

Supplementary Information

Molecular basis for CSB stimulation of the SNM1A DNA repair nuclease

Rebecca Roddan¹, Matthew A. Schaich^{2,3}, William J. Nathan⁴, Iulia Kis¹, Andrea Pierangelini⁵, Lucy R. Henderson¹, Raman van Wee^{6,7}, Louise Martin¹, Afaf El-Sagheer⁸, Philipp Kukura⁶, Tom Brown⁹, Darragh O'Brien⁵, Joseph A. Newman¹⁰, David M. Wilson III¹¹, Christopher J. Schofield^{9,12}, Bennett Van Houten^{2,3}, and Peter J. McHugh^{1*}.

¹Department of Oncology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, OX3 9DS, U.K.

²Department of Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.

³UPMC-Hillman Cancer Center, Pittsburgh, PA, 15232, USA.

⁴Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA.

⁵Target Discovery Institute, Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7FZ, U.K.

⁶Physical and Theoretical Chemistry, Department of Chemistry, Kavli Institute for Nanoscience Discovery, Dorothy Crowfoot Hodgkin Building, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

⁷Department of Physiology, Anatomy and Genetics, University of Oxford, OX1 3QX, U.K.

⁸School of Chemistry, University of Southampton, SO17 1BJ, U.K.

⁹Chemistry Research Laboratory, 12 Mansfield Road, Department of Chemistry, University of Oxford, Oxford, OX1 3TA, U.K.

¹⁰Centre for Medicines Discovery, University of Oxford, NDMRB, Oxford, OX3 7DQ, U.K.

¹¹Biomedical Research Institute, Hasselt University, 3590 Diepenbeek, Belgium.

¹²Ineos Oxford Institute for Antimicrobial Research, University of Oxford, Oxford, OX1 3TA, U.K.

*Corresponding author: Peter J. McHugh

Table of Contents

Figure S1: SDS-PAGE of purified proteins used.....	4
Table S1: Parameters from single-cycle SPR fitting.	4
Table S2: Parameters from single-cycle SPR fitting when wild-type N-His-ZB-CSB-Avi-Biotin was immobilized on the chip.	5
Figure S2: Multi-cycle and dose response curves for CSB (a.a. 1187-1493), CSB (a.a. 1308-1493) and CSB (a.a. 1424-1493).	5
Fig S3: Peptide coverage obtained for CSB in HDX-MS experiments.	6
Figure S4: CSB peptides with changes in deuterium uptake in HDX-MS experiments upon interaction with SNM1A.....	7
Fig S5: Changes in deuterium uptake in the CSB protein upon interaction with SNM1A are localised to the CSB-WHD.	8
Fig S6: Multi-cycle and dose response curves for CSB (a.a. 1424-1493, wild-type T1447A.	9
Fig S7: Circular dichroism spectroscopy of isolated CSB-WHD. Wild-type and T1447A.....	9
Table S3: Oligonucleotide sequences used for biochemical assays. P refers to monophosphate.	9
Figure S8: DNA structures used in biochemical assays.	9
Figure S9: The substrate scope of CSB stimulation of SNM1A activity.....	10
Table S4: Oligonucleotides used to prepare ICL-containing DNA repair intermediate for single-molecule studies.	10
Figure S10: Isolation of Cu(I)-click ICL-containing oligonucleotides. A.....	11
Figure S11: The ATPase activity of CSB is not involved in the stimulation of SNM1A activity.....	11
Figure S12: E647Q-CSB is ATPase-dead, and this does not affect the DNA binding ability of CSB.	11
Figure S14: EMSA of C-terminal domains of CSB binding to DNA.....	12
Figure S16: Akta trace of purified full-length CSB with corresponding calibration curve.....	13
Figure S17: Mass photometry of CSB lacking N or C terminal domains. Measurements were performed in the absence/presence of MgCl₂ and it was determined that Mg²⁺ does not influence CSB dimerisation.....	14
Figure S18: SEC-MALS of ΔN-CSB	14
Figure S19: Co-purifications of CSB and SNM1A.....	15
Figure S20: Size exclusion calibration curves.....	16
Figure S21: SDS-PAGE gels of nuclear extracts used for single-molecule studies	16
Figure S22: Western blot analysis of CSB-transfected nuclear extracts.....	17

<u>Table S5: Parameters determined for protein:DNA binding events on lambda DNA in the LUMICKS' C-Trap.</u>	<u>18</u>
<u>Table S6: Parameters determined for protein:DNA binding events on the ICL lesion in the LUMICKS' C-Trap.</u>	<u>18</u>
<u>Figure S23: Kymograph showing 2 different colours of CSB protein diffusing together on lambda DNA.</u>	<u>19</u>
<u>Figure S24: Histograms of average CSB photon counts along a kymotrack</u>	<u>19</u>
<u>Table S7: Statistical analysis (AICc test) to determine whether binding event lifetimes (CTRD plots) exhibited one- or two-phase decay.</u>	<u>19</u>
<u>Table S8: Statistical analysis (AICc test) of SNM1A recruitment times to DNA-bound CSB</u>	<u>19</u>

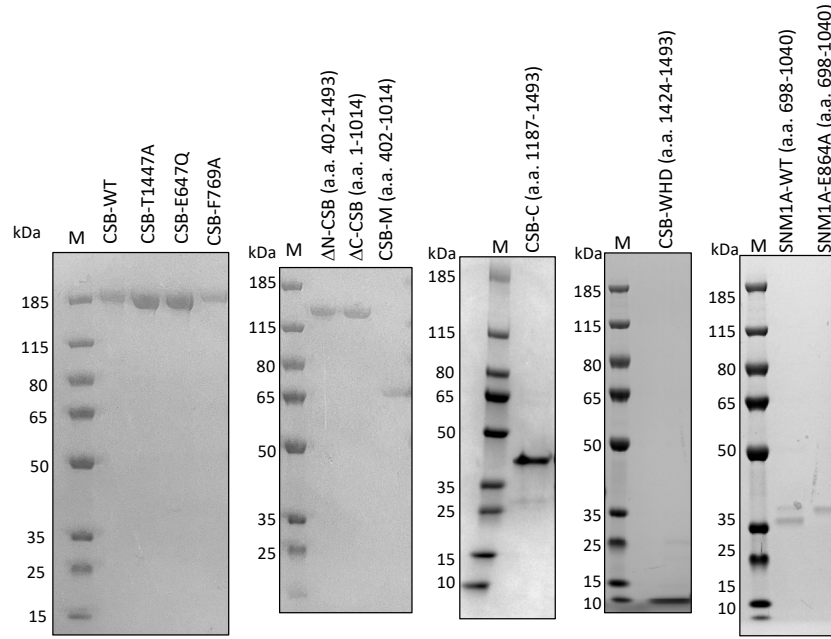


Figure S1: SDS-PAGE of purified proteins used. M = Prestained PageRuler Plus protein ladder. Predicted protein masses are: FL-CSB (WT, T1447A, E647Q, F796A) – 180 kDa, Δ C-CSB – 136 kDa, Δ N-CSB – 132 kDa, CSB-M – 72 kDa, CSB-C – 42 kDa, CSB-WHD – 8 kDa, SNM1A (WT + E864A) – 37 (untagged) and 42 kDa (tagged).

Table S1: Parameters from single-cycle SPR fitting. Experiments performed where N-Avi-Biotin- Δ N-SNM1A proteins (WT or E864A) were immobilized on the chip. See Methods section for experimental details.

