace the blunt-end-interacting DNA-B from its dsDNA-interaction site B. Instead, the extra single-strand bases protrude out of the catalytic triad pocket (Fig. 2b), and (Supplementary Fig. 2), suggesting that the dsDNA-interaction site B has specifically evolved to bind duplex DNA and may form the binding site for 5′ overhang DNA structures as well. Nevertheless, given that DNA is a mediator of the crystal lattice in this crystal form, we cannot entirely rule out that a longer ssDNA might occupy the dsDNA-interaction site B in the absence of a competing duplex DNA.

HMCES prefers ssDNA and 3-nucleotide gap DNA over intact dsDNA. In SRAPd_3nt, the distance between the 3′ end of DNA-A and the 5′ end of DNA-B at the catalytic triad is around 3.2 Å, which is sufficient to accommodate a phosphate group linking the two substrates together (Fig. 2c). Consistent with our observations, the affinity of SRAPd to dsDNA with a 3-nucleotide gap is approximately sevenfold higher than intact dsDNA of the same sequence (Fig. 2d). These data suggest the potential for binding other types of gapped DNA structures that form during DNA repair (Fig. 2e), such as nucleotide excision repair intermediates. The Trp81Glu substitution at dsDNA-interaction site A, and Arg4Ala substitution at dsDNA-interaction site B, severely hindered SRAPd binding to 3-nucleotide gap DNA, suggesting that both interaction sites are crucial for substrate DNA binding (Supplementary Fig. 1).

Crystal structure of the human HMCES SRAP domain cross-linked to DNA abasic site. To further investigate the covalent interaction of HMCES with a DNA abasic site, we crystallized

Fig. 2 | SRAPd interaction with potential DNA-damage repair substrates. a, Crystal structure of Apo_SRAPd in magenta, superposed with SRAPd_3nt in cyan, at the dsDNA-interaction site A. DNA-A is shown as a ribbon representation in orange. b, Crystal structure of SRAPd_6nt in complex with a six-nucleotide 3′ overhang. SRAPd is shown in surface representation in gray. The six-nucleotide overhang at the dsDNA-interaction site A is shown as a stick model in brown superposed with the SRAPd_3nt three-nucleotide 3′ overhang, in green. The electron density for the last two nucleotides in SRAPd_6nt was not resolved and not modeled (indicated by dashed lines). DNA-B at the dsDNA-interaction site B is shown in yellow. c, The 3′ and 5′ ends of two DNA molecules at the catalytic triad site of SRAPd_3nt are in close enough proximity to be linked by a phosphate group, shown in black color for the purposes of illustration. d, Fluorescence polarization (FP) DNA-binding affinities of HMCES SRAPd to ssDNA, dsDNA and dsDNA containing a 3-nucleotide gap (3-nt gap DNA). Experiments were performed in triplicates and data are represented as mean ± s.d. e, A model illustrating the potential DNA-damage substrates that can be recognized by HMCES. Red line represents the abasic site in DNA. Source data for panel d are available online.
the HMCES SRAP domain with 3’ overhang DNA (similar to that in SRAPd_6nt) having an abasic site at position 9 (AP9) of the longer strand (referred to here as SRAPd_DPC) (Fig. 3a). To generate a physiologically relevant, reactive aldehydic form of the abasic site capable of cross-linking with HMCES, we designed the 3’ overhang to have a deoxyuridine (dU) at position 9 of the longer DNA strand, which is predicted to bind immediately adjacent to the Cys 2. We then treated this DNA with UDG to generate AP9 DNA. For the cross-link reaction to proceed between the AP9 and Cys 2 of HMCES, the N-terminal methionine needs to be removed to expose the NH₂ of Cys 2. For the cross-link reaction to proceed between the AP9 and Cys 2 of HMCES, the N-terminal methionine needs to be removed to expose the NH₂ of Cys 2. Mass spectrometric analysis of our purified C-terminally His-tagged SRAPd protein showed that the N-terminal methionine (Met 1) was cleaved, either cotranslationally, catalyzed by *Escherichia coli* methionyl-aminopeptidase, or due to its predicted autopeptidase activity (Supplementary Fig. 3 and Supplementary Table 1), yielding a catalytically active form of HMCES. Incubation of this active SRAPd with AP9 DNA yielded cross-linked SRAPd_DPC (Supplementary Fig. 3 and Supplementary Table 1).

The SRAPd_DPC structure was refined to 2.2-Å resolution and is isomorphous with both the SRAPd_3nt and SRAPd_6nt structures. Continuous electron density was observed between Cys 2 of HMCES and the AP9, confirming the covalent cross-link (Fig. 3a,b). The Cys 2-AP9 cross-link was modeled at full occupancy and shows approximately 20% higher B-factor values compared with its surrounding residues (Table 1). The electron density map for the remaining three nucleotides downstream of the abasic site at the 3’ end was not resolved. Notably, the SRAPd_DPC structure suggests a model in which the terminal Cys 2 of HMCES reacts with the ring-opened aldehyde form of the abasic deoxyribose (AP9) to form a thiazolidine DNA-protein cross-link (Fig. 3a–c). This cross-link entails an approximately 120° rotation around AP9 and possible hydrogen bonding between the 4′-hydroxyl of AP9 and the tau-nitrogen of His 210 (Fig. 3b,d).

**Discussion**

HMCES was recently reported to promote genome stability by shielding abasic sites from error-prone repair pathways at stalled
replication forks. In particular, Mohni et al.\(^4\) showed that HMCES forms DPC intermediates with abasic sites in ssDNA generated by UDG, which is a monofunctional glycosylase that cannot cleave ssDNA. However, other variants of damaged bases require the use of bifunctional glycosylases with both glycosylase and lyase activities, such as NEIL3, which is a single-strand-specific glycosylase with a limited lyase activity able to cleave ssDNA 3′ to an abasic site to generate a 3′ overhang\(^{11}\). Our structures confirm this cross-link and reveal it to be a thiazolidine ring that involves both the side chain sulfur and the NH\(_2\) of Cys 2 that is exposed by the peptidolytic removal of N-terminal methionine. Cys 2, along with Glu 127 and His 210, forms a predicted catalytic triad that is situated in a pocket characteristic of SRAP domains\(^5\). Thus, our structure reveals that, in addition to predicted autopeptidase activity, this pocket is also required for accommodation of the DPC. It also suggests that HMCES can recognize and covalently cross-link to DNA abasic sites at cleaved 3′ ends, which may shield them from further processing by exonucleases and regulate the choice between different DNA-repair pathways. The presence of two distinct dsDNA-interaction sites provides HMCES with flexibility to interact with abasic sites located at both 5′ and 3′ overhangs (Fig. 2e). Because HMCES binds ssDNA, 3′ and 5′ overhangs, and 3-nucleotide gap DNA, it could be involved in a variety of DNA-repair pathways other than at stalled replication forks. For example, it could potentially be involved in physiological genome rearrangements such as class switch recombination in lymphocytes.

Our SRAPd structures also shed light on other proposed activities of HMCES. Proteomics studies using dsDNA baits with modified cytosines identified HMCES as a reader for oxidized 5-methyl-cytosines (oxi-mC) containing duplex DNA\(^{12}\). The SRAPd only contacts one base pair at the ssDNA-dsDNA junction (Fig. 1b); hence SRAPd of HMCES could potentially recognize a single oxi-mC either at this junction or alternatively in single-strand regions. Taken together, our structures support an important role for HMCES in recognizing and sensing flapped and gapped DNA-damage products and describe the covalent interaction of HMCES with DNA abasic sites, revealing its broad substrate recognition spectrum.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at [https://doi.org/10.1038/s41594-019-0246-6](https://doi.org/10.1038/s41594-019-0246-6).

