



Published in final edited form as:

*Nat Struct Mol Biol.* 2019 July ; 26(7): 607–612. doi:10.1038/s41594-019-0246-6.

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

Levon Halabelian<sup>1</sup>, Mani Ravichandran<sup>1</sup>, Yanjun Li<sup>1</sup>, Hong Zeng<sup>1</sup>, Anjana Rao<sup>2,3,4,5</sup>, L. Aravind<sup>6,7</sup>, and Cheryl H Arrowsmith<sup>\*,1,8,9</sup>

<sup>1</sup>Structural Genomics Consortium, University of Toronto, Toronto, ON, Canada.

<sup>2</sup>Division of Signaling and Gene Expression, La Jolla Institute for Immunology, La Jolla, California, USA.

<sup>3</sup>Department of Pharmacology, University of San Diego, La Jolla, California, USA.

<sup>4</sup>Moores Cancer Center, University of San Diego, La Jolla, California, USA.

<sup>5</sup>Sanford Consortium for Regenerative Medicine, La Jolla, California, USA.

<sup>6</sup>National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland, USA.

<sup>7</sup>National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA.

<sup>8</sup>Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada.

<sup>9</sup>Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada.

### Abstract

HMCES can covalently crosslink to abasic sites in single-stranded DNA at stalled replication forks to prevent genome instability. Here, we report crystal structures of the Human HMCES SRAP domain in complex with DNA-damage substrates, including HMCES crosslinked 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 crosslink structure provides structural insights into a novel thiazolidine covalent interaction between the DNA abasic site and conserved Cys2 of HMCES. Collectively,

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Corresponding author: Cheryl H Arrowsmith, [cheryl.arrowsmith@uhnresearch.ca](mailto:cheryl.arrowsmith@uhnresearch.ca), Tel: +1 416 946 0881; Fax: +1 416 946 0880. 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 financial interests.

#### Reporting Summary

Further information on experimental 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: 5K09 for Apo\_SRAPd, 6OEB for SRAPd\_3nt, and 6OEA for SRAPd\_6nt, 6OE7 for SRAPd\_DPC. LC-MS data underlying 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.

our structures demonstrate the capacity for the SRAP domain to interact with 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.

## Introduction

DNA bases are constantly damaged by factors such as reactive oxygen species (ROS), chemotoxic agents, ionizing radiation (IR) and UV radiation<sup>1</sup>, and are subject to physiological modification by enzymes such as AID, DNA methylases and the TETs<sup>2,3</sup>. These types of DNA alterations are primarily repaired by the base excision repair (BER) pathway, which is initiated by DNA glycosylases that recognize and cleave damaged or modified bases creating apurinic or apyrimidinic sites (AP sites) or abasic sites<sup>1</sup>. HMCES was recently reported to recognize and covalently crosslink to abasic sites in single-stranded DNA (ssDNA), generated by uracil DNA glycosylase (UDG) at stalled replication forks<sup>4</sup>. The authors suggested that these DNA-protein crosslink (DPC) intermediates prevented ssDNA breaks that may consequently occur upon cleavage by AP endonucleases, which could subsequently be repaired through error-prone pathways<sup>4</sup>.

Human HMCES has a highly conserved N-terminal SOS Response-Associated Peptidase domain (SRAPd) that is widely found in bacteria and eukaryotes, with a sporadic presence in certain bacteriophages and archaea<sup>5</sup>. Animal SRAP proteins have an additional C-terminal disordered extension with multiple copies of the PCNA-interacting motif (PIP)<sup>4,6</sup>. Gene-neighborhood analysis in prokaryotes identified SRAPd as a novel component of the bacterial SOS response, associated with multiple components of the DNA repair machinery<sup>5</sup>. The SRAPd contains a triad of predicted catalytic residues, namely Cys2, Glu127, and His210, which are highly conserved across SRAP domains from all the three superkingdoms of life and viruses. This triad is predicted to support autoproteolytic activity acting on the peptide bond N-terminal to Cys2 as all characterized SRAP proteins show a cleavage of residues N-terminal to this cysteine<sup>5</sup>. Moreover, Cys2 was recently shown to mediate the formation of the DPC via a covalent linkage to the deoxyribose at the abasic site of the DNA<sup>4</sup>. 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 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  $\pi$ -stacking interactions with the duplex segment of DNA at the ssDNA-dsDNA junction and 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 directing it towards the catalytic triad. The ssDNA-binding cleft includes conserved Arg98 and Arg212, which form salt-bridges with the phosphate backbone of ssDNA (Fig. 1b, d). Alanine substitutions 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.<sup>4</sup>