Analyte		Immobilized	
		WT-SNM1A	E864A-SNM1A
	FL-CSB	$k_a = 4.16 \times 10^5 \pm 1.1 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ $k_d = 0.001991 \pm 4.6 \times 10^{-5} \text{ s}^{-1}$ $K_D = 4.79 \times 10^{-9} \text{ M}$ Drift = 0.3 RU s^{-1} (fixed) $t_c = 1.88 \times 10^7 \pm 6.5 \times 10^4$	Not determined
	CSB (a.a. 1187-1493)	$k_a = 3.82 \times 10^5 \pm 6.3 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ $k_d = 0.004048 \pm 6.5 \times 10^{-5} \text{ s}^{-1}$ $K_D = 1.06 \times 10^{-8} \text{ M}$ Drift = 0.3894 RU s^{-1} (fitted) $t_c = 3.01 \times 10^7 \pm 9.2 \times 10^4$	$k_a = 5.61 \times 10^5 \pm 3.3 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$ $k_d = 0.007121 \pm 4.2 \times 10^{-6} \text{ s}^{-1}$ $K_D = 1.27 \times 10^{-8} \text{ M}$ Drift = 0.3894 RU s^{-1} (fitted) $t_c = 4.07 \times 10^7 \pm 1.9 \times 10^4$
	CSB (a.a. 1424-1493)	$k_a = 1.02 \times 10^5 \pm 3.0 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ $k_d = 0.009041 \pm 2.5 \times 10^{-4} \text{ s}^{-1}$ $K_D = 8.90 \times 10^8 \text{ M}$ Drift = 0.1231 RU s^{-1} (fitted) $t_c = 3.12 \times 10^6 \pm 1.3 \times 10^4$	$k_a = 3.76 \times 10^5 \pm 1.0 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ $k_d = 0.01098 \pm 3.0 \times 10^{-4} \text{ s}^{-1}$ $K_D = 2.92 \times 10^8 \text{ M}$ Drift = 0.845 RU s^{-1} (fitted) $t_c = 4.54 \times 10^7 \pm 5.9 \times 10^5$
		$k_a = 644.7 \pm 1.1$ $k_d = 7.04 \times 10^{-4} \pm 2.9 \times 10^{-6} \text{ s}^{-1}$ $K_D = 1.09 \times 10^6 \text{ M}$ Drift = 0.06 RU s^{-1} (fixed) $t_c = 3.00 \times 10^6$ (fixed)	

Table S2: Parameters from single-cycle SPR fitting when wild-type N-His-ZB-CSB-Avi-Biotin was immobilized on the chip. See Methods section for experimental details.

Analyte	WT-SNM1A	E864A-SNM1A
Parameters determined	$k_a = 6.88 \times 10^5 \pm 5.8 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ $k_d = 0.002993 \pm 1.8 \times 10^{-5} \text{ s}^{-1}$ $K_D = 4.35 \times 10^{-9} \text{ M}$ Drift = 0.3 RU s^{-1} (fixed) $t_c = 6.80 \times 10^7$	$k_a = 1.71 \times 10^5 \pm 1.1 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ $k_d = 0.001386 \pm 5.7 \times 10^{-6} \text{ s}^{-1}$ $K_D = 8.11 \times 10^{-9} \text{ M}$ Drift = 0.3 RU s^{-1} (fixed) $t_c = 6.80 \times 10^7$

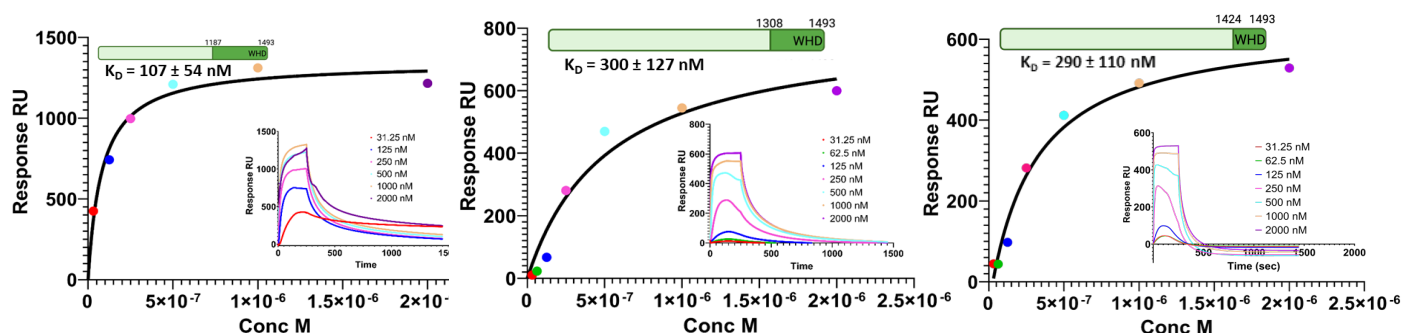


Figure S2: Multi-cycle and dose response curves for CSB (a.a. 1187-1493), CSB (a.a. 1308-1493) and CSB (a.a. 1424-1493). N-Avi-Biotin-SNM1A was immobilised on the chip. Dissociation constants are given. For CSB a.a 1187-1493, data for analyte at 62.5 nM was omitted due to the presence of an air bubble during the injection.



Fig S3: Peptide coverage obtained for CSB in HDX-MS experiments. 376 peptides were observed with 80% sequence coverage, 3.1 average redundancy.

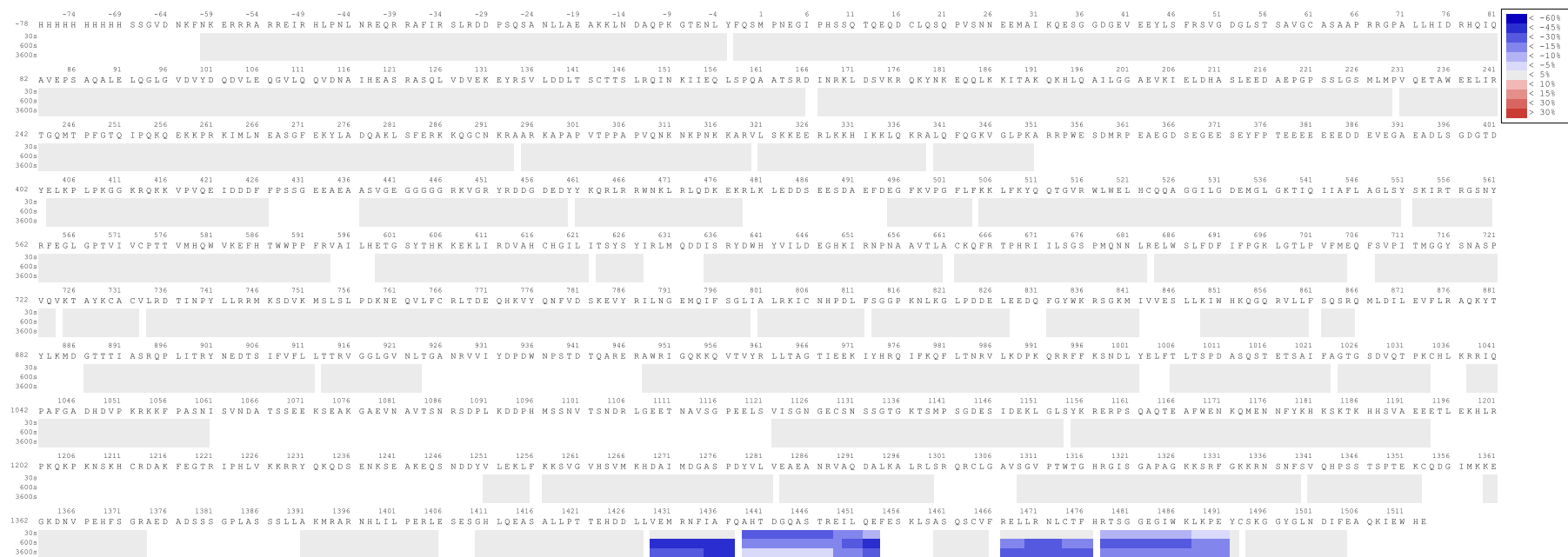


Figure S4: CSB peptides with changes in deuterium uptake in HDX-MS experiments upon interaction with SNM1A.

