Supporting Information

A structure-specific nucleic acid binding domain conserved among DNA repair proteins

Aaron C. Mason, Robert P. Rambo, Briana H. Greer, Michael Pritchett, John A. Tainer, David Cortez, Brandt F. Eichman

Experimental Procedures

HARP1 purification

Mus musculus (m) HARP1 (amino acids 197-268) was cloned into the BamHI and NotI restriction sites of pBG101 (Vanderbilt Center for Structural Biology) using primers BamH1-mHARP1 and Not1-mHARP1 (Table S3) to create a His_6-GST-fusion protein cleavable by rhinovirus 3C protease. Protein was expressed in HMS174(DE3) cells for 3 h at 37°C after induction with isopropyl-1-thio-β-D-galactopyranoside (IPTG). Cells were harvested by centrifugation and resuspended in buffer containing 20 mM Tris (pH 8), 300 mM NaCl, 2 mM $MgCl_2$, 0.5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 20% glycerol, 0.2 mM phenylmethylsulfonyl fluoride (PMSF), and 10 mM imidazole supplemented with 10 µg/ml aprotinin and 1 mM leupeptin. Following Ni-NTA chromatography, peak fractions were digested with 3C protease overnight at 4°C to remove the His_6-GST tag and loaded onto a 5 ml HiTrap SP HP column (GE Healthcare) equilibrated with buffer containing 20 mM Tris (pH 8), 50 mM NaCl, 2 mM $MgCl_2$, 0.5 mM TCEP, and 10% glycerol. Protein was eluted with a linear gradient from 50 mM to 1M NaCl. Peak fractions were concentrated to 5 ml and loaded onto a HiLoad 26/600 Superdex 200 (S200) column (GE Healthcare) equilibrated with buffer containing 20 mM Tris (pH 8), 200 mM NaCl, 2mM $MgCl_2$, 0.5 mM TCEP, and 10% glycerol. Purified HARP1 was buffer exchanged into 20 mM Tris (pH 8), 100 mM NaCl, 2mM $MgCl_2$, and 0.5 mM TCEP and concentrated to ~650 µM. SeMet-labeled HARP1 was produced in HMS174(DE3) cells by metabolic inhibition of Met biosynthesis (1). SeMet-HARP1 was expressed in minimal media overnight at 16°C and purified similar to the native protein.

SMARCAL1 purification

Wild-type and mutant human SMARCAL1 and $SMARCAL1^{CD}$ (aa 325-870) proteins were purified as N-terminal His-tagged proteins from baculovirus-infected High Five cells using the Bac-to-Bac expression system and pFastBac-HTb vectors (Invitrogen). $SMARCAL1^{CD}$ was overexpressed as amino acids 1-870 with a 3C protease cleavage site inserted immediately before residue 325. Viruses were amplified in Sf9 cells cultured in SF-900 II SFM, and High Five cells were cultured in Express Five SFM. Cells were lysed 45-48 h post infection by homogenization in buffer containing 20 mM HEPES (pH 7.6), 300 mM NaCl, 2 mM $MgCl_2$, 0.5 M TCEP, 20% glycerol, 0.2 mM PMSF, 10 mM imidazole, 1 mM leupeptin, and 10 µg/ml aprotinin. Clarified lysate was loaded onto pre-equilibrated Ni-NTA resin. Following a three-column volume wash with lysis buffer, protein was eluted with 20 mM HEPES (pH 7.6), 50 mM NaCl, 2 mM $MgCl_2$, 0.5 M TCEP, 20% glycerol, 0.2 mM PMSF, and 250 mM imidazole. Peak fractions were pooled, and in the case of $SMARCAL1^{CD}$, treated with 3C protease overnight at 4°C to remove the N-terminal 324 amino acids. Peak fractions were diluted threefold with buffer containing 20 mM HEPES (pH 7.6), 50 mM NaCl, 2 mM $MgCl_2$, 0.5 M TCEP, and 20% glycerol and loaded onto a 5 ml

HiTrap SP HP cation exchange column (GE Healthcare). Protein was eluted with a linear gradient from 0.05 to 1 M NaCl. Peak fractions were pooled and concentrated to 2 ml and loaded onto a HiLoad 16/600 S200 column that had been equilibrated with buffer containing 20 mM HEPES (pH 7.6), 200 mM NaCl, 2 mM $MgCl_2$, 0.5 M TCEP, and 20% glycerol.

Mutant SMARCAL1^{CD} expression vectors were constructed as follows. DNA encoding a chimeric HARP1-ATPase (residues 230-302 and 398-954 with a stop codon at aa 871) was synthesized by DNA2.0 in pJExpress411 with an *NcoI* restriction site at the 5' end, an *SpeI* restriction site at aa 398, and a *KpnI* restriction site at the 3' end. HARP1-ATPase was subcloned into pFastBac-HTb using *Nco*I and *KpnI* restriction sites. Human HARP2 (aa 325-397) patch mutant cassettes were synthesized by DNA2.0 in pJExpress411 with 5' *NcoI* and 3' *SpeI* restriction sites and subsequently subcloned into the *NcoI* and *SpeI* sites of the HARP1-ATPase/pFastBac-HTb vector, resulting in HARP2(mutant)-ATPase/FastBac-HTb constructs.

UvsW and HARP-UvsW chimera purification

T4 Phage UvsW was subcloned from pMRW7 (gift from K. Kreuzer, Duke University) into a pET27 (Novagen)-derived expression vector pBG101 (Vanderbilt Center for Structural Biology) to produce an N-terminal His_6-GST fusion protein. Protein was overexpressed in *E. coli* HMS174 (DE3) cells (Novagen) in LB medium supplemented with 30 ug/ml kanamycin, and 0.5 mM IPTG for 16 hours at 16 °C. The cells were lysed in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 20% (v/v) glycerol, 0.2% Triton X-100, and protease inhibitor cocktail tablets EDTA-free (Roche). Protein was isolated using HisPur (Thermo Scientific) affinity chromatography followed by 3C protease cleavage of the His_6-GST tag and purified further by heparin affinity (HiTrap Heparin HP, GE Healthcare).

The HARP-UvsW chimera was constructed in pBG101 by restriction free cloning of UvsW residues 83-503 (ATPase) and human SMARCAL1 residues 237-302 (HARP1) using 5′-BamHI-UvsW, 3′-NotI-UvsW, 5′-HARP1Swap, and 3′-HARP1Swap primers (Table S3). The HARP-UvsW^{K141R} point mutant was produced using HARP-UvsW/pBG101 as a template for the QuikChange kit (Stratagene). The UvsW^{ATPase} deletion mutant construct was subcloned from the wild-type UvsW/pBG101 vector using primers 5′-BamHI-UvsW^{ATPase} and 3′-NotI-UvsW (Table S3). All UvsW mutants were purified the same as UvsW but with an additional S200 gel filtration step.