Received: 10 January 2019; Accepted: 8 May 2019; Published online: 24 June 2019

---

**Table 1 | Data collection and refinement statistics**

|                     | Apo_SRAPd (PDB 5KO9) | SRAPd_3nt (PDB 6OEB) | SRAPd_6nt (PDB 6OEA) | SRAPd_DPC (PDB 6O7) |
|---------------------|----------------------|----------------------|----------------------|----------------------|
| **Data collection** |                      |                      |                      |                      |
| Space group         | I2                   | I2                   | I2                   | I2                   |
| a, b, c (Å)         | 80.10, 44.74, 82.90  | 55.72, 51.15, 149.21 | 55.86, 52.06, 148.30 | 55.51, 51.53, 149.72 |
| α, β, γ (°)         | 90.00, 107.15, 90.00 | 90.00, 92.76, 90.00  | 90.00, 93.11, 90.00  | 90.00, 92.72, 90.00  |
| Resolution (Å)      | 48.37-1.5 (1.53-1.50) | 48.38-2.10 (2.16-2.10) | 49.11-2.10 (2.16-2.10) | 48.72-2.20 (2.27-2.20) |
| Rmerge              | 0.039 (0.61)         | 0.069 (0.71)         | 0.059 (0.74)         | 0.045 (0.70)         |
| I/σ(I)              | 15.9 (1.9)           | 10.2 (1.4)           | 15.0 (2.2)           | 12.3 (1.3)           |
| CC\(_{1/2}\)         | 0.667                | 0.814                | 0.926                | 0.818                |
| Completeness (%)    | 97 (93.2)            | 99.8 (99.7)          | 99.9 (99.9)          | 96.8 (98.0)          |
| Redundancy          | 3.8 (3.4)            | 4.4 (4.3)            | 6.6 (6.6)            | 3.2 (3.4)            |
| **Refinement**      |                      |                      |                      |                      |
| Resolution (Å)      | 48.37-1.5            | 48.38-2.10           | 49.11-2.10           | 48.72-2.20           |
| No. reflections     | 41504                | 24718                | 25066                | 20951                |
| R\(_{work}\)/R\(_{free}\) | 0.181/0.209        | 0.205/0.235          | 0.208/0.242          | 0.194/0.216          |
| No. atoms           | 2,292                | 2,492                | 2,457                | 2,402                |
| Protein             | 2,062                | 2,078                | 2,056                | 2,046                |
| DNA                 | -                    | 302                  | 324                  | 282                  |
| Cys2-AP9 cross-link | -                    | -                    | -                    | 17                   |
| Water               | 186                  | 77                   | 46                   | 33                   |
| B factors           | 25.3                 | 50.8                 | 60.5                 | 62.5                 |
| Protein             | 24.6                 | 50.4                 | 59.6                 | 62.4                 |
| DNA                 | -                    | 53.7                 | 66.7                 | 62.7                 |
| Cys2-AP9 cross-link | -                    | -                    | -                    | 72.9                 |
| Water               | 31.5                 | 47.5                 | 54.8                 | 54.4                 |
| R.m.s. deviations   |                      |                      |                      |                      |
| Bond lengths (Å)    | 0.011                | 0.008                | 0.008                | 0.007                |
| Bond angles (°)     | 1.499                | 1.526                | 1.563                | 1.509                |

\(^*\)Data are from one crystal. \(^*\)Values in parentheses are for highest-resolution shell.
Acknowledgements

We are grateful to H. Wyatt for fruitful discussions, advice from P.J. Brown and W. Tempel on interpretation of DNA-protein cross-link chemistry and S. Ackloo on mass spectrometry data analysis. Special thanks to S. Duan for preparing the DNA abasic site digestion. We also thank U. Chinte, J. Chrzas, N. Duke and Z. Jin from SERCAT 221D-D beamline for collecting the initial SRAPD_DPC datasets. This research used resources of the Advanced Light Source, which is a Department of Energy Office of Science User Facility under contract no. DE-AC02-05CH11231. Results shown in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center (SBC) at the Advanced Photon Source. SBC-CAT is operated by UChicago Argonne, LLC, for the US Department of Energy, Office of Biological and Environmental Research under contract no. DE-AC02-06CH11357. This work is based on research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (no. P41 GM103403). The Pilatus 6M detector on beamline 24-ID-C is funded by a NIH Office of Research Infrastructure Programs High End Instrumentation grant (no. S10 RR029205). The Structural Genomics Consortium is a registered charity (no: 1097737) that receives funds from AbbVie; Bayer Pharma AG; Boehringer Ingelheim; Canada Foundation for Innovation; Eshelman Institute for Innovation, Genome Canada through Ontario Genomics Institute (no. OGI-055); Innovative Medicines Initiative (EU/EFPIA) (no. ULTRA-DD: 115766); Janssen, Merck & Co.; Novartis Pharma AG; Ontario Ministry of Research Innovation and Science; Pfizer, São Paulo Research Foundation-FAPESP, Takeda and the Wellcome Trust. This research is also supported by the Canadian Institutes of Health Research (no. FDN154328) and Natural Sciences and Engineering Research Council (no. RGPIN-2015-05939) to C.H.A., intramural funds of the National Library of Medicine, NIH, to L.A. and the National Cancer Institute (no. R35 CA210043) to A.R.

Author contributions

L.H. performed the experiments. Y.L., M.R. and H.Z. cloned, expressed and purified the proteins. A.R. and C.H.A. conceived the project. L.H., L.A., A.R. and C.H.A. contributed to experimental design and review of data. L.H. and C.H.A. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0246-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.H.A.