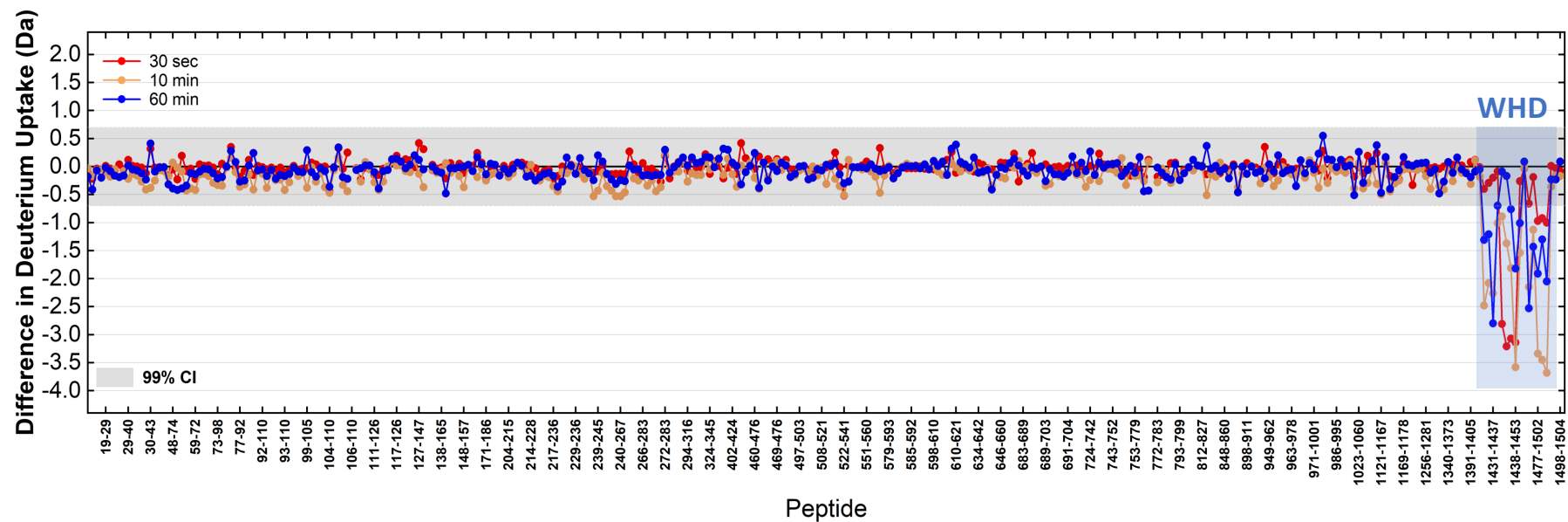


Fig S5: Changes in deuterium uptake in the CSB protein upon interaction with SNM1A are localised to the CSB-WHD.

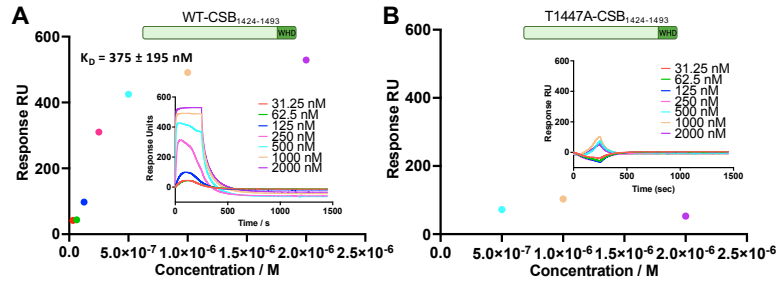


Fig S6: Multi-cycle and dose response curves for CSB (a.a. 1424-1493, wild-type T1447A. N-Avi-Biotin-SNM1A was immobilised on the chip. Dissociation constants are given.

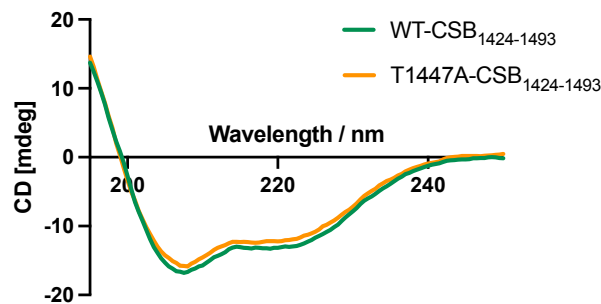


Fig S7: Circular dichroism spectroscopy of isolated CSB-WHD. Wild-type and T1447A.

Table S3: Oligonucleotide sequences used for biochemical assays. P refers to monophosphate.

Code	Sequence of oligonucleotide (5' to 3')
WN1	P-CAATTAGCTGAGGGTCACGTTTAATTAATTATTACGGAAGAGTGTGAAG-Biotin
WN10	Biotin-TCTTCACACTCTCCGTTAATAATTAATAAACGTGACCCTCAGCTAATT
WN9	P-AACGGAAGAGTGTGAAG
WN2	Biotin-TCTTCACACTCTCCGTAGGACCGACGGATTGGTGACCCTCAGCTAATT
WN10 short	Biotin-TCTTCACACTCTCCGTTAATAATT
'0-nt' ICL	Top strand: YACGGAAGAGTGTGAAG (Y = azide) Bottom strand: TCTTCACACTCTCCGT X GGACCGACGGATTGGTGACCCTCAGCTAATT (X = alkyne)

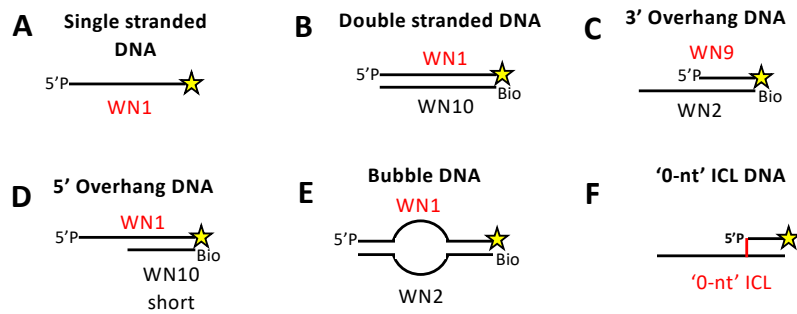


Figure S8: DNA structures used in biochemical assays. The annealing protocol is described in the Methods section. 5'P denotes a 5' phosphate and Bio represents a biotin modification. The radiolabeled oligonucleotide is given in red, and the location of labelling annotated by a yellow star.

