Human CtIP forms a tetrameric dumbbell-shaped particle which binds and bridges complex DNA end structures for double-strand break repair

Wilkinson, O.J.[1], Martín-González, A.[2], Kang, H.J.[3], Northall, S.J.[1], Wigley, D.B.[3], Moreno-Hererro, F.[2] and Dillingham, M.S.[1*]

[1]DNA:Protein Interactions Unit, School of Biochemistry, Life Sciences, University of Bristol, UK. [2]Department of Macromolecular Structures, Centro Nacional de Biotecnologia, Consejo Superior de Investigaciones Cientificas, Madrid, Spain. [3]Section of Structural Biology, Dept. Medicine, Imperial College London, London, UK.

*To whom correspondence should be addressed: mark.dillingham@bristol.ac.uk

Lead contact: Mark Dillingham

Summary

CtIP is involved in the resection of double-stranded DNA breaks during the S and G2 phases of the cell cycle for repair by homologous recombination. Acting in concert with the MRN complex, it plays a particularly important role in handling complex DNA end structures by localised nucleolytic processing of DNA termini in preparation for longer range resection. Here we show that human CtIP is a tetrameric protein adopting a dumbbell architecture in which DNA binding domains are connected by long coiled-coils. The protein complex binds two short DNA duplexes with high affinity and bridges DNA molecules *in trans*. DNA binding is potentiated by dephosphorylation and is not specific for DNA end structures *per se*. However, the affinity for linear DNA molecules is increased if the DNA terminates with more complex structures including forked ssDNA overhangs and model nucleoprotein conjugates. This work provides a biochemical and structural basis for the function of CtIP at complex DNA breaks.

Introduction

DNA double-strand breaks (DSBs) are a potentially lethal form of DNA damage associated with genomic instability, gross chromosomal rearrangements and apoptosis (Ranjha et al., 2018). They are caused by exogenous agents such as ionising radiation (IR) and chemotherapeutic drugs, but also by normal cell metabolism such as during V(D)J/class-switch recombination, Spo11-mediated meiotic recombination, or replication fork collapse. DSBs are repaired by two major pathways: homologous recombination (HR) and non-homologous end-joining (NHEJ). The predominance of either pathway is regulated in a cell cycle-dependent manner owing to the requirement for a sister chromatid to act as a template for error-free repair. HR is promoted in S and G2 phases of the cell cycle by a variety of mechanisms including activation of CtIP (Huertas and Jackson, 2009; Shibata and Jeggo, 2014).

The process of DNA end resection is a critical regulatory step, which commits a cell to repair by HR. During end resection, DSBs are preferentially degraded by nucleases to reveal long 3′-ssDNA overhangs (Cejka, 2015; Daley et al., 2015). These are stabilised by RPA and then bound by Rad51, which forms nucleoprotein filaments that undergo strand invasion with the sister chromatid, eventually leading to DNA repair without loss of genetic information (Kowalczykowski, 2015). DNA end resection can be broadly separated into three mechanistic stages: initial recognition of the DSB, short-range resection and long-range resection. Recognition of the DSB is achieved by binding of the Mre11-Rad50-Nbs1 complex (MRN) (Paull and Gellert, 1999) which activates the DNA damage response through its interaction with ATM, and subsequently recruits other repair factors to the damage site including CtIP, which is itself activated by cyclin-dependent kinase in an Mre11-mediated process (Buis et al., 2012; Huertas and Jackson, 2009). Human CtIP, which is mutated in Seckel2 and Jawad syndromes (Qvist et al., 2011), has been shown to be crucial for the promotion of end resection and HR-mediated repair in many genetic studies (Huertas and Jackson, 2009; Sartori et al., 2007; You et al., 2009), but exactly how MRN and CtIP co-operate to achieve short-range resection is not fully understood. MRN harbours both 5′-3′ endonuclease and 3′-5′ exonuclease activities in the Mre11 subunit, and these activities are mediated by CtIP (Cannavo and Cejka, 2014; Paull and Gellert, 1999). Some reports suggest that CtIP and its orthologues possess intrinsic endonuclease activity (Lengsfeld et al., 2007; Makharashvili et al., 2014; Wang et al., 2014), but the primary structure of the protein is not obviously related to any known class of nuclease, and other groups have reported that their preparations are devoid of nuclease activity

(Andres and Williams, 2017). Whatever the case, the combined activities of MRN and CtIP are especially important for processing complex DNA break structures, where the ends are composed of damaged nucleotides, non-canonical DNA secondary structures, or covalent nucleoprotein complexes (Andres and Williams, 2017; Aparicio et al., 2016; Aparicio and Gautier, 2016; Hartsuiker et al., 2009a; Hartsuiker et al., 2009b; Paudyal et al., 2017; Quennet et al., 2011). Short range nucleolytic processing of these ends by MRN-CtIP facilitates further long-range resection by processive helicases and nucleases including Exo1, DNA2, BLM, and WRN (Cejka, 2015; Daley et al., 2017).

In vitro experiments with recombinant human proteins show that Mre11 can cut the 5′-strand of a DNA duplex downstream of a break. This activity is dependent on the presence of Nbs1 and enhanced by a nucleoprotein block at the 5′-terminus (Anand et al., 2016; Reginato et al., 2017; Wang et al., 2017). In Xenopus egg cell-free extracts, it was shown that CtIP-MRN is essential for the removal of Top2-DNA adducts and the subsequent resection of these breaks (Aparicio et al., 2016). Furthermore, the interaction between CtIP and BRCA1 (mediated by phosphorylation of S327 on CtIP) is required for the processing of Top2-adducted 'complex' DSBs but dispensable for the resection of endonuclease-generated 'simple' breaks (Aparicio et al., 2016). This is reminiscent of the finding that the CtIP mutant N289A/H290A cannot rescue the end resection-deficient phenotype for topoisomerase-induced breaks but is proficient for simple breaks, although here it was postulated that this deficiency was based on a disruption of CtIP endonuclease activity (Makharashvili et al., 2014). In any case, once a nick is made, Mre11 uses its 3′-5′ exonuclease function to strip back the 5′ strand towards the blocked end. In these experiments, wild type CtIP did not enhance the endonuclease activity of MRN, although a stimulation of activity was observed when a phosphomimic mutant (T847E/T859E) was used (Deshpande et al., 2016). In mammalian cells, it was shown that both phosphorylation sites are important for CtIP function. In contrast, studies using recombinant *S.cerevisiae* proteins show that the CtIP orthologue Sae2 appears to be absolutely required for unlocking a cryptic endonuclease activity of Mre11, possibly highlighting a difference in regulation between the yeast and human systems (Cannavo and Cejka, 2014).

Structural information for CtIP is limited to a small N-terminal region of the *S.pombe* homologue Ctp1 (amino acids 5-60) and a similar region of human CtIP (amino acids 18-52) (Andres et al., 2015; Davies et al., 2015). In both cases, there is an interlocking arrangement of two antiparallel coiled coils which leads to a 'dimer of dimers' arrangement. In addition, it

4

has been shown that a truncated form of the human protein comprising residues 18-145 forms a stable homotetramer *in vitro* (Davies et al., 2015). A mutation (L27E) that prevented tetramerization also rendered CtIP non-functional *in vivo*. The proposed model that arises from these reports is one where two rigid coiled coils (aa1-145) protrude in opposite directions from the tetramerisation domain, spanning a distance of about 30 nm, with the unstructured C-termini presumably placed at the opposing ends (Forment et al., 2015). It has been shown that CtIP binds to DNA; for S. *pombe* Ctp1 a number of DNA substrates have been interrogated by EMSA and were all shown to be bound with similar affinities (Andres et al., 2015). This study also used scanning alanine mutagenesis to identify residues that are important for binding. A C-terminal 'RHR' motif (equivalent to conserved residues R837 to R839 in the human protein) was shown to be crucial for DNA recognition, but it was also shown that an N-terminal part of the protein binds DNA. For human CtIP, it has been shown that the C-terminal region (aa769-897) interacts with a 200 bp duplex, whereas the N-terminal region (aa1-145) does not (Davies et al., 2015). In the case of the full length human protein, DNA binding activity has only been demonstrated using gel-based crosslinking assays which gives limited information about equilibrium binding affinities (Makharashvili et al., 2014). An important outstanding question is whether the affinity of CtIP for different DNA end structures can explain its relative importance in the processing of "complex" versus "simple" ends *in vivo*. Moreover, since the DNA binding domains are thought to reside in the opposing C-termini, this may be a way in which a CtIP tetramer could bridge two distant broken DNA ends to promote DSB repair.

In this report, we show that full length human CtIP is a tetrameric protein that forms a dumbbell-shaped particle consisting of two polar globular domains separated by a thin and flexible 'rod'. Semi-quantitative measurements of CtIP affinity for a range of DNA substrates show that, somewhat surprisingly, CtIP binding does not require DNA ends *per se*. However, CtIP binds much more effectively to structures with DNA ends that contain model nucleoprotein blocks or single-stranded DNA fork structures. Our results are rationalised in terms of a "slide and capture" model for the recognition of complex DNA ends. Furthermore, we provide direct evidence for CtIP-dependent DNA bridging activity and show that both DNA binding and bridging are reduced by mutations which target the tetramerization interface or the C-terminal RHR motif.