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 (WT) 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 Gly3, Arg4, Pro46, Asp47, Trp128 (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 affinities<sup>4</sup>. This dsDNA-interaction site B accounts for the remaining residues, which are highly conserved in SRAP domains across all superkingdoms of life and phages<sup>5</sup>, 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 (referred to here as SRAPd\_6nt). Although SRAPd has nearly 10-fold 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 3.2Å, 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 7-fold 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-nt gap DNA, suggesting that both interaction sites are crucial for substrate DNA binding (Supplementary Fig. 1).

### Crystal Structure of the Human HMCES SRAP domain Crosslinked to DNA Abasic site

To further investigate the covalent interaction of HMCES with a DNA abasic site, we crystallized 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 crosslinking 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 Cys2. We then treated this DNA with Uracil-DNA glycosylase (UDG) to generate AP9 DNA<sup>7</sup>. For the crosslink reaction to proceed between the AP9 and Cys2 of HMCES, the N-terminal methionine needs to be removed in order to expose the NH<sub>2</sub> of Cys2. Mass spectrometry analysis of our purified C-terminally His-tagged SRAPd protein showed that the N-terminal methionine (Met1) was cleaved, either co-translationally, catalyzed by *E.Coli* Methionyl-aminopeptidase<sup>8</sup> (MAP), or due to its predicted autopeptidase activity (Supplementary Fig. 3; Supplementary Table 1) yielding a catalytically active form of HMCES. Incubation of this active SRAPd with AP9 DNA yielded crosslinked SRAPd\_DPC (Supplementary Fig. 3; Supplementary Table 1).

The SRAPd\_DPC structure was refined to 2.2Å resolution and is isomorphous with both SRAPd\_3nt and SRAPd\_6nt structures. Continuous electron density was observed between Cys2 of HMCES and the AP9 confirming the covalent crosslink (Fig. 3a, b). The Cys2-AP9 crosslink was modeled at full occupancy, and shows an approximately 20% higher B-factor values compared to 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 Cys2 of HMCES reacts with the ring-opened aldehyde form of the abasic deoxyribose (AP9) to form a thiazolidine DNA-protein crosslink<sup>9</sup> (Fig. 3a, b, c). This crosslink entails an approximately 120° rotation around AP9  $\gamma^{10}$  and possible hydrogen bonding between the 4'-hydroxyl of AP9 and the N $\tau$  position of His210 (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.<sup>4</sup> showed that HMCES forms DPC intermediates with abasic sites in ssDNA generated by uracil-DNA glycosylase (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<sup>11</sup>. Our structures confirm this crosslink and reveal it to be a thiazolidine ring that involves both the sidechain sulfur and the NH<sub>2</sub> of Cys2 that is exposed by the peptidolytic removal of N-terminal methionine. Cys2 along with Glu127, and His210 form a predicted catalytic triad that is situated in a pocket characteristic of SRAP domains<sup>5</sup>. Thus, our structure reveals that in addition to predicted autopeptidase activity this pocket is also required for the accommodation of the DPC. It also suggests that HMCES can recognize and covalently crosslink 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-nt 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<sup>12</sup>. 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, for the first time, the covalent interaction of HMCES with DNA abasic sites, revealing its broad substrate recognition spectrum.

## 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 TEV cleavable N-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 S200 column. Purified SRAPd was concentrated to ~20 mg/mL 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 weight for all purified constructs were verified by liquid chromatography-mass spectrometry.