Peer review information: Anke Sparmann was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Protein expression and purification. Wild-type and mutant HMCES variants were subcloned into pNIC-CH vector by modifying the C-terminal tag with a tobacco etch virus (TEV) cleavable C-terminal His6-tag, and were expressed in E. coli Rosetta. The recombinant proteins were first purified by nickel-affinity chromatography and, after TEV cleavage of the His6-tag, by anion exchange and gel-filtration chromatography using an S200 column. Purified SRAPd was concentrated to ~20 mg ml⁻¹ in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM tris(2-carboxyethyl)phosphine (TCEP). The sequences for all cloned constructs were verified by sequencing, and the corresponding molecular weights for all purified constructs were verified by liquid chromatography–mass spectrometry (LC–MS).

Crystalization and structural determination. Apo_SRAPd was crystallized using the sitting-drop vapor-diffusion method by mixing a 1:1 ratio of protein and reservoir solution containing 0.1 M Bis-Tris propane, 2% taccimide and 20% (w/v) PEG 3350. DNA used for co-crystallization was purchased from Integrated DNA Technologies, Inc. For SRAPD_DPC structure, a 12-residue ssDNA containing a deoxyuridine (dU) at position 9 (5′-CCACAGCTTGGT-3′) was first incubated with UDG for 1 h at 37 °C, followed by heat treatment at 95 °C for 10 min to inactivate the enzyme. Reaction products were immediately extracted with phenol chloroform to remove the UDG enzyme. The abasic site containing ssDNA was then annealed with the complementary strand (5′-GTCTTG-3′) by mixing equal amounts at 95 °C, followed by cooling to room temperature. The UDG enzyme was purchased from New England BioLabs (catalog no. M0280L). SRAPD at 10 mg ml⁻¹ was mixed, at a molar ratio of 1:1.2, with DNA containing site-specific abasic site and incubated for 0.5 h on ice. The mixture was then crystallized by setting 24-well vapor-diffusion sitting drops at room temperature, under conditions containing 20% (w/v) PEG 3350, 0.1 M KCl, 0.1 M Bis-Tris pH 5.5, 0.05 M MgCl₂. Diffraction-quality crystals were obtained by streak seeding the drops using a Hampton Research seeding tool (catalog no. HR8-133) with previously generated SRAPd_6nt crystals and incubating them at 20°C for 1 week. SRAPD_DPC crystals were cryo-protected by using reservoir solution supplemented with 15% ethylene glycol and cryo-cooled in liquid nitrogen.

For SRAPD_3nt and SRAPD_6nt co-crystallization, purified SRAPD protein at 12 mg ml⁻¹ was mixed, at a molar ratio of 1:1.2, with different 3′ overhang DNAs prepared by annealing equimolar amounts of two oligonucleotides, 5′-CCACAGCTTGGT-3′ and 5′-GTCTTG-3′ for DNA_3nt; 5′-GTCTTG-3′ and 5′-CCACAGCTTGGT-3′ for DNA_6nt, and incubated for 0.5 h on ice. The mixture was then crystallized by setting 24-well vapor-diffusion sitting drops at room temperature, under conditions containing 20% (w/v) PEG 3350, 0.1 M KCl, 0.1 M Bis-Tris pH 5.5, 0.05 M MgCl₂. Diffraction-quality crystals were obtained by streak seeding the drops using a Hampton Research seeding tool (catalog no. HR8-133) with previously generated SRAPD_6nt crystals and incubating them at 20°C for 1 week. SRAPD_DPC crystals were cryo-protected by using reservoir solution supplemented with 15% ethylene glycol and cryo-cooled in liquid nitrogen.

Fluorescence-based DNA-binding assay. All fluorescence polarization DNA-binding assays were performed in a final volume of 20 μl in a buffer containing 20 mM HEPES pH 7.4, 140 mM KCl, 5 mM NaCl, 0.1 mM EDTA, 0.01% Triton X-100 and 0.2 mM TCEP in 384-well black polypropylene PCR plates. Fluorescence polarization (nF) measurements were performed at room temperature using a BioTek Synergy 4 reader (BioTek). The Kᵥ values were calculated by fitting the curves in GraphPad Prism v.7.04 using non-linear regression, one site-specific binding. The sequences of 6-carboxyfluorescein-labeled DNA oligonucleotides are listed in (Supplementary Table 2). All DNAs were purchased from Integrated DNA Technologies, Inc.

LC–MS analyses of HMCES. SRAPD was incubated with the abasic site containing 3′ overhang DNA (the same as used in SRAPD_DPC co-crystalization) at 1:1.2 ratio in a buffer containing 150 mM NaCl, 20 mM HEPES pH 7.5, 10 mM MgCl₂ at room temperature overnight. All LC–MS data were acquired according to the previously published protocol on an Agilent 6545 Q-TOF equipped with a Dual Agilent Jet Stream ESI source coupled with an Agilent 1260 Infinity HPLC system. The analytical column utilized was a 300 StableBond Poroshell (Agilent, part number 883750-909) 2.1 x 100 mm i.d. internal diameter reversed-phase C5 (5 μm particle size). The mobile phase (A) consisted of 97% HPLC grade water with 0.5% formic acid and 2.5% acetonitrile, while the mobile phase (B) was 96% acetonitrile with 0.5% formic acid and 3.5% HPLC-grade water. A gradient profile was utilized at a flow rate of 500 μl min⁻¹. The mobile phase was held for 2 min at 5% B (with eluant going to waste) and then switched to the mass spectrometer from 2 to 6 min, during which time solvent B increased from 5 to 95%. Two microliters of a 30-μM solution of each sample was injected. Raw data files were analyzed by Agilent MassHunter BioConfirm software (v.8.07.00). Mass spectra between 4 and 6 min were extracted, averaged and deconvoluted using the MaxEnt algorithm.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes: 5KO9 for Apo_SRAPD, 6OEB for SRAPD_3nt, 6OEA for SRAPD_6nt, 6OEG for SRAPD_DPC. LC–MS data underlying Supplementary Fig. 3 have been deposited in Zenodo (https://doi.org/10.5281/zenodo.2662532). Source data for Fig. 2d and Supplementary Fig. 1 are available with the paper online.

References

15. Kabsch, W. XDS. Acta Crystallogr. D 66, 125–132 (2010).
16. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? Acta Crystallogr. D 69, 1204–1214 (2013).
17. Winn, M. D. et al. Overview of the CCP 4 suite and current developments. Acta Crystallogr. D 67, 235–242 (2011).
18. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).
19. Rychlewski, L., Li, W., Jaroszewski, L. & Godzik, A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Sci. 9, 232–241 (2000).
20. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and maximum-likelihood macromolecular crystallographic refinement. Acta Crystallogr. D 59, 2114–2124 (2003).
21. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315 (2018).
22. Chalk, R. in Heterologous Gene Expression in E. coli Vol. 1586 (ed. Burgess-Brown, N. A.) 373–395 (Humana Press, 2017).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|-----|-----------|
| ☑ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☑ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☑ | The statistical test(s) used AND whether they are one- or two-sided |
| ☑ | Only common tests should be described solely by name; describe more complex techniques in the Methods section. |
| ☑ | A description of all covariates tested |
| ☑ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☑ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☑ | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted Give P values as exact values whenever suitable. |
| ☑ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☑ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☑ | Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated |

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Synchrotron beamline specific softwares were used for data collection.

Data analysis

X-ray diffraction data were processed with XDS (version Oct 15, 2015 for PDB:5KO9, and Jan 26, 2018 for the rest) and merged with Aimless (versions 0.5.2.6 for PDB:5KO9 and 0.7.4 for the rest), Phaser-MR (versions 2.6.1 for PDB:5KO9 and 2.8.2 for the rest) for molecular replacement. Models were built with COOT (version 0.8.9.2), and refined with refmac5 (version 5.8.0151 for PDB:5KO9 and 5.8.0238 for the rest). Figures were generated with PyMOL (http://pymol.org). Electrostatic surface potentials were calculated using APBS. Evolutionary conservation was assessed by ConSurf web server. IC-MS raw data files were analyzed by Agilent MassHunter BioConfirm software (v8.07.00).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes: 5KO9 for Apo_SRAPd, 6OEB for SRAP_d_3nt, 6OEA for SRAP_d_6nt, and 6OEF for SRAP_d_DPC. Source data for Supplementary figure 3 have been deposited in Zenodo with DOI:10.5281/zenodo.2662532. Source data for figure 2d, and Supplementary figure 1 are available with the paper online.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences       ☐ Behavioural & social sciences       ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
- one crystal was used for determination of each crystal structure.

Data exclusions
- Resolution cutoffs were applied to X-ray diffraction data.

Replication
- Different batches of expressed and purified wild-type proteins were used for protein-DNA binding assays. Only one batch of purified mutant variants were used for protein-DNA binding assays. All attempts for replicating protein-DNA bindings with Fluorescence polarization were successful.

Randomization
- Randomization was not applicable for protein-DNA crystals.

Binding
- Subsets of X-ray diffraction amplitudes were withheld for calculation of Rfree values.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a
- Involved in the study
  ☒ Antibodies
  ☒ Eukaryotic cell lines
  ☒ Palaeontology
  ☒ Animals and other organisms
  ☒ Human research participants
  ☒ Clinical data

Methods

n/a
- Involved in the study
  ☒ ChIP seq
  ☒ Flow cytometry
  ☒ MRI-based neuroimaging