Methods

Protein expression and purification

A synthetic gene codon-optimised for *S. frugiperda* (Geneart, Invitrogen) encoding wild type human CtIP was cloned into the pACEBac1 vector using the BamHI and XbaI restriction sites for use in the MultiBac system (Geneva Biotech). This was then screened for expression and purification using affinity tags in different positions. Following cleavage of a construct with a C-terminal 3C-cleavable StrepII tag, we obtained CtIP in good yield and purity. The full length recombinant protein as prepared (i.e. post-3C cleavage) contains N- and C-terminal extensions: MELL- and -SGLEVLFQ respectively (MW = 103171 Da monomer), with the rest of the protein identical to UniProt entry Q99708. A contaminant band of approximately 75 kDa is also routinely present following purification, and this was identified as a CtIP degradation product by mass spectrometry (data not shown). Mutagenesis (Quikchange XL, Agilent) was performed using this construct to create the L27E and R839A mutants. Bacmids were prepared by transposition of these plasmids and were used to transfect Sf9 insect cells in Insect Express media (Lonza) before viral amplification in the same cell line using standard techniques. For large scale expression, 500 mL of Hi5 cells at density 2×10^6 /mL were infected with 25 mL of P3 virus and harvested by centrifugation after 70 hours at 27°C with shaking. The pellets were lysed into buffer containing 50 mM Tris pH8.0, 500 mM NaCl, 2 mM DTT, 10% glycerol, 1 mM Na_3VO_4, 20 mM β-glycerophosphate, protease inhibitor cocktail (Roche) and then sonicated on ice for a total of 3 minutes. After centrifugation at 4°C for 30 mins at 50000g, the cleared lysate was applied to Streptactin beads (GE Healthcare) in batch and incubated for 1h at 4°C with rotation. After washing five times in batch with buffer containing 20 mM Tris HCl pH8.0, 500 mM NaCl, 5% glycerol, 1 mM DTT, the protein was then eluted in the same buffer containing 2.5 mM desthiobiotin. The CtIP-containing fractions were diluted to approximately 125 mM NaCl and applied to a 5 mL Heparin column (GE Healthcare). After washing, CtIP was eluted with a gradient from 150 mM to 1 M NaCl over 20 CV in buffer containing 20 mM Tris HCl pH8.0, 5% glycerol, 1 mM DTT. The CtIP-containing fractions were pooled and digested overnight at 4°C with 40 units of 3C protease to remove the StrepII tag. The cleavage reaction was run over a 5 mL Streptactin column (Qiagen) to remove any uncleaved CtIP and free StrepII peptide and the cleaved CtIP-containing flow-through collected. This was diluted to give a NaCl concentration of approximately 125 mM and then loaded onto a 1mL MonoQ column (GE Healthcare). CtIP was eluted with a gradient from 125 mM to 500 mM NaCl over 20 CV in buffer containing 20 mM Tris HCl pH8.0, 5% glycerol, 1 mM DTT. For some

6

preparations, the most concentrated 1 mL from the MonoQ elution was applied to a Superdex 200 column in buffer containing 20 mM Tris HCl pH8.0, 200 mM NaCl, 5% glycerol, 1 mM DTT, followed by concentration of the CtIP peak using a centrifugal filter unit (Millipore) and storage at -80°C. Protein concentration was determined using a theoretical extinction coefficient of 37360 M^{-1} (monomer) cm^{-1}. All CtIP concentrations are stated as tetramer (CtIP$_4$). Analysis of the purified protein by Orbitrap LC-MS/MS spectrometry was performed by the University of Bristol Mass Spectrometry Facility.

Preparation of DNA substrates for binding assays

HPLC-purified single-stranded oligonucleotides (Eurofins, ATD Bio) were used as supplied. For dsDNA substrates, oligonucleotides were annealed by heating at 50 μM in 50 mM Tris pH7.5, 150 mM NaCl, 1 mM EDTA for 10 mins at 95°C and allowed to cool slowly overnight. For streptavidin-blocked DNA substrates, a 6 μM biotinylated DNA solution was blocked with 50 μM streptavidin (8.5-fold molar excess) to promote a 1:1 stoichiometry of biotin:streptavidin. For substrates with two biotinylated ends, a 6 μM DNA solution was blocked with 100 μM streptavidin (8.5-fold molar excess). The blocking reactions were incubated for 5 mins at 25°C before performing titrations. The experiments examining the effect of ends on CtIP binding used: φX174 Virion (circular ssDNA), φX174 Virion cut with HaeIII (15 fragments of ssDNA), pSP73JY10 (supercoiled dsDNA, (Yeeles et al., 2011)), pSP73JY10 cut with BamHI (linear dsDNA) and pSP73JY10 cut with Hin1II (45 fragments of dsDNA).

Size exclusion chromatography coupled to multiple angle light scattering (SEC-MALS)

SEC-MALS was used to determine the absolute molecular masses of full-length CtIP WT and mutants. Approximately 50 μg samples of CtIP were loaded at 0.5 ml/min onto a Superose 6 10/300 size-exclusion chromatography column (GE Healthcare) in 20 mM Tris, pH 8.0, 200 mM NaCl, 1 mM TCEP using an Agilent HPLC. The eluate from the column was coupled to a DAWN HELEOS II MALS detector (Wyatt Technology) and an Optilab T-rEX differential refractometer (Wyatt Technology). ASTRA 6 software (Wyatt Technology) was used to collect and analyse light scattering and differential refractive index data according to the

manufacturer's instructions. Molecular masses and estimated errors were calculated across individual eluted peaks.

Atomic Force Microscopy

Sample preparation: The forks for the ends of the DNA substrate were assembled by annealing two complementary oligonucleotides (Eurofins) at 50 µM in 50 mM Tris pH7.5, 150 mM NaCl, 1 mM EDTA. The forks were designed to have two dT_{20} tails to comprise the forked region, a 25bp duplex region and a 4-nucleotide overhang compatible with specific restriction endonuclease-generated sticky ends. The 'front' fork had a 3′-overhang (SphI) and the 'end' fork a 5′-overhang (HindIII). Both forks also had a 5′-phosphate group at one end for downstream ligation. 100 µg of plasmid pSP73 JY10 (Yeeles et al., 2011) was cut with SphI and HindII (NEB) and the 500 bp band was gel purified (1xTAE 1% agarose). The two forks (10 µM each) and 500bp fragment (65 nM) were then ligated overnight at 16°C with T4 DNA ligase. The product ran at a position equivalent to a duplex of ~600 bp, and was gel purified (1xTAE 1% agarose) with approximately 1 µg being recovered.

Imaging: For imaging of protein complexes, the stock protein solution was first diluted in storage buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM DTT) to a final concentration of 4 nM. Then, the protein solution was deposited onto a freshly cleaved mica surface. After 30 sec, the mica surface was washed with 3 ml Milli-Q water (Millipore, Billerica, MA) and gently dried under nitrogen air flow. DNA bridging reactions were performed with 10 nM DNA and increasing concentrations of $CtIP_4$ (0-250 nM) for different $DNA:CtIP_4$ ratios (1:0, 1:6, 1:13, 1:19, 1:25) in buffer A (100 mM HEPES 7.5, 100 mM NaCl). The binding reaction mixture was incubated for 15 min in a 2 µl reaction volume. Then, the mica was pretreated with a solution of 7.5 mM $MgCl_2$ and washed with 3 ml MilliQ water. A droplet of 18 µl of buffer B (4 mM HEPES 7.5, 10 mM NaCl) were added onto the pretreated mica and the 2 µl of binding reaction added to the droplet. Therefore the final solution deposited on the mica was 10 times less concentrated than in the binding reaction (Fuentes-Perez et al., 2012). The mica samples were incubated for 30 sec, washed with 3 ml MilliQ water, and dried under nitrogen air flow. Samples were imaged in air at room temperature with Point-Probe-Plus tips (PPP-NCH, 42 N/m, 330 kHz, Nanosensors, Neuchâtel, Switzerland) in amplitude modulation mode

AFM (Nanotec Electrónica, Madrid, Spain). Image flattening and in-plane substraction were performed with WSxM freeware (Horcas et al., 2007).

DNA length analysis: AFM images of DNA molecules in the presence or absence of CtIP protein have been analyzed using the WSxM freeware and Origin Pro 8. A profile along the DNA molecules present in 9-10 images from different samples (typically 250 molecules) was used to obtain the value of the contour length for the analysis. The data from DNA molecules are classified depending on the concentration ratio of DNA with respect to CtIP tetramer and analyzed by histogram representation including data from all the molecules. The histograms were fitted to a Gaussian distribution and the mean contour length and error were extracted from the Gaussian parameters.

Protein volume characterization with the atomic force microscope: Direct calculation of volumes by AFM images is affected by a tip convolution effect that may distort the shape and size of the molecules imaged. Correction to the volume measured by the AFM was performed by considering the volume of a fiducial marker, namely a piece of DNA, and following the procedure described in (Fuentes-Perez et al., 2012). Images of 1000x1000 nm^2 at 512x512 pixels were selected from CtIP-DNA samples. The volume of CtIP was calculated from the heights and area of variable size windows depending on the protein size and subsequent subtraction of a basal volume of a similar blank window near the CtIP window. The volume of the fiducial DNA was calculated from a fixed 20x20 nm^2 window. The normalized CtIP/DNA values ("relative volume") were calculated for each protein species and displayed as histograms. Proteins were classified by the overall shape of the molecules in 5 different classes.

Negative stain electron microscopy and image analysis

Purified CtIP was diluted to approximately 0.15 mg/ml in 20 mM Tris HCl pH8.0, 1 mM DTT, adsorbed onto glow-discharged carbon-coated grids, and negatively stained with 1 % uranyl acetate. The grids were observed using a FEI T12 Spirit electron microscope operating at 120 kV and the images were recorded using FEI 2K eagle camera at a magnification of 52,000. Image processing was performed using RELION 2.1 (Scheres, 2012). Initially 15,236 particles were picked from 96 micrographs and reference-free 2D classification was used to remove poor particles, resulting in 6,924 particles with approximately a 35/65 ratio divided between the two classes α and β.

Electrophoretic mobility shift assays

5′-Cy5-labelled DNA substrates (2.5 nM final) were mixed with increasing amounts of CtIP protein (up to 250 nM for WT and up to 1 µM for mutants L27E and R839A, as indicated) in a total volume of 10 µL 1X EMSA buffer (20 mM Tris HCl pH8.0, 100 mM NaCl, 1 mM DTT, 0.1 mg/mL BSA, 5% glycerol) and then incubated for 10min at 25˚C. The samples were then loaded onto a 6% polyacrylamide (29:1) native 1xTBE gel and separated by electrophoresis in 1xTBE at 150V for 40 mins. The gels were visualised using a Typhoon scanner and analysed using ImageQuant software.

Fluorescence anisotropy

Fluorescence anisotropy measurements were made at 25˚C on a Horiba Jobin Yvon FluoroMax fluorimeter. The HEX dye was excited at wavelength 530 nm (excitation slit width 5 nm), and the emission was detected at 550 nm (emission slit width 5nm). For the direct titrations, a 5 nM solution of DNA substrate in binding buffer (20 mM Tris–HCl (pH 8.0), 20 mM NaCl, 1 mM DTT) was used to which was added increasing amounts of concentrated CtIP protein. Note that the stated CtIP concentrations are as a tetramer. The contents of the cuvette were mixed thoroughly with a pipette and left for 1 minute, which was sufficient to reach equilibrium, before each anisotropy reading was taken in duplicate and the numbers averaged. Fluorescence intensity changes were negligible (<10% at saturation). The titrations were carried out three times independently to generate the error bars shown. The competition assays were carried out in the same way but with the pre-formation of the CtIP-Forked DNA complex at 5 nM DNA: 50 nM CtIP$_4$ (yielding about 75% bound DNA in the absence of competitor) followed by the addition of increasing amounts of unlabelled competitor DNA.

Anisotropy is given by equation 1:

$$A = \frac{(I_{VV} - I_{VH})}{(I_{VV} + 2I_{VH})}$$

where I_{VV} and I_{VH} are the intensities of the vertical and horizontal components of the emitted light using vertical polarised excitation. Anisotropy values were normalised (to percent of

maximum change) between repeats to account for small differences in the initial and final absolute readings according to equation 2:

$$A(\%) = 100 \left(1 - \frac{(A_{\max} - A)}{(A_{\max} - A_{\min})} \right)$$

The error bars shown are calculated from normalised values and represent the standard error of the mean calculated from 3 independent titrations. For the direct titrations of labelled DNA with CtIP, the binding isotherms were generated by plotting protein concentration against normalised anisotropy and were fit using GraphPad Prism to equation 3:

$$B_{obs} = \left(\frac{B_{max} \times [CtIP]}{K_d + [CtIP]} \right) + C$$

where B_{obs} is the observed binding, B_{max} is the maximal binding signal at saturation, C is an offset value and K_d is the equilibrium dissociation constant.

For competition assays, the binding data were normalised and then fit to a hyperbolic decay function to obtain an IC_{50} value, with which to compare the efficacy of different competitors using equation 4:

$$B_{obs} = \left(\frac{C \times [DNA_{comp}]}{IC_{50} + [DNA_{comp}]} \right) + B_{ini}$$

Where $[DNA_{comp}]$ is the concentration of the competitor unlabeled DNA (the dependent variable), B_{ini} is the normalized starting anisotropy, C is a scaling value equivalent to total loss of binding signal, and IC_{50} is the total concentration of competitor DNA required to half maximally inhibit binding of the labelled DNA by CtIP. Note that, under the conditions used here, the relationship between IC_{50} and K_{comp} (the equilibrium dissociation constant for CtIP interaction with the unlabeled competitor) is complex (Huang, 2003). For weak binding competitors (relative to the probe DNA), the measured IC_{50} value is proportional to K_{comp} whereas for tight binding inhibitors the measured IC_{50} is limited by the affinity of CtIP for the probe DNA.

For experiments with the dephosphorylated CtIP protein where the interaction with DNA was under tight binding conditions, the stoichiometry and affinity of interaction between CtIP and DNA was determined using equation 5:

11

$$B_{obs} = \left(B_{max} \times \frac{\left(([CtIP] \times K_d \times n[DNA]) - \sqrt{([CtIP] \times K_d \times n[DNA])^2 - (4 \times [CtIP] \times n[DNA])} \right)}{2n[DNA]} \right) + B_{ini}$$

where B_{obs} is the observed binding (fluorescence anisotropy), B_{ini} is the starting anisotropy, B_{max} is the maximal anisotropy signal at saturation, K_d is the equilibrium dissociation constant and n is the number of CtIP tetramers binding to each DNA molecule. In these fits, the K_d value was held constant at the measured value of 1.5 nM, and the value of n was shared in a global fit of all three data sets collected at different values of [DNA].

Nuclease assays

Nuclease assays were carried out in 10 µL volumes and were assembled on ice. $CtIP_4$ was titrated against 5 nM Cy5 labelled DNA substrate at 16, 31, 63, 125 and 250 nM in 10 mM Tris-HCl pH 8.0, 1 mM DTT, 250 µg/mL BSA and 0.5 mM $MnCl_2$. Reactions were incubated at 37^OC for 120 minutes and were stopped by the addition of 10 µL 4 mg/mL proteinase K and 1% w/v SDS. Reaction products were separated on 16% w/v acrylamide TBE gels at 200 V for 45 minutes. Gels were imaged using a Typhoon FLA 9500 scanner (GE Healthcare Life Sciences) and were quantified using ImageQuantLE software (GE Healthcare Life Sciences). To assess divalent cation preference, 1 µM CtIP was incubated with 5 nM Cy5 labelled DNA substrate in 10 mM Tris-HCl pH 8.0, 1 mM TCEP, 250 µg/mL BSA and 0.5 mM divalent cation (as indicated) and were then treated as before.

Results

Full length human CtIP is a tetrameric protein that forms a dumbbell-shaped particle

Full length recombinant human CtIP and mutant variants were overexpressed in insect cells and purified to homogeneity using a cleavable C-terminal strep-tag (**Figure 1**; see Materials and Methods for details). Analysis by SEC-MALS showed that wild type CtIP ran as a single symmetrical peak with a predicted molecular weight that was equal within error to a tetrameric species (391 ±23 kDa; **Figure 1A**), as we had expected based on previous work with the isolated N-terminal region of the protein (Andres et al., 2015; Davies et al., 2015). However, the CtIP peak elutes earlier in the size exclusion column than would be expected for a ~400 kDa protein, which suggests that it adopts an unusual shape. Indeed, direct observation of purified CtIP using negative stain electron microscopy revealed an extraordinary dumbbell-like structure in which polar globular domains are held about 30 nm apart by a central "rod" (**Figure 1B,** β class; **Supplementary Figure 1**). The length of the rod is broadly consistent with the predicted length of the coiled-coil domains, which suggests that the centre of the dumbbell is the site of the tetramerization interface as proposed in a previous model (Forment et al., 2015). In addition to the dumbbell structures, we less frequently observed structures that we call "tadpoles" containing a single globular domain and a tail, which presumably represent a dimeric form of the protein formed during EM sample preparation and deposition (**Figure 1B**, α class; **Supplementary Figure 1**).

Broadly similar structures were observed by atomic force microscopy, although the sample appeared more heterogenous when applied to a mica surface, and five classes of particles were distinguished based on their volume and morphology (**Figure 1C**). The simplest and smallest structures consisted of a single bright spot or a single spot with a tail (class I and II). Somewhat larger structures were observed in which two or three spots are connected by rods (class III and IV respectively). Finally, the largest class of particle contained up to four connected spots (class V). These different particles were associated with different volume profiles, with the largest particle class (class V) displaying a relative volume consistent with a tetrameric species (~380 kDa; **Figure 1D**) (Fuentes-Perez et al., 2012). These either resembled the dumbbells we had observed by EM (β class particles) or showed a morphology that suggested that the polar globular domains had separated ("splayed dumbbells"; see **Supplementary Figure 2**). Other particle classes presumably arise from dissociation of the tetramer into smaller species, either upon dilution for the AFM analysis or because of deposition onto the mica surface. To aid

13

interpretation of the different particle classes and for further biochemical analysis (see below) we also analysed two mutant proteins (**Figure 2**). The putative DNA binding mutant CtIP^{R839A} (Andres et al., 2015), which disrupts the C-terminal RHR motif, retained a tetrameric structure based on SEC-MALS analysis (**Figure 2A**) and produced very similar results to the wild type protein in the AFM (**Figures 2B and 2D**). In contrast, the CtIP^{L27E} mutant, which disrupts the N-terminal tetramerization interface (Davies et al., 2015), ran as an apparently smaller particle in SEC-MALS with predicted molecular weights across the peak ranging from ~270 to 190 kDa (most similar to a dimer molecular weight). Importantly, when imaged using AFM, the CtIP^{L27E} preparation was completely devoid of the largest class of particle (Class V; **Figures 2C and 2E**) which, in combination with the SEC-MALS results, suggests again that the class V particle represents tetrameric CtIP. Possible structural interpretations of the other CtIP particle classes we have observed using AFM are shown alongside further examples in **Supplementary Figure 2**.

For all particle classes, the distance between the highest points of the connected dots was on the order of 30 nm, although significant variability was observed consistent with flexibility in the connecting rod and/or the central region of the protein which is predicted to be disordered (see **Supplementary Figure 3** for examples and discussion).

The CtIP tetramer binds tightly to DNA in a manner dependent on both the N-terminal tetramerization and C-terminal DNA binding domains.

We next used electrophoretic mobility shift assays (EMSA) to compare the binding of CtIP to different DNA substrates (**Figure 3**). CtIP bound tightly to a forked DNA molecule containing a 25 bp duplex region flanked by 20 base ssDNA tails (**Figure 3A**). The binding to a completely single-stranded DNA substrate was weaker and resulted in the formation of more poorly defined complexes in the gel. Binding to a completely duplex DNA substrate was considerably weaker than to a forked substrate. To facilitate a more quantitative approach, we also developed a fluorescence anisotropy assay using HEX-labelled DNA substrates. Firstly, we analysed directly the binding of CtIP to a HEX-labelled DNA fork using anisotropy (**Figure 3D**). The binding isotherm was well fit to a hyperbola yielding a value for the equilibrium dissociation constant (K_d) of 16 nM, which is somewhat tighter than suggested by using the semi-quantitative EMSA technique with the same substrate. Comparative analysis of the mutant

14

proteins CtIP^{R839A} and CtIP^{L27E} using both assays showed that they were severely defective for DNA binding (K_d = 160 nM and 230 nM respectively) (**Figures 3B-D**).

Dephosphorylation potentiates the binding of the CtIP tetramer to DNA, allowing determination of the binding stoichiometry

Phosphorylation is crucial for activation of CtIP during the S and G2 phases of the cell cycle. Somewhat surprisingly however, CtIP treated with phosphatases has shown evidence for an *increased* DNA binding affinity although this effect was not extensively investigated (Anand et al., 2016). This implies a possible regulation of DNA binding by phosphorylation at one or more undefined sites. Analysis by mass spectrometry showed that our standard CtIP preparation is hyper-phosphorylated (see **Supplementary Figure 4** for details). Therefore, to test how phosphorylation status affects DNA binding quantitatively, we next prepared dephosphorylated CtIP (CtIP^λ) by treatment with λ phosphatase (**Figure 4**). CtIP^λ retained a tetrameric state (**Figure 1A**) and bound to forked DNA significantly more tightly than did untreated CtIP as judged crudely by EMSA (**Figure 4A**). Furthermore, it displayed >10-fold tighter affinity for forked DNA based on fluorescence anisotropy measurements (**Figure 4B**). The best fit to these data using the tight-binding equation (Equation 5) gave K_d = 1.5 nM and stoichiometry (n) = 0.58 CtIP$_4$ per DNA fork. Note that this K_d value should be regarded as an upper limit, because the binding is too tight to measure accurately at the lowest probe concentration we used here (5 nM). The enhanced affinity observed for dephosphorylated CtIP is not the result of an artefactual change in solution conditions imposed by the λ phosphatase or its reaction buffer as the increase in binding affinity can be time-resolved, occurring slowly after the addition of the phosphatase (**Supplementary Figure 5**). The molecular basis for this increased affinity presumably relates to the phosphorylation status of one or more amino acids in CtIP (see Discussion section). Although this very tight binding precluded an accurate determination of K_d, it instead allowed us to rigorously investigate the stoichiometry of the CtIP-DNA interaction by performing anisotropy assays with different DNA concentrations, all of which were significantly above the value of the dissociation constant (**Figure 4C**). Under such conditions, the binding isotherms are approximately linear until saturation for a range of DNA concentrations as would be expected. The binding stoichiometry was determined as 0.56 CtIP$_4$ per DNA fork molecule by global fitting of the data to the tight-binding equation

15

(Equation 5). This is equivalent to two DNA molecules bound to each CtIP tetramer at saturation (assuming that our CtIP preparation is 100% active).

CtIP binds preferentially to DNA substrates with blocked DNA ends

Several lines of evidence have suggested that CtIP function is especially important for the recognition and repair of DNA breaks containing complex DNA end structures. To explore the biochemical basis for this phenomenon we tested whether the affinity of CtIP for different DNA substrates depended on the structure of the DNA ends. To compare the relative affinity of a wider range of DNA substrates, including those with complex DNA end structures such as DNA conjugated to other proteins, we performed competition unbinding assays using fluorescence anisotropy (**Figure 5 and Supplementary Figures 6 and 7;** further details for the structures of the DNA molecules used and their construction can be found in the **Supplementary Information**). CtIP was equilibrated with the HEX-fork DNA at a concentration equivalent to 3 x K_d yielding a high starting anisotropy signal. We then titrated increasing concentrations of unlabelled competitor DNA molecules and monitored the steady-state anisotropy (**Figure 5A**). For each competitor DNA tested, the loss of anisotropy was well fit to a hyperbolic unbinding curve (Equation 4) yielding an IC_{50} value; the concentration of competitor at which the binding signal is reduced by 50%. This is consistent with a simple competition between the labelled probe DNA and the competitor DNA for the CtIP DNA binding loci. Note that the measured IC_{50} value is limited at low values by the K_d of the interaction between CtIP and the reference DNA (Huang, 2003). Therefore, this method is most useful for comparing the relative affinities of DNA substrates that bind weaker to CtIP than does the reference (HEX-fork) probe, as is generally the case here. Accordingly, the difference in affinity between two competitor DNAs is larger than the difference implied by their IC_{50} values, especially if the values approach the K_d for the probe.

In our first competition experiment, we found that titration of the unlabelled fork DNA against the reference probe yielded an IC_{50} value of 55 nM (**Figure 5B**). In comparison, fully single- and double-stranded DNA substrates were relatively poor competitors (IC_{50} values of 314 nM and 392 nM respectively) in qualitative agreement with the EMSA analysis which had also shown a clear preference for forked DNA (**Figure 5C**). Interestingly, this effect required the presence of both the 3′- and 5′- ssDNA overhangs in the fork (i.e. a Y-junction) as neither polarity overhang alone significantly improved the observed binding relative to duplex (**Figure**

16

5C). We next titrated a variety of much larger competitor DNA molecules which differed in terms of their average length, topology, and whether they were single- or double-stranded (**Figure 5D**). In terms of the total nucleotide concentration, the IC_{50} value was broadly similar for single- and double-stranded DNA. Moreover, the IC_{50} was largely unaffected by whether the DNA was circular, linearised or fragmented into many pieces. This demonstrates that CtIP does not bind preferentially to DNA ends compared to internal sites because the efficacy of the competitor is independent of the total concentration of free DNA ends. Further experiments with a range of short single-stranded oligonucleotides showed that binding was length dependent (in terms of absolute molecule concentrations) and very poor on substrates shorter than around 30 nucleotides (**Supplementary Figure 7**).

Next, to investigate how modified "complex" DNA ends might affect the affinity of CtIP for DNA, we end-labelled different DNA competitor molecules with a model nucleoprotein block based on the biotin:streptavidin interaction (**Figure 5E**). The addition of biotin labels alone to the termini of the DNA competitor molecules had no significant effect or slightly weakened the IC_{50} values. However, the additional binding of streptavidin to the terminal biotin moieties dramatically increased the efficacy of the DNA molecules as competitors. For example, a fully duplex 45 base pair DNA molecule displayed an IC_{50} value of 392 nM and biotinylation of either one, or both of the ends, resulted in values of 413 and 448 nM respectively. However, the IC_{50} value reduced to 146 nM with a single biotin:streptavidin block and further to 80 nM with a biotin:streptavidin block at either end (compare the red bars in **Figure 5E**). Moreover, similar effects were observed on both single-stranded and forked DNA substrates (blue and green bars respectively; **Figure 5E**). The enhanced binding afforded by streptavidin blocking was apparent regardless of whether the biotin moiety was placed at the 5′ or 3′ end of DNA strands (**Supplementary Figure 7**). In these competition experiments, streptavidin was added at a large excess to prevent multimerization of the competitor DNA. Importantly, the addition of streptavidin to either single- or double-stranded competitor DNAs that lacked biotin moieties had no effect on the observed IC_{50} value (**Figure 5F**). This shows that the observed effect is dependent on binding of a streptavidin molecule to the biotinylated DNA end, as opposed to a non-specific effect of free streptavidin. Interestingly, we also found that when a biotin was placed in the middle of a duplex DNA molecule, the IC_{50} value was *weaker* in the presence of streptavidin (**Supplementary Figure 7**). This could indicate either that the streptavidin is blocking the association of CtIP with internal sites and/or that the positive effect of streptavidin binding on CtIP affinity requires it to be positioned at a DNA end.

CtIP bridges DNA *in vitro*

We next used atomic force microscopy to directly image CtIP:DNA interactions and potential bridging interactions. We engineered a DNA substrate containing short ssDNA tails at both ends because EMSA assays had shown that this was a favourable structure for binding (see Methods for details). Initially we used AFM to image the DNA substrate alone (**Figure 6A**). The DNA was homogenous, consisting virtually entirely of molecules with a single contour length as would be expected (i.e. the DNA substrate was monomeric). The length of the substrate DNA with single-strand overhangs is 587 bp, which we expect to be ~200 nm long based on a value of 0.34 nm/bp for B-DNA. In good agreement with this estimate, the mean contour length measured for the control bare DNA substrate was 195±22 nm. We next repeated this imaging experiment, but in the presence of increasing concentrations of CtIP. In the presence of CtIP, we observed free protein on the mica surface alongside the DNA (**Figure 6B**). However, we also observed many instances where the protein was apparently interacting with the DNA. Furthermore, we saw an increase in the number of DNA molecules with contour lengths equal to 2x, 3x or even 4x the monomer length (**Figure 6B**). This is illustrated most clearly in **Figure 6C**, which shows the cumulative frequency of DNA molecules as a function of DNA length for increasing doses of CtIP. Considering a threshold length of 195+22 nm (mean control length + SD), the percentage of molecules above this threshold is 4% in the control experiment rising to about 30% in samples containing the highest concentrations of CtIP. Note that this quantification method does not score for intramolecular bridging events as these would result in single contour length products, although these were observed in our experiments (data not shown). As would be expected based on our observation that the binding specificity is not strictly for ends, we observed a range of intermolecular connections between two DNA duplexes including end-to-end, end-to-centre and centre-to-centre (**Figure 6B**, zoomed-in panels). In the presence of CtIP, a small proportion of the DNA substrate was less than a single contour length, which could be explained by the presence of a weak nuclease activity in the preparation (see **Supplementary Discussion 1 and Supplementary Figure 8** for further discussion and analysis of this activity). We next compared the bridging ability of dephosphorylated CtIP and the two mutant proteins L27E and R839A, as well as performing an additional negative control with BSA in place of CtIP (**Figure 6D; Supplementary Figure 9**). In comparison to wild type, dephosphorylated CtIP was marginally more effective at bridging DNA molecules. The total percentage of bridged molecules was nearly identical, but

18

there was a larger proportion of molecules displaying greater than two or three DNA contour lengths. In contrast, both mutant proteins were severely defective in DNA bridging as would be expected based on their weak DNA binding affinities. Although, both mutant proteins displayed a somewhat larger proportion of bridged DNA molecules than a "no protein" control, they both displayed similar bridging activity to a BSA control which may suggest that their low level of activity mainly reflects a crowding effect as opposed to *bona fide* DNA bridging activity.

Bridging experiments performed with a substrate that lacked the terminal ssDNA forks showed a reduced level of bridging, suggesting that the tighter DNA binding afforded by modifications to the DNA ends also facilitates DNA bridging as would be expected (data not shown). We also attempted to measure DNA bridging by exploiting a pulldown assay that has been used to study intermolecular DNA bridging by the *S. pombe* orthologue Ctp1 (Andres et al., 2015). However, we found that human CtIP bound non-specifically to the beads used for the pulldown under a wide range of conditions, and so we were unable to distinguish between DNA binding and bridging activity using that assay (data not shown).

Discussion

In this study we purified human CtIP protein and characterised its structure and DNA binding properties *in vitro*. The purified protein is tetrameric and adopts a striking "dumbbell" architecture, which we have observed with both negative-stain EM and AFM. This shape is expected based on the hypothesis that CtIP adopts a "dimer of dimers" arrangement, in which the tips of long N-terminal parallel coiled coils interact to form the tetrameric interface observed in crystal structures (Andres et al., 2015; Davies et al., 2015; Forment et al., 2015) (**Figure 7**). We found that CtIP bound tightly to DNA using both EMSA and fluorescence anisotropy assays. Evidence that a major DNA binding locus resides in the distal C-terminal domains is provided by the inability of the R839A mutant to bind effectively to DNA, despite retaining a normal oligomeric state (this work; (Andres et al., 2015); **Figure 7**). A surprising result was that the L27E point mutant also displayed a dramatically reduced DNA binding affinity. This mutation was designed on the basis of crystal structures of the extreme N-terminal coiled-coil to disrupt protein:protein interactions responsible for tetramer formation, and renders CtIP non-functional *in vivo* (Davies et al., 2015). Indeed, analysis by both SEC-MALS and AFM suggested that this mutant was completely unable to form tetrameric assemblies. Our work suggests therefore, that mutations in the tetramer interface may have a more broadly destabilising effect on the biochemical activity of CtIP than simply interfering with the oligomeric state. It may be that association with DNA requires communication between two distant binding loci, although we have observed no evidence for co-operativity in our binding experiments. Alternatively, the failure to tetramerise may simply have an unfavourable effect on the global architecture of CtIP that disfavours DNA binding.

There was no evidence to suggest that CtIP bound preferentially to DNA ends *in vitro*, as might be expected given its well-defined role in the early stages of DSB repair. However, we did find that modifications to the ends of DNA substrates, including ssDNA forks and model nucleoprotein blocks, strongly increased their affinity for CtIP. This provides a biochemical basis for the observation that the activity of CtIP is especially important for the repair of "complex" or "dirty" DSBs *in vivo*, such as those programmed during meiosis by Spo11 or formed during topoisomerase poisoning by etoposide (Aparicio and Gautier, 2016). It can also help to explain why MRN-dependent nicking of DNA neighbouring blocked DNA ends requires (or is at least strongly stimulated by) CtIP, which interacts physically and functionally with MRN (Anand et al., 2016; Williams et al., 2009).

The mechanism by which modifications to DNA ends increase the affinity of CtIP for DNA remains to be determined. One simple possibility would be that CtIP interacts directly with the protein or nucleic acid conjugates that we have appended to our DNA substrates. However, we do not favour such a model: it is difficult to see how a direct protein-mediated interaction could recognise the fundamentally different types of ends that we have used *in vitro* or which must presumably be recognised *in vivo*. Alternatively, the protein might act via a "slide and capture" mechanism in which the interaction is initially with internal sites with the DNA ends acting as exit sites. In that case the dissociation may be disfavoured if the end is modified with any bulky blocking moiety, including Y-junctions and nucleoprotein complexes, such that complex ends would act to capture CtIP (**Figure 7**).

CtIP was found to be hyperphosphorylated as prepared from insect cells lysed in the presence of dephosphorylation inhibitors. This is as expected based on several previous reports of regulatory phosphorylation events on CtIP mediated by kinases linked to the DNA damage response or cell cycle (Huertas and Jackson, 2009; Peterson et al., 2013; Wang et al., 2013). Intriguingly, we observed a strong potentiation of DNA binding affinity (>10-fold tighter) when CtIP was dephosphorylated either during- or post-purification. This is consistent with a previous report that highlighted an apparent stimulation of DNA binding by dephosphorylation (Anand et al., 2016). It remains unclear whether this stimulatory effect is manifest by one or a small number of dephosphorylation events at specific residues, and therefore may represent a regulatory mechanism, or whether this simply represents a general effect of removing extensive negative charge from the surface of the protein, and this will be the subject of further study. Nevertheless, the regulation of DSB repair factors by downregulation of DNA binding via post-translational modification is certainly not unprecedented (Blanco et al., 2014; Honda et al., 2011; Lee et al., 2016).

The very tight binding mode of dephosphorylated CtIP facilitated experiments to determine the binding stoichiometry between the CtIP tetramer and short DNA fork substrates. We found that each tetramer can bind two such DNA molecules. Given the dumbbell architecture that we have observed and the location of a key DNA binding motif in the distal C-terminus, we speculate that these two DNA molecules bind to the opposing ends of the tetramer, where a major DNA binding locus might be formed by dimerization of the C-terminal domains including the RHR motif. This arrangement may help to co-ordinate the repair and eventual re-joining of two DSBs (**Figure 7**). In this respect, it is interesting to note that our AFM experiments observed significant flexibility in the connecting rod between the globular domains, which may help

facilitate such co-ordination. Consistent with this model, we also showed that CtIP was able to bridge DNA segments intermolecularly. Such a bridging activity might also help to tether the homology donor DNA in proximity to the broken DNA molecules to promote strand exchange. Given the very large number of reported CtIP interaction partners, it is possible that CtIP acts as a binding platform for co-ordinating two ended DSBs with the recruitment of signalling factors of the DNA damage response and for the assembly of a "resectosome" complex to promote repair by recombination. Higher resolution structural information for full length CtIP protein and its interactions with partner proteins will be instrumental in understanding its important role in DNA break recognition and the regulation of homologous recombination.

Acknowledgements

Author Contributions

Conceptualization, M.S.D., F.M.H., D.B.W. and O.J.W.; Methodology, Investigation and Formal Analysis, all authors; Writing – Original Draft, M.S.D. and O.J.W.; Writing – Review & Editing, all authors; Funding Acquisition, M.S.D., F.M.H. and D.B.W.

Conflicts of Interest statement

The authors declare no competing interests.

References

Anand, R., Ranjha, L., Cannavo, E., and Cejka, P. (2016). Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. Mol Cell *64*, 940-950.

Andres, S.N., Appel, C.D., Westmoreland, J.W., Williams, J.S., Nguyen, Y., Robertson, P.D., Resnick, M.A., and Williams, R.S. (2015). Tetrameric Ctp1 coordinates DNA binding and DNA bridging in DNA double-strand-break repair. Nat Struct Mol Biol *22*, 158-166.

Andres, S.N., and Williams, R.S. (2017). CtIP/Ctp1/Sae2, molecular form fit for function. DNA Repair (Amst) *56*, 109-117.

Aparicio, T., Baer, R., Gottesman, M., and Gautier, J. (2016). MRN, CtIP, and BRCA1 mediate repair of topoisomerase II-DNA adducts. J Cell Biol *212*, 399-408.

Aparicio, T., and Gautier, J. (2016). BRCA1-CtIP interaction in the repair of DNA double-strand breaks. Mol Cell Oncol *3*, e1169343.

Blanco, M.G., Matos, J., and West, S.C. (2014). Dual control of Yen1 nuclease activity and cellular localization by Cdk and Cdc14 prevents genome instability. Mol Cell *54*, 94-106.

Buis, J., Stoneham, T., Spehalski, E., and Ferguson, D.O. (2012). Mre11 regulates CtIP-dependent double-strand break repair by interaction with CDK2. Nat Struct Mol Biol *19*, 246-252.

Cannavo, E., and Cejka, P. (2014). Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. Nature *514*, 122-125.

Cejka, P. (2015). DNA End Resection: Nucleases Team Up with the Right Partners to Initiate Homologous Recombination. J Biol Chem *290*, 22931-22938.

Daley, J.M., Jimenez-Sainz, J., Wang, W., Miller, A.S., Xue, X., Nguyen, K.A., Jensen, R.B., and Sung, P. (2017). Enhancement of BLM-DNA2-Mediated Long-Range DNA End Resection by CtIP. Cell Rep *21*, 324-332.

Daley, J.M., Niu, H., Miller, A.S., and Sung, P. (2015). Biochemical mechanism of DSB end resection and its regulation. DNA Repair (Amst) *32*, 66-74.

Davies, O.R., Forment, J.V., Sun, M., Belotserkovskaya, R., Coates, J., Galanty, Y., Demir, M., Morton, C.R., Rzechorzek, N.J., Jackson, S.P.*, et al.* (2015). CtIP tetramer assembly is required for DNA-end resection and repair. Nat Struct Mol Biol *22*, 150-157.

Deshpande, R.A., Lee, J.H., Arora, S., and Paull, T.T. (2016). Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. Mol Cell *64*, 593-606.

Forment, J.V., Jackson, S.P., and Pellegrini, L. (2015). When two is not enough: a CtIP tetramer is required for DNA repair by Homologous Recombination. Nucleus *6*, 344-348.

Fuentes-Perez, M.E., Gwynn, E.J., Dillingham, M.S., and Moreno-Herrero, F. (2012). Using DNA as a fiducial marker to study SMC complex interactions with the atomic force microscope. Biophys J *102*, 839-848.

Hartsuiker, E., Mizuno, K., Molnar, M., Kohli, J., Ohta, K., and Carr, A.M. (2009a). Ctp1CtIP and Rad32Mre11 nuclease activity are required for Rec12Spo11 removal, but Rec12Spo11 removal is dispensable for other MRN-dependent meiotic functions. Mol Cell Biol *29*, 1671-1681.

Hartsuiker, E., Neale, M.J., and Carr, A.M. (2009b). Distinct requirements for the Rad32(Mre11) nuclease and Ctp1(CtIP) in the removal of covalently bound topoisomerase I and II from DNA. Mol Cell *33*, 117-123.

Honda, M., Okuno, Y., Yoo, J., Ha, T., and Spies, M. (2011). Tyrosine phosphorylation enhances RAD52-mediated annealing by modulating its DNA binding. EMBO J *30*, 3368-3382.

Horcas, I., Fernandez, R., Gomez-Rodriguez, J.M., Colchero, J., Gomez-Herrero, J., and Baro, A.M. (2007). WSXM: a software for scanning probe microscopy and a tool for nanotechnology. Rev Sci Instrum *78*, 013705.

Huang, X. (2003). Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. J Biomol Screen *8*, 34-38.

Huertas, P., and Jackson, S.P. (2009). Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. J Biol Chem *284*, 9558-9565.

Kowalczykowski, S.C. (2015). An Overview of the Molecular Mechanisms of Recombinational DNA Repair. Cold Spring Harb Perspect Biol *7*.

Lee, K.J., Saha, J., Sun, J., Fattah, K.R., Wang, S.C., Jakob, B., Chi, L., Wang, S.Y., Taucher-Scholz, G., Davis, A.J.*, et al.* (2016). Phosphorylation of Ku dictates DNA double-strand break (DSB) repair pathway choice in S phase. Nucleic Acids Res *44*, 1732-1745.

Lengsfeld, B.M., Rattray, A.J., Bhaskara, V., Ghirlando, R., and Paull, T.T. (2007). Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. Mol Cell *28*, 638-651.

Makharashvili, N., Tubbs, A.T., Yang, S.H., Wang, H., Barton, O., Zhou, Y., Deshpande, R.A., Lee, J.H., Lobrich, M., Sleckman, B.P.*, et al.* (2014). Catalytic and noncatalytic roles of the CtIP endonuclease in double-strand break end resection. Mol Cell *54*, 1022-1033.

Paudyal, S.C., Li, S., Yan, H., Hunter, T., and You, Z. (2017). Dna2 initiates resection at clean DNA double-strand breaks. Nucleic Acids Res *45*, 11766-11781.

Paull, T.T., and Gellert, M. (1999). Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. Genes Dev *13*, 1276-1288.

Peterson, S.E., Li, Y., Wu-Baer, F., Chait, B.T., Baer, R., Yan, H., Gottesman, M.E., and Gautier, J. (2013). Activation of DSB processing requires phosphorylation of CtIP by ATR. Mol Cell *49*, 657-667.

Quennet, V., Beucher, A., Barton, O., Takeda, S., and Lobrich, M. (2011). CtIP and MRN promote non-homologous end-joining of etoposide-induced DNA double-strand breaks in G1. Nucleic Acids Res *39*, 2144-2152.

Qvist, P., Huertas, P., Jimeno, S., Nyegaard, M., Hassan, M.J., Jackson, S.P., and Borglum, A.D. (2011). CtIP Mutations Cause Seckel and Jawad Syndromes. PLoS Genet *7*, e1002310.

Ranjha, L., Howard, S.M., and Cejka, P. (2018). Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. Chromosoma.

Reginato, G., Cannavo, E., and Cejka, P. (2017). Physiological protein blocks direct the Mre11-Rad50-Xrs2 and Sae2 nuclease complex to initiate DNA end resection. Genes Dev *31*, 2325-2330.

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. Nature *450*, 509-514.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol *180*, 519-530.

Shibata, A., and Jeggo, P.A. (2014). DNA double-strand break repair in a cellular context. Clin Oncol (R Coll Radiol) *26*, 243-249.

Wang, H., Li, Y., Truong, L.N., Shi, L.Z., Hwang, P.Y., He, J., Do, J., Cho, M.J., Li, H., Negrete, A.*, et al.* (2014). CtIP maintains stability at common fragile sites and inverted repeats by end resection-independent endonuclease activity. Mol Cell *54*, 1012-1021.

Wang, H., Shi, L.Z., Wong, C.C., Han, X., Hwang, P.Y., Truong, L.N., Zhu, Q., Shao, Z., Chen, D.J., Berns, M.W.*, et al.* (2013). The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR-mediated double-strand break repair. PLoS Genet *9*, e1003277.

Wang, W., Daley, J.M., Kwon, Y., Krasner, D.S., and Sung, P. (2017). Plasticity of the Mre11-Rad50-Xrs2-Sae2 nuclease ensemble in the processing of DNA-bound obstacles. Genes Dev *31*, 2331-2336.

Williams, R.S., Dodson, G.E., Limbo, O., Yamada, Y., Williams, J.S., Guenther, G., Classen, S., Glover, J.N., Iwasaki, H., Russell, P.*, et al.* (2009). Nbs1 flexibly tethers Ctp1 and Mre11-Rad50 to coordinate DNA double-strand break processing and repair. Cell *139*, 87-99.

Yeeles, J.T., van Aelst, K., Dillingham, M.S., and Moreno-Herrero, F. (2011). Recombination hotspots and single-stranded DNA binding proteins couple DNA translocation to DNA unwinding by the AddAB helicase-nuclease. Mol Cell *42*, 806-816.

You, Z., Shi, L.Z., Zhu, Q., Wu, P., Zhang, Y.W., Basilio, A., Tonnu, N., Verma, I.M., Berns, M.W., and Hunter, T. (2009). CtIP links DNA double-strand break sensing to resection. Mol Cell *36*, 954-969.

Figure Legends and Figures

Figure 1 Wild type CtIP is a tetrameric protein that forms a dumbbell-shaped particle

(A) SDS-PAGE and SEC-MALS analysis of wild type CtIP as prepared (lane 1, black trace) and following dephosphorylation post-purification (lane 2, grey trace). Horizontal lines on the SEC-MALS graph show the expected molecular weights for monomeric, dimeric and tetrameric CtIP species. (B) 2D class averages for wild type CtIP observed by negative stain EM. Scale bar = 20nm. Further examples are shown in **Supplementary Figure 1**. (C) AFM imaging of purified wild type CtIP protein. Five classes of particles (I-V) were detected based on morphology and volume analysis (see text for details). Scale bar = 100nm. Examples are marked with yellow arrows and further images at higher magnification are shown in **Supplementary Figures 2 and 3**. Representative height profiles are shown which are derived from the boxed particle. (D) Relative volume distributions for the five AFM particle classes shown in C. The volume is determined using a technique in which a fiducial DNA marker is used as a reference (Fuentes-Perez et al., 2012). There exists a linear relationship between the relative volume and absolute molecular weight for a wide range of proteins. The molecular weight of the large class V particle imaged here is 380 kDa (i.e. close to the value expected for a tetramer) based on a mean relative volume of 4.4x the marker.

Figure 1 Wild type CtIP is a tetrameric protein that forms a dumbbell-shaped particle
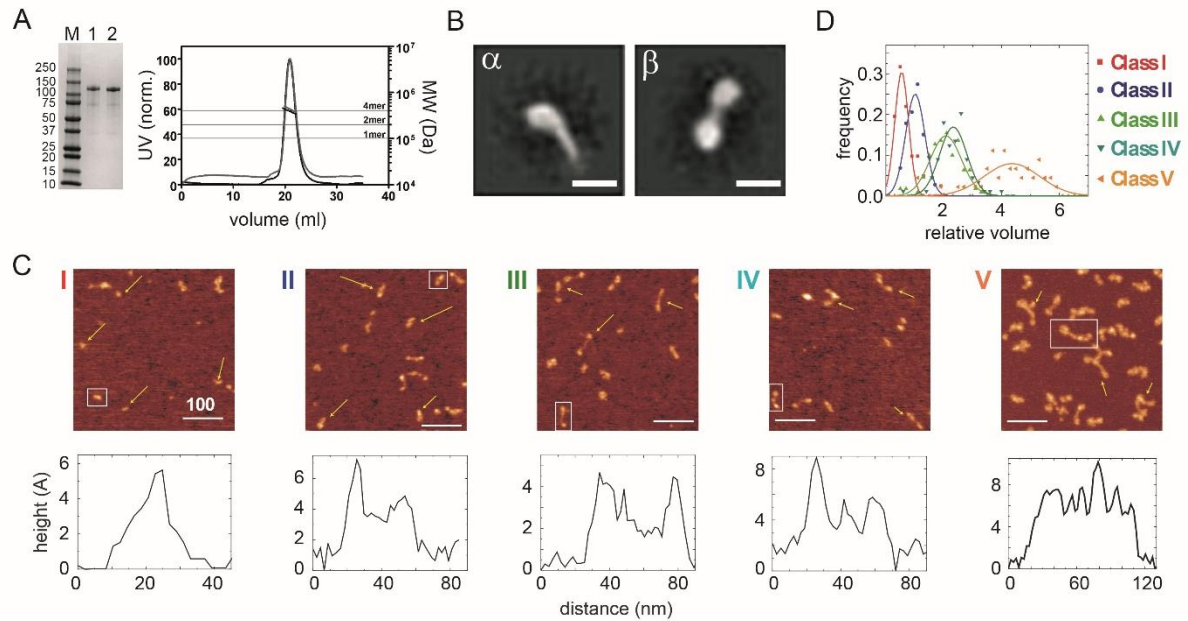
Figure 2 A mutation in the N-terminal coiled-coil domain prevents CtIP tetramerisation

(A) SDS-PAGE and SEC-MALS analysis of CtIP L27E (lane 1, blue trace) and CtIP R839A (lane 2, red trace). Data for wild type CtIP is also shown for comparison (dotted black lines). (B) AFM imaging of purified CtIP R839A protein. Five classes of particles (I-V) were detected based on morphology and volume analysis. Representative height profiles are shown below the images which are derived from the boxed particle. (C) AFM imaging of purified CtIP L27E protein. Four classes of particles (I-IV) were detected based on morphology and volume analysis. Representative height profiles are shown which are derived from the boxed particle. (D) Relative volume distributions for the AFM particle classes detected in the CtIP R839A and (E) CtIP L27E preparations. Note that the class V particle is completely absent from the CtIP L27E preparation.

Figure 2 A mutation in the N-terminal coiled-coil domain prevents CtIP tetramerisation
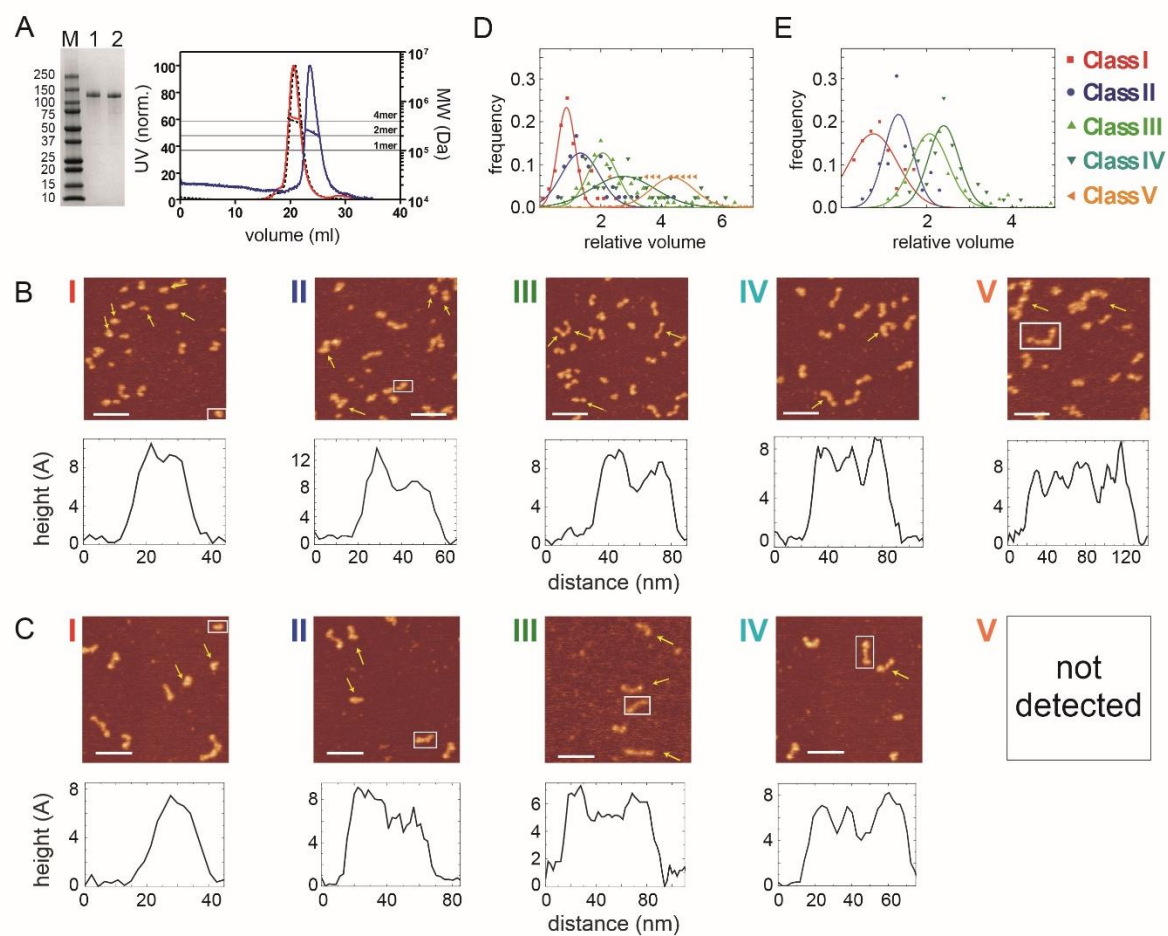
Figure 3 CtIP binds preferentially to ss-dsDNA Y-junctions in a manner dependent on both the N-terminal tetramerisation and C-terminal DNA binding motifs.

(A) Electrophoretic mobility shift assay. Radiolabelled DNA molecules with the different structures (indicated) were incubated with increasing concentrations of CtIP tetramer and run on non-denaturing PAGE gels as described in the Methods section. (B) Binding of the fork DNA substrate by the L27E mutant protein. (C) Binding of the fork DNA substrate by the R839A mutant protein. (D) Quantitative comparison of wild type and mutant CtIP proteins binding to a HEX-labelled DNA fork structure using fluorescence anisotropy.

Figure 3 CtIP binds preferentially to ss-dsDNA Y-junctions in a manner dependent on both the N-terminal tetramerisation and C-terminal DNA binding motifs.
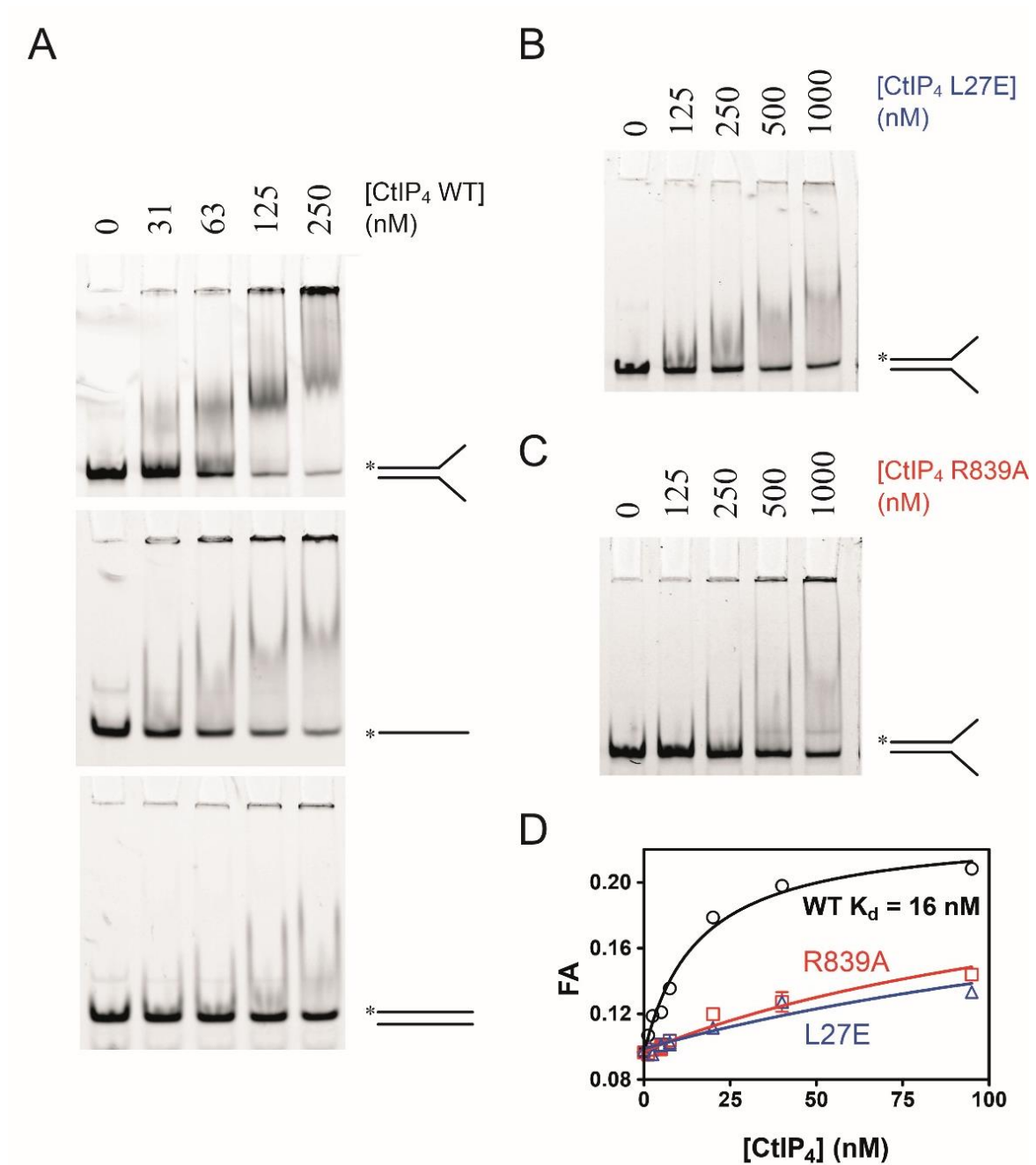
Figure 4. Dephosphorylation of CtIP potentiates DNA binding and facilitates determination of the DNA binding stoichiometry.

(A) EMSA assays comparing the binding of wild type CtIP as prepared and following treatment with λ phosphatase (denoted CtIP$^{\lambda}$) as described in the Methods. (B) Fluorescence anisotropy assay monitoring the binding of CtIP to 5 nM HEX-labelled fork DNA. The data were fit to the tight binding equation. (C) Assay as in B, but using 50, 100 and 200 nM HEX-labelled fork DNA as indicated. The data were fit globally to the Equation 5 to determine the binding stoichiometry (n) which measures the number of CtIP tetramers bound to each DNA fork.

Figure 4. Dephosphorylation of CtIP potentiate s DNA binding and facilitates determination of the DNA binding stoichiometry.
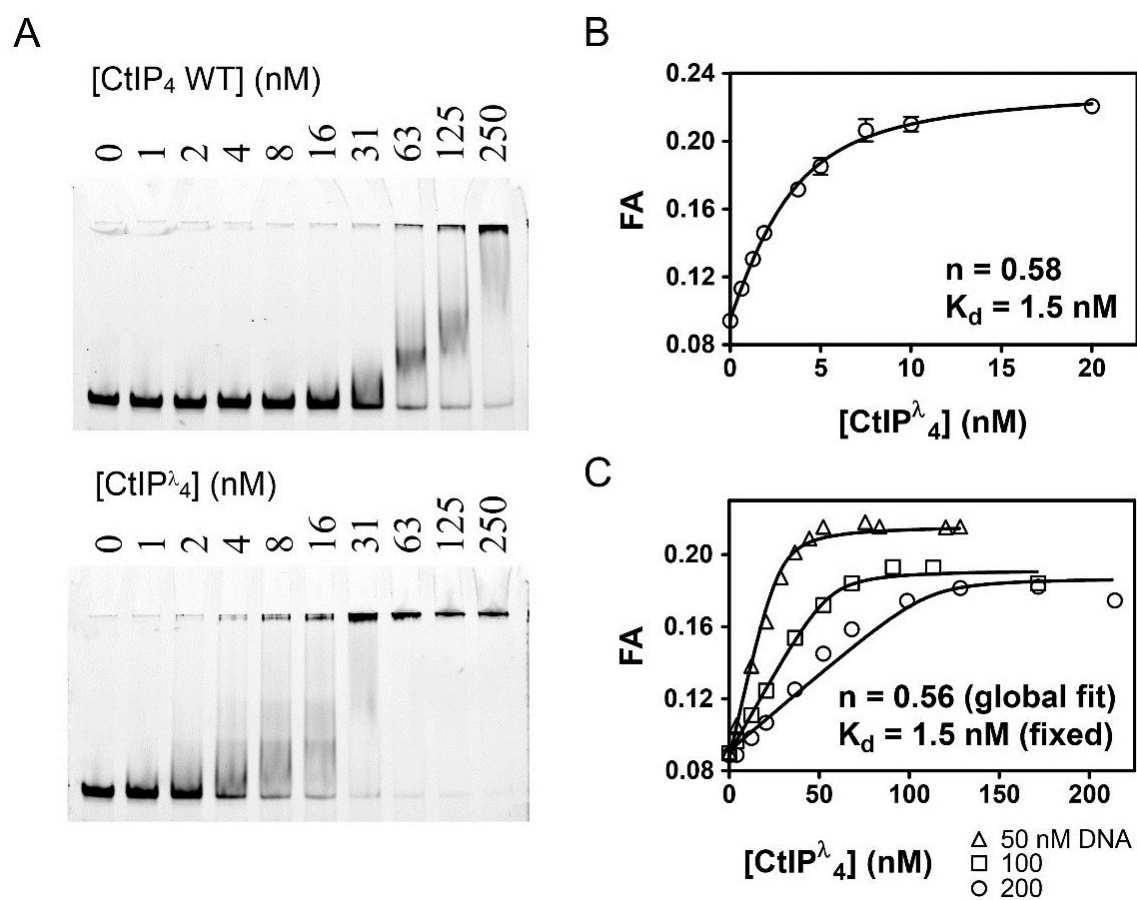
Figure 5. CtIP binding to DNA is stabilised by DNA Y-junctions and DNA end blocks.