### Crystallization and structural determination

Apo\_SRAPd was crystallized using sitting drop vapor-diffusion method by mixing 1:1 ratio of protein and reservoir solution containing 0.1 M Bis-Tris Propane, 2% Tacsimate, 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'-CCAGACGTUGTT-3') was first incubated with Uracil-DNA glycosylase (UDG) for 1 hour 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. 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 (Cat # M0280L). SRAPd at 10 mg mL<sup>-1</sup> 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, in a condition containing 20% (w/v) PEG 3350, 0.1 M KCl, 0.1 M Bis-Tris pH 5.5, 0.05 M MgCl<sub>2</sub>. Diffraction quality crystals were obtained by streak seeding the drops using Hampton research seeding tool (Cat # HR8-133) with previously generated SRAPd\_6nt crystals and incubating them at 20°C for one 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<sup>-1</sup> was mixed, at a molar ratio of 1:1.2, with different 3' overhang DNA prepared by annealing equimolar amounts of two oligonucleotides, 5'-CCAGACGTT-3' and 5'-GTCTTG-3' for DNA\_3nt; 5'-GTCTTG-3' and 5'-CCAGACGTTGTT-3' for DNA\_6nt, and incubated for 0.5 h on ice. The mixture was then crystallized using sitting drop vapor-diffusion method in a condition containing 25% (w/v) PEG 3350, 0.2 M ammonium sulphate, 0.1 M Hepes pH 7.5 for SRAPd\_3nt; and 20% (w/v) PEG 3350, 0.1 M KCl, 0.1 M Bis-Tris pH 5.5, 0.05 M MgCl<sub>2</sub> for SRAPd\_6nt. Apo\_SRAPd, SRAPd\_3nt and SRAPd\_6nt crystals were cryo-protected using reservoir solution supplemented with 20–30% glycerol and 20–30% ethylene-glycol, respectively, and cryo-cooled in liquid-nitrogen.

Diffraction data for the Apo\_SRAPd and SRAPd\_DPC was collected at the 19ID and 24ID-C beamlines of the Advanced Photon Source (APS), respectively. Diffraction data for SRAPd\_3nt and SRAPd\_6nt was collected at the 5.0.1 beamline of the Advanced Light Source (ALS) Berkeley Lab. Datasets were processed with XDS<sup>15</sup> and merged with Aimless<sup>16,17</sup>. Initial phases for the Apo\_SRAPd was obtained by molecular replacement with Phaser-MR<sup>18</sup>, using a combination of models (PDB ID: 2F20, 2BDV, 2ICU, 1ZN6) generated by FFAS<sup>19</sup>. Initial phases for the SRAPd\_3nt were obtained by molecular replacement with Phaser-MR<sup>18</sup>, using the Apo\_SRAPd (PDB ID: 5KO9) as a search model. Whereas, initial phases for SRAPd\_6nt and SRAPd\_DPC were obtained by Fourier transform using SRAPd\_3nt structure as a starting model. Models were built with COOT<sup>20</sup>, and refined with refmac5<sup>21</sup>. Structures were validated with Molprobit<sup>22</sup>. Data collection and refinement statistics are shown in (Table 1). Figures were generated with PyMOL (<http://pymol.org>).

### 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 Ethylenediaminetetraacetic acid (EDTA), 0.01% Triton X-100 and 0.2 mM TCEP in 384-well black polypropylene PCR plates. Fluorescence polarization (mP) measurements were performed at room temperature using a BioTek Synergy 4 (BioTek, Winooski, VT). The  $K_D$

values were calculated by fitting the curves in GraphPad Prism 7.04 using nonlinear regression, one site-specific binding, equation  $Y=B_{max} * X / (K_d + X)$ . The sequences of 6-carboxyfluorescein (FAM)-labeled DNA oligonucleotides are listed in (Supplementary Table 2). All DNAs were purchased from Integrated DNA Technologies, Inc.

### Liquid chromatography mass spectrometry analyses of HMCES

SRAPd was incubated with the abasic site containing 3' overhang DNA (same as used in SRAPd\_DPC co-crystallization) at 1:1.2 ratio in a buffer containing 150 mM NaCl, 20 mM Hepes pH 7.5, 10 mM MgCl<sub>2</sub> at room temperature overnight. All LC-MS data were acquired according to the previously published protocol<sup>23</sup>, on an Agilent 6545 Q-TOF (Santa Clara, CA) equipped with a Dual Agilent Jet Stream ESI source coupled with an Agilent 1260 Infinity HPLC system (Santa Clara, CA). The analytical column utilized was a 300 StableBond Poroshell (Agilent, part number 883750-909) 2.1 × 100-mm-i.d. reversed-phase C<sub>3</sub> (5 μm particle size). Mobile phase (A) consisted of 97% HPLC grade water with 0.5% formic acid and 2.5% ACN, while mobile phase (B) was 96% ACN with 0.5% formic acid and 3.5% HPLC grade water. A gradient profile was utilized at a flow rate of 500 μL/min. The mobile phase was held for 2 min at 5% B (with eluant going to waste) and then switched to the mass spectrometer from 2–6 min during which time solvent B increased from 5–95%. Two microliters of a 30 μM solution of each sample was injected. Raw data files were analyzed by Agilent MassHunter BioConfirm software (vB.07.00). Mass spectra between 4 and 6 minutes were extracted, averaged and deconvoluted using the MaxEnt algorithm.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgements

We are grateful to H. Wyatt for fruitful discussions, advice from P.J. Brown and W. Tempel on interpretation of DNA-protein crosslink 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 22ID-D beamline for collecting the initial SRAPd\_DPC datasets. This research used resources of the Advanced Light Source, which is a DOE 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 U.S. Department of Energy, Office of Biological and Environmental Research under contract 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 (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 (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 [OGI-055]; Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD: 115766]; Janssen, Merck & Co.; Novartis Pharma AG; Ontario Ministry of Research Innovation and Science (MRIS); Pfizer, São Paulo Research Foundation-FAPESP, Takeda and the Wellcome Trust. This research is also supported by the Canadian Institutes of Health Research [FDN154328] and Natural Sciences and Engineering Research Council [RGPIN-2015-05939] to CHA, intramural funds of the National Library of Medicine, NIH, USA to LA, and the National Cancer Institute [R35 CA210043] to AR.

## References

1. Maynard S, Schurman SH, Harboe C, de Souza-Pinto NC & Bohr VA Base excision repair of oxidative DNA damage and association with cancer and aging. *Carcinogenesis* 30, 2–10 (2008). [PubMed: 18978338]
2. Nabel CS, Manning SA & Kohli RM The Curious Chemical Biology of Cytosine: Deamination, Methylation, and Oxidation as Modulators of Genomic Potential. *ACS Chem. Biol* 7, 20–30 (2012). [PubMed: 22004246]
3. Iyer LM, Zhang D, Maxwell Burroughs A & Aravind L Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res.* 41, 7635–7655 (2013). [PubMed: 23814188]
4. Mohni KN et al. HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell* (2018). doi:10.1016/j.cell.2018.10.055
5. Aravind L, Anand S & Iyer LM Novel autoproteolytic and DNA-damage sensing components in the bacterial SOS response and oxidized methylcytosine-induced eukaryotic DNA demethylation systems. *Biol. Direct* 8, (2013).
6. Srivastava M et al. Replisome Dynamics and Their Functional Relevance upon DNA Damage through the PCNA Interactome. *Cell Rep.* 25, 3869–3883.e4 (2018). [PubMed: 30590055]
7. Lindahl T DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*. *J. Biol. Chem* 252, 3286–3294 (1977). [PubMed: 324994]
8. Hirel PH, Schmitter MJ, Dessen P, Fayat G & Blanquet S Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci* 86, 8247–8251 (1989). [PubMed: 2682640]
9. Roberts JC, Charyulu RL, Zera RT & Nagasawa HT Protection Against Acetaminophen Hepatotoxicity by Ribose-Cysteine (RibCys). *Pharmacol. Toxicol* 70, 281–285 (1992). [PubMed: 1608914]
10. Abbreviations and symbols for the description of conformations of polynucleotide chains (Recommendations 1982). *Pure Appl. Chem* 55, 1273–1280 (1983).
11. Liu M et al. The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. *Proc. Natl. Acad. Sci* 107, 4925–4930 (2010). [PubMed: 20185759]
12. Spruijt CG et al. Dynamic Readers for 5-(Hydroxy)methylcytosine and Its Oxidized Derivatives. *Cell* 152, 1146–1159 (2013). [PubMed: 23434322]
13. Landau M et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* 33, W299–W302 (2005). [PubMed: 15980475]
14. Jurrus E et al. Improvements to the APBS biomolecular solvation software suite: Improvements to the APBS Software Suite. *Protein Sci.* 27, 112–128 (2018). [PubMed: 28836357]
15. XDS Kabsch W.. *Acta Crystallographica Section D Biological Crystallography* 66, 125–132 (2010). [PubMed: 20124692]
16. Evans PR & Murshudov GN How good are my data and what is the resolution? *Acta Crystallographica Section D Biological Crystallography* 69, 1204–1214 (2013). [PubMed: 23793146]
17. Winn MD et al. Overview of the CCP 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* 67, 235–242 (2011). [PubMed: 21460441]
18. McCoy AJ et al. Phaser crystallographic software. *Journal of Applied Crystallography* 40, 658–674 (2007). [PubMed: 19461840]
19. Rychlewski L, Li W, Jaroszewski L & Godzik A Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Science* 9, 232–241 (2008).
20. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* 66, 486–501 (2010). [PubMed: 20383002]
21. Steiner RA, Lebedev AA & Murshudov GN Fisher's information in maximum-likelihood macromolecular crystallographic refinement. *Acta Crystallographica Section D Biological Crystallography* 59, 2114–2124 (2003). [PubMed: 14646069]

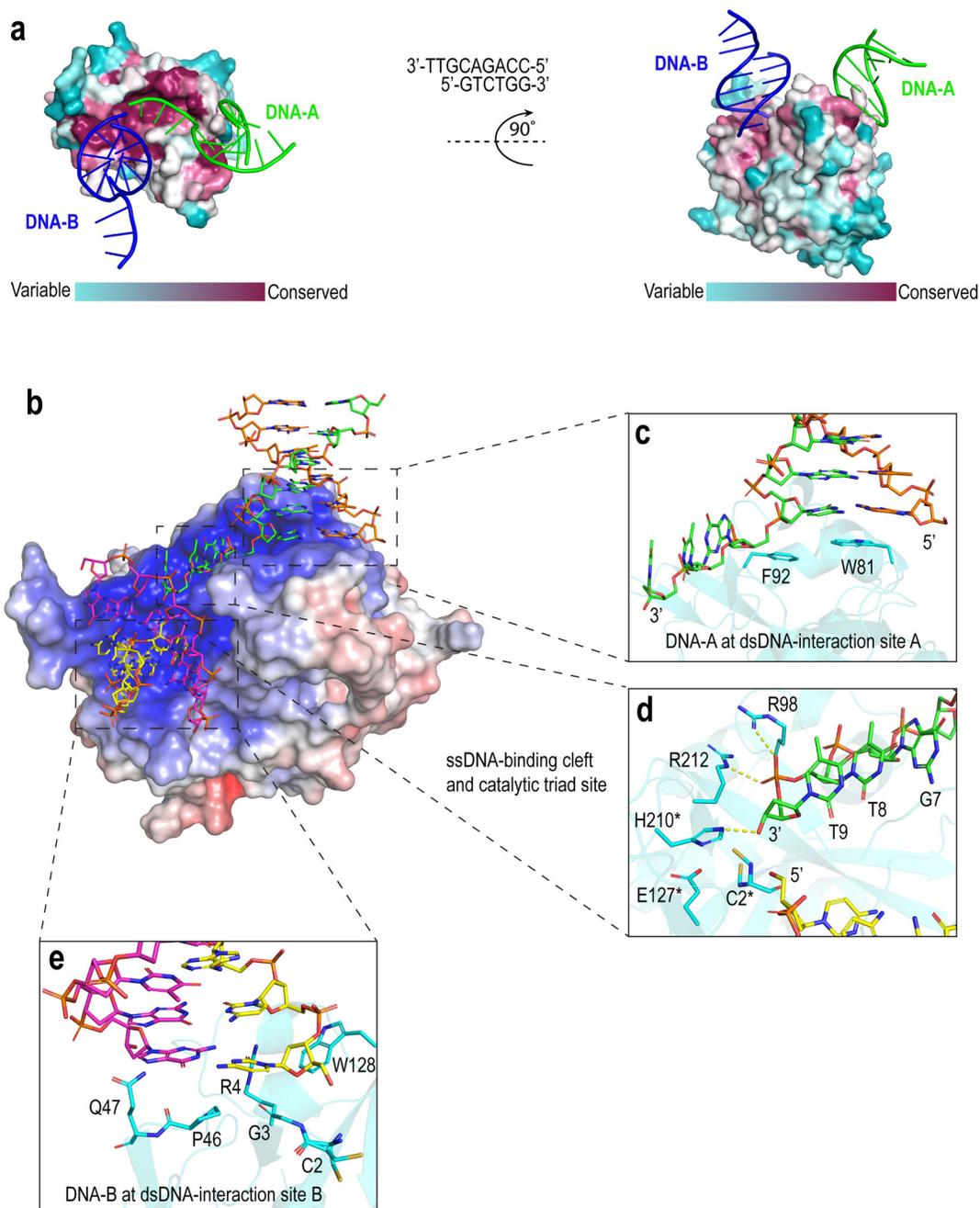
22. Williams CJ et al. MolProbity: More and better reference data for improved all-atom structure validation: PROTEIN SCIENCE.ORG. *Protein Science* 27, 293–315 (2018). [PubMed: 29067766]
23. Chalk R *Mass Spectrometric Analysis of Proteins in Heterologous Gene Expression in E.coli* (ed. Burgess-Brown NA.) 1586, 373–395 (Springer New York, 2017).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 1. Interactions between SRAPd and 3'overhang DNA.**