Crystallization and X-ray Data Collection

All mHARP1 crystals were grown by sitting-drop vapor diffusion at 21°C. For native crystals, 1 µl protein containing 20 mM Tris (pH 8), 100 mM NaCl, 2mM $MgCl_2$, and 0.5 mM TCEP was mixed with 1 µl reservoir solution containing 0.1 M MES (pH 6.5), 0.2 M ammonium sulfate, and 30% PEG 5K MME,

and drops equilibrated against 0.5 ml reservoir solution. Native crystals were flash cooled in mother liquor containing 20% glycerol prior to data collection. SeMet-mHARP1 crystals were grown in the same manner as wild-type against reservoirs containing 0.1 M sodium acetate trihydrate (pH 4.5) and 2 M ammonium sulfate, and were flash cooled in mother liquor containing 20% glycerol and 5% sucrose. X-ray diffraction data were collected at the 21-ID beamline at the Advanced Photon Source (Argonne, IL) and processed using HLK2000 (2). X-ray phases were obtained from a SAD experiment using SeMet-HARP1 crystals belonging to space group $P3_222_1$. Eight Se positions in the asymmetric unit were located using autoSHARP (3). Polypeptide chains for all four HARP1 molecules in the asymmetric unit were built into the electron density using Coot (4). This model was then used as a molecular replacement search model against native diffraction data (space group $P2_12_12_1$). The model, containing four HARP1 molecules in the asymmetric unit, was refined against a maximum likelihood target using PHENIX (5). Changes to the model were made in Coot after each round of refinement using $2F_o$-F_c composite omit and F_o-F_c maps. All crystallographic software was configured by SBGrid (6).

Fork Regression Assay

Fork regression assays were performed as described previously (7) with minor modifications. Oligonucleotides #48, #50, #52, and #53 (Table S3) were used to create the DNA fork substrate. Oligonucleotide #48 was 5'-^{32}P-labeled and annealed to oligonucleotide #50. Separately, oligonucleotides #52 and #53 were annealed, and the ^{32}P-48/50 and 52/53 duplexes annealed into a fork. Fork regression reactions (10 µl) contained 40 mM HEPES (pH 7.6), 20 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP, 0.1 mg/ml BSA, 2 mM ATP, and 1 nM DNA fork substrate. The reaction was incubated for 20 min at 37°C and terminated by adding proteinase K (Sigma-Aldrich) to a final concentration of 1 mg/ml and incubating 10 min. Reactions were brought to 5% glycerol, 10 mM EDTA, and 0.1% bromophenol blue and run on 8% polyacrylamide/TBE gels at 400V for 1.5 hrs. Gels were imaged with a Molecular Dynamics Typhoon Variable Mode Imager and quantified using ImageJ (8).

ATPase Assay

DNA-dependent ATPase activity was determined using a High Throughput Colorimetric ATPase Assay (Innova Biosciences) according to the manufacture's protocol. Activity was determined using 8 nM protein and 2-80 nM DNA in 100 mM Tris (pH 7.4), 5 mM $MgCl_2$, and 1 mM ATP. DNA overhang substrate was formed by annealing oligonucleotides #25.1 and #25.3 (Table S3). The fork DNA substrate was formed from oligonucleotides #48, #50, #52, and #53 by separately annealing #48 with #50 and #52 with #53, followed by annealing 48/50 and 52/53 duplexes.

DNA Binding

DNA binding was assayed by fluorescence polarization using 6-carboxyfluorescein (FAM) labeled DNA substrates. Oligonucleotides used to generate overhang, splayed arm, flap, fork, and 4-way junction DNA structures are shown in Table S3, and were annealed as follows: Fam40 and F20.20 (overhang); Fam40 and F20.40 (splayed arm); Fam40, Lead20.20, and F20.40 (flap); Fam40, Lead20.20, Lag20.20, and F20.40 (fork); and Fam40, HJ.IM.2, HJ.IM.3, and HJ.IM.4 (immobile 4-way junction). Reactions (15 µl) contained 10-3000 nM protein, 25 nM DNA, and 20 mM Hepes pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 0.25 mg/mL BSA, 0.05 mM EDTA, 0.5 mM DTT, 0.01% Tween20, and incubated in a 384-well plate for 20 min at 25°C. Fluorescence polarization was measured using a BioTek Synergy H1 Hybrid Reader with a filter cube containing 485/20 nm excitation and 528/20 nm emission filters.

References to the Supporting Information

1. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, & Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of molecular biology* 229(1):105-124.
2. Otwinowski Z MW (1997) Processing of X-ray Diffraction Data Collected in the Oscillation Mode. *Methods in Enzymology* 276:307-326.
3. Vonrhein C, Blanc E, Roversi P, & Bricogne G (2007) Automated structure solution with autoSHARP. *Methods in molecular biology* 364:215-230.
4. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
5. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
6. Morin A*, et al.* (2013) Collaboration gets the most out of software. *Elife* 2:e01456.
7. Bétous R*, et al.* (2012) SMARCAL1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. *Genes & development* 26(2):151-162.
8. Schneider CA, Rasband WS, & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671-675.

Table S1. Crystallographic data collection and refinement statistics

	Native	SeMet
Data collection		
Wavelength (Å)	0.9787	0.9786
Space group	$P2_12_12_1$	$P3_222_1$
Cell dimensions		
a, b, c (Å)	56.5, 56.6, 104.3	103.0, 103.0, 103.9
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution (Å)	52.2-1.90 (1.97-1.90)	89.18-3.30 (3.42-3.30)
R_{sym}	0.093 (0.464)	0.097 (0.290)
I / σ_I	26.2 (4.4)	43.4 (9.6)
Completeness (%)	97 (96.4)	100 (100)
Redundancy	9.5 (9.6)	11.2 (9.9)
Refinement		
Resolution	1.90	
No. reflections	26,465	
R_{work} / R_{free}	0.169 / 0.197	
No. of atoms		
Protein	2181	
Solvent	198	
B-factors (Å^2)		
Protein	18.4	
Solvent	24.1	
R.m.s. deviations		
Bond lengths (Å)	0.014	
Bond angles (°)	1.306	

Values in parentheses refer highest resolution shell.