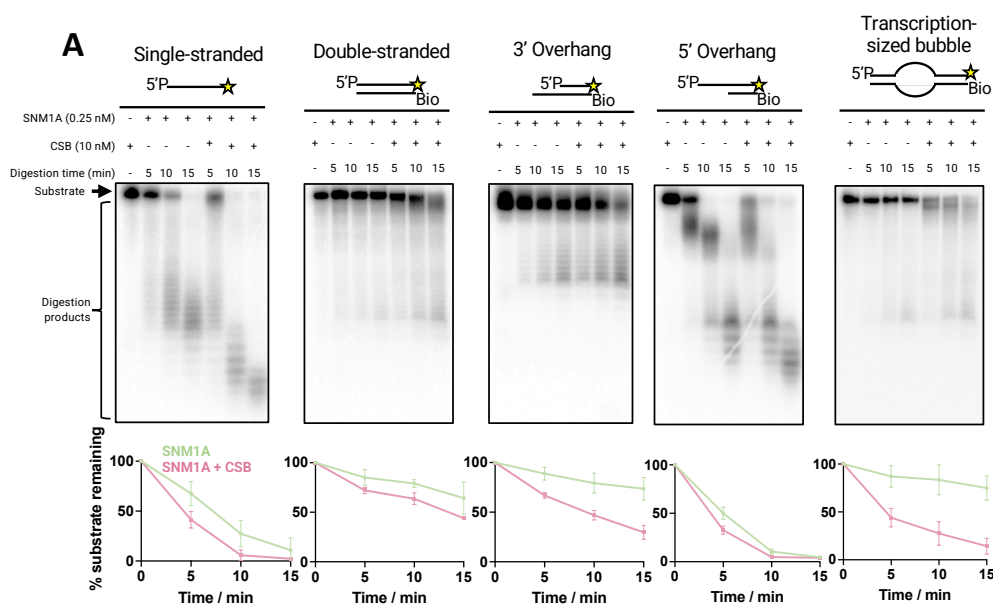


Figure S9: The substrate scope of CSB stimulation of SNM1A activity. Oligonucleotides used (with preparation method) are given in Table S3 and Fig. S8. The location of the 3' 32P is given by a yellow star, Bio represents the location of 5' biotins on the oligo and 5'P the location of a 5' phosphate group. Quantification was based upon assays conducted in triplicate; error bars represent one standard deviation. Time course assays of SNM1A activity on a panel of DNA repair intermediate substrates in the absence of WT-CSB.

Table S4: Oligonucleotides used to prepare ICL-containing DNA repair intermediate for single-molecule studies.

Top strand 1	CAAC GCTCTTC AATTAGCTGAGGGTCAC
Top strand 2	P-GTTTAATTAATTATT Y ACGGAAGAGTGTGAAG (Y = azide)
Bottom strand	ACCA CTCACACTCTTCCGT X GGACCGACGGATTG GTGACCCTCAGCTAATT GAAGAGC (X = alkyne)

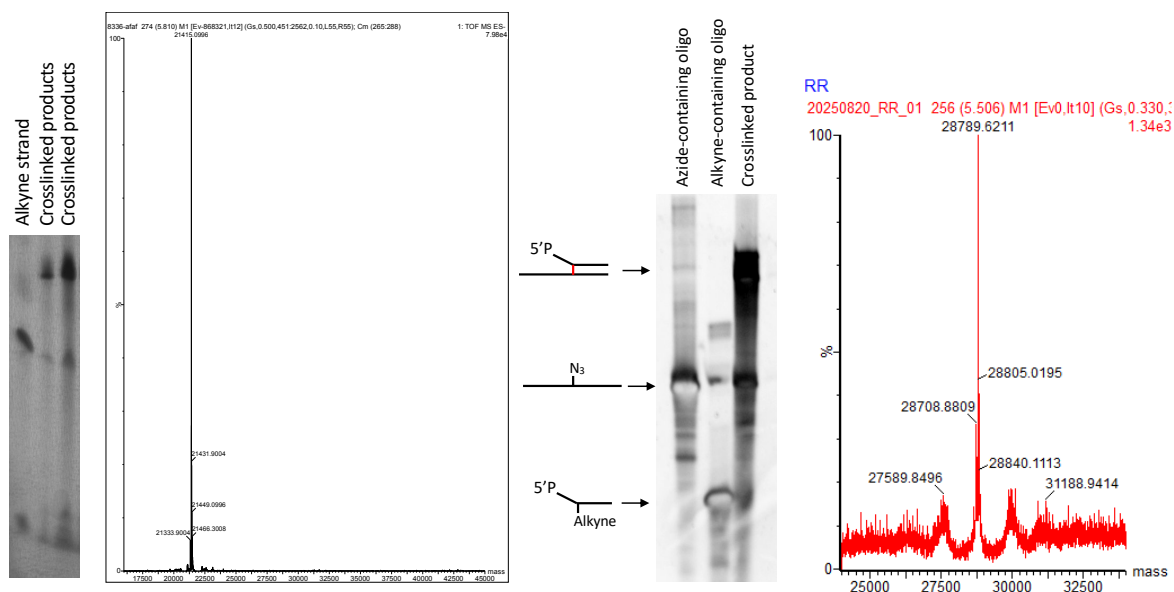


Figure S10: Isolation of Cu(I)-click ICL-containing oligonucleotides. **A.** Gel purification of the '0-nucleotide' ICL substrate used for biochemical assays. 8% denaturing PAGE gel showing the un-reacted alkyne-containing strand (left lane) and the reaction products from the Click-iT reaction between the azide- and alkyne-containing strands (all other lanes). The larger, 'clicked' products were excised from the gel and purified as described in the Methods section. See Supplementary Figure 1 for oligonucleotide structure. **B.** Q-TOF MS of the '0-nt ICL' substrate. Expected $[M+H]^-$ m/z = 21414.5, observed $[M+H]^-$ m/z = 21415.0996 Da. **C.** 10% denaturing PAGE gel of the ICL-containing substrate used for single-molecule studies, showing the azide and alkyne-containing oligos alone, and post crosslinking. The crosslinked product was run out on a 10% denaturing PAGE gel and the band corresponding to the higher molecular weight product was excised from the gel and purified as detailed in the Methods section. **D.** Q-TOF MS of the ICL-containing oligonucleotide tethered in the LUMICKS' C-Trap. Expected $[M+H]^-$ m/z = 28789.1, observed $[M+H]^-$ m/z = 28789.6211.

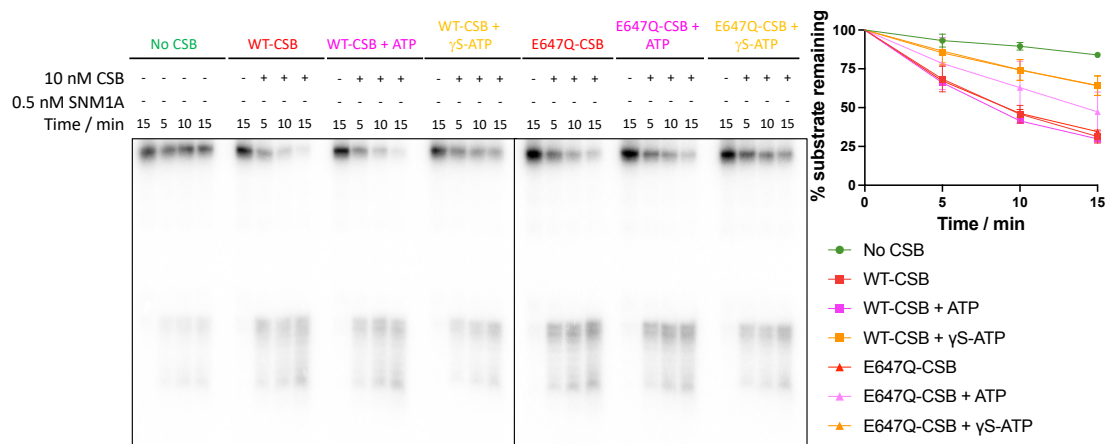


Figure S11: Evidence that the ATPase activity of CSB is not involved in the stimulation of SNM1A activity. The presence of ATP does not affect the enhancement of SNM1A activity by CSB. ATP was used at 50 μ M final and hydrolysis observed on the '0-nucleotide ICL' substrate used.