**(a)** Surface representation of SRAPd colored by degree of sequence conservation, bound to two symmetry-related DNA molecules shown in green (DNA-A) and blue (DNA-B). Evolutionary conservation was assessed by ConSurf web server<sup>13</sup>. The 3' overhang DNA sequence that was used for co-crystallization is shown in the middle. **(b)** Electrostatic surface potential representation of SRAPd interacting with two symmetry-related DNA molecules: DNA-A in green and orange, and DNA-B in magenta and yellow. SRAPd surface color indicates electrostatic potential ranging from  $-7kT/e$  (red) to  $+7kT/e$  (blue). Electrostatic surface potentials were calculated using APBS<sup>14</sup>. **(c)** Close-up view of the

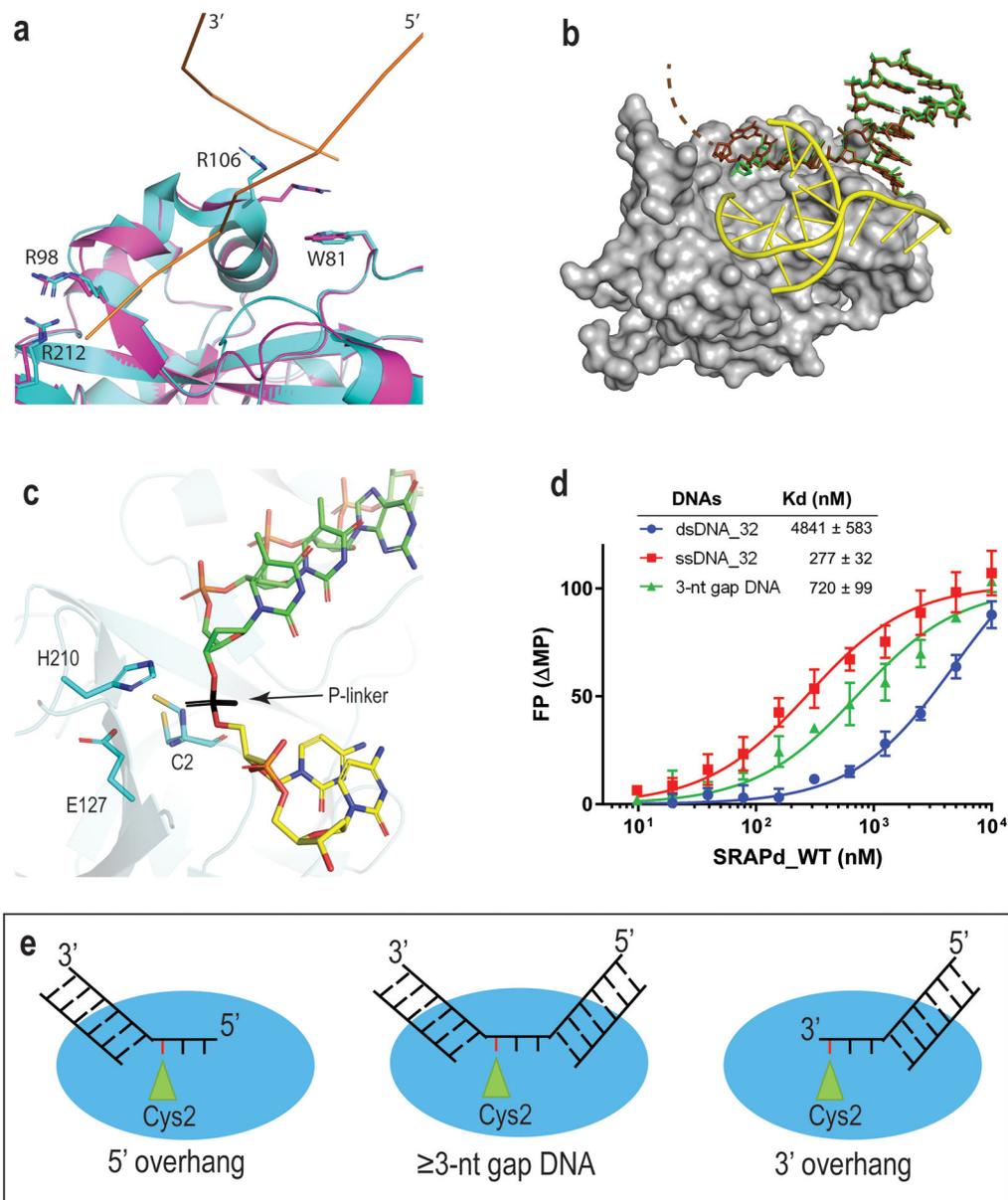
SRAPd dsDNA-interaction site A in stacked conformation with the duplex segment of DNA-A. **(d)** Close-up view of the SRAPd catalytic triad site and ssDNA-binding cleft bound to the phosphate backbone of the single-strand segment of DNA-A. The catalytic triad residues as well as R98 and R212 in the ssDNA-binding cleft are shown as stick models in cyan. The ssDNA segment of DNA-A is colored green and DNA-B colored yellow. The catalytic triad residues are marked with asterisk. **(e)** Close-up view of the SRAPd dsDNA-interaction site B, which stacks with the blunt-end of DNA-B.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 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 ribbon representation in orange. (b) Crystal structure of SRAPd\_6nt in complex with six-nucleotide 3' overhang. SRAPd is shown in surface representation in grey. The six-nucleotide overhang at the dsDNA-interaction site A is shown as a stick model in brown superposed with 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 DNA-binding

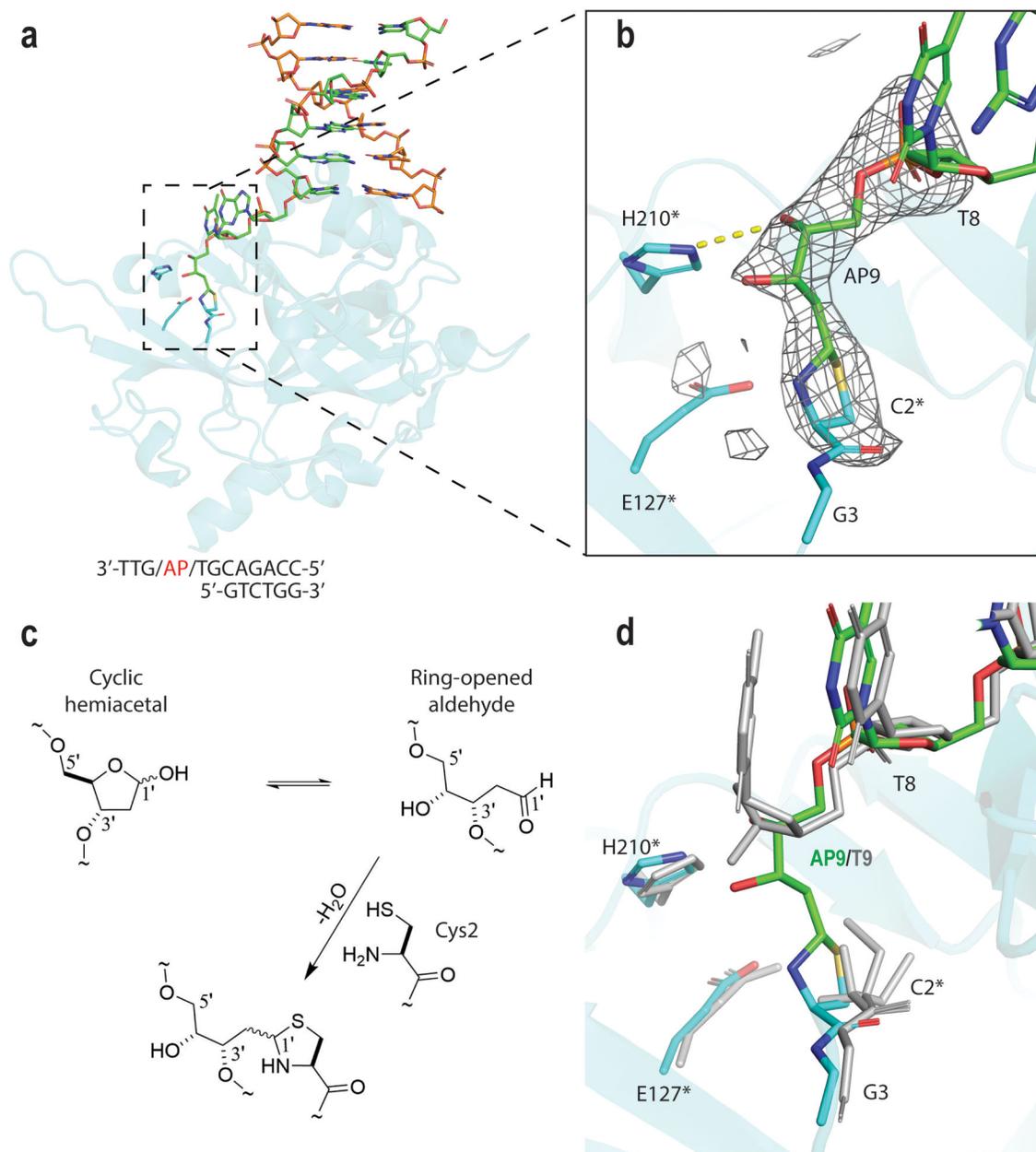
affinities of HMCES SRAPd to ssDNA, dsDNA, and dsDNA containing a three-nucleotide gap (3-nt gap DNA). Experiments were performed in triplicates and data are represented as mean  $\pm$  SD. **(e)** A model illustrating the potential DNA damage substrates that can be recognized by HMCES. Red line represents the abasic site in DNA. Source Data for panel d are available online.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 3. Crystal structure of the Human HMCES SRAPd crosslinked to a DNA abasic site.**

