Models for the architecture of the human inner kinetochore CCAN complex on centromeric α-satellite CENP-A nucleosome arrays

Cong Yu[1], Kyle W. Muir[1,3], Jing Yang[1], Ziguo Zhang[1], Stephen H. McLaughlin[1] and David Barford[1,2]

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, UK

[2]Corresponding author: dbarford@mrc-lmb.cam.ac.uk

[3]Present address: Institute of Cell Biology, University of Edinburgh, Edinburgh, UK.

Abstract

Human kinetochores assemble onto centromeric DNA defined by tandem copies of thousands of the 171 bp α-satellite repeat sequence. The centromere-specific CENP-A nucleosome (CENP-A^{Nuc}) recruits the inner kinetochore constitutive centromere associated network (CCAN) complex. A previous cryo-EM structure of CCAN bound to a CENP-A^{Nuc} reconstituted with a 171 bp α-satellite repeat sequence showed how extranucleosomal linker DNA threads through an internal tunnel in the CCAN complex. To understand the higher-order architecture of the inner kinetochore assembled onto α-satellite repeat arrays, we have determined cryo-EM structures of CCAN with longer DNA sequences. These include free DNA and single and dimeric α-satellite repeats with CENP-A nucleosomes. We show from the structures of CCAN bound to both free DNA and monomeric CENP-A^{Nuc} that CCAN engages 65-70 bp of DNA comprising 30-35 bp of an upstream α-satellite repeat. This upstream DNA interacts with the histone-fold domain subunits of the CENP-TWSX module in a manner resembling how DNA is wrapped in nucleosomes. A complex of CCAN assembled onto a dimeric α-satellite repeat with two CENP-A nucleosomes showed that CCAN can only be accommodated on the linker DNA by unwrapping DNA from the CENP-TWSX module together with 25 bp of DNA from the upstream nucleosome. We discuss the implications of these results for models of CCAN assembly on arrays of α-satellite chromatin containing CENP-A^{Nuc}.

Introduction

The faithful inheritance of genetic information is crucial for life. Central to this process in eukaryotes are the roles fulfilled by kinetochores – large protein complexes that assemble

at centromeric chromatin. Kinetochores function to attach sister chromatids to the mitotic spindle, and couple the work of microtubule depolymerization to chromatid segregation [1-3]. Kinetochores also act as signalling hubs for the error correction and spindle assembly checkpoints that together coordinate the correct attachment of sister chromatids to the mitotic spindle with the initiation of sister chromatid segregation [4,5].

Human regional centromeres are megabase-sized structures, comprising thousands of contiguous 171 bp α-satellite repeat sequences. Almