Molecular insights into the stimulation of SNM1A nuclease activity by CSB during interstrand crosslink processing

Rebecca Roddan[1], William J. Nathan[2], Joseph A. Newman[3], Afaf El-Sagheer[4,5], David M. Wilson III[6], Tom Brown[4], Christopher J. Schofield[4], and Peter J. McHugh[1].

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

[2]Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA.

[3]Centre for Medicines Discovery, University of Oxford, NDMRB, OX3 7DQ, U.K.

[4]Chemistry Research Laboratory, 12 Mansfield Road, Department of Chemistry and the Ineos Oxford Institute for Antimicrobial Research, University of Oxford, Oxford OX1 3TA, U.K.

[5]School of Chemistry, University of Southampton, SO17 1BJ, U.K.

[6]Biomedical Research Institute, Hasselt University, 3590 Diepenbeek, Belgium.

*Corresponding author: Peter J. McHugh

Abstract

The SNM1A exonuclease plays a key role in repair of interstrand crosslinks (ICLs) which represent a particularly toxic class of DNA damage. Previous work suggests that the SWI/SNF family ATP-dependent, chromatin remodeler, Cockayne Syndrome B protein (CSB) interacts with SNM1A, during transcription-coupled DNA interstrand crosslink repair (TC-ICL repair). Here, we validate this interaction using purified proteins and demonstrate that the ubiquitin-binding and winged-helix domains of CSB are required for interaction with the catalytic domain of SNM1A. The winged helix domain is essential for binding, although high-affinity SNM1A binding requires the entire CSB C-terminal region (residues 1187-1493), where two copies of the C-terminal domain of CSB are necessary for a stable interaction with SNM1A. CSB stimulates SNM1A nuclease activity on varied model DNA repair intermediate substrates. Importantly, CSB was observed to stimulate digestion through ICLs *in vitro*, implying a key role of the interaction in 'unhooking' during TC-ICL repair. AlphaFold3 models of CSB constructs complexed with the SNM1A catalytic domain enabled mapping of the molecular contacts required for the CSB-SNM1A interaction. This identified specific protein-protein interactions necessary for CSB's stimulation of SNM1A's activity that we confirmed experimentally. Additionally, our studies reveal the C-terminal region of CSB as a novel DNA binding region that also is involved in stimulation of SNM1A-mediated ICL repair. Moreover, targeting protein-protein interactions that are vital for specific nuclease activities, such as CSB's stimulation of SNM1A's nuclease activity, may be a productive alternative therapeutic strategy to nuclease active site inhibition.

Introduction

Cellular DNA is under constant threat from genotoxic agents, which modify DNA and interfere with fundamental cellular processes, in particular replication and transcription. Chemicals that covalently join the two strands of the DNA duplex produce interstrand crosslinks (ICL), lesions that prevent strand separation and thus are particularly problematic (Hashimoto *et al*, 2016). ICL induction *via* exogenous agents is a widely used cytotoxic chemotherapeutic strategy that targets mainly rapidly dividing cells. ICLs can also be formed by endogenous aldehydes produced during normal cellular metabolism or via abasic sites created spontaneously or during DNA repair (Clauson *et al*, 2013; Housh *et al*, 2021). To remove ICL damage, efficient DNA repair pathways have evolved, employing different factors, dependent on the genomic context and cell cycle phase e.g., global-genome repair, transcription-coupled (TC) repair (both of which are theoretically operational throughout the cell cycle), and replication-coupled repair (acting in S-phase) (Hashimoto *et al*, 2016).

To date, most studies on mammalian ICL repair mechanisms have focused on replication-coupled ICL repair by the Fanconi Anaemia (FA) pathway, in which the key ICL 'unhooking' incision step is performed by the heterodimeric, structure-specific endonuclease, XPF-ERCC1 (McHugh, 2020). After ICL 'unhooking', XPF-ERCC1 is proposed to collaborate with the 5′-to-3′ exonuclease SNM1A, a metallo-β-lactamase (MBL) fold nuclease, to further process the ICL, though alternative nucleases including FAN1 have been proposed to be involved (Smogorzewska *et al*, 2010). SNM1A-deficient cells exhibit ICL sensitivity that is epistatic with XPF-ERCC1 (Wang *et al*, 2011) and *in vitro* an incision by XPF-ERCC1 results in a nick to the DNA backbone, forming a 5′ phosphate group that presents a loading site for SNM1A exonuclease activity. SNM1A can then processively hydrolyse the phosphodiester backbone of the DNA, and uniquely, digest past the ICL in the 5′-3′ direction (Wang *et al*, 2011). Downstream repair is performed by translesion synthesis and homologous recombination.

How ICL repair occurs outside of S-phase is not well understood, although because most cells in the body are non-dividing (Iyama & Wilson III, 2013), understanding this is crucial to fully grasp how ICL-inducing agents impact mostly non-dividing tissues (e.g. stem-like cancer cells), for example, during chemotherapy (Basu *et al*, 2022). Beyond its role in replication-coupled ICL repair, there is emerging evidence for a role of SNM1A in ICL repair outside S-phase involving factors from the nucleotide excision repair (NER) pathway, through which helix-distorting (mainly bulky) DNA base lesions are removed. Most notably, SNM1A has been shown to directly interact

with the TC-DNA repair factor, Cockayne Syndrome Protein B (CSB) (Iyama *et al*, 2015). CSB (ERCC6) is a large (168 kDa), SWI/SNF family ATPase (Fan & Lake, 2013; Tiwari *et al*, 2021) linked to a rare segmental progeria, termed Cockayne Syndrome (CS) which features neurodevelopmental defects/neurodegeneration, growth failure, photosensitivity, and a premature ageing phenotype (Vessoni *et al*, 2020). CSB has a fundamental role in recognising DNA damage-stalled RNAPII during transcription, with CS phenotypes thought to be caused by defects in DNA repair and transcription recovery (Nieto Moreno *et al*, 2023). The CSB-RNAPII recognition event initiates recruitment of downstream DNA repair factors that act to remove and replace the damaged DNA. The role of CSB in TC repair has been extensively studied in the context of UV irradiation-induced damage (via NER), although there is emerging evidence for CSB guiding resolution of other types of DNA damage including DNA-protein crosslinks, oxidative DNA damage and ICLs (Oka *et al*, 2024; Carnie *et al*, 2024; Sluis *et al*, 2024; Aamann *et al*, 2014; Chakraborty *et al*, 2021; Iyama *et al*, 2015; Enoiu *et al*, 2012).

The physical interaction between CSB and SNM1A was initially discovered in a yeast-two hybrid screen. Purified recombinant SNM1A and CSB were also shown to interact *in vitro*, with CSB enhancing the exonuclease activity of SNM1A on simple single-stranded (ss) and double-stranded (ds) DNA oligonucleotides (Iyama *et al*, 2015). Additionally, using confocal microscopy, fluorescently tagged CSB and SNM1A co-localise at sites of ICLs in a transcription-dependent manner (Iyama *et al*, 2015). CSB deficient cells exhibit sensitivity to ICL-inducing agents, particularly in G_0/G_1 where replication-coupled pathways are not available (Iyama *et al*, 2015). Comet assay analysis revealed that CSB-deficiency resulted in reduced ICL 'unhooking' in non-cycling, differentiated neural cells, where there was an association with persistence of γ-H2AX DNA damage foci. Further yeast two-hybrid studies suggested that the C-terminal domain of CSB (amino acids 1187-1493) interacts with the catalytic domain of SNM1A (amino acids 698-1040) in a manner similar to full-length CSB, thus narrowing down the interacting regions. Overall, this prior work proposed that CSB co-ordinates ICL resolution through SNM1A in a transcription-dependent manner (Iyama *et al*, 2015).

Our work addresses the molecular basis of CSB's stimulation of SNM1A's nuclease activity. We describe the structural and molecular-level requirements for CSB to interact with SNM1A, defining the nature of the interaction between these two proteins and relevant DNA substrates. Importantly, we show biochemically that CSB dramatically stimulates SNM1A's nuclease activity

during biochemically reconstituted ICL unhooking reactions, providing new molecular insights into the possible TC-dependent resolution of such DNA damage forms.

Methods

Plasmids for recombinant protein expression

Full-length CSB (*ERCC6*) was purchased as a synthetic gene from Genscript in the pFastBac1 vector, codon optimized for *Spodoptera frugiperda* (*Sf*9) expression with N-terminal His_{10} and ZBasic tags with a TEV cleavage site and a C-terminal Avi tag. For preparation of biotinylated protein, the full-length CSB gene (including tags) was subcloned from the pFastBac1 vector by InFusion cloning into a pFastBac Dual vector, with CSB expressed under the polyhedrin promoter and BirA (with C-terminal mCherry tag) co-expressed under the p10 promoter. Expression of N-Avi-SNM1A$_{698-1040}$ was performed using the same pFastBac Dual vector (co-expressing BirA-mCherry), with the SNM1A sequence codon optimized for baculovirus expression.

The pNIC28-Bsa4 vector (with N-terminal His_6 followed by TEV cleavage site) was used for expression of all C-terminal and N-terminal CSB domains, with the CSB sequence codon optimized for expression in *E. coli*. For expression of the ATPase domain of CSB (residues 402-1040), codon optimization for baculovirus expression and ligation independent cloning was used to transfer to the pFB-LIC-Bse expression vector, giving a fusion protein with N-terminal His_6 followed by a TEV cleavage site. Alanine mutations (SNM1A-E864A and CSB-T1447A) were generated from the wild-type plasmids, using InFusion cloning and verified by Sanger sequencing.

Expression, and purification of recombinant proteins

For bacmid preparation, full-length and ATPase domain (a.a. 402-1040) CSB plasmids were transformed into DH10Bac *E. coli* cells (Life Technologies). Bacmid DNA was isolated and used to transfect *Sf*9 insect cells (Gibco™, cultured in Sf-900 II SFM). Viral amplifications were performed following the manufacturer's instructions. For expression, *Sf*9 cells were prepared at 2-3 x 10^6 cells/mL, 7 mL/L culture of P2 virus was added and incubated at 27 °C for 72 h at 180 rpm. For expression of the biotinylated full-length CSB (in pFastBac Dual vector co-expressing BirA-mCherry), 50 μM D-biotin was added (1 in 1000 of 50 mM stock prepared in 0.1 M NaOH). The cells were harvested by centrifugation (1,500 x *g*, 10 min, 4 °C) and stored at -80 °C until further processing. For the CSB ATPase domain, purification was performed as per the C- and N-terminal fragments (expressed in *E. coli* and described later). For full-length CSB, lysis buffer (80 mL per L of initial culture; 50 mM HEPES pH 8, 1 M NaCl, 20 mM imidazole, 1 mM TCEP,

10% glycerol + 1 Roche cOmplete EDTA-free protease inhibitor tablet) was used to resuspend the pellet on ice. Cells were lysed by sonication (10 A, 6 x 30 s ON, 60 s OFF) then centrifuged (74,000 x *g*, 60 min, 4 °C). The supernatant was filtered (0.45 μm), then loaded onto Ni-NTA agarose beads (1 mL beads per L of expression culture pre-equilibrated with lysis buffer). The beads were washed with lysis buffer (20 CV), lysis buffer + 750 mM NaCl + 40 mM imidazole (20 CV) and the protein eluted in 5 CV of lysis buffer + 500 mM NaCl + 300 mM imidazole with 5% *v/v* glycerol (vs. 10% *v/v* glycerol previously). Protein-containing fractions (the 300 mM imidazole elution) were pooled, desalted into gel filtration buffer (50 mM HEPES pH 8, 300 mM NaCl, 1 mM TCEP) then loaded onto a Superdex S200 GL 10/300 column (pre-equilibrated with gel filtration buffer). Protein-containing fractions were pooled (protein elutes 9.5-13.5 mL) and diluted 2-fold with a salt-free buffer (50 mM Tris pH 8, 1 mM TCEP). The protein was loaded onto a 1 mL HiTrap SP FF column; an elution gradient was performed from 150 mM NaCl to 1.5 M NaCl over 8 column volumes at 1 mL/min. Protein was pooled, concentrated, aliquoted and stored at -80 °C.

N-Avi-SNM1A$_{698-1040}$ (WT and E864A) bacmid DNA was prepared as per CSB expressions and was produced in *Sf9* insect cells prepared at 2-3 x 10^6 cells/mL then 7 mL/L of culture of P2 virus and 500 μL/L of culture of D-biotin (50 mM in 0.1 M NaOH) were added. Cells were incubated at 27 °C for 72 h at 180 rpm. The cells were harvested by centrifugation (1,500 x *g*, 10 min, 4 °C), washed with 1 x PBS (4 °C), then stored at -80 °C until further processing. The pellet was resuspended in lysis buffer (80 mL per L of initial culture: 20 mM HEPES pH 8, 500 mM NaCl, 5% v/v glycerol, 0.5 mM TCEP + 1 Roche cOmplete EDTA-free protease inhibitor tablet) on ice. Cells were lysed by sonication (10 A, 6 x 30 s ON, 60 s OFF) then centrifuged (74,000 x *g*, 60 min, 4 °C). The supernatant was incubated with Streptavidin mutein matrix beads (1 mL, Roche) for 1 h, 4 °C, with gentle rolling. The unbound protein was removed by centrifugation (3,000 x g, 5 min, 4 °C) and the beads transferred to a gravity column. The beads were washed with lysis buffer (50 mL) and then the protein eluted in 5 mL lysis buffer + 5 mM D-biotin. The elution was concentrated and injected onto a Superdex S75 GL 10/300 column (in 20 mM HEPES pH 8, 300 mM NaCl, 0.5 mM TCEP). Protein containing fractions were pooled, concentrated, and stored at -80 °C. Expression and purification of N-His-SNM1A$_{698-1040}$ has already been reported (Allerston *et al*, 2015).

Plasmids for CSB$_{1187-1493}$, CSB$_{1308-1493}$, CSB$_{1424-1493}$, CSB$_{1187-1401}$, and CSB$_{1187-1308}$ were expressed in BL21(DE3) pLysS cells. A single colony from a fresh transformant was used to inoculate LB media (+ 50 μg/mL kanamycin sulfate final). This culture was incubated overnight at 37 °C, 220

rpm then added at 2% *v/v* final to 1 L of LB media (+ 50 μg/mL kanamycin sulfate final). The culture was incubated at 37 °C, 220 rpm until OD_{600} = 10-20 then cooled to 18 °C and induced by addition IPTG (to 0.5 mM final). The culture was incubated at 18 °C for 16-18 h. Cells were harvested by centrifugation (5,000 x *g*, 15 min, 4 °C), then stored at –20 °C until purification. For protein purification (for all C-terminal, N-terminal and ATPase domain fragments), cells were resuspended in lysis buffer (40 mL per L of expression culture; 50 mM HEPES pH 8, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP + 1 cOmplete EDTA-free protease inhibitor tablet) and sonicated (16 A, 3 x 1 min ON, 1 min OFF) on ice then centrifuged (50,000 x *g*, 4 °C, 30 min). The supernatant was incubated with 2 mL Ni-NTA agarose beads for 30 min, 4 °C, with gentle rolling. The unbound protein was removed by centrifugation (3,000 x *g*, 5 min, 4 °C) and the beads transferred to a gravity column. The beads were washed with lysis buffer (30 mL), lysis buffer + 40 mM imidazole (30 mL) then the protein eluted with lysis buffer + 300 mM imidazole (5 mL). The protein was desalted into gel filtration buffer (50 mM HEPES pH 8, 300 mM NaCl, 0.5 mM TCEP) using Zeba Spin 7K MWCO desalting columns and injected onto a Superdex S75 GL 10/300 column (pre-equilibrated with gel filtration buffer). Protein-containing fractions were analysed by SDS-PAGE and those containing the desired protein were pooled, concentrated, flash-frozen and stored at -80 °C.

For preparation of the $CSB_{1187-1493}$ or full-length CSB and N-Avi-Biotin-$SNM1A_{698-1040}$ complexes, the two proteins were individually expressed and incubated with affinity chromatography columns as previously described (Ni-NTA agarose and streptavidin mutein matrix, respectively with all buffers supplemented with 1 mM Mg^{2+}). The SNM1A-bound streptavidin mutein matrix beads (1 mL per 500 mL expression culture) were washed with SNM1A lysis buffer. CSB bound to Ni-NTA was washed, eluted, and desalted (as previously described for the protein alone) and incubated with SNM1A bound to streptavidin mutein matrix beads for 30 min, 4 °C. Unbound protein was removed from the beads by centrifugation (3,000 x *g*, 4 °C, 5 min) and the beads were washed with gel filtration buffer (20 CV, 20 mM HEPES pH 8, 500 mM NaCl, 0.5 mM TCEP, 1 mM Mg^{2+}) in a gravity column. The protein was eluted with 5 CV gel filtration buffer + 5 mM D-biotin. The protein was concentrated and injected onto a Superdex S200 GL 10/300 column. Protein-containing fractions were analysed by SDS-PAGE and those containing the desired protein were pooled, concentrated, flash-frozen and stored at -80 °C.

Oligonucleotide preparations

High-performance liquid chromatography (HPLC)-purified oligonucleotides used for biochemical assays were purchased from Eurofins Genomics. Stocks were prepared at 100 μM in nuclease-free water and stored at -20 °C. Oligo sequences and annealing partners to give different secondary structures are described in Supplementary Table 1 and Fig. S1.

Preparation of '0-nucleotide' ICL substrate

ICL-containing DNA substrate synthesis was performed as reported (Abdullah *et al*, 2017). Briefly, both strands of the respective DNA substrates were synthesised independently using an Applied Biosystems 394 DNA/RNA Synthesizer. Solid-phase synthesis was performed using a standard 1 μmol phosphoramidite cycle of acid-catalysed detritylation, coupling, capping, and iodine oxidation. The coupling time for normal nucleoside phosphoramidite monomers (A, T, G, C) was 1 minute. For the eventual azide-containing nucleotide, 5'-(4,4'-dimethoxytrityl)-N6-benzoyl-N8-[6-(trifluoroacetylamino)-hex-1-yl]-8-amino-2'-deoxyadenosine,3'-[(2-cyanoethyl)- (N,N-diisopropyl)]-phosphoramidite ('amino-modified C6 dA,' Glen Research) was incorporated (coupling time: 14 minutes). For the alkyne-containing nucleotide, 5'-dimethoxytrityl-5-(octa-1,7-diynyl)-2'-deoxyuridine,3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite ('C8-Alkyne-dT-CE Phosphoramidite', Glen Research) was incorporated (coupling time: 14 minutes). Following synthesis completion, oligonucleotides were cleaved from the solid support and deprotected by exposure to concentrated aqueous ammonia solution (1 h, rt), then heated within a sealed tube (5 h, 55°C). Oligonucleotides were purified using HPLC with triethyl ammonium bicarbonate buffer (TEAB), as reported (Kočalka *et al*, 2008).

To convert the amino-modified C6 dA amino group into an azide, the oligonucleotide (200 nmol) was prepared with 6-azidohexanoic acid NHS ester (1 mg, 3.9 mmol) in 1:1 dimethyl sulfoxide: 0.5 M $Na_2CO_3/NaHCO_3$, pH 8.75 (160 μL) for 4 h at rt. The azide-labelled oligonucleotide was then purified by reversed-phase HPLC, as previously reported (Kočalka *et al*, 2008). Under an inert atmosphere, the alkyne and azide strands were prepared in 0.2 M NaCl solution (110 μL, 50 μM) and annealed by heating to 95 °C then allowing to cool to room temperature (2-3 h). Under an inert atmosphere, $CuSO_4.5H_2O$ (0.03 mg, 0.12 μmol from an 8 mM stock in H_2O), Tris(3-hydroxypropyltriazolylmethyl)amine ligand (0.5 mg, 1.15 μmol) and sodium ascorbate (0.2 mg, 1.0 μmol from a 500 mM stock in H_2O) were prepared in water (15 μL). This was added to the annealed DNA and incubated for 2 h at rt. The completed click reaction was then desalted using a NAP-25 DNA purification column (Cytiva) and lyophilised.

Desalted click reactions were purified by denaturing PAGE using an 8% 19:1 acrylamide:bis-acrylamide and 7 M urea gel. Successfully crosslinked DNA substrate migrated more slowly compared to non-crosslinked strands (Fig. S2). Crosslinked substrates were cut out from the gel and recovered by the crush and soak method. The gel pieces were crushed and soaked in 30 mL of water for 18 hours at 37 °C while shaking. Gel particles were filtered out and the resulting solution was desalted using a NAP-25 column, lyophilised and resuspended in water. The correct identity of the final crosslinked product was confirmed by quadrupole time-of-flight (Q-TOF) mass spectrometry (Fig. S3).

Generation of 3' radiolabelled oligonucleotides

All substrates, except the '0-nucleotide' (herein referred to as '0-nucleotide') ICL-containing substrate (Fig. S1 and Table S1), were 3′- labelled using terminal deoxynucleotide transferase (TdT) (NEB), then annealed with the complementary oligonucleotide. For labelling, oligonucleotide (10 pmol), TdT (20 U) and α-^{32}P-dATP (3.3 pmol) were prepared in 1 x TdT buffer (NEB) and incubated at 37 °C for 1 h. The reaction mixture was then passed through a P6 Micro Bio-Spin chromatography column (Bio-Rad, 1,000 x *g*, 4 min, rt), to give the labelled oligonucleotide at ~500 nM. To generate duplex-containing substrates, the annealing reactions were performed with radiolabelled substrate: unlabelled substrate(s) 1:1.5 molar ratio by heating to 100 °C for 5 min in annealing buffer (10 mM Tris pH 7.5, 100 mM NaCl, 0.1 mM EDTA) then allowing to cool to room temperature for 3-18 h. Stocks were prepared at 100 nM final based on the radiolabelled oligonucleotide. For the ss oligonucleotide, the 500 nM stock was diluted to 100 nM in annealing buffer.

For labelling of the '0-nucleotide' ICL substrate, the oligonucleotide (10 pmol) was first annealed by boiling to 100 °C for 5 min, then allowed to cool to rt in annealing buffer. The oligonucleotide (10 pmol) was incubated with Klenow (exo-) (12.5 U, NEB) and α-^{32}P-dATP (3.3 pmol) in 1 x NEB buffer 2 at 37 °C for 45 min. dGTP (1 mM) and dATP (1 mM) were added to 0.1 mM final concentrations and the mixture incubated for 30 min at 37 °C. The reaction was then passed through a P6 Micro Bio-Spin chromatography column (Bio-Rad, 1,000 x g, 4 min, rt), annealed in 1 x annealing buffer and prepared at 100 nM oligonucleotide concentration.

Nuclease assays

Reactions were prepared in 20 mM HEPES KOH pH 7.5, 50 mM KCl, 4 mM $MgCl_2$, 0.05% *v/v* Triton X-100, 0.5 mM TCEP, 5% *v/v* glycerol. Protein was added first, followed by 10 nM

oligonucleotide (all concentrations given are final) on a 10 μL scale. Reactions were incubated at 37 °C for the time indicated, then reactions quenched by addition of stop solution (10 μL; 95% formamide, 10 mM EDTA, 0.25% *v/v* xylene cyanol, 0.25% *v/v* bromophenol blue) then incubated at 100 °C for 5 min. Samples were then immediately incubated at 4 °C for 5 min, then centrifuged (1 min, 13,000 x g, 4 °C). Analysis was performed by 20% denaturing PAGE (50% *v/v* 40% 19:1 acrylamide:bis-acrylamide, 7 M urea, 1 x Tris-borate EDTA, 0.12% *w/v* ammonium persulfate, 0.12% *v/v* tetramethylethylenediamine). The gel was preheated (700 V, 1 h) then electrophoresis performed at 700 V, 1 h. Gels were fixed for 1 h in 50% *v/v* methanol, 10% *v/v* acetic acid in water, dried at 80 °C for 2 h under vacuum, then exposed to a phosphor imager screen (Kodak). Screens were scanned with a GE Typhoon FLA 9500. Quantification was performed in ImageJ, measuring the amount of intact substrate, compared to that which had been digested. Assays were quantified based upon duplicate reactions, with the end of each error bar representing each data point.

Electrophoretic mobility shift assays (EMSAs)

Reactions were prepared on ice in 20 mM HEPES KOH pH 7.5, 50 mM KCl, 0.05% *v/v* Triton X-100, 0.5 mM TCEP, 5% *v/v* glycerol, 1 nM oligonucleotide and the indicated concentration of protein on a 10 μL scale. All concentrations given are final. The metal ion or EDTA concentration used, if applicable, is given for each figure. For the binding of a single protein to DNA, protein was added last to the reaction whereas to study potential for co-operative binding of proteins to DNA, DNA was added last. Reactions were incubated at 4 °C for 10 min then immediately loaded onto a 0.5% agarose gel (in 0.5x TBE, pre-cooled to 4 °C). The gel was run at 100 V for 90 min at 4 °C then fixed (30 min - 1 h in in 50% methanol, 10% acetic acid) and dried under vacuum (50 °C, 1 h) on three pieces of Whatman filter paper. The gel was then exposed to a phosphor imager screen (Kodak) and imaged using a GE Typhoon FLA 9500.

Pull-down assays

N-Biotin-Avi-SNM1A (200 nM, 250 μL) was prepared in binding buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 50 mM KCl, 0.5 mM TCEP, 0.05% *v/v* Triton X-100, 5 mM $MgCl_2$, 5% *v/v* glycerol, 1 mM phenylmethylsulfonyl chloride, 0.05% *v/v* NP-40) and immobilised on Pierce™ Streptavidin magnetic beads (10 μL, ThermoFisher Scientific) by incubation at 4 °C for 1 h. The beads were then washed with 3 x 1 mL binding buffer with centrifugation at 1,000 x *g*, 4 °C, 2 min between each wash. Potential SNM1A binding partners were prepared at 400 nM, in 250 μL of binding buffer and incubated with the SNM1A-beads at 4 °C for 1 h. The unbound protein was isolated. 10 μL was sometimes retained for SDS-PAGE. The beads were washed 5 x 1 mL, binding buffer

+ 300 mM NaCl, then 5 x 1 mL binding buffer + 500 mM NaCl, then 1 x 200 mL binding buffer + 500 mM NaCl with 10 μL of the final wash retained for SDS-PAGE. Protein was eluted with 100 μL binding buffer + 100 mL 2 x SDS loading buffer with boiling for 5 min. Samples were analysed by SDS-PAGE using 4-12% Bis-Tris gels run in 1 X MOPS buffer, with Thermo Scientific™ PageRuler™ Plus Prestained Protein Ladder as the molecular weight standards, given in kDa.

Surface plasmon resonance (SPR) spectroscopy

Protein-protein interactions were measured using a Biacore S200 machine using a Series S SA sensor surface. SNM1A was immobilised via a biotinylated N-terminal Avi-tag at an immobilisation level of ~4000 RU. Binding analysis using CSB analytes was performed in a buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP. Single cycle analysis was performed using either 5 or 6 two-fold dilutions starting from 500 nM or 1 mM. CSB protein was injected for 90s with the final dissociation phase of at least 900 s. Full-length CSB was immobilised *via* a biotinylated C-terminal Avi tag in 20 mM HEPES pH 7.5, 500 mM NaCl, 0.5 mM TCEP, 0.01% TWEEN-100. SNM1A was injected for 90 s with the final dissociation phase of at least 900 s.

For multi cycle analysis SNM1A was immobilised *via* a biotinylated N-terminal Avi-tag at an immobilization level of approximately 4000 RU. Two-fold serial dilutions starting from 2 mM were injected for 300 seconds with a dissociation phase of 1200 seconds. Regeneration was performed between each injection using the mobile phase buffer containing 1 M NaCl.

Data were fitted using Biacore S200 evaluation software. Single cycle curves were fitted using a 1:1 Langmuir binding model with accommodation for drift as a single parameter when necessary (based upon an estimation of the curve before the first injection). A table of fitting parameters is shown in Supplementary Tables 4 and 5.

Circular Dichroism (CD) spectroscopy

Samples were prepared at 0.1 mg/mL (determined by Nanodrop spectrophotometer) in 10 mM sodium phosphate pH 7, 20 mM NaCl in a 0.1 cm quartz cuvette. Measurements were performed using a JASCO-815 spectropolarimeter with a Peltier temperature control unit. The CD signal (in mdeg) was measured at wavelength 195-250 nm at 20 ^{o}C.

AlphaFold3

AlphaFold3 used the opensource online server with standard presets (Abramson *et al*, 2024). USCF ChimeraX1.15 was used to visualize the results and generate figures (Meng *et al*, 2023). PICKLUSTER (via USCF Chimera X1.5 plug-in) was used to analyse AlphaFold3 model interfaces (Genz *et al*, 2023).

Results

Evidence of a 2:1 CSB:SNM1A complex stoichiometry

Previous yeast two-hybrid technology work found that CSB and SNM1A interact *via* their respective C-terminal regions (Iyama *et al*, 2015). We sought to (i) replicate this finding biochemically, (ii) to validate that the two proteins could directly interact in the absence of other repair factors, and (iii), if relevant, to investigate the stoichiometry of CSB and SNM1A, both alone and in complex to gain insight into CSB oligomerization in DNA repair. Previous reports suggest that CSB is a dimer in solution based on size-exclusion chromatography and immunoprecipitation experiments with over-expressed individual CSB domains (Christiansen *et al*, 2005). However, structural investigations of CSB-containing, TC-NER initiation complexes indicate the presence of only one copy. It is unknown whether this is due to a lack of resolution (due to high flexibility of its numerous unstructured domains) or is functionally relevant (Kokic *et al*, 2021, 2024; Mevissen *et al*, 2024; Heuvel *et al*, 2024).

To address the items above, recombinant full-length CSB and C-terminal SNM1A (amino acids (a.a.) 698-1040) were purified from insect cell expression systems (*Sf*9), with C-terminal CSB (a.a. 1187-1493) expressed in *E. coli* (see Methods for construct and preparation details). Purification of full-length CSB alone produced a primarily dimeric protein (~400 kDa by gel filtration, see Fig. S4 and S5), although some monomeric form was observed, consistent with previous reports (Christiansen *et al*, 2005). The catalytic, C-terminal domain of SNM1A (a.a. 698-1040), was used rather than full-length protein due to stability issues. This form of SNM1A has been previously shown to have the same exonuclease activity as the full-length version and provides the minimal interaction motif with CSB (Iyama *et al*, 2015).

Initially, we sought to validate the previously reported interaction between the C-terminal domains of CSB and SNM1A, by performing a pull-down assay with immobilised, N-terminally-biotinylated $SNM1A_{698\text{-}1040}$. This experiment confirmed that the two proteins do indeed interact *in vitro* (Fig. 1A). Co-purification of the C-terminal constructs of CSB and SNM1A resulted in a size-exclusion

chromatography peak at 12.9 mL (Fig. 1B), which, based on calibration (Fig. S12), corresponds to ~120 kDa. SNM1A is 41 kDa and CSB is 37 kDa (see Fig. S6), indicating a trimeric complex, although with unknown stoichiometry.

Co-purification of full-length CSB (360 kDa) and the catalytic domain of SNM1A (41 kDa) resulted in a complex of ~400 kDa (Fig. 1D). The presence of both proteins was confirmed by SDS-PAGE (Fig. 1E) and gel-based MS-MS (data not shown). Although the stoichiometry of SNM1A within the complex is unclear, work here (see Fig. S4 and S5) and prior investigations (Baddock *et al*, 2020) show that the catalytic MBL-β-CASP domain of SNM1A alone exists in a monomeric form in solution, suggesting that one molecule of SNM1A likely resides in the complex with two copies of CSB. To experimentally validate the previously reported CSB stoichiometries (Christiansen *et al*, 2005; Beerens *et al*, 2005), truncated C-terminal CSB domains were purified separately or in complex with SNM1A (Fig. S4). Based upon size-exclusion chromatography and associated calibration curves, $CSB_{1308-1493}$ alone is present as a dimer, whereas, as predicted, the WHD of CSB (a.a. 1424-1493) exists as a monomer. These findings indicate that amino acids within 1308-1424 are involved in CSB self-dimerization, although this region was predicted to be unfolded in the Alphafold3 (AF3) model (see below). When combined with a CSB fragment, a single copy of SNM1A was found to co-purify with one copy of the $CSB_{1424-1493}$ WHD (see Fig. S4). The stoichiometry of the complex between $CSB_{1308-1493}$ and SNM1A was, unfortunately, unclear due to resolution issues.

At present, there is limited understanding of the structural composition and oligomerization state of the C-terminal region of CSB, although α-helices are predicted in amino acid stretches, 1247-1269, 1279-1305, 1386-1399 based upon FoldIndex (Prilusky *et al*, 2005), and recent cryo-EM TC-NER initiation complexes containing CSB (Mevissen *et al*, 2024; Heuvel *et al*, 2024). To understand the CSB-SNM1A interaction further, AF3 was used to generate a model of the complex of the catalytic domain of SNM1A (a.a. 698-1040) and two copies of full-length CSB (Fig. 1F). While statistical analysis of the entire model gave only an ipTM of 0.32 and pTM of 0.33, good confidence for the model (pIDDT > 70%) was observed in most folded regions. Large, unfolded regions of CSB could only be poorly predicted i.e. a.a. 1-76, 296-412 and 1025-1250. The predicted CSB dimerisation interface within the trimer complex encompasses regions within the ATPase, C-terminal, and N-terminal domains (Table S3) and symmetry was noted in the CSB dimer. Interestingly, CSB dimer contacts were observed between the N-terminal α-helix (a.a. 270-296) and the C-terminal winged-helix domain (WHD) (a.a. 1424-1493).

Comparison of the predicted structure of the CSB dimer alone with that of the CSB dimer complexed with SNM1A (Fig. S9) revealed a different dimer orientation, with significant structural rearrangements occurring. Symmetry was noted in the CSB dimer in both models. PICKLUSTER was also used to analyze the interface between the two copies of CSB in the absence and presence of SNM1A to delineate the structural differences between the two models (given in Supplementary Tables 2 and 3). Within the CSB dimer, an interface is formed between the two C-terminal α-helices (a.a. 1280-1287). There are also interactions between the N-terminal domain of one copy (a.a. 153-177) and the ATPase domain (a.a. in region 751-758 and 956) of the other dimer partner. There are also interactions between the two copies within the N-terminal domains (a.a. 139-184 and 233-251), and there is a smaller interface between the ATPase domain of one protein copy and the C-terminal domain of the other. Conversely, when SNM1A is modelled with the CSB dimer, a large interface was observed between the C-terminal domain of one CSB (a.a. 1285-1314) and the ATPase domain (a.a. 655, 683 and 696-727) of the other CSB with fewer interactions observed between the N-terminal and C-terminal domains.

In addition to the above differences, detailed interrogation of the two models indicates that the orientation and interdomain molecular interactions of the WHDs in CSB change drastically when SNM1A is present. In the AF3 model of the CSB dimer alone, a single intramolecular hydrogen-bonding interaction is observed between D1442 and K696 of one copy (Fig. 1G). However, when SNM1A is present, additional interactions were observed, for example, a salt bridge between the backbone carbonyl of Q1438 and the sidechain of R1298 (intramolecular), the D1442 sidechain and R1301 (intramolecular), the D1442 sidechain and K696 (intermolecular) and the backbone carbonyl of D1442 and K262 (intramolecular) (Fig. 1H). No significant structural differences in SNM1A were observed between the model and the reported apo structure (PDB: 4B87 vs. AF3 model with CSB shown in Fig. 1F and 1H).

Analysis of our model indicates that the WHD of one copy of CSB interacts with the DNA-binding, β-CASP domain of SNM1A with a hydrogen bonding interaction identified between T1447-CSB and E864-SNM1A and a salt bridge formed between E864-SNM1A and R1478-CSB. There is also a hydrogen bonding interaction between CSB-R1467 and the backbone amide carbonyl of SNM1A-L894. We also noted good shape complementarity at the CSB:SNM1A interface.

Further AF3 modelling of the interface between just the catalytic domain of SNM1A and the WHD of CSB supported the above modeling observations. In the refined analysis, only one copy of each protein was used, as this stoichiometry was suggested by both the modelling with full-length CSB and our co-purification experiments (Fig. S4). AF3 modelling of the catalytic domain of $SNM1A_{698-1040}$ with $CSB\text{-}WHD_{1424-1493}$ resulted in a predicted structure with mean pLDDT of 94.04 and iPTM of 0.94, suggesting the generation of a very high confidence model (Fig. S9). Indeed, AF3 modelling of two copies of the WHD of CSB with one copy of the SNM1A catalytic domain resulted in poor prediction for the second WHD copy, and a dimer formed between the two WHDs (Fig. S10). There is no experimental evidence here or in the literature that the WHD of CSB alone is a dimer (Fig. S4) (Takahashi *et al*, 2019). SNM1A is predicted to interact with the opposite face of the CSB-WHD compared with a structure of the CSB-WHD and monoubiquitin (PDB: 6A6I). The functional implications of this are unclear, although it is possible that the WHD of CSB can interact with SNM1A and ubiquitin simultaneously (AF3 modelling given in Fig. S11).

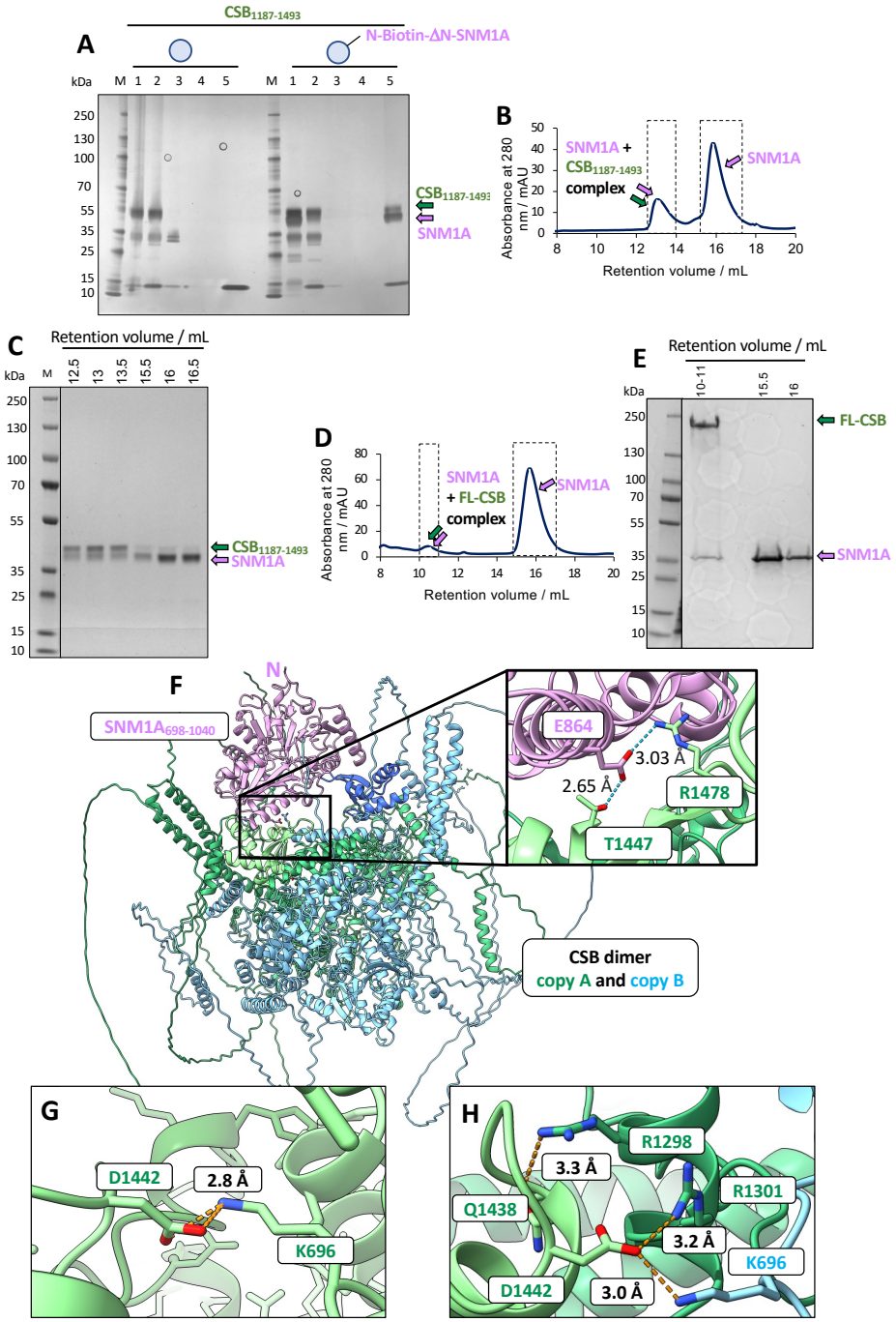

Figure 1: The C-terminal domains of CSB and SNM1A interact *in vitro* with two copies of C-terminal and full-length CSB. A. Pulldown assay of $CSB_{1187-1493}$ to N-Avi-Biotin-ΔN-SNM1A (42 kDa). Binding to Streptactin mutein matrix beads was used as a control. Beads were washed with buffer containing 0.5 M NaCl, then incubated with $CSB_{1187-1493}$ (appears at ~52 kDa). Unbound protein was removed, the beads washed, and protein eluted. Samples were run on SDS-PAGE (4-12% Bis-Tris) then analyzed by silver staining. Lanes 0: Protein standards in kDa, 1: CSB protein alone, 2: CSB protein unbound to Streptactin mutein matrix beads (with or without immobilised SNM1A), 3: wash 5/5 with 0.3 M NaCl, 10/10 with 0.5 M NaCl, 4: elution in 5 mM D-biotin buffer. **B.** Size-exclusion chromatography trace from co-purification of $CSB_{1187-1493}$ and $SNM1A_{698-1040}$. The peak at 12.5-13.5 mL corresponds to a CSB:SNM1A 2:1 complex as

supported by SDS-PAGE, MS analysis (for protein identities) and size-exclusion calibration curve from Superdex S200 GL 10/300. See Supplementary Information for calibration curve. **C.** Fractions from size-exclusion chromatography from co-purification of $CSB_{1187-1493}$ and N-Avi-Biotin-$SNM1A_{698-1040}$ analysed by SDS-PAGE. **D.** Size-exclusion chromatography trace from co-purification of FL-CSB and $SNM1A_{698-1040}$. The peak at 10.5 mL corresponds to the CSB:SNM1A 2:1 complex as supported by SDS-PAGE and MS analysis (for protein identification) and quantitative size-exclusion chromatography (Superdex S200 GL 10/300 column). **E.** Fractions from size-exclusion chromatography from co-purification of full-length CSB and N-Avi-Biotin-$SNM1A_{698-1040}$ analysed by SDS-PAGE. **F.** AlphaFold3 modelling of two copies of CSB with one copy of the catalytic domain of SNM1A, with ipTM of 0.32 and pTM of 0.33. CSB is in green (one copy in green, the other in blue, with the winged-helix domain in a different shade) with SNM1A (a.a. 698-1040) in pink. **G.** Key intermolecular interactions observed between the WHD of CSB and other CSB domains using an AF3 model of the CSB dimer. **H.** Intermolecular interactions observed in an AF3 model of CSB and SNM1A, focusing on interactions of the CSB WHD with the CSB N-terminal and ATPase domains.

The C-terminal WHD of CSB is required for interaction with SNM1A

We next sought to experimentally define the molecular nature of the interaction between SNM1A and CSB. As noted earlier, previous yeast two-hybrid work indicates that the C-terminal domain of CSB (a.a. 1187-1493) interacts with the C-terminal catalytic domain of SNM1A (residues 608-1040) (Iyama *et al*, 2015). Additionally, our AF3 modelling (Fig. 1F and Fig. S9) suggests that within the C-terminal region, the WHD of CSB is crucial for the SNM1A interaction. To test the model, three new C-terminal variants of CSB were generated i.e., a.a. 1187-1401, 1308-1493 and 1424-1493, the latter of which primarily consists of the WHD (Takahashi *et al*, 2019; Batenburg *et al*, 2018). Pull-down assays were performed with the immobilised catalytic domain of SNM1A (N-Biotin-$SNM1A_{698-1040}$), and only the CSB variants containing amino acids 1401-1493 were stably captured by the SNM1A fragment (Fig. 2A).

To further characterize the interaction between the C-terminal domain of CSB and SNM1A, surface plasmon resonance (SPR) spectroscopy was performed, again using immobilized N-Avi-Biotin-$SNM1A_{698-1040}$ and the three CSB constructs as analytes. Consistent with the pull-down assays, no binding was observed with CSB constructs that lacked the WHD (a.a. 1424-1493) and the ubiquitin binding domain (a.a. 1401-1424) (Fig. 2B and 2C; CSB C-terminal constructs consisting of amino acids 1187-1401 and 1187-1308). Conversely, the full C-terminal domain of CSB (a.a. 1187-1493) exhibited high affinity binding to SNM1A (K_D = 10.7 nM), with a comparable binding affinity as measured when using full-length CSB protein (apparent dissociation constant (K_D) = 4.0 nM). This data verifies that the C-terminal portion of CSB is the major binding region for SNM1A. . Notably, we observed that the WHD of CSB alone (a.a. 1424-1493) interacts with SNM1A with a K_D of 89 nM, indicating that high affinity binding is only conferred by the full C-terminal region (a.a. 1187-1493). This finding is consistent with the AF3 modelling, where the

WHD is predicted to physically interact with other C-terminal residues in CSB (R1298 and R1301) during the interaction with SNM1A. This complex network likely results in increased stabilization of the CSB:SNM1A intermolecular interaction and thus a higher affinity binding.

Dose response curves were obtained for immobilised SNM1A with $CSB_{1187-1493}$, $CSB_{1308-1493}$ and $CSB_{1424-1493}$ by multicycle SPR analysis (Fig. S12). Analysis could not be performed with full-length CSB as we could not obtain sufficient dissociation from the chip between each injection. Apparent K_D values obtained were of the same trend as single cycle values (Fig. 3C), but weaker than those determined by single cycle analyses likely due to slow and/or insufficient dissociation.

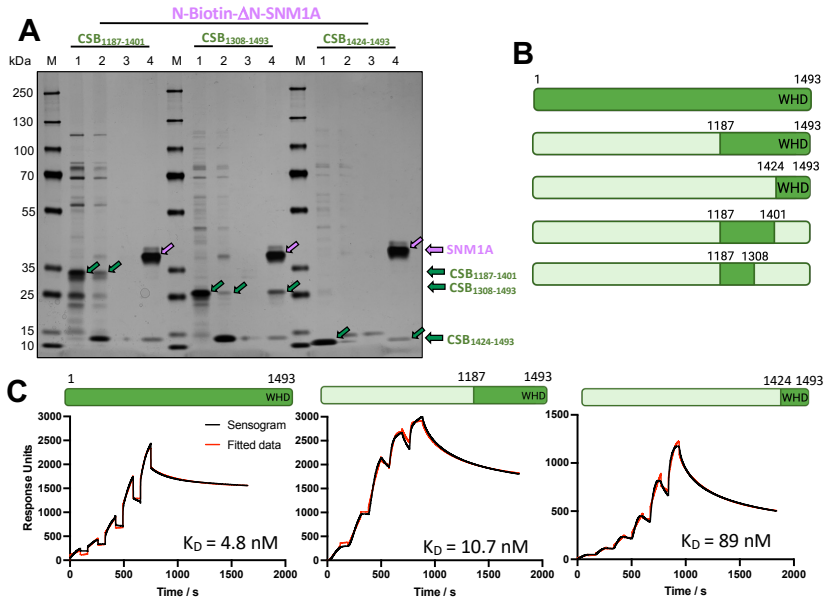

Figure 2: Validating the physical interaction between the C-terminal domain of CSB and SNM1A. A. N-Avi-Biotin-ΔN-SNM1A (42 kDa) was immobilised on Streptavidin mutein matrix beads, washed with 0.5 M NaCl, then incubated with CSB protein: $CSB_{1187-1401}$ (26 kDa, appears 32 kDa), $CSB_{1308-1493}$ (23 kDa, appears 26 kDa) or $CSB_{1424-1493}$ (11 kDa). Unbound protein was then removed, the beads washed, and protein eluted. Samples were separated by SDS-PAGE (4-12% Bis-Tris). then analysed by silver staining. Lanes 0: Protein standards in kDa, 1: CSB protein alone, 2: CSB protein unbound to SNM1A, 3: wash 10/10 with 0.5 M NaCl, 4: elution. **B**. Apparent equilibrium dissociation (K_D) constants obtained for CSB domains for interaction with N-Avi-Biotin-SNM1A. The dark green region of the protein sequence map represents the amino acid ranges used for experiments. **C.** Single cycle response curves from SPR analyses, with apparent equilibrium dissociation constants. Parameters determined from fitting are given in Supplementary Table 4.

Based our AF3 modelling, we predicted alanine substitutions, i.e., CSB-T1447A and SNM1A-E864A, would abrogate the interaction between CSB's WHD and SNM1A. We first isolated the WHD of CSB with and without an alanine substitution (T1447A) by recombinant expression in *E. coli* expression. Circular dichroism (CD) spectroscopy measurements revealed that both protein

versions had identical folds (Fig. S7), indicating that the mutation had no effect on the overall protein structure. Initially, pull-down assays were performed (as previously described) with immobilised N-Avi-Biotin-SNM1A$_{698-1040}$, and either wild-type or T1447A-CSB C-terminal fragments (a.a. 1187-1493, 1308-1493 and 1424-1493) as the analytes. In all cases, the interaction with SNM1A was abrogated in the case of the T1447A-substituted CSB C-terminal constructs (Fig. 3A). This result was confirmed by SPR with the T1447A-WHD fragment alone (a.a. 1424-1493) (Fig. 3B and 3C). Specifically, by multi-cycle SPR analysis using immobilized N-Avi-Biotin-SNM1A, a K_D of 290 nM ± 110 nM was obtained for the wild-type WHD of CSB (a.a. 1424-1493) as the analyte (Fig. 3B), whereas no binding was observed with T1447A-WHD as the analyte (Fig. 3C). These results further support that the WHD of CSB interacts with SNM1A in a manner that requires an intermolecular interaction involving T1447 of CSB, consistent with the modelling prediction.

To define the interacting elements of SNM1A, the mutant protein, E864A was generated, as this residue was predicted to facilitate a key hydrogen bonding connection. By pull-down assay, E864A-SNM1A could interact with full-length CSB protein with a similar affinity as wild-type SNM1A (Fig. 3A). There was a diminished interaction of the mutant protein with the C-terminal domains of CSB, CSB$_{1187-1493}$ and CSB$_{1308-1493}$ in comparison with wild-type SNM1A, and no or a weak interaction was observed between E864A-SNM1A and the WHD of CSB (Fig. 3D). In addition, with either full-length CSB or CSB$_{1187-1493}$, a similar albeit slightly reduced K_D was measured for wild-type SNM1A relative to E864A in SPR experiments (12 vs. 29 nM) (Fig. 3E). With the WHD of CSB, the interaction with WT-SNM1A had a K_D of 89 nM, whereas immobilized E864A exhibited a much weaker interaction in the low micromolar range, although the fitting was poor (Fig. 3E). A similar affinity for WT and E864A-SNM1A was measured when using immobilized full-length CSB (4 vs. 8 nM, Fig. 3F). The above observations are consistent with the results of the pull-down assays and with our hypothesis that even though the WHD of CSB is necessary for the physical interaction with SNM1A, the full C-terminal domain is required for high affinity binding. It is also likely that a single alanine substitution in SNM1A is insufficient to completely abrogate the interaction with the full-length CSB protein, considering that other intermolecular interactions are predicted.

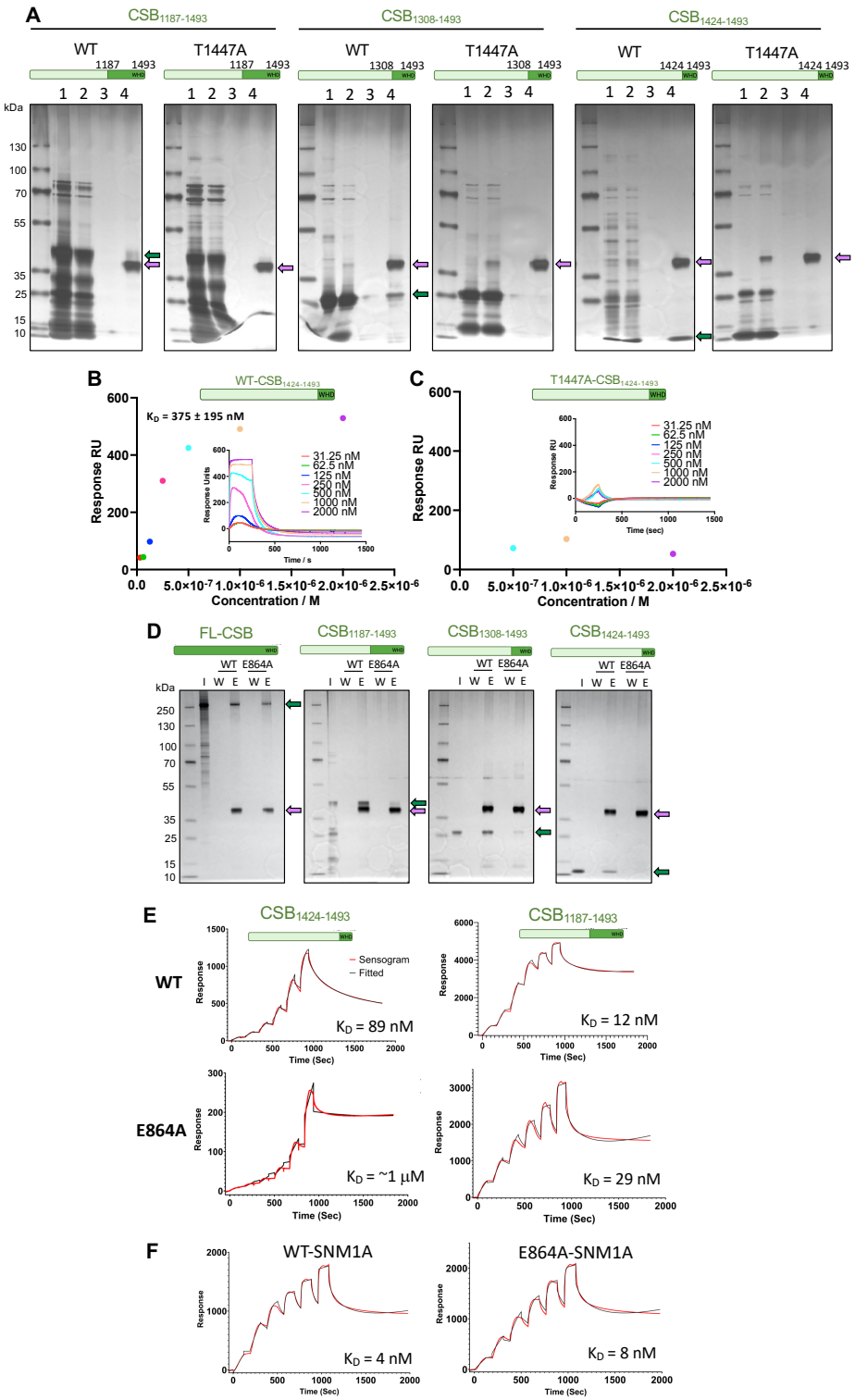

Figure 3: Validation of the physical interaction between SNM1A and the winged-helix domain of CSB. A. N-Avi-Biotin-ΔN-SNM1A (42 kDa) was immobilised on Streptavidin mutein matrix beads, washed with 0.5 M NaCl, then incubated with $CSB_{1424-1493}$ (WT or T1447A, 11 kDa). Unbound protein was removed, the beads washed, and protein eluted. Samples were separated by SDS-PAGE (4-12% Bis-Tris) then analyzed by silver staining. Lanes 0: Protein standards in kDa, 1: CSB protein alone, 2: CSB protein

unbound to SNM1A, 3: wash 10/10 with 0.5 M NaCl, 4: elution. **B.** SPR (dose response and multi-cycle analysis) with immobilized N-Avi-Biotin-SNM1A with WT-$CSB_{1424-1493}$ as the analyte. Response units are given. **C.** SPR (dose response and multi-cycle analysis) with immobilized N-Avi-Biotin-SNM1A with T1447A-$CSB_{1424-1493}$ as the analyte. Response units are given. **D.** N-Avi-Biotin-ΔN-SNM1A (WT or E864A, appears 42 kDa) was immobilised on Streptavidin mutein matrix beads, then incubated with $CSB_{1424-1493}$ (WT or T1447A, 11 kDa). Unbound protein was removed, the beads washed, and protein eluted. Samples were separated by SDS-PAGE (4-12% Bis-Tris) then analysed by silver staining. Lanes 0: Protein standards in kDa, 1: CSB protein alone, 2: CSB protein unbound to SNM1A, 3: wash 10/10 with 0.5 M NaCl, 4: elution. **E.** Single cycle SPR spectroscopy with N-Avi-Biotin-ΔN-SNM1A (WT or E864A) immobilised on the chip and $CSB_{1424-1493}$ or $CSB_{1187-1493}$ as the analyte. Dissociation constants with other parameters are given in Supplementary Table 4. **F.** Single cycle SPR spectroscopy with N-His-ZB-CSB-Avi-Biotin immobilized on the chip with untagged SNM1A (WT or E864A) as the analyte. Dissociation constants are given with other parameters determined in Supplementary Table 5.

CSB DNA binding and stimulation of SNM1A nuclease activity

Previous work has demonstrated that CSB stimulates SNM1A's exonuclease activity with simple ss and ds DNA substrates (Iyama *et al*, 2015). We sought here to expand our understanding of the substrate scope for this functional interaction. Thus, building on prior work, we examined the stimulation of SNM1A's exonuclease activity by CSB across a panel of potentially relevant DNA repair substrates (Fig. 4A). All substrate oligos were designed with a phosphate at the 5′ end of the 3′-^{32}P-labelled DNA strand (5′P), a feature which is required for SNM1A's nuclease activity, while the 5′ end of the strand not under analysis was capped with a biotin group (Bio) to prevent competing digestion that would confound interpretation. As previously reported (Sengerová *et al*, 2012), SNM1A alone has a clear preference for ssDNA, with significantly reduced hydrolysis observed on dsDNA (27% vs. 79% digestion on ssDNA vs. dsDNA after 10 min for SNM1A alone), on a substrate with a recessed 5' end (3' overhang substrate), or on DNA containing a 5' overhang substrate, where SNM1A can be observed slowing when entering the region of dsDNA (Fig. 4A). Notably, the presence of full-length CSB enhanced the nuclease activity of SNM1A on all substrates, albeit with only a minor effect on the 5' overhang substrate. The bubble DNA used corresponds to a 17 bp bubble, which mimics the size of a transcription intermediate. SNM1A alone has poor activity on this substrate (83% substrate remaining after 10 min), but in the presence of CSB, an ~5-fold stimulation was observed with only 12% initial substrate remaining at 10 min.

We next examined whether the interacting, C-terminal domain of CSB alone exerts a similar influence on the catalytic activity of SNM1A as the full-length CSB protein. SNM1A nuclease assays were initially performed without CSB or with full-length or individual domains of CSB (Fig. 4B). Importantly, prior work suggested that CSB stimulates the nuclease activity of SNM1A

through a direct physical interaction that increases the affinity of SNM1A for DNA rather than by modifying the DNA structure via its chromatin remodelling activity (Iyama *et al*, 2015). Using the ssDNA as the substrate, full-length CSB again showed stimulation of SNM1A nuclease activity (only 27% of substrate remaining after 10 min vs. 45% with SNM1A alone). Unexpectedly, the presence of neither the CSB N-terminal, ATPase, nor C-terminal domain alone had a measurable effect on SNM1A activity, even though the latter was observed to stably interact with SNM1A. Thus, it appears that CSB's stimulatory effects on SNM1A activity go beyond the binding of its WHD to SNM1A and any associated conformational shifts that this induces.

In an attempt to address the apparent discrepancy above, we assessed DNA binding activities of the different CSB domains. Using EMSAs, we observed that CSB, i.e., full-length, the ATPase domain and C-terminal domain, bind to ssDNA (Fig. 4C) a feature of the protein that has not been reported, specifically for the C-terminal region (1187-1493). The DNA binding substrate scope of the C-terminal domain of CSB (1187-1493) was thus more extensively characterized by EMSA (Fig. 4E). In these experiments, CSB-DNA binding was observed with all DNA substrates tested, consistent with the substrate scope and DNA binding activity previously reported for the full-length protein (Berquist & Wilson, 2009).

To identify the precise C-terminal region of CSB that binds DNA, EMSAs were performed with increasingly shorter C-terminal fragments. The C-terminal region of CSB that interacts with SNM1A (a.a. 1187-1493) was found to bind DNA, while removal of the WHD (a.a. 1187-1401) reduced the stable complex observed, and no DNA binding was detected with $CSB_{1187\text{-}1308}$. These results indicate that the UBD and WHD domains of CSB are not essential for DNA binding in conjunction with the 1187-1401 region but instead confer higher affinity binding (Fig. S13). Other CSB, C-terminal peptide fragments (a.a. 1187-1308, 1424-1493, 1308-1493) were found to not bind DNA in EMSAs (data not shown).

We next determined whether the C-terminal domain of CSB and SNM1A can form a complex together with DNA. Firsty, we formed a CSB:SNM1A complex *in vitro* (by mixing the two proteins), and second, incubated the complex with either ssDNA or a reactive intermediate mimic (a minimal version of the substrate likely formed post XPF-ERCC1 incision but lacking an ICL) before analysis via an EMSA. Binding of CSB (a.a. 1187-1493) alone to both substrates showed a small molecular weight shift in the gel (see Fig. 4D, lanes 4-6 vs. lane 1 and lanes 10-12 vs. lane 7), consistent with stable association. SNM1A bound well to ssDNA well, with evidence of two copies

of SNM1A bound to DNA when 2:1 protein:ssDNA ratio was used. No 2:1 SNM1A:DNA complex was observed to the reactive intermediate substrate, with a 1:1 SNM1A:DNA complex observed even with 2 molar equivalents of protein to DNA. With both substrates, when both SNM1A and CSB were incubated together on ssDNA, a small molecular weight shift of the SNM1A band is observed and a maximum of one copy of SNM1A is present in the complex. Less of the CSB:DNA complex is also observed. The DNA binding substrate scope of the C-terminal domain of CSB (1187-1493) was assessed by EMSA (Fig. 4E), with CSB-DNA binding observed with all DNA substrate structures tested, consistent with the substrate scope and DNA binding affinity previously reported for the full-length protein (Berquist & Wilson, 2009). Our results indicate that the full-length CSB is required for optimal SNM1A activity stimulation, although the C-terminal region (a.a. 1187-1493) is required for physical interaction with SNM1A.

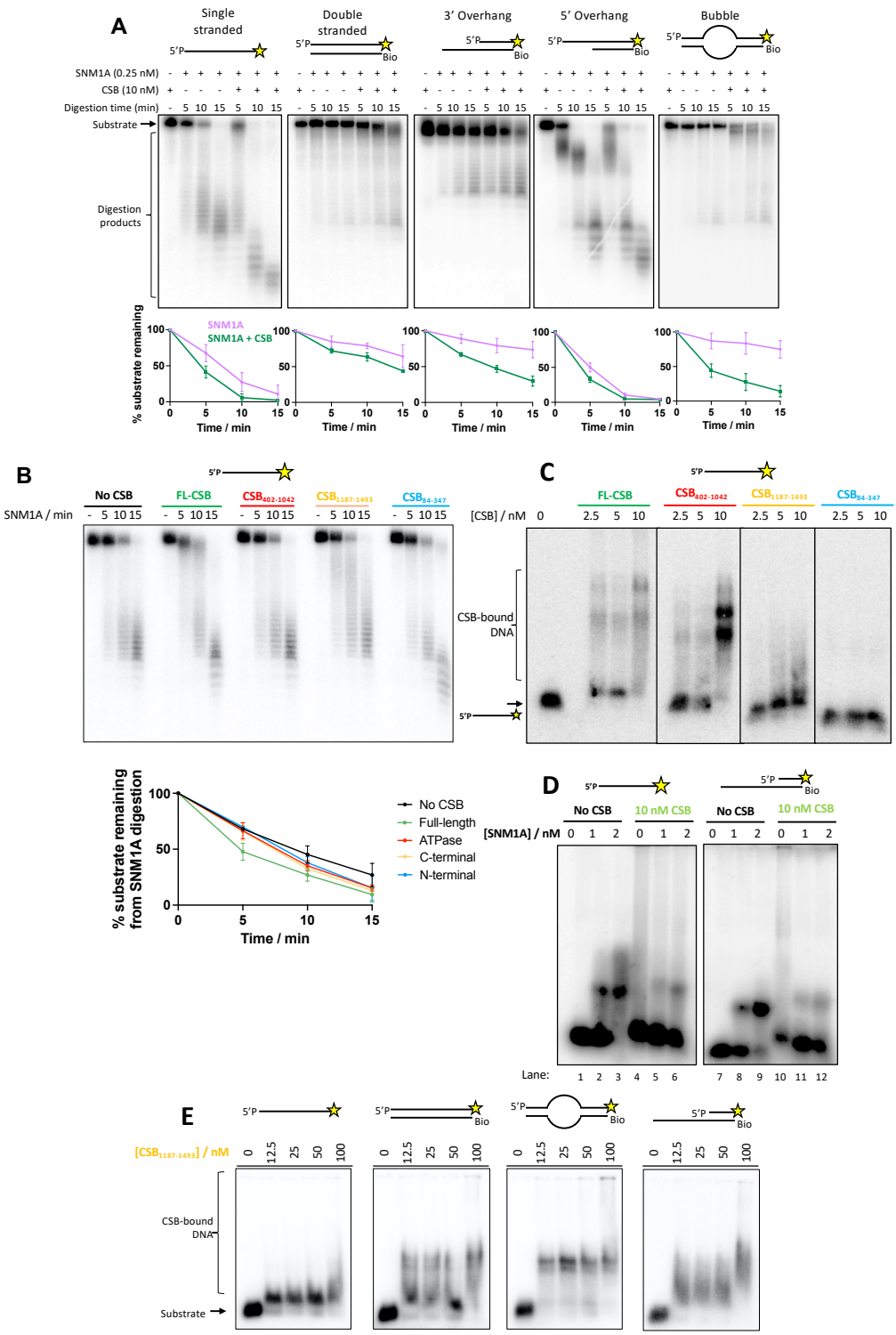

Figure 4: CSB stimulates SNM1A's nuclease activity on a range of DNA repair intermediate substrates. Oligonucleotides used (with preparation) are given in Supplementary Table 1. The location of the 3′ ^{32}P is given by a yellow star. Quantification was based upon assays conducted in duplicate; error bars represent one standard deviation. **B.** Assays of SNM1A's activity on ssDNA in the absence or presence of CSB (full-length or shorter domains), with quantification from assays performed in duplicate; error bars represent one standard deviation. **C.** EMSA of full-length CSB, CSB ATPase domain (a.a. 402-

1014), CSB C-terminal domain (a.a., 1187-1493) and CSB N-terminal domain (a.a. 84-347) binding to single stranded DNA in the presence of 4 mM EDTA. **D.** EMSA observed SNM1A binding to DNA substrates in the presence/absence of $CSB_{1187-1493}$ in the presence of 4 mM EDTA**. E**. EMSA of CSB C-terminal domain (a.a. 1187-1493) binding to different DNA structures in the presence of 4 mM EDTA.

CSB stimulates SNM1A's nuclease degradation through DNA interstrand crosslinks

Cellular studies indicate that CSB is required for efficient unhooking of ICLs in non-dividing cells (Iyama *et al*, 2015). To specifically and quantitatively assess whether CSB has a role in stimulating SNM1A resection past an ICL oligonucleotide duplex to achieve unhooking, a substrate (herein referred to as the '0-nucleotide' substrate) was designed whereby SNM1A encounters a DNA interstrand crosslink in its first hydrolytic event, i.e., there are no nucleotides preceding (5' to) the crosslink. The '0-nucleotide' substrate was synthesized by phosphoramidate chemistry with a triazole crosslink formed by a Cu(I)-catalysed 'click chemistry' reaction between azide and alkyne groups incorporated on opposing bases in the duplex. Previous work has shown that this crosslink produces only minor helical distortions (Kočalka *et al*, 2008; Abdullah *et al*, 2017), meaning that enhanced nuclease activities cannot be falsely attributed to more favorable substrate conformations in the presence of the ICL. Hydrolysis of the substrate by SNM1A alone was partial, with ~40% of the substrate digested after 5 min. However, with the addition of CSB, only ~15% of the substrate remained, suggesting that CSB has a substantial role in aiding SNM1A's digestion past ICLs (Fig. 5A). EMSAs comparing CSB binding to ICL-containing and control substrates indicate higher affinity binding to the former, consistent with a role in facilitating ICL repair (Fig. 5B vs. 5C).

We next examined whether the stimulation of SNM1A at the ICL was due to enhanced recruitment of SNM1A to DNA by CSB and/or whether SNM1A and CSB form a complex together on DNA that might facilitate nuclease degradation. Initially, we performed EMSAs with CSB and SNM1A together on a range of DNA substrates (ssDNA, a 3' overhang substrate and a substrate with no nucleotides 5' prior to the ICL) in the presence of 4 mM EDTA to abrogate SNM1A's divalent cation-dependent nuclease activity and substrate digestion. When 2 molar equivalents of SNM1A are mixed with ssDNA or the 0-nucleotide substrate (Fig. 2D, lanes 4-6, 10-12 and 16-18), two copies of SNM1A appear to bind (Fig. 5D), providing insight into the mode by which SNM1A binds DNA. Additionally, a maximum of 2 copies of SNM1A bind to a 50-mer oligonucleotide, 3 copies to a 75-mer, and 4 copies to a 100-mer, suggesting that the DNA binding footprint of SNM1A is approximately 25 nucleotides (Fig. S14). Less binding to the 100-mer oligonucleotide was observed compared with the 75-mer perhaps due to decreased accessibility of the 5'-phosphate

to the protein, as a 5'-phosphate is required for SNM1A nuclease function. Increased binding of SNM1A to the '0-nucleotide' ICL substrate compared with the substrate lacking ICL was also observed (Fig. 2D). Since the ICL present is 'native-like' with minor helical distortions, this enhanced binding is likely attributed to increased accessibility of the 5'-phosphate and perhaps increased affinity of SNM1A for ICL-DNA.

Examination of potential co-operative binding of CSB (dimeric protein of 400 kDa) with SNM1A to DNA is challenging, since the molecular weight of SNM1A (38 kDa) is comparatively small. Moreover, in previous work attempting to visualize CSB binding to DNA, only a weak affinity was observed in gel-based assays, perhaps reflective of unstable DNA association due to intrinsic protein flexibility (Sarker *et al*, 2005). Nevertheless, for all three substrates tested (ssDNA, 0-nucleotide intermediate and 0-nucleotide intermediate with an ICL), there was reduced binding of SNM1A alone to DNA in the presence of CSB, suggesting that the two bind together, although a stable complex is not detected. In addition, a higher molecular weight DNA complex was seen with CSB alone (lane 4 for all substrates, Fig. 4D), with increased intensity of this complex band when SNM1A was present (lanes 4 vs. lanes 5 and 6, Fig. 4D), consistent with cooperative substrate binding. Finally, as CSB is a member of the SWI/SNF family of ATP-dependent chromatin remodelers, we tested the effect of ATP on SNM1A activation. Nuclease assays were performed on ssDNA and the '0-nucleotide' ICL-containing substrate, with SNM1A alone, with SNM1A and ATP, with SNM1A and CSB, and with SNM1A, CSB and ATP (Fig. S15). On both substrates, ATP had no effect on SNM1A nuclease activity or the CSB stimulatory effect.

Due to the potential for accumulation of R-loop intermediates during TC-ICL repair, the effect of the presence of an R-loop on CSB's stimulation of SNM1A's exonuclease activity on the '0-nucleotide ICL' substrate was examined (Sollier *et al*, 2014). Substrates were generated by annealing the '0-nucleotide ICL' substrate to a 10 nucleotide RNA oligo (Fig. S1 and Table S1) which is complementary to the nucleotides directly 5' of the ICL on the complementary strand to that possessing the 5'-phosphate (Fig S16). Nuclease assays comparing SNM1A's digestion at different timepoints in the presence or absence of CSB, with or without an R-loop, revealed slightly increased substrate hydrolysis in the presence of the RNA intermediate (Fig. S17). By contrast, the presence of a D-loop had little effect on SNM1A's activity. There was no notable difference in the stimulation of SNM1A nuclease activity by CSB when comparing the three substrates (no loop, D-loop, R-loop), although increased CSB binding to the R-loop containing substrate was suggested by EMSA (Fig. S18).

Finally, we determined whether abrogating the key intermolecular interaction (CSB-T1447 and CSB-R1478 with SNM1A-E864) between CSB and SNM1A would affect the stimulation of nuclease activity by CSB. We performed nuclease assays with wild-type and E864A-SNM1A in the presence/absence of CSB, on the '0-nt' ICL substrate (Fig. 5E). E864A-SNM1A exhibited similar, although slightly less efficient exonuclease activity compared with WT-SNM1A, implying that the alanine mutation has minimal effect on activity and thus similar protein folding can be expected. Unlike wild-type SNM1A, no stimulation of E864A-SNM1A nuclease activity by CSB was observed and SPR indicated that full-length CSB binds to SNM1A with a similar but slightly weaker affinity (K_D = 4 nM vs. 8 nM for WT vs. E864A-SNM1A, Fig. 3F) with this mutation present. These results thus imply that an intermolecular interaction between E864 on SNM1A and the WHD of CSB is required for the observed stimulation of nuclease activity.

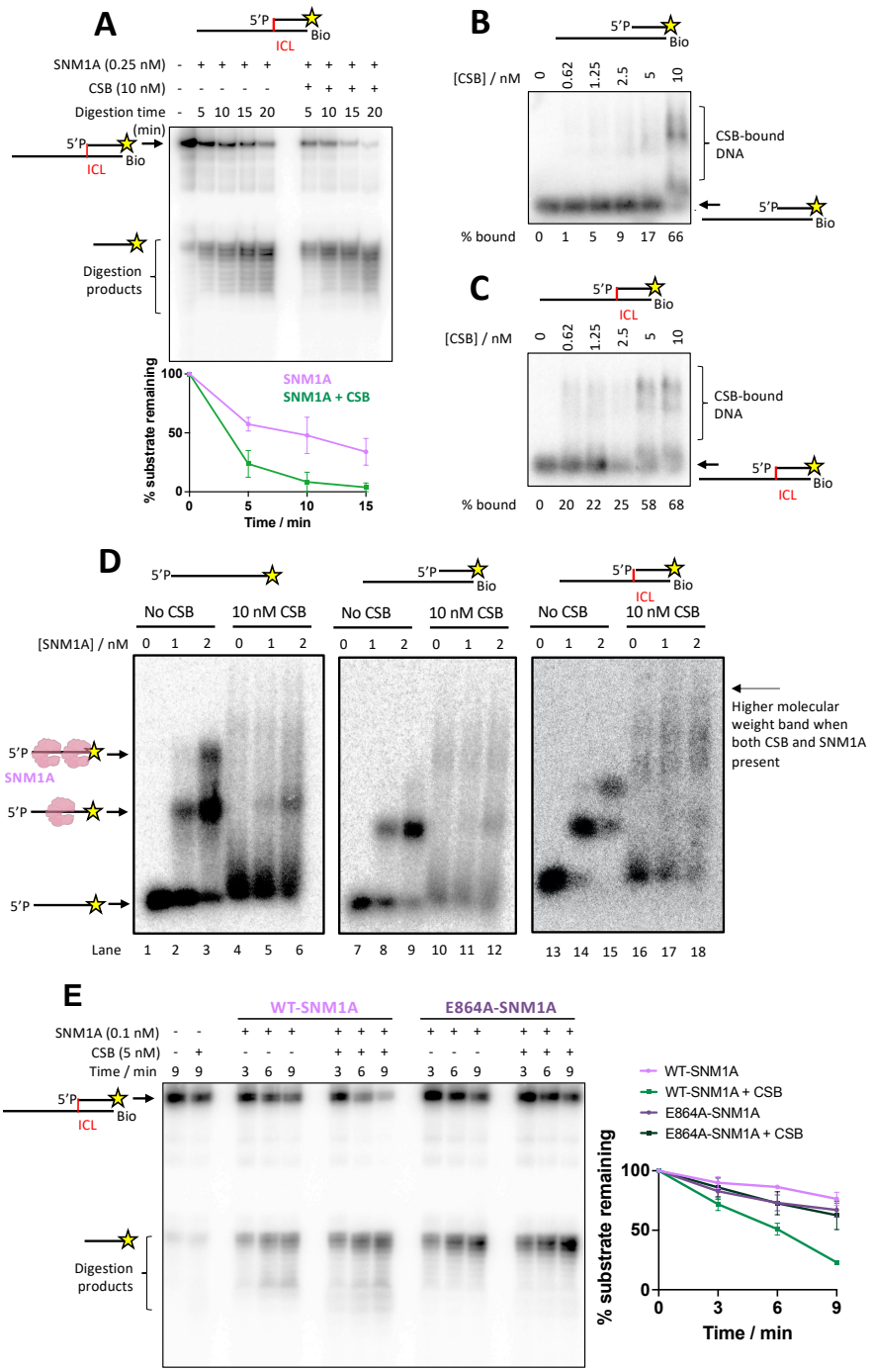

Figure 5: CSB binding to DNA is promoted by the presence of an ICL and CSB enhanced SNM1A's exonuclease activity on DNA ICLs. A. Time-course nuclease assay of SNM1A's activity on the '0-nucleotide' substrate (Fig. S1) in the presence or absence of CSB, with quantification based on assays in duplicate; error bars represent one standard deviation. **B**. EMSA with CSB titration of binding to the '0-nucleotide' substrate (1 nM) without an ICL, in the presence of 1 mM $CaCl_2$. The percentage of protein-bound DNA was determined by ImageJ quantification, correcting for the negative control. **C**. EMSA with CSB titration of binding to the '0-nucleotide' substrate (1 nM) with an ICL, in the presence of 1 mM $CaCl_2$. The percentage of protein-bound DNA was determined by ImageJ quantification, correcting for the negative control. **D.** EMSA of CSB and SNM1A binding to 1 nM of substrate in the presence of 4 mM EDTA. **E.**

Assay of N-Avi-Biotin-ΔN-SNM1A (WT or E864A, 0.1 nM) exonuclease activity in the presence/absence of full-length CSB (5 nM) at different timepoints. Quantification is based on reactions performed in duplicate; error bars represent one standard deviation. AF3 modelling suggests the activity requires a specific hydrogen-bonding interaction between CSB and SNM1A.

Discussion

Human SNM1A plays a role in both replication-dependent and replication-independent ICL repair (Baddock *et al*, 2020). In yeast, and likely other eukaryotes, most ICL repair depends upon the ancestral SNM1A homolog, Pso2, and occurs in a manner that is genetically epistatic and therefore predicted to operate in the same pathway as NER factors (Cattell *et al*, 2010). By contrast, replication-coupled ICL repair in mammals is controlled through the FA pathway and a number of key NER factors (with the notable exception of XPF-ERCC1) are apparently dispensable within this pathway (Semlow & Walter, 2021). Prior work identified a physical interaction between SNM1A and CSB, suggesting a potential role for SNM1A in TC-ICL repair (Iyama *et al*, 2015), a pathway that has been minimally explored. In support of this concept, CSB is involved in the recruitment of SNM1A to localised, UV-induced psoralen ICL damage 'stripes' in a manner dependent upon active transcription. Moreover, in non-dividing neuronal cells, CSB is important for achieving normal levels of ICL unhooking, as determined by a modified Comet assay. Together, these observations provide robust evidence for a coupled role between SNM1A and CSB in TC-ICL repair, but the nature and molecular details of the interaction between these two proteins and how this interaction influences ICL processing activities of SNM1A has remained unknown.

Our work supports the proposal that CSB dimerises *in vitro* and provides evidence for the involvement of the C-terminal region of CSB (a.a.1308-1424) in this process. A homodimeric CSB complex is consistent with previous gel filtration studies and complex trapping studies where full-length CSB was shown to interact with the ATPase domain of CSB alone (Christiansen *et al*, 2005). Scanning force microscopy experiments of CSB binding to circular, nicked DNA also suggests dimerization upon binding with a single conformational state observed (Beerens *et al*, 2005), although conformational changes upon DNA binding cannot be excluded. However, in recent cryogenic electron microscopy (cryoEM) structures of CSB in complex with RNA polymerase II and other TC-DNA repair factors, only one copy of CSB was present (Kokic *et al*, 2021, 2024). These structures resolved only the ATPase domain of CSB and the complexes studied were crosslinked with glutaraldehyde, providing potential complications that could have resulted in only one copy being present.

A striking aspect of our biochemical analysis was the direct stimulation of a DNA repair enzyme (SNM1A) by CSB in a manner aligning with its predicted cellular role. While CSB has been reported to interact with a broad range of DNA repair and transcription factors, influencing their localisation and turnover (Nieto Moreno *et al*, 2023), an active involvement for CSB in modulating DNA repair enzyme activity has been restricted to several factors involved in BER (NEIL1, NEIL2 and APE1) (Muftuoglu *et al*, 2009; Wong *et al*, 2007; Aamann *et al*, 2014). Notably, the molecular basis of these CSB-mediated effects are not presently understood. Using biochemical, biophysical, and structural methods to define the nature of the interaction between CSB and SNM1A, we reveal that CSB promotes SNM1A's activity during the key ICL unhooking step. In brief, the DNA binding catalytic domain of SNM1A (β-CASP domain) interacts with the C-terminal portion of CSB, where the WHD of CSB provides a minimal interacting region, but high-affinity binding (equivalent to that observed with full-length CSB) is conferred by the presence of more extensive C-terminal region of CSB, where two CSB molecules are involved. CSB and SNM1A were also shown to bind together on DNA repair intermediate oligonucleotides.

The WHD of CSB has been shown to mediate other CSB protein-protein interactions, e.g., with RIF1, a downstream effector of 53BP1 that promotes non-homologous end joining in G1 (Batenburg *et al*, 2017). The WHD also non-covalently interacts with monoubiquitin, for which the biological role is unclear but hypothesised to be required for association with ubiquitinated repair factors (e.g. CSA or RNAPII) (Batenburg *et al*, 2021; Takahashi *et al*, 2019). We note here that the affinity between SNM1A and the CSB WHD is over two orders of magnitude greater than the reported interaction between the WHD and monoubiquitin (89 nM vs. 30 μM) (Takahashi *et al*, 2019), supporting a biologically relevant role. WHDs in proteins have also been shown to have functional roles beyond mediating protein-protein interactions, such as in facilitating specific protein-DNA interactions and in aiding strand separation by DNA helicases (such as for RecQ family helicases) (Harami *et al*, 2013).

AF3 structural predictions support an interaction between the WHD and SNM1A catalytic domain, and we confirmed the contributions of key residues in the CSB WHD fragment (T1447A) and SNM1A catalytic domain (E864A), where a dramatic increase in K_D was observed with mutant protein forms. Most notably, the interaction between the WHD and SNM1A is necessary, but not sufficient, for stimulation of SNM1A nuclease activity, implying that further communication, either between proteins or with the DNA substrate, are induced by full-length CSB. Indeed, our AF3

modelling predictions of the CSB dimer alone and in the presence of SNM1A suggest structural rearrangements of full-length CSB upon SNM1A binding. Previous co-localisation experiments in cells over-expressing CSB domain constructs demonstrate that the C- and N-terminal CSB domains interact with the core ATPase domain and that disengagement of the WHD from the ATPase domain of CSB is required for efficient NER (Batenburg *et al*, 2018). The observations in full suggest significant structural rearrangements of CSB during DNA repair.

The observation that the interaction with full-length CSB substantially increases the nuclease activity of SNM1A on a variety of substrates, including a substrate where the first digestion event is a crosslinked nucleotide, strongly suggests a biochemically functional role for the CSB-SNM1A complex (as opposed to an exclusive role in localising the nuclease at sites of damage). This idea is fully consistent with the cellular studies showing (Iyama *et al*, 2015) a key role for CSB in recruiting SNM1A and in achieving normal rates of ICL unhooking. Moreover, our work reveals that CSB contains a previously unassigned DNA binding domain within its C-terminal region and that CSB and SNM1A bind together to DNA substrates, potentially bringing SNM1A into the vicinity of its substrate and potentially positioning the enzyme for its hydrolysis reaction.

Notably, we found no evidence of ATP-dependent roles of CSB in enhancing SNM1A activity, although experiments here are performed in a minimal system and not in the context of chromatin. Nevertheless, CSB is well-known to possess DNA-dependent ATPase activity and to carry out single-strand annealing (Citterio *et al*, 2000; Berquist & Wilson, 2009; Muftuoglu *et al*, 2006), and prior studies have also described DNA binding of full-length CSB via its ATPase domain in structures of CSB within the NER initiation apparatus (Kokic *et al*, 2021). It has been shown previously that stable association of CSB with chromatin requires physical co-operation between the CSB ATPase and C-terminal domains, based upon *in vivo* chromatin association experiments with a CSB construct lacking a.a. 1010-1493 (Lake *et al*, 2010). There is also a computationally predicted nucleotide binding fold (NBF) in the CSB C-terminal domain from residues 1134 to 1138, although a DNA-binding activity of this domain has not been explored experimentally (Iyama & Wilson, 2016). The lack of DNA binding by the N-terminal domain (a.a. 84-347) is consistent with prior work (Figure 4C).

It is interesting to note that SNM1A, a potent and potentially damaging 5′-3 exonuclease, has a relatively low affinity for all DNA substrates examined to date, but has high affinity for a number of DNA binding ligands (e.g., CSB) that can direct its activity to relevant repair intermediates.

However, in replication-coupled ICL repair, the binding to monoubiquitinated PCNA through interaction of the SNM1A PIP box and an N-terminal UBZ (ubiquitin binding domain) are required (Yang *et al*, 2010). Moreover, for the recently identified role of SNM1A in the repair of complex DNA double-strand breaks, a PAR-binding zinc finger (PBZ) appears to act in concert with the PIP box and UBZ motif to control recruitment of SNM1A in anticipation of its repair role (Swift *et al*, 2024). How SNM1A is subsequently released from sites of repair and kept in a functionally dormant state, where it is not able to access non-pathological, free DNA 5′ termini, remains unknown and is a topic of substantial interest for future investigation.

While the interaction between SNM1A and the CSB C-terminus is of high affinity and is required for efficient localisation of SNM1A to sites of TC-ICL repair, it remains unknown what other factors are involved in the process of TC-ICL repair (Enoiu *et al*, 2012). By analogy with TC-NER, it will be important to assess the involvement of other core NER factors and accessory TC-NER factors (e.g., CSA, UVSSA, ELOF1, STK19, etc.) reviewed in (Nieto Moreno *et al*, 2023; van den Heuvel *et al*, 2021; Boeing *et al*, 2016). The recent evidence for the involvement of UVSSA, CSA and TFIIH components in TC-ICL repair is also noteworthy (Liebau *et al*, 2023). Additionally, canonical TC-NER involves numerous post-translational modifications (PTMs) and other interactions, including ubiquitination of CSB by CSA (Groisman *et al*, 2006) and RNA Pol II ubiquitination (Tufegdžić Vidaković *et al*, 2020; Nakazawa *et al*, 2020; Boeing *et al*, 2016). As the study of TC-ICL repair is in its infancy, building a comprehensive picture of the cellular requirements is needed, including the spatial and temporal relationship with other transcription and repair factors and the impact of PTMs. Moreover, it's worth emphasizing that the barrier to RNA Pol II progression presented by an ICL is structurally distinct from lesions more widely studied (e.g. UV-induced damage) that are repaired by TC-NER. ICLs will likely arrest RNA Pol II progress at a greater distance from the lesion due to their ability to prevent strand separation during transcription, and therefore assumptions of similarity with classical TC-NER should be made with caution.

The potential physiological importance of TC-ICL repair is underscored by recent studies examining the interplay between aldehyde detoxification and transcription-associated DNA repair (Mulderrig *et al*, 2021). In these studies, the cellular accumulation of formaldehyde (genetically induced, *via ADH5* knockout) produced endogenous lesions that blocked transcription (noting that formaldehyde can produce a wide range of DNA lesions, including both ICLs and DNA protein crosslinks (Benedict *et al*, 2024)). Subsequent elimination of CSB in these *ADH5*$^{-/-}$ mice induced features reminiscent of Cockayne syndrome. It is of interest to determine whether, at a

physiological level, the key TC-ICL repair unhooking step we have begun to define here, catalysed by CSB and SNM1A, plays a key role in removing the endogenous DNA damage that contributes to Cockayne syndrome, and therefore, potentially, premature aging features and neurodegeneration seen in the general population.

DDR nucleases have long been targeted for cancer treatment and some advances have been made; however, sufficiently safe, selective and potent active site targeting inhibitors suitable for *in vivo* target validation have yet to be identified. Our results suggest that targeting protein-protein interactions that are vital for specific nuclease activities, such as is the case with CSB mediated stimulation of SNM1A's nuclease activity, may be a productive alternative to active site targeting (Barakat *et al*, 2009, 2012). Since many nucleases have multiple context-dependent functions, such an approach may be particularly useful for achieving selectivity with respect to disruption of specific aspects of nuclease mediated biological processes, e.g., only the TC-ICL repair role of SNM1A.

Acknowledgements

We thank Dr. David Staunton in the Molecular Biophysics Suite, Department of Biochemistry, University of Oxford for assistance with circular dichroism spectroscopy and Dr. Rod Chalk from the Centre for Medicines Discovery, University of Oxford for mass spectrometry analysis. This work was supported by Cancer Research UK Discovery Programme awards CRUK/A24759 and DRCRPG-Nov23/100005 to PJM and CJS and an NIH-OxCam student award to WJN. Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

References

Aamann MD, Hvitby C, Popuri V, Muftuoglu M, Lemminger L, Skeby CK, Keijzers G, Ahn B, Bjørås M, Bohr VA, *et al* (2014) Cockayne Syndrome group B protein stimulates NEIL2 DNA glycosylase activity. *Mech Ageing Dev* 135: 1–14

Abdullah UB, McGouran JF, Brolih S, Ptchelkine D, El-Sagheer AH, Brown T & McHugh PJ (2017) RPA activates the XPF - ERCC 1 endonuclease to initiate processing of DNA interstrand crosslinks . *EMBO J* 36: 2047–2060

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L,

Ballard AJ, Bambrick J, *et al* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630: 493–500

Allerston CK, Lee SY, Newman JA, Schofield CJ, McHugh PJ & Gileadi O (2015) The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res* 43: 11047–11060

Baddock HT, Yosaatmadja Y, Newman JA, Schofield CJ, Gileadi O & McHugh PJ (2020) The SNM1A DNA repair nuclease. *DNA Repair (Amst)* 95: 102941

Barakat KH, Jordheim LP, Perez-Pineiro R, Wishart D, Dumontet C & Tuszynski JA (2012) Virtual Screening and Biological Evaluation of Inhibitors Targeting the XPA-ERCC1 Interaction. *PLoS One* 7: e51329

Barakat KH, Torin Huzil J, Luchko T, Jordheim L, Dumontet C & Tuszynski J (2009) Characterization of an inhibitory dynamic pharmacophore for the ERCC1-XPA interaction using a combined molecular dynamics and virtual screening approach. *J Mol Graph Model* 28: 113–130

Basu S, Dong Y, Kumar R, Jeter C & Tang DG (2022) Slow-cycling (dormant) cancer cells in therapy resistance, cancer relapse and metastasis. *Semin Cancer Biol* 78: 90–103

Batenburg NL, Cui S, Walker JR, Schellhorn HE & Zhu XD (2021) The winged helix domain of CSB regulates RNAPII occupancy at promoter proximal pause sites. *Int J Mol Sci* 22: 3379

Batenburg NL, Qin J, Walker JR & Zhu XD (2018) Efficient UV repair requires disengagement of the CSB winged helix domain from the CSB ATPase domain. *DNA Repair (Amst)* 68: 58–67

Batenburg NL, Walker JR, Noordermeer SM, Moatti N, Durocher D & Zhu XD (2017) ATM and CDK2 control chromatin remodeler CSB to inhibit RIF1 in DSB repair pathway choice. *Nat Commun* 8: 1921

Beerens N, Hoeijmakers JHJ, Kanaar R, Vermeulen W & Wyman C (2005) The CSB protein actively wraps DNA. *J Biol Chem* 280: 4722–4729

Benedict B, Kristensen SM & Duxin JP (2024) What are the DNA lesions underlying formaldehyde toxicity? *DNA Repair (Amst)* 138: 103667

Berquist BR & Wilson DM (2009) Nucleic Acid Binding Activity of Human Cockayne Syndrome B Protein and Identification of Ca2+ as a Novel Metal Cofactor. *J Mol Biol* 391: 820–832

Boeing S, Williamson L, Encheva V, Gori I, Saunders RE, Instrell R, Aygün O, Rodriguez-Martinez M, Weems JC, Kelly GP, *et al* (2016) Multiomic Analysis of the UV-Induced DNA Damage Response. *Cell Rep* 15: 1597–1610

Carnie CJ, Acampora AC, Bader AS, Erdenebat C, Zhao S, Bitensky E, van den Heuvel D,

Parnas A, Gupta V, D'Alessandro G, *et al* (2024) Transcription-coupled repair of DNA–protein cross-links depends on CSA and CSB. *Nat Cell Biol* 26: 797–810

Cattell E, Sengerova B & McHugh PJ (2010) The SNM1/Pso2 Family of ICL Repair Nucleases: From Yeast to Man. *Environ Mol Mutagen* 51: 635–645

Chakraborty A, Tapryal N, Islam A, Mitra S & Hazra T (2021) Transcription coupled base excision repair in mammalian cells: So little is known and so much to uncover. *DNA Repair (Amst)* 107: 103204

Christiansen M, Thorslund T, Jochimsen B, Bohr VA & Stevnsner T (2005) The Cockayne syndrome group B protein is a functional dimer. *FEBS J* 272: 4306–4314

Citterio E, Van Den Boom V, Schnitzler G, Kanaar R, Bonte E, Kingston RE, Hoeijmakers JHJ & Vermeulen W (2000) ATP-Dependent Chromatin Remodeling by the Cockayne Syndrome B DNA Repair-Transcription-Coupling Factor. *Mol Cell Biol* 20: 7643–7653

Clauson C, Schärer OD & Niedernhofer L (2013) Advances in understanding the complex mechanisms of DNA inter strand cross-link repair. *Cold Spring Harb Perspect Med* 3: 1–25

Enoiu M, Jiricny J & Schärer OD (2012) Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res* 40: 8953–8964

Fan YH & Lake RJ (2013) Structure, function and regulation of CSB: A multi-talented gymnast. *Mech Ageing Dev* 134: 202–211

Genz LR, Mulvaney T, Nair S & Topf M (2023) PICKLUSTER: a protein-interface clustering and analysis plug-in for UCSF ChimeraX. *Bioinformatics* 39: 29–31

Groisman R, Kuraoka I, Chevallier O, Gaye N, Magnaldo T, Tanaka K, Kisselev AF, Harel-Bellan A & Nakatani Y (2006) CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes Dev* 20: 1429–1434

Harami GM, Gyimesi M & Kovács M (2013) From keys to bulldozers: Expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends Biochem Sci* 38: 364–371

Hashimoto S, Anai H & Hanada K (2016) Mechanisms of interstrand DNA crosslink repair and human disorders. *Genes Environ* 38: 1–8

Heuvel D Van Den, Rodríguez-martínez M, Meer PJ Van Der & Moreno NN (2024) STK19 facilitates the clearance of lesion-stalled RNAPII during transcription-coupled DNA repair. 1–66

van den Heuvel D, van der Weegen Y, Boer DEC, Ogi T & Luijsterburg MS (2021) Transcription-Coupled DNA Repair: From Mechanism to Human Disorder. *Trends Cell Biol*

31: 359–371

Housh K, Jha JS, Yang Z, Haldar T, Johnson KM, Yin J, Wang Y & Gates KS (2021) Formation and Repair of an Interstrand DNA Cross-Link Arising from a Common Endogenous Lesion. *J Am Chem Soc* 143: 15344–15357

Iyama T, Lee SY, Berquist BR, Gileadi O, Bohr VA, Seidman MM, McHugh PJ & Wilson DM (2015) CSB interacts with SNM1A and promotes DNA interstrand crosslink processing. *Nucleic Acids Res* 43: 247–258

Iyama T & Wilson DM (2016) Elements That Regulate the DNA Damage Response of Proteins Defective in Cockayne Syndrome. *J Mol Biol* 428: 62–78

Iyama T & Wilson III DM (2013) DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)* 18: 620–636

Kočalka P, El-Sagheer AH & Brown T (2008) Rapid and efficient DNA strand cross-linking by click chemistry. *ChemBioChem* 9: 1280–1285

Kokic G, Wagner FR, Chernev A, Urlaub H & Cramer P (2021) Structural basis of human transcription – DNA repair coupling. *Nature* 598: 368–372

Kokic G, Yakoub G, van den Heuvel D, Wondergem AP, van der Meer PJ, van der Weegen Y, Chernev A, Fianu I, Fokkens TJ, Lorenz S, *et al* (2024) Structural basis for RNA polymerase II ubiquitylation and inactivation in transcription-coupled repair. *Nat Struct Mol Biol* 31: 536–547

Lake RJ, Geyko A, Hemashettar G, Zhao Y & Fan HY (2010) UV-Induced Association of the CSB Remodeling Protein with Chromatin Requires ATP-Dependent Relief of N-Terminal Autorepression. *Mol Cell* 37: 235–246

Liebau RC, Waters C, Ahmed A, Soni RK & Gautier J (2023) Transcription-Coupled Repair of DNA Interstrand Crosslinks by UVSSA. *bioRxiv*

McHugh PJ (2020) XPF–ERCC1: Linchpin of DNA crosslink repair. *PLoS Genet* 16: 1–5

Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH & Ferrin TE (2023) UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* 32: 1–13

Mevissen TET, Kümmecke M, Schmid EW, Farnung L & Walter JC (2024) STK19 positions TFIIH for cell-free transcription-coupled DNA repair. *bioRxiv*

Muftuoglu M, Sharma S, Thorslund T, Stevnsner T, Soerensen MM, Brosh RM & Bohr VA (2006) Cockayne syndrome group B protein has novel strand annealing and exchange activities. *Nucleic Acids Res* 34: 295–304

Muftuoglu M, de Souza-Pinto NC, Dogan A, Aamann M, Stevnsner T, Rybanska I, Kirkali G, Dizdaroglu M & Bohr VA (2009) Cockayne syndrome group B protein stimulates repair of

formamidopyrimidines by NEIL1 DNA glycosylase. *J Biol Chem* 284: 9270–9279

Mulderrig L, Garaycoechea JI, Tuong ZK, Millington CL, Dingler FA, Ferdinand JR, Gaul L, Tadross JA, Arends MJ, O'Rahilly S, *et al* (2021) Aldehyde-driven transcriptional stress triggers an anorexic DNA damage response. *Nature* 600: 158–163

Nakazawa Y, Hara Y, Oka Y, Komine O, van den Heuvel D, Guo C, Daigaku Y, Isono M, He Y, Shimada M, *et al* (2020) Ubiquitination of DNA Damage-Stalled RNAPII Promotes Transcription-Coupled Repair. *Cell* 180: 1228-1244.e24

Nieto Moreno N, Olthof AM & Svejstrup JQ (2023) Transcription-Coupled Nucleotide Excision Repair and the Transcriptional Response to UV-Induced DNA Damage. *Annu Rev Biochem* 92: 81–113

Oka Y, Nakazawa Y, Shimada M & Ogi T (2024) Endogenous aldehyde-induced DNA – protein crosslinks are resolved by transcription-coupled repair. *Nat Cell Biol* 26: 784–796

Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I & Sussman JL (2005) FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21: 3435–3438

Sarker AH, Tsutakawa SE, Kostek S, Ng C, Shin DS, Peris M, Campeau E, Tainer JA, Nogales E & Cooper PK (2005) Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: Insights for transcription-coupled repair and Cockayne syndrome. *Mol Cell* 20: 187–198

Semlow DR & Walter JC (2021) Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair. *Annu Rev Biochem* 90: 107–135

Sengerová B, Allerston CK, Abu M, Lee SY, Hartley J, Kiakos K, Schofield CJ, Hartley JA, Gileadi O & McHugh PJ (2012) Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J Biol Chem* 287: 26254–26267

Sluis M Van, Yu Q, Woude M Van Der, Gonzalo-Hansen C, Dealy SC, Janssens RC, Somsen HB, Ramadhin AR, Dekkers DHW, Wienecke HL, *et al* (2024) Transcription-coupled DNA – protein crosslink repair by CSB and CRL4 CSA -mediated degradation. *Nat Cell Biol* 26: 770–783

Smogorzewska A, Desetty R, Saito TT, Schlabach M, Lach FP, Sowa ME, Clark AB, Kunkel TA, Harper JW, Colaiacovo MP, *et al* (2010) A genetic screen identifies FAN1, a Fanconi anemia associated nucleases necessary for DNA interstrand crosslink repair. *Mol Cell* 39: 36–47

Sollier J, Townsend Stork C, Garcia-Rubio ML, Paulsen RD, Aguilera A & Cimprich KA (2014) Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome

instability. *Mol Cell* 56: 777–785

Swift LP, Lagerholm BC, Baddock HT, Ratnaweera M, Sengerova B, Lee SY, Ulrich HD, Renz C, Waithe D, Schofield CJ, *et al* (2024) SNM1A (DCLRE1A) is required for the efficient repair of complex DNA breaks in human cells. *Nat Commun* 15: 5392

Takahashi TS, Sato Y, Yamagata A, Goto-Ito S, Saijo M & Fukai S (2019) Structural basis of ubiquitin recognition by the winged-helix domain of Cockayne syndrome group B protein. *Nucleic Acids Res* 47: 3784–3794

Tiwari V, Baptiste BA, Okur MN & Bohr VA (2021) Current and emerging roles of Cockayne syndrome group B (CSB) protein. *Nucleic Acids Res* 49: 2418–2434

Tufegdžić Vidaković A, Mitter R, Kelly GP, Neumann M, Harreman M, Rodríguez-Martínez M, Herlihy A, Weems JC, Boeing S, Encheva V, *et al* (2020) Regulation of the RNAPII Pool Is Integral to the DNA Damage Response. *Cell* 180: 1245–1261

Vessoni AT, Guerra CCC, Kajitani GS, Nascimento LLS & Garcia CCM (2020) Cockayne syndrome: The many challenges and approaches to understand a multifaceted disease. *Genet Mol Biol* 43: e20190085

Wang AT, Sengerová B, Cattell E, Inagawa T, Hartley JM, Kiakos K, Burgess-Brown NA, Swift LP, Enzlin JH, Schofield CJ, *et al* (2011) Human SNM1a and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev* 25: 1859–1870

Wong HK, Muftuoglu M, Beck G, Imam SZ, Bohr VA & Wilson DM (2007) Cockayne syndrome B protein stimulates apurinic endonuclease 1 activity and protects against agents that introduce base excision repair intermediates. *Nucleic Acids Res* 35: 4103–4113

Yang K, Moldovan GL & D'Andrea AD (2010) RAD18-dependent recruitment of SNM1A to DNA repair complexes by a ubiquitin-binding zinc finger. *J Biol Chem* 285: 19085–19091