Table S2. Dali search results				
Protein	Ranking	Z-score	RMSD (Å)	PDB id
A.fulgidus **XPB**	1 (2,4,6)	6.9	2.1	2fwr (2fz4)
T4 phage **UvsW**	3 (5,7)	6.9	2.1	1rif (2oca)
T.thermophila **eIF1**	8 (10)	5.5	2.7	2xzm (2xzn)
H.sapiens **MSH6**	9 (11,12)	5.4	2.1	2o8e (2o8c, 2o8d)
A. tumefaciens **Auto741**	13	5.0	2.3	1zhv
E. coli **Ycih**	14	5.0	2.6	1d1r
T4 phage **MotA**	15	5.0	2.3	1kaf
H.sapiens **MSH3**	18 (19,25)	4.9	2.2	3thw
E.coli **MutS**	26	4.8	2.2	1w7a
H.sapiens **MSH2**	45	4.5	2.6	3thy

Table S3. Oligonucleotide sequences used	
Name	Sequence
25.1	TCCGTGAGACCCCTTGACAGCGATG
25.3	TCATCTGATACATCGCTGTCAAGGGGTCTCACGGA
48	ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC
50	GGGTGAACCTGCAGGTGGGCAAAGATGTCC
52	GGGTGAACCTGCAGGTGGGCAAAGATGTCCCAGCAAGGCACTGGTAGAATTCGGCAGCGTC
53	GGACATCTTTGCCCACCTGCAGGTTCACCC
Fam40	(FAM)CTCAGGACTCAGTTCGTCAGCCCTTGACAGCGATGGAAGC
F20.20	CTGACGAACTGAGTCCTGAG
F20.40	CGAAGGTAGCGACAGTTCCCCTGACGAACTGAGTCCTGAG
Lead20.20	GCTTCCATCGCTGTCAAGGG
Lag20.20	GGGAACTGTCGCTACCTTCG
HJ.IM.2	GCTTCCATCGCTGTCAACCCGACCAGGTGCAGGTTCACCC
HJ.IM.3	GGCATAAAGCTTGACGACCCGACACGAACTGAGTCCTGAG
HJ.IM.4	GGGTGAACCTGCACCTGCAGCCCTCGTCAAGCTTTATGCC
BamH1-mHARP1	TGATCCGGATCCCCCCAAAACACAGGC
NotI-mHARP1	TTATATGCGGCCGCTTACAGTGGCTTCAGGGA
5'-BamHI-UvsW	CGGCAGGGATCCATGGATATTAAAGTACAT
5'-BamHI-UvsW^{ATPase}	GCGAGGGGATCCCCACAAATTAACGAA
3'-NotI-UvsW	TGACGTGCGGCCGCTTATAAATTAACTGTTTT
5'-HARP1Swap	TCTGTTCCAGGGGCCCGGATCCCAGAAGGGAAAGTGCGTAAGG
3'-HARP1Swap	TTTCTTGATAATTCTTCTTTTTCGTTAATTTGTGGCAGAGGCTGCAGGTTGAC
DNA Structure	Oligonucleotides Annealed
Overhang	25.1 + 25.3 (fork regression, ATPase and SAXS); Fam40 + F20.20 (DNA binding)
Splayed Arm	Fam40 + F20.40 (DNA binding)
Flap	Fam40 + Lead20.20 + F20.40 (DNA binding)
Fork	48 + 50 + 52 + 53 (fork regression, ATPase); Fam40 + Lead20.20 + Lag20.20 + F20.40 (DNA binding)
4-way Junction	Fam40 + HJ.IM.2 + HJ.IM.3 + HJ.IM.4 (DNA binding)

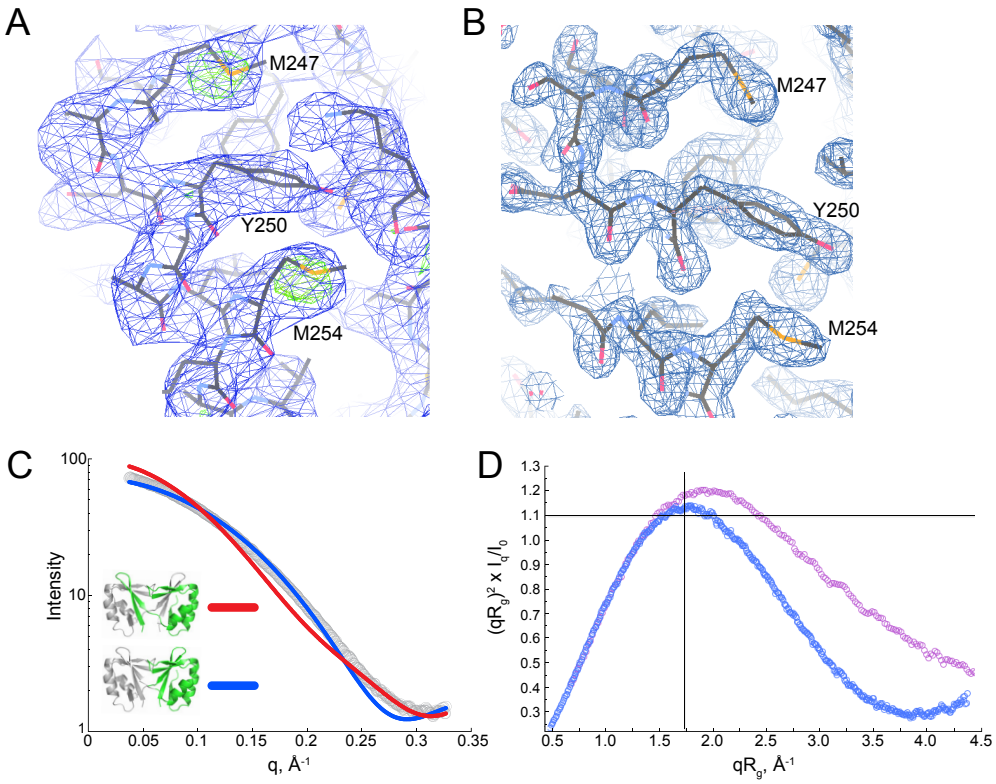


Fig S1. HARP1 structure determination. (A) The refined HARP1 atomic model (gray sticks) superimposed against the 3.3 Å experimental Se-SAD electron density maps (blue, Fo; green, ΔF_{anom}). The region shown corresponds to helix α6, with Met247, Tyr250, and Met254 in the foreground. (B) Final, refined 2Fo-Fc density contoured at 1σ for the same region shown in panel A. (C) Theoretical SAXS curves for models of a single subunit of mouse HARP1 in the dimeric (red) and monomeric forms, superimposed on the experimental SAXS scattering data (grey). (D) Dimensionless Kratky plot of HARP1 domain (blue) and SMARCAL1^{CD} (purple). Convergence of data at the cross-hairs is an indicator of compactness in globular proteins; deviation from the cross-hairs indicates unfolded or asymmetric proteins.

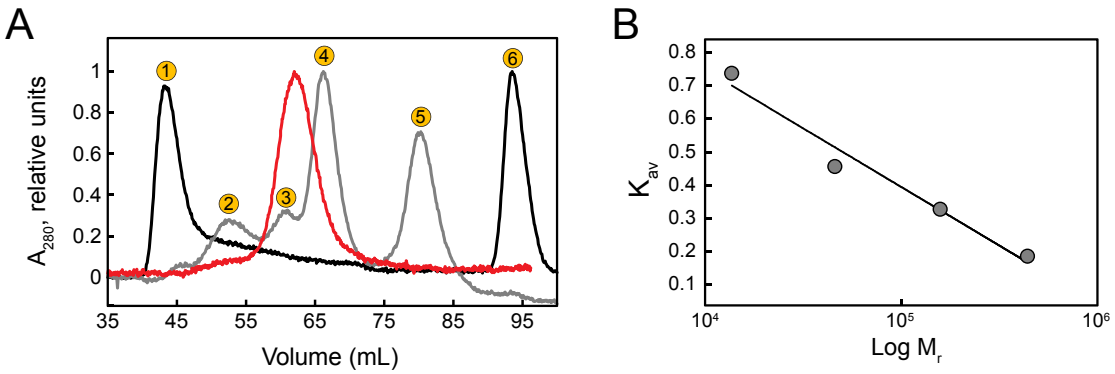
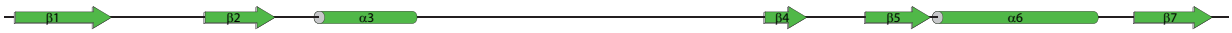


Fig S2. SMARCAL1 is a monomer in solution. (A) Chromatogram of SMARCAL1 (red) and molecular weight standards (black, grey) eluted from a Superose12 size exclusion column running at 0.5 ml/min in 20 mM HEPES (pH 7.6), 200 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, and 20% glycerol. Molecular weight standards: 1, blue dextran 2000 (2 MDa); 2, ferritin (440 kDa); 3, aldolase (158 kDa); 4, ovalbumin (43 kDa); 5, ribonuclease A (13.7 kDa); 6, acetone (58 Da). (B) Standard curve derived from the chromatograms in panel A. The linear regression equation was fit as $K_{av} = -0.154 \ln(M_r) + 2.1645$ with an R^2 of 0.9695. From these data, SMARCAL1 is calculated to have a molecular weight of 112.7 kDa (expected = 105.9 kDa).

A

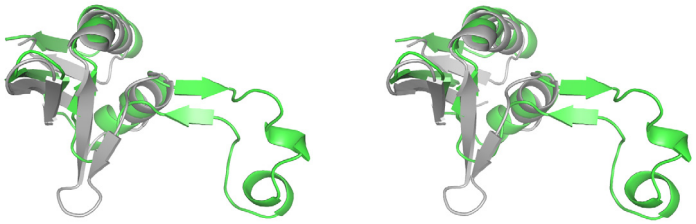



```
mHARP1  202 FLRG-ACIKT-G-----DRFRVKI-GYNQELIAVFKSL----------------------P-----SRHYDSF---TKTWDFSMS--DYRALMKAVERLSTV---SLKPL 268
hHARP1  236 SQKG-KCVRN-G-----DRFQVLI-GYNAELIAVFKTL----------------------P-----SKNYDPD---TKTWNFSMN--DYSALMKAAQSLPTV---NLQPL 302
hHARP2  330 FVKG-RCMLISR-----AYFEADI-SYSQDLIALFKQM----------------------D-----SRRYDVK---TRKWSFLLE--EHSKLIAKVRCLPQV---QLDPL 397
afXPB   003 -MIA-EIYYE-R-----GTIVVKG------dAHVP------------------------------HAKFDSR---SGTYRALAF--RYRDIIEYFESNG-I---EFVDN 058
T4UvsW  001 -MDI-KVHFH-D----fSHVRIDC--EESTFHELRDFFsfeadgyrfnprfrygnwdgriR-----LLDY--------NRLLPFG--LVGQIKKFCDNFG-Y---KAWID 082
tteIF1  016 QTHI-HIRVE-QrrgrkCFTTVEGiPPEFdyEKIMKYW----------------------Kkwlscnativeedegkkvikingd--HRNQIQQFLSEEGIAavdNITIH 099
hMSH6   422 fDLV-ICYKV-G-----KFYELYH-MDaligVSELG------------------------------LVFMKG--nWAHSGFPEI--AFGRYSDSLVQKG-Y---KVARV 483
T4MotA  146 TNGIrNFEINnN-----GNMRIFGyKMMEHHIQKFTDI----------------------G------MSCKIAkngnVYLDIKRSaeNIEAVITVASEL----------- 211
```

B

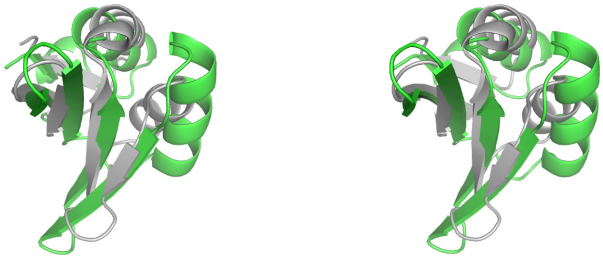
XPB

C



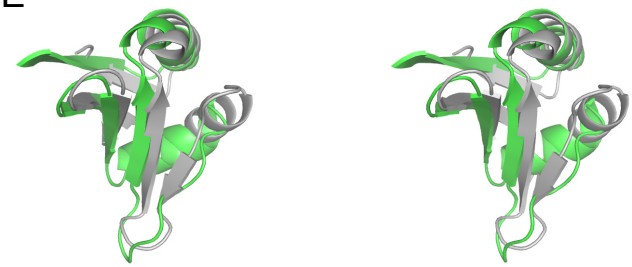
UvsW

D



eIF1

E



MSH6

F

MotA

Fig S3. Substrate recognition domains. Structure based sequence alignment (A) and stereo-diagrams (B-F) of substrate recognition domains from XPB (PDB ID 2FWR), UvsW (2OCA), eIF-1 (2XZM), MSH6 (2O8E), and MotA (1KAF) in green superimposed on SMARCAL1 HARP (silver). h, human; m, mouse; af, *Archaeoglobus fulgidus*; tt, *Tetrahymena thermophila*.

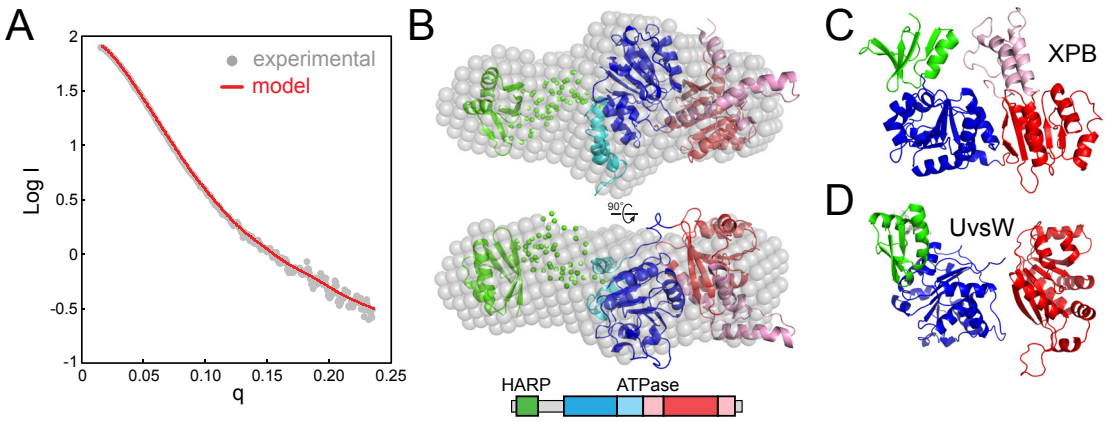


Fig. S4. Similarity between SMARCAL1^{CD} and other structure-specific translocases. (A) Experimental (grey) and calculated (red) SAXS curves for SMARCAL1^{CD} (residues 325-870). (B) Molecular envelope generated from the scattering data using the program BUNCH. Atomic models for HARP and the ATPase region were fit to the scattering data independently using BUNCH as described (7). (C) Crystal structure of archaeal XPB (PDB ID 2FWR). (D) Crystal structure of T4 UvsW (PDB ID 2OCA).

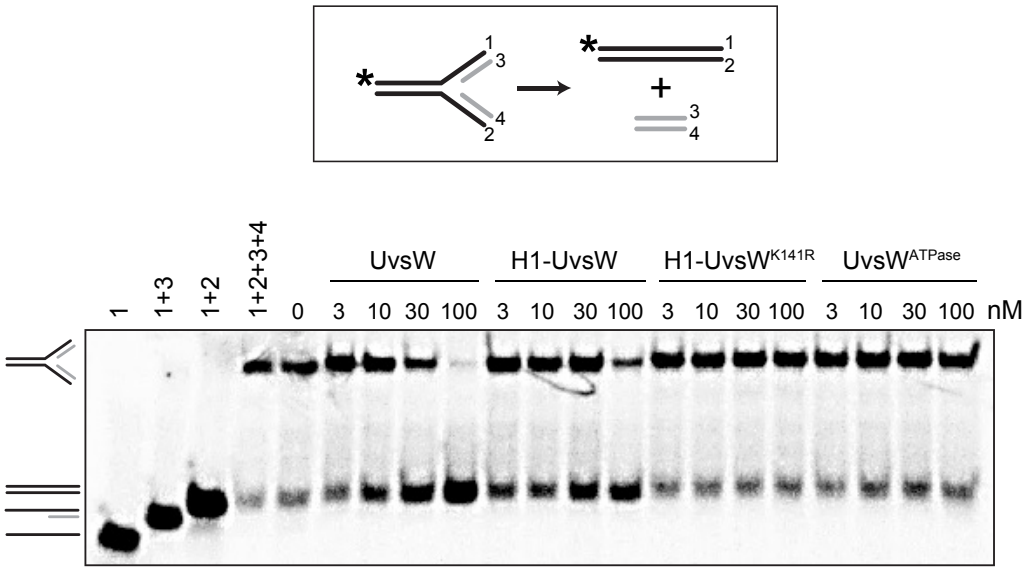
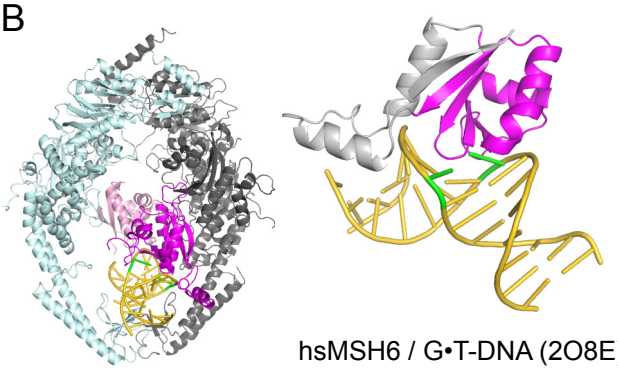


Fig S5. Fork regression activity of UvsW and SMARCAL1-UvsW chimera. *Top*, reaction schematic. *Bottom*, raw data for the experiment shown in Fig. 2. Fork regression substrate and products were separated by native gel electrophoresis. Lanes 1-4 are standards showing the migration of possible DNA structures. The numbers at the top of the gel refer to the nM concentration of protein.

A

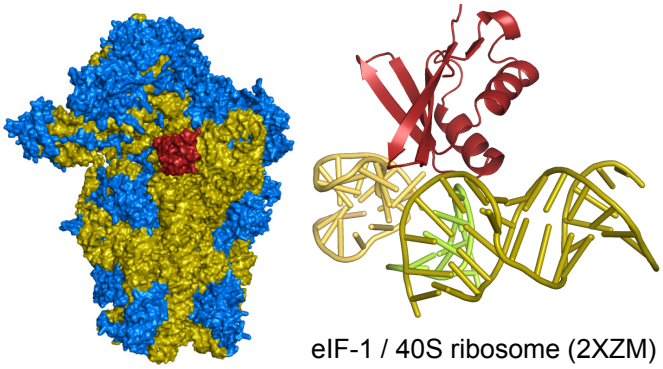
```
mmHARP1 202 FLRGACIKT-GDRFRVKIGYNQELIAVFKSLPSRHYDS------F-------TKTWDFSMSDYRALMKAVERLSTVSLKPL 268
hsHARP1 236 SQKGKCVRN-GDRFQVLIGYNAELIAVFKTLPSKNYDP------D-------TKTWNFSMNDYSALMKAAQSLPTVNLQPL 302
hsHARP2 330 FVKGRCMLISRAYFEADISYSQDLIALFKQMDSRRYDV------K-------TRKWSFLLEEHSKLIAKVRCLPQVQLDPL 397
hsMSH6  422 fDLVICYKV-GKFYELYHMDaligVSELG----LVFMK------G------nWAHSGFPEIAFGRYSDSLVQKG-YKVARV 483
hsMSH3  245 -DAVLCVEC-GYKYRFFGEDaeiAARELN----IYCHL------D-----hnFMTASIPTHRLFVHVRRLVAKG-YKVGVV 307
ecMutS   25 -EILLFYRM-GDFYELFYDDakrASQLLD----ISLT---------------PMAGIPYHAVENYLAKLVNQG-ESVAIC 092
hsMSH2   32 -TTVRLFDR-GDAYTAHGEDallaAREVFK-tqgVIKY------MgpagaknLQSVVLSKMNFESFVKDLLLVRQYRVEVY 103
taMutS   29 -DYLLLFQV-GDFYECFGEDaerlARALG----LVLTHktskdfT-------TPMAGIPLRAFEAYAERLLKMG-FRLAVA 095
```

B



hsMSH6 / G•T-DNA (2O8E)

C



eIF-1 / 40S ribosome (2XZM)

Fig S6. Structural comparison of HARP1 to known DNA and RNA binding domains. (A) Structure based sequence alignment of HARP1, HARP2, and the mismatch recognition domain (MRD) from the MSH family of proteins. mm, *Mus musculus*, hs, *Homo sapiens*; ec, *Escherichia coli*; ta, *Thermus aquaticus*. (B) Left, crystal structure of human MutSα heterodimer bound to DNA containing a G•T mismatch (2O8E). MSH6 is colored grey and magenta, MSH2 is colored cyan and pink, and the DNA is gold with the mismatch green. Right, close-up of the MSH6 MRD (magenta) in contact with the mismatched nucleotide (green). (C) Left, structure of eIF-1 bound to the 40S ribosome (2XZM). Ribosomal proteins and RNA are colored blue and olive, respectively, and eIF-1 is red. Right, close-up of eIF-1 (red) bound to 18S rRNA, colored by RNA strand.