**(a)** Overall structure of SRAPd\_DPC. SRAPd is shown as cartoon representation in cyan, 3' overhang DNA in green and orange. DNA-B at dsDNA-interaction site B is not shown for clarity. **(b)** The  $mF_o-DF_c$  electron density omit-map for the Cys2 crosslink with DNA abasic site (AP9) in the crystal structure of SRAPd\_DPC, displayed as grey mesh and contoured at  $3.0\sigma$ . The catalytic triad residues are marked with an asterisk. **(c)** Reaction scheme for formation of the covalent crosslink between Cys2 of HMCES and the ring-opened aldehyde form of abasic deoxyribose. **(d)** SRAPd\_DPC structure in cyan (protein) and green (DNA), superposed with SRAPd\_3nt in grey.

Table 1

## Data collection and refinement statistics

	Apo_SRAPd (PDB: 5KO9)	SRAPd_3nt (PDB: 6OEB)	SRAPd_6nt (PDB: 6OEA)	SRAPd_DPC (PDB: 6OE7)
<b>Data collection<sup>a</sup></b>				
Space group	<i>I</i> 2	<i>I</i> 2	<i>I</i> 2	<i>I</i> 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (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
$\alpha$ , $\beta$ , $\gamma$ (°)	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) <sup>b</sup>	48.38–2.10(2.16–2.10)	49.11–2.10(2.16–2.10)	48.72–2.2(2.27–2.20)
<i>R</i> <sub>merge</sub>	0.039 (0.61)	0.069 (0.71)	0.059 (0.74)	0.045 (0.70)
<i>I</i> / $\sigma$ ( <i>I</i> )	15.9 (1.9)	10.2 (1.4)	15.0 (2.2)	12.3 (1.3)
<i>CC</i> <sub>1/2</sub>	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)
<b>Refinement</b>				
Resolution (Å)	48.37–1.5	48.38–2.10	49.11–2.10	48.72–2.20
No. reflections	41504	24718	25066	20951
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.181/0.209	0.205/0.235	0.208/0.242	0.194/0.216
No. atoms	2292	2492	2457	2402
Protein	2062	2078	2056	2046
DNA	-	302	324	282
Cys2-AP9 crosslink	-	-	-	17
Water	186	77	46	33
<i>B</i> 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 crosslink	-	-	-	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

<sup>a</sup>A single crystal was used for all structures.

<sup>b</sup>Values in parentheses are for highest-resolution shell.