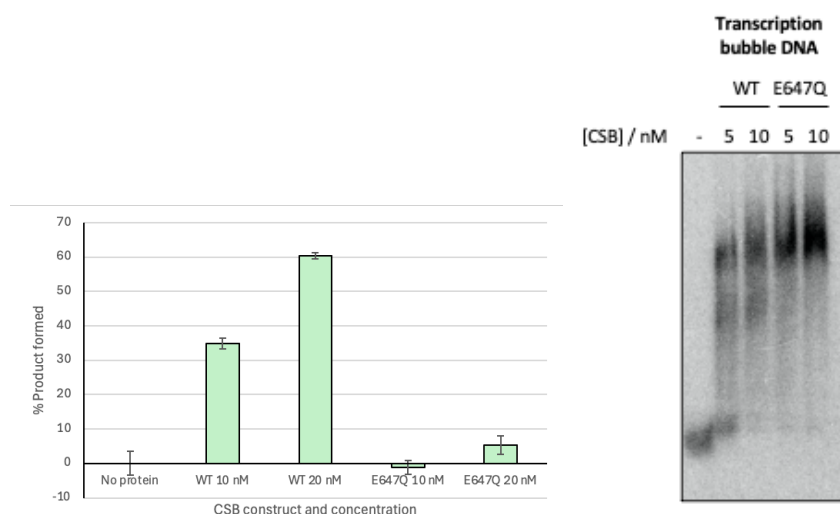


Figure S12: E647Q-CSB is ATPase-dead, and this does not affect the DNA binding ability of CSB. ATPase reactions were performed as detailed in the methods section using the Malachite Green assay, detecting for release of orthophosphate. EMSA of WT and E647Q.

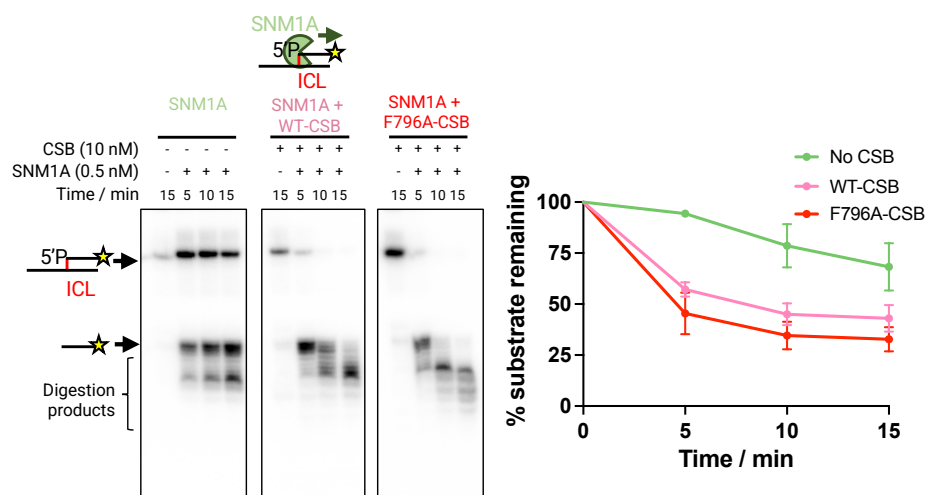


Figure S13: Pulling hook dead CSB (F796A) is not involved in the stimulation of SNM1A activity

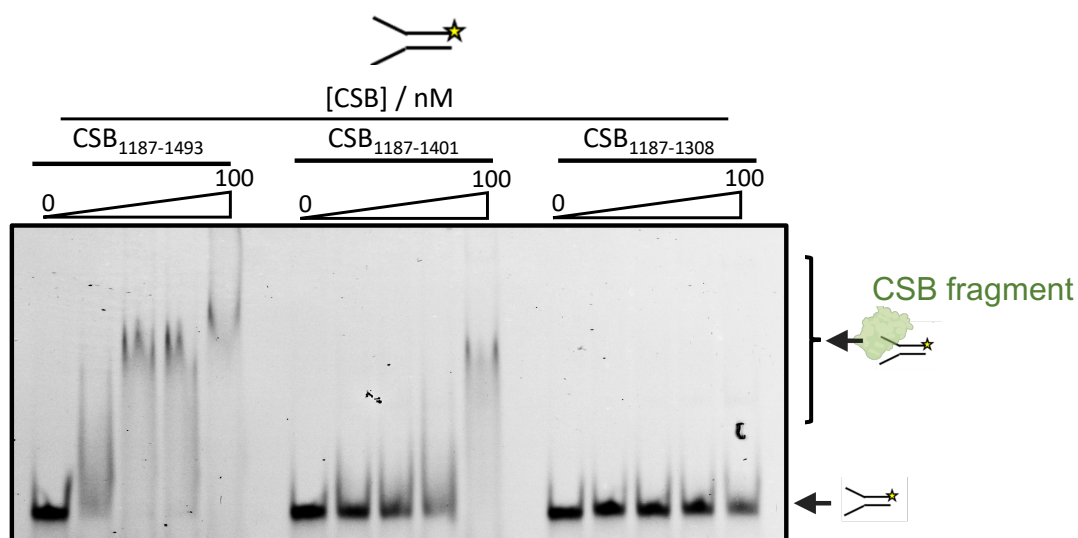


Figure S14: EMSA of C-terminal domains of CSB binding to DNA. Fork DNA (as reported in Abdullah *et al.* 2017) was chosen as the substrate because it possesses both double stranded and single stranded character. CSB fragments CSB₁₃₀₈₋₁₄₉₃ and CSB₁₄₂₄₋₁₄₉₃ were not found to bind DNA (data not shown). The results are representative of three repeats.

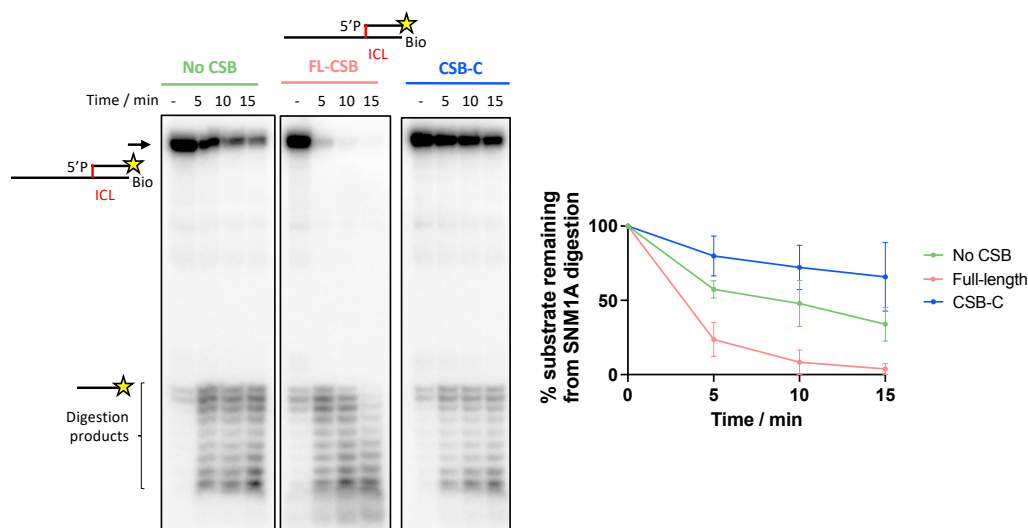


Figure S15: Nuclease assay of SNM1A activity in the presence of different CSB domains. Quantification is performed from three repeats with the standard deviation given in the error bars.