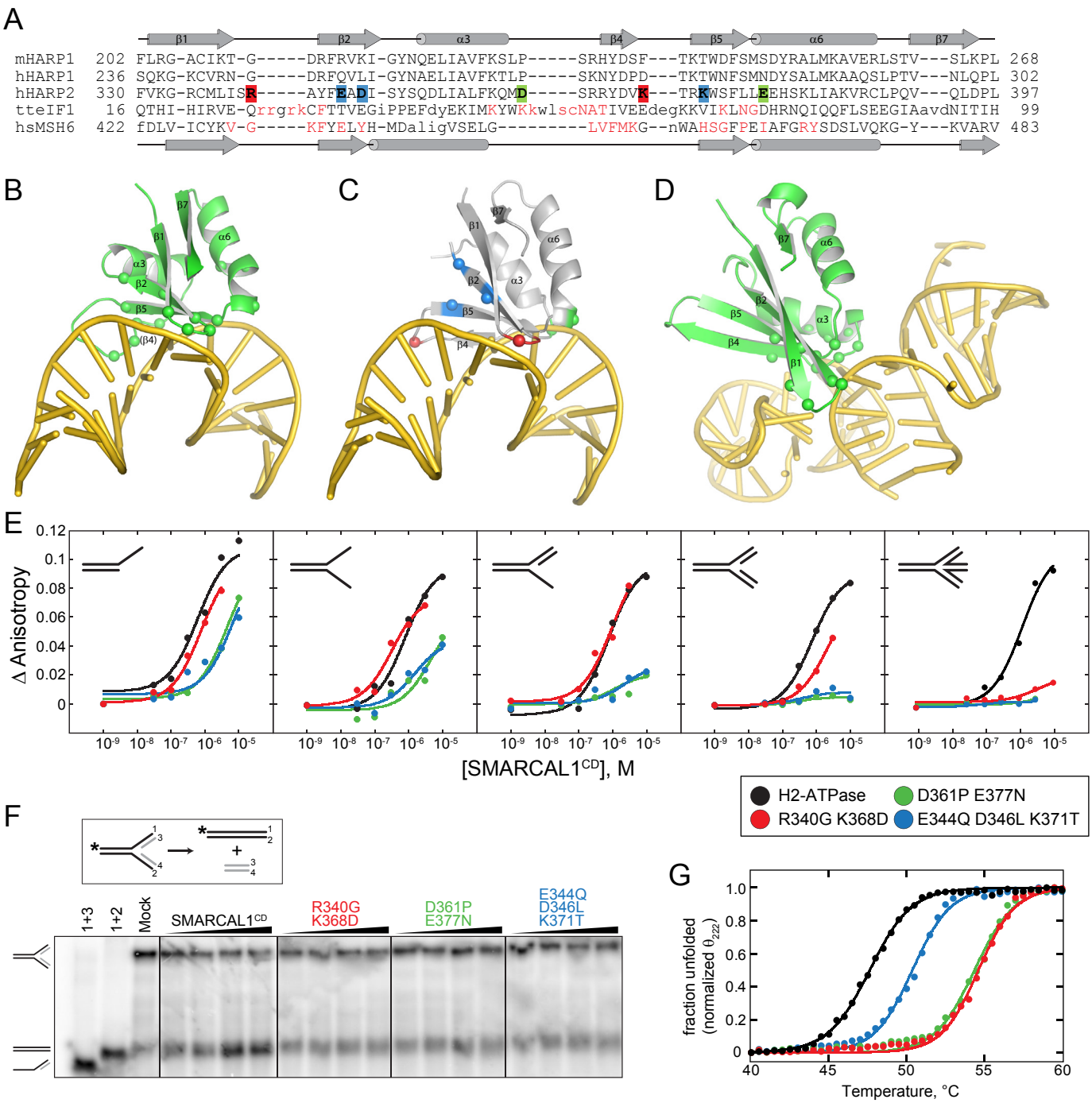


Fig S7. Design of HARP DNA binding mutants. (A) Structure based sequence alignment of mHARP1, eIF-1 and MSH6, with hHARP sequences aligned to mHARP1. Residues mutated in SMARCAL1^{CD} are shaded red, green and blue. Residues that contact nucleic acid in MutS and eIF-1 are in red text. (B) MRD from human MSH6 bound to mismatch DNA (2O8E). DNA contacting residue Cα atoms are shown as spheres. (C) A model of HARP bound to DNA was created by superposition of HARP1 onto MSH6 (panel B). Residues predicted to contact DNA are shown as spheres and colored according to mutation. (D) eIF-1 bound to RNA (2XZM). (E) DNA binding isotherms for various DNA structures quantified in Fig. 3C. (F) Fork regression data for SMARCAL1^{CD} cluster mutants, quantified in Fig. 3F. (G) Protein unfolding was monitored by CD spectroscopy, in which the change in molar ellipticity at 222 nm was followed as a function of temperature. Data were fit using the equation, $f_u = 1/(1+e^{(T_m-T)/k})$, in which f_u is the fraction unfolded, T_m is the temperature at 50% denaturation, and k is the cooperativity of the transition.

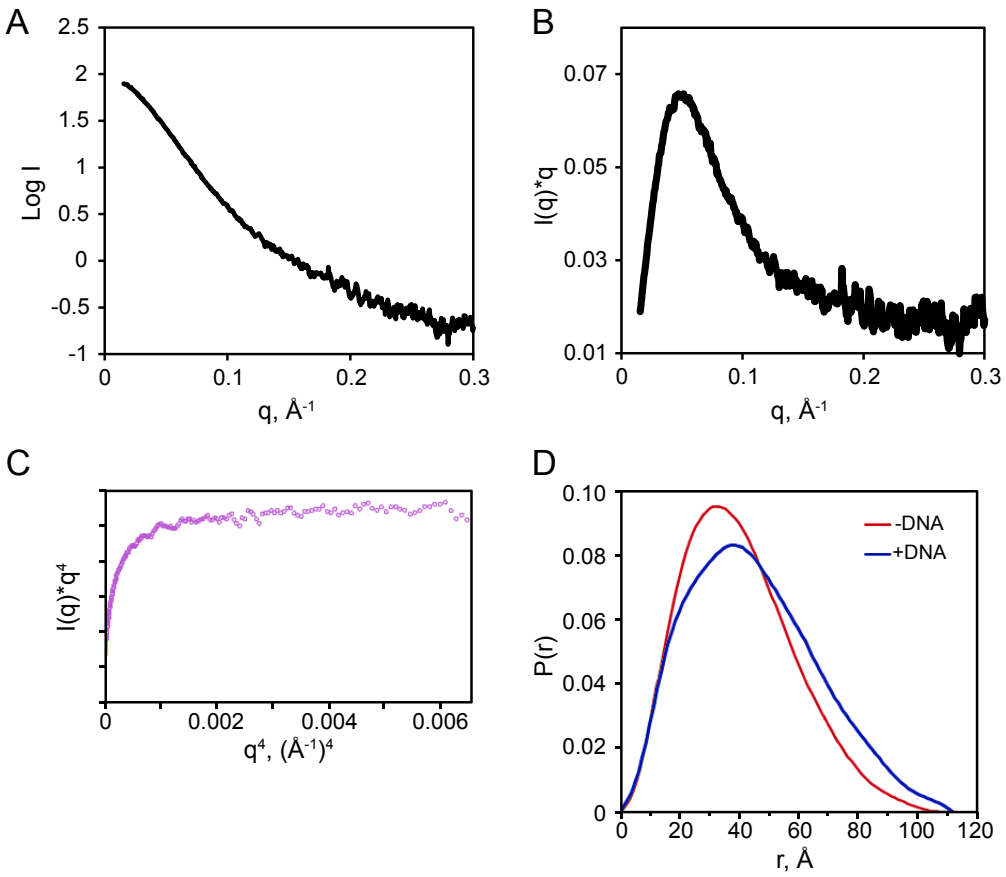


Fig S8. SAXS analysis of SMARCAL1^{CD} bound to overhang DNA. The DNA substrate contained 25 base pairs and a 10-nucleotide 5′ ssDNA overhang. (A) Raw X-ray scattering curve. (B) Kratky plot. (C) Porod-Debye plot of SMARCAL1^{CD} alone exhibits the typical hyperbolic asymptote of a compact particle. (D) P(r) distributions of SMARCAL1^{CD} in the absence (red) and presence (blue) of overhang DNA.