(A) Principle of the competition DNA unbinding assay monitored by fluorescence anisotropy. HEX-labelled and unlabelled DNA fork molecules are shown in purple and black respectively. CtIP is in blue. (B) Increasing concentrations of unlabelled fork DNA were titrated into a pre-formed complex between CtIP and a HEX-labelled DNA. The data were fit to a hyperbolic decay curve to yield an IC_{50} value for the inhibition of binding. (C) Comparison of the IC_{50} values for different DNA structures. Full details of the DNA sequences and structures used in these experiments can be found in the Supplementary Information. (D) Comparison of IC_{50} values for single- and double-stranded DNA competitors with different topologies and containing different densities of DNA breaks (bracketed number). Note that the IC_{50} value is quoted in terms of the total nucleotide concentration. (E) Comparison of the IC_{50} values for duplex (red), single-stranded (blue) and forked (green) DNA substrates with and without the ends blocked by biotin:streptavidin. (F) Comparison of IC_{50} values for single- or double-stranded DNA substrates which lack biotin, determined both in the presence and absence of streptavidin. Details for the substrate construction and data for a wider range of substrates are available in the **Supplementary Information**.

36

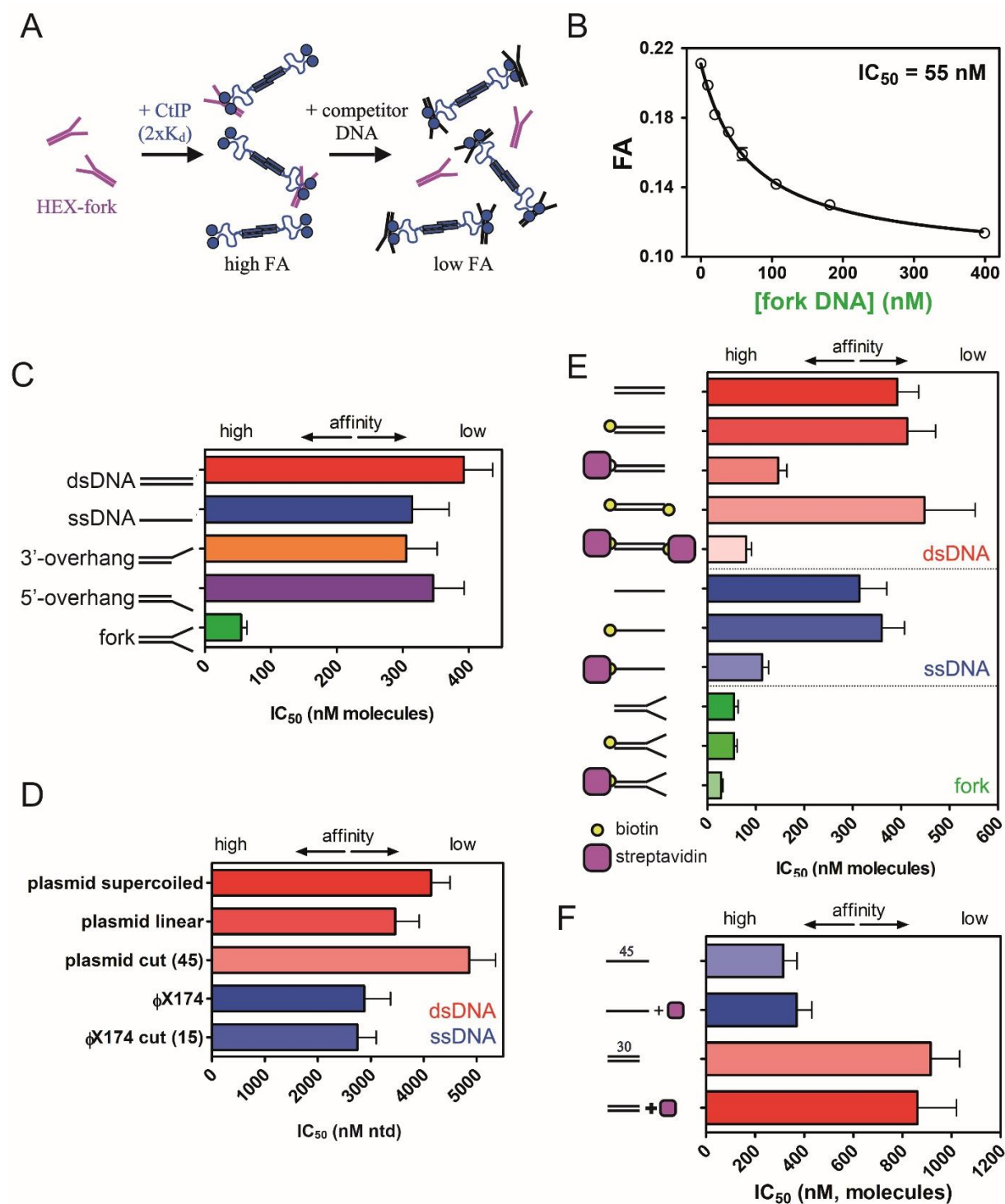**Figure 5. CtIP binding to DNA is stabilised by DNA Y-junctions and DNA end blocks.**

Figure 6. CtIP promotes intermolecular DNA bridging

(A) Representative AFM images of forked DNA substrates in the absence of CtIP. The contour length histogram shows a single gaussian peak centred on a value equivalent to a single contour length. (B) Representative AFM images of forked DNA substrates in the presence of CtIP. The contour length histogram shows multiples peaks for single-, double-, triple and even quadruple contour length peaks. The solid line shows the gaussian fit for data collected in the absence of CtIP for comparison. Note that the size of the single-contour peak is substantially reduced. The zoomed-in panels show examples of bridged DNA molecules (yellow arrows). (C) Cumulative frequency of DNA molecules longer than a given contour length, highlighting the CtIP-dependent increase in 2mer and 3mer DNA substrates. The vertical dashed lines are equivalent to theoretical single- and double- contour lengths. (D) Histogram showing percentage of molecules with contour length (x) greater than 1 for wild type and mutant CtIP proteins in addition to controls.

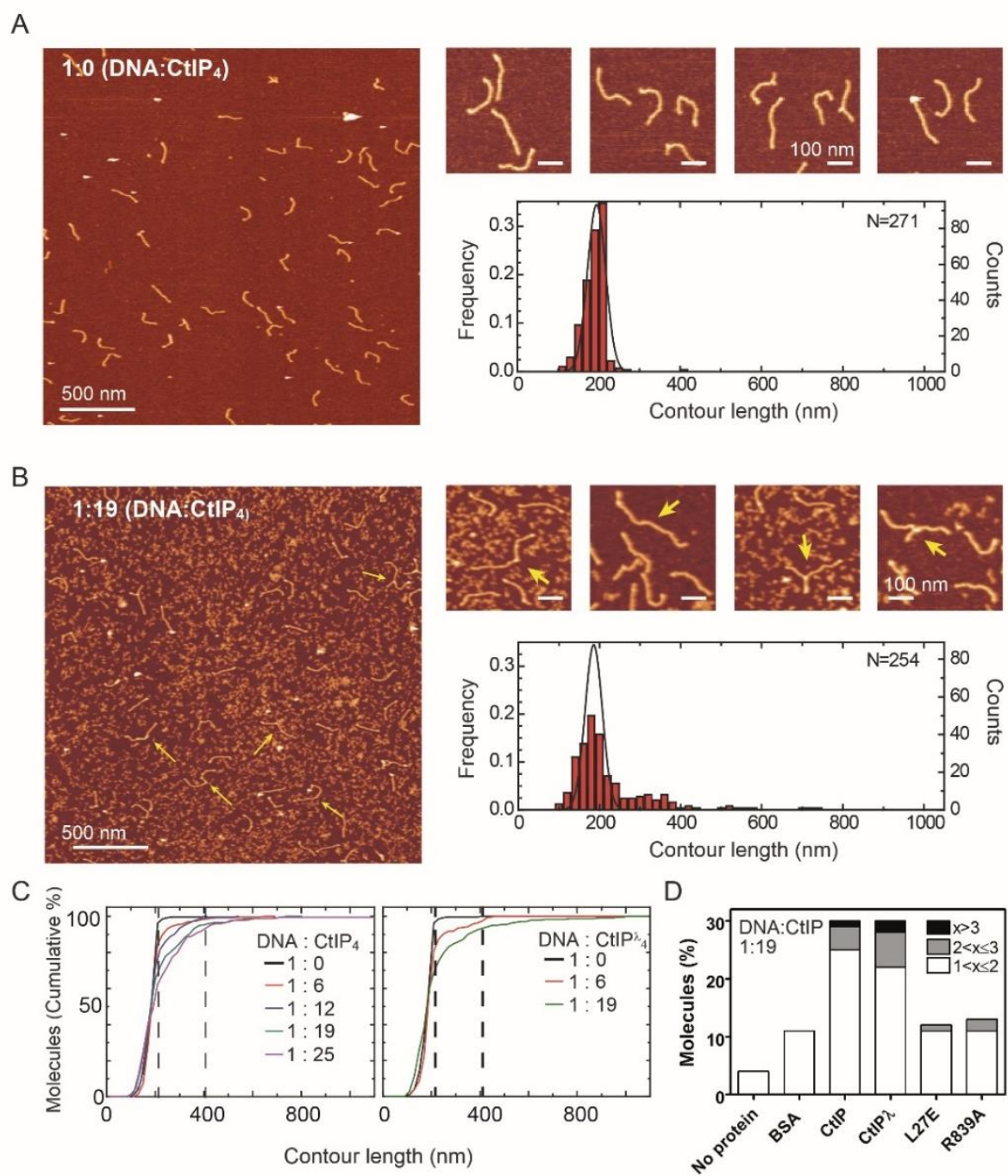## Figure 6. CtIP promotes intermolecular DNA bridging

Figure 7. A working model for binding and bridging of broken DNA ends by CtIP

The CtIP monomer comprises at least three functional regions; an N-terminal tetramerization domain, a central region of predicted disorder, and a C-terminal DNA binding domain. Wild type CtIP binds tightly to two DNA molecules via C-terminal RHR motifs, but this is disfavoured by hyperphosphorylation. Mutation within this motif (R839A) prevents DNA binding and bridging. Mutation within the N-terminal coiled-coil domains prevents assembly into a tetramer, DNA binding and DNA bridging. The grey coloured domains are intended to indicate their inactivation by mutation. The inability of L27E to bind to DNA implies that mutations in the coiled coils have effects on the distal DNA binding domains (see text for details). Wild type CtIP binds and bridges DNA segments and may then slide and capture DNA ends possessing forked structures (indicated) or nucleoprotein conjugates, possibly because they prevent dissociation from the DNA ends (see main text for discussion). We speculate that this arrangement might allow the dimeric sub-assemblies within CtIP to slide away from each DNA end to which they are bound, facilitating resection and/or access to the DSBs for resection and signalling factors while maintaining their pairing.

Figure 7. A working model for binding and bridging of broken DNA ends by CtIP