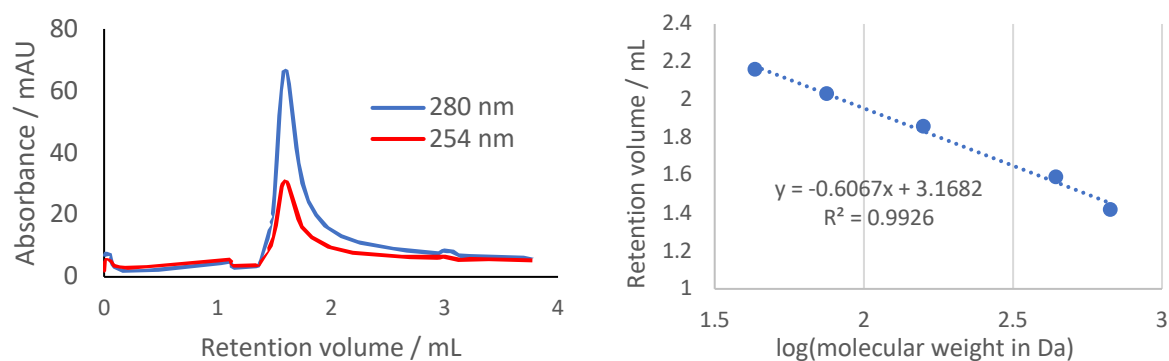


Figure S16: Akta trace of purified full-length CSB with corresponding calibration curve. The peak corresponds to a retention volume of 1.56 mL on Superdex S200 GL 5/150, corresponding to a molecular weight of 440 kDa (predicted dimer = 336 kDa) based on the calibration curve given. The standards used were thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (43 kDa).

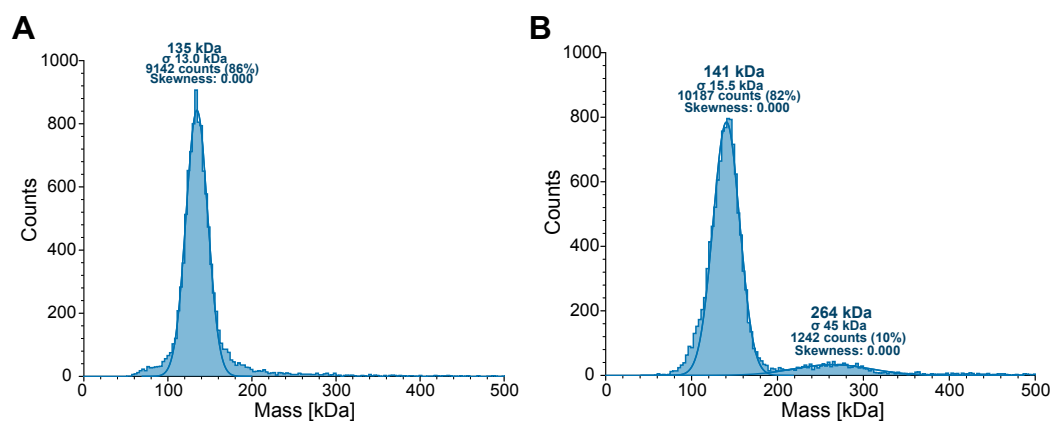


Figure S17: Mass photometry of Δ C-CSB and Δ N-CSB. **A.** MP of Δ C-CSB. Expected monomer = 136 kDa, expected dimer = 272 kDa. **B.** MP of Δ N-CSB. Expected monomer = 132 kDa, expected dimer – 264 kDa. Experiments were performed as described in Materials and Methods.

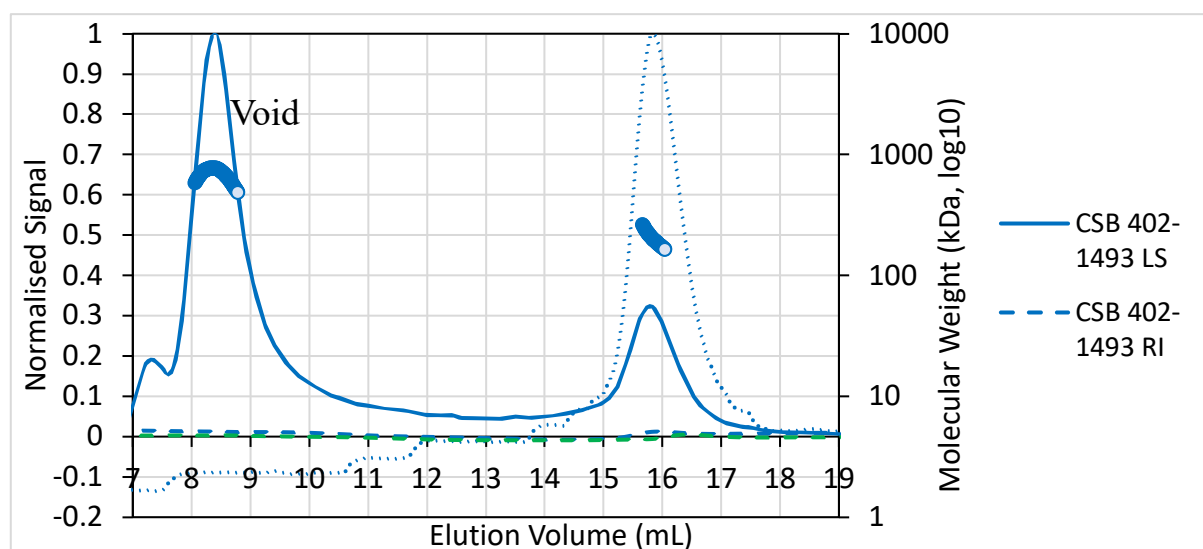


Figure S18: SEC-MALS of Δ N-CSB. The peak at 15.8 mL corresponds to Δ N-CSB of molecular weight 199 kDa. Expected monomer = 132 kDa, expected dimer – 264 kDa.