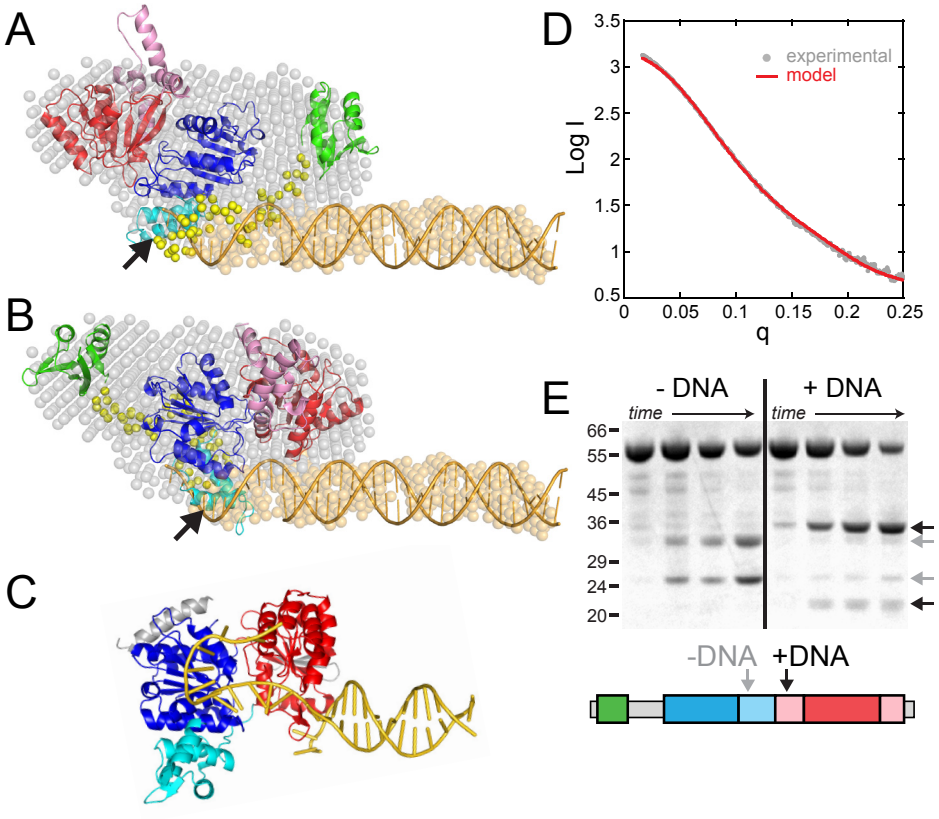
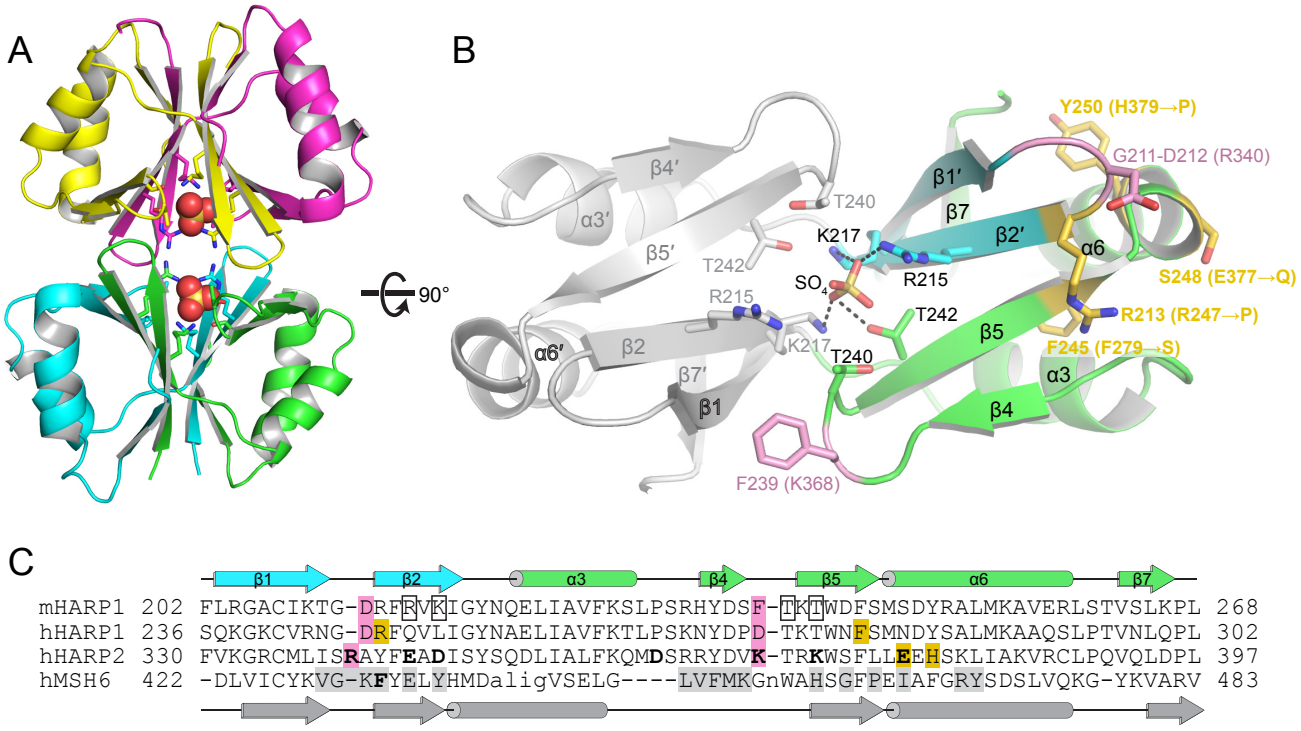


Fig. S9. SAXS model for HARP2-ATPase binding ss/dsDNA junctions. (A,B) SAXS molecular envelopes determined by MONSA for SMARCAL1^{CD} (gray spheres) and 5′-DNA overhang substrate (gold spheres), superimposed against ideal B-DNA (gold) and two independent SMARCAL1^{CD} models (colored as in Fig. 1B), using Sso1653 (PDB ID 1Z63) as the SNF2 ATPase domain. The molecular envelopes are the same and are rotated ~90° about a horizontal axis relative to the view as that shown in Figure 4B. Envelopes are the average of 7 independent MONSA runs and aligned with DAMAVER. SUPCOMB was used to place the molecular model of HARP2-ATPase and the overhang DNA into the SAXS envelope. SUPCOMB returned two possible orientations of the protein (A and B), with goodness-of-fits (χ^2) of 1.18 (A) and 1.22 (B). In both orientations, the HD1 domain (cyan) at the C-terminus of ATPase-N domain (blue) contacts the DNA. Black arrows indicate sites of proteolytic sensitivity that are protected in the presence of DNA (see panel E). (C) Crystal structure of *Archaeoglobus fulgidus* Hel308 bound to DNA containing a ssDNA overhang (PDB ID 2P6R). (D) MONSA fit (red) of the models shown in panels A and B superimposed on the experimental SAXS data (gray circles). (E) Coomassie stained SDS-PAGE gel showing limited proteolysis of SMARCAL1^{CD} in the absence (-) and presence (+) of 5′-overhang DNA. Protein and 1.2-fold molar excess of DNA were incubated with trypsin in a 12:1 (SMARCAL1^{CD}:trypsin) molar ratio at 37°C for 0, 2, 10, and 20 min. Proteolytically sensitive sites were determined by mass spectrometry of bands marked by black and grey arrows in the gel and are shown against the SMARCAL1^{CD} schematic at the bottom.



C

Fig. S10. Location of SIOD mutations on the HARP1 crystal structure. (A) The crystallographic asymmetric unit, colored by polypeptide chain. Sulfate ions and interacting side chains are rendered as CPK spheres and sticks, respectively (B) Close-up of the sulfate binding region. The view is rotated 90° about a horizontal axis from the view shown in panel A. Only two protomers are shown for clarity, colored according to the monomeric HARP1 fold (one subunit is blue/green and the other is silver). Noteworthy residues are rendered as sticks and colored according to function. Sulfate binding residues are green, blue, and grey. Residues mutated in SIOD patients are colored gold, and positions that affect DNA binding in HARP2 are pink. Residue numbers refer to mouse HARP1 positions, and numbers in parentheses refer to human HARP1 or HARP2. (C) HARP structure-based sequence alignment. Positions of SIOD mutations are highlighted gold. Residues shown by mutagenesis to affect DNA binding are in boldface, with the putative DNA binding residues R340/K368 in HARP2 highlighted pink and MSH6 DNA binding residues highlighted grey. Sulfate binding residues are boxed.