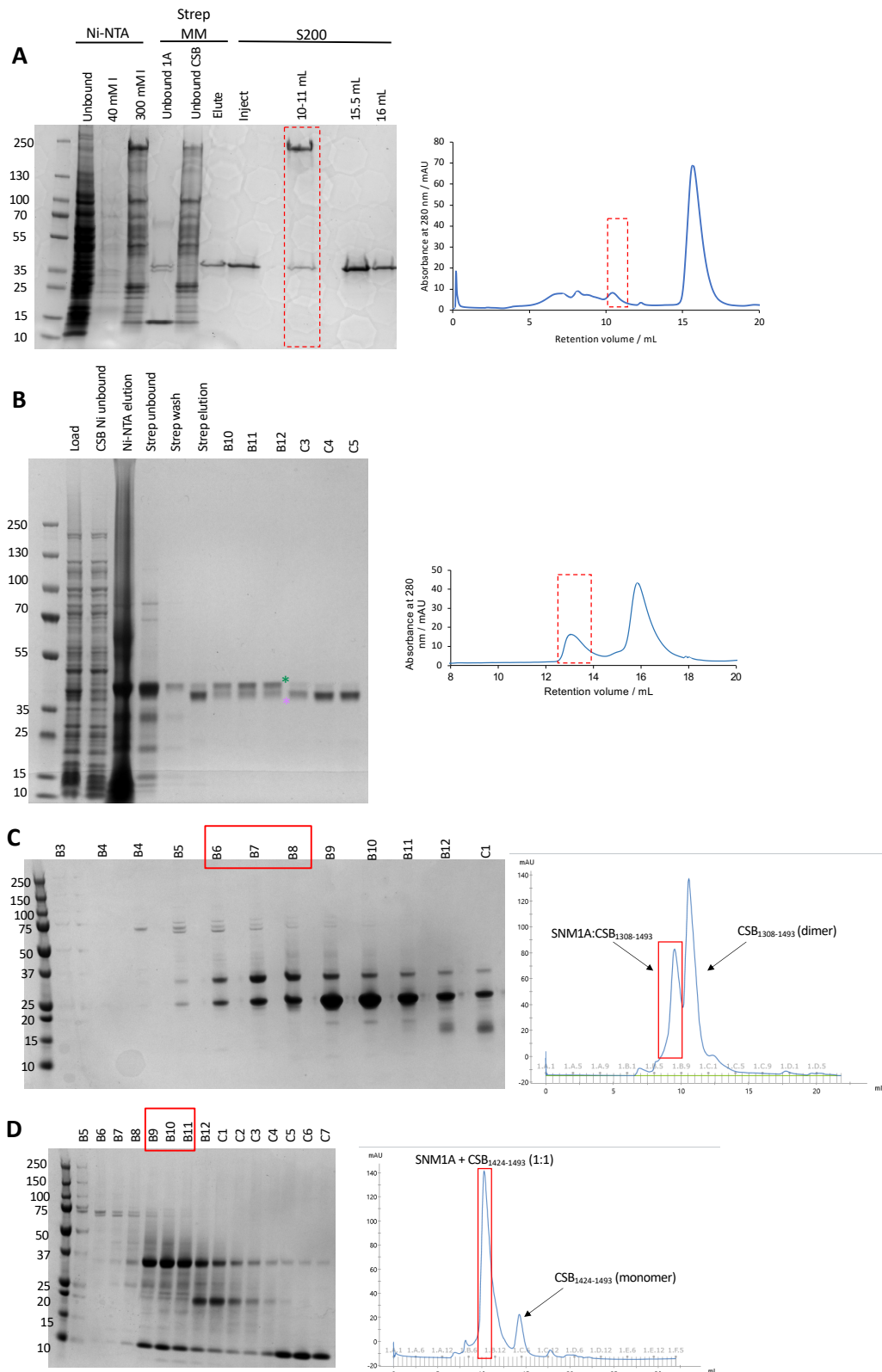


Figure S19: Co-purifications of CSB and SNM1A. Methods for purifying each complex are detailed in the Methods section. Figures shown from size-exclusion chromatography traces of eluted protein complexes on Superdex S75 GL 10/300 or Superdex S200 GL 10/300.

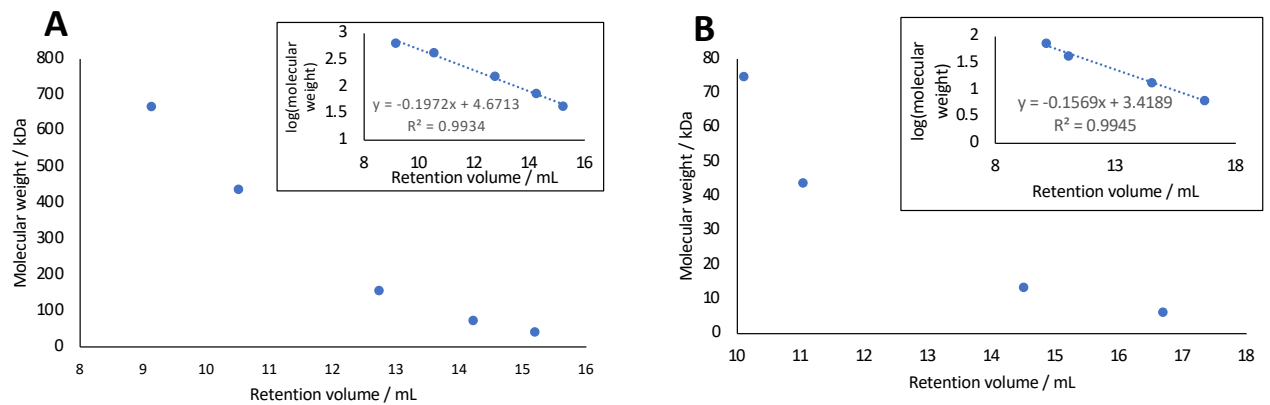


Figure S20: Size exclusion calibration curves. **A.** Superdex S200 GL 10/300. The standards used were thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (43 kDa). **B.** Superdex S75 GL 10/300. The standards used were conalbumin (75 kDa), ovalbumin (43 kDa), ribonuclease A (13.7 kDa), aprotinin (6.5 kDa).

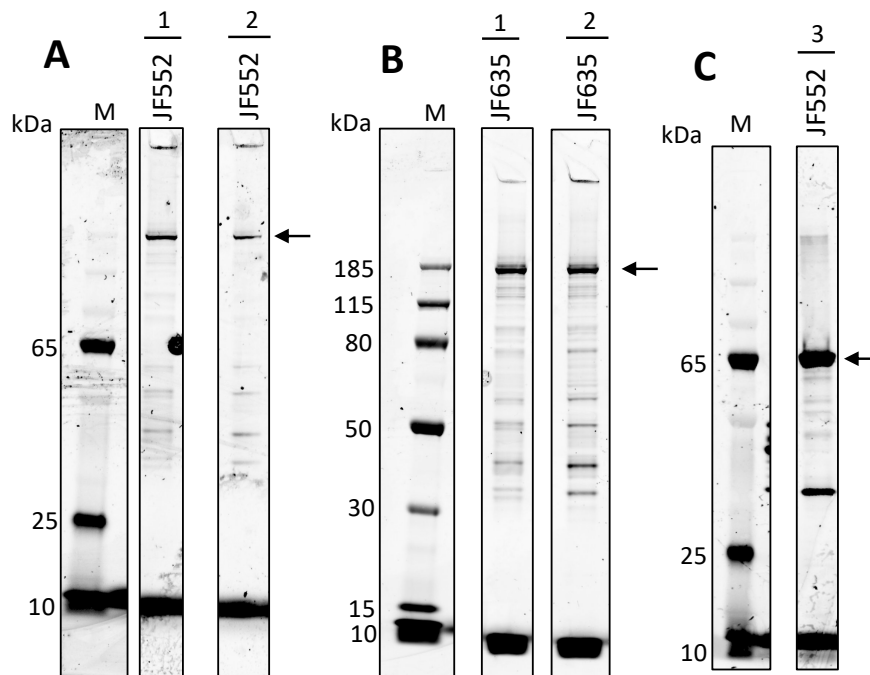


Figure S21: SDS-PAGE gels of nuclear extracts used for single-molecule studies. **A.** Halotag-CSB constructs labelled with JF552, imaged using a Typhoon imager at 532 nm. **B.** Halotag-CSB constructs labelled with JF635, imaged using a Typhoon imager at 647 nm. **C.** Partially purified Halo-SNM1A (a.a. 698-1040) labelled with JF552, imaged using a Typhoon imager at 532 nm.

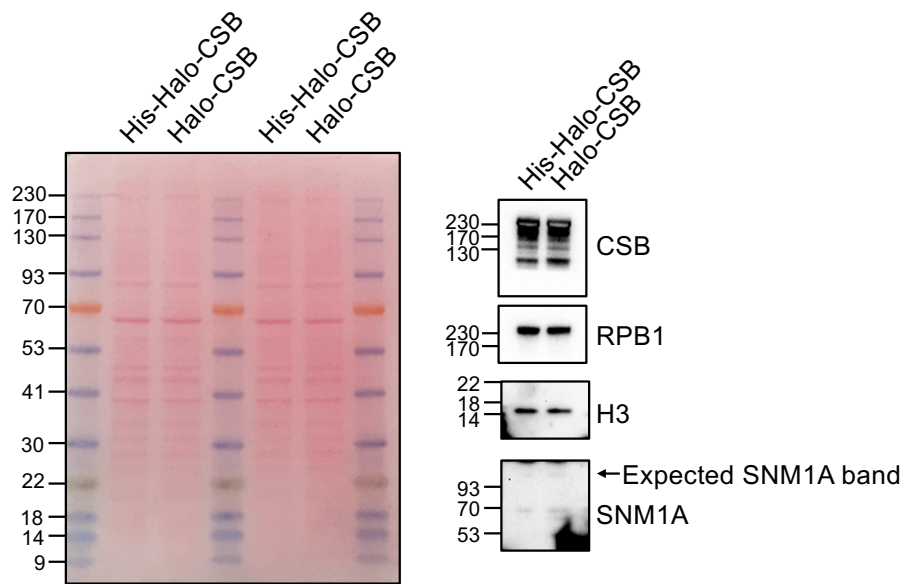


Figure S22: Western blot analysis of CSB-transfected nuclear extracts. Ponceau staining of membrane with western blotting for CSB, the major RNAPII subunit, RPB1, H3 and SNM1A.

Table S5: Parameters determined for protein:DNA binding events on lambda DNA in the LUMICKS' C-Trap.

Dataset	Final protein concentration (nM)	Recruitment time of SNM1A to bound CSB (s)	Apparent k_{on} (s^{-1})	Corrected k_{on} ($M^{-1} s^{-1}$)	Lifetime(s) and percentages (s)	Binding lifetime t_{avg} (weighted average, s)	Weighted k_{off} (s^{-1})	K_D (nM)	Oligomerisation state
SNM1A	2.23 ± 0.426	-	0.0403 ± 0.0405	$1.81 \pm 1.81 \times 10^7$	8.64 ± 0.548 (23 $\pm$ 0.6%) 0.706 ± 0.0107 (77 $\pm$ 0.6%)	2.53	0.395	21.8 ± 10.9	Monomer
CSB	0.547 ± 0.0351	-	0.0156 ± 0.00882	$2.86 \pm 1.05 \times 10^7$	73.2 ± 24.1 (34 $\pm$ 5%) 7.12 ± 0.926 (66 $\pm$ 5%)	29.6	0.0338	1.18 ± 0.667	CSB 73:27 monomer:dimer
Co-localised SNM1A + CSB	SNM1A: 2.23 ± 0.426 CSB: 2.38 ± 0.268	71.9 ± 5.52 (79 $\pm$ 3%) 3.56 ± 0.95 (21 $\pm$ 3%) Average: 57.5 ± 1.42 s	0.017 ± 0.000420	$7.80 \pm 0.72 \times 10^6$	16.2 ± 2.52 (42 $\pm$ 2%) 0.825 ± 0.0955 (58 $\pm$ 2%)	7.28	0.137	17.5 ± 1.62	CSB 27:73 monomer:dimer

Table S6: Parameters determined for protein:DNA binding events on the ICL lesion in the LUMICKS' C-Trap.

Dataset	Final protein concentration (nM)	Recruitment time of SNM1A to bound CSB (s)	Apparent k_{on} (s^{-1})	Corrected k_{on} ($M^{-1} s^{-1}$)	Lifetime(s) and percentages (s)	Binding lifetime t_{avg} (weighted average, s)	Weighted k_{off} (s^{-1})	K_D (nM)	Oligomerisation state
SNM1A	2.99 ± 0.626	-	$9.36 \pm 7.57 \times 10^{-3}$	$3.13 \pm 3.19 \times 10^6$	0.901 ± 3.19	0.901	1.11	355 ± 361	Monomer
Co-localised SNM1A + CSB	SNM1A: 13.2 ± 1.13 CSB: 2.95 ± 0.463	40.6 ± 7.3 s	0.0246 ± 0.00372	$1.87 \pm 0.16 \times 10^6$	31.1 ± 17.9 (33 $\pm$ 12%) 2.26 ± 0.783 (67 $\pm$ 12%)	11.8	0.0847	45.3 ± 3.88	CSB 8:92 monomer:dimer

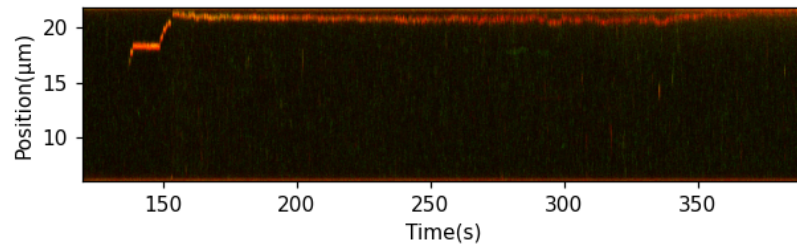


Figure S23: Kymograph showing 2 different colours of CSB protein diffusing together on lambda DNA.

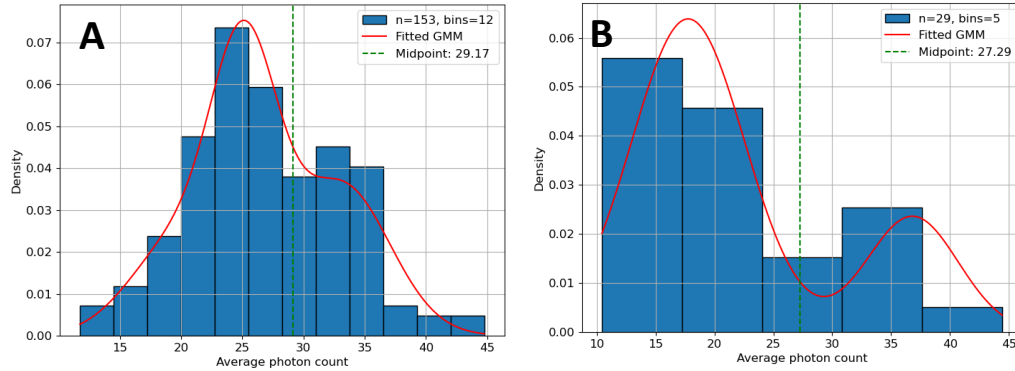


Figure S24: Histograms of average CSB photon counts along a kymotrack. To determine whether a CSB:DNA binding event was with a monomeric or dimeric CSB, the average photon count across an entire kymotrack was determined by taking the average of a line of best fit across the plotted photon count intensities for the track. The values for each CSB binding event were combined and shown as histograms. **A.** CSB average photon counts on lambda DNA. Events with an average photon count >29.17 were denoted as dimers, and <29.17 , monomeric. **B.** CSB average photon counts for binding events localised at the ICL. Events with an average photon count >27.29 were denoted as dimers, and <27.29 , monomeric.

Table S7: Statistical analysis (AICc test) to determine whether binding event lifetimes (CTRD plots) exhibited one- or two-phase decay. Highlighted in grey are the most negative values i.e. the most appropriate fit for the data.

DNA substrate	Protein	AICc for one phase decay	AICc for two phase decay
Lambda	SNM1A	-5180	-6973
Lambda	CSB	-399.9	-474.9
Lambda	SNM1A + CSB	-550.8	-767
ICL	SNM1A	-165.6	-164.3
ICL	SNM1A + CSB	-122.9	-137.8

Table S8: Statistical analysis (AICc test) of SNM1A recruitment times to DNA-bound CSB. Highlighted in grey are the most negative values i.e. the most appropriate fit for the data.

DNA substrate	Protein	AICc for one phase decay	AICc for two phase decay
Lambda	SNM1A + CSB	-612.4	-699.8
ICL	SNM1A + CSB	-107.2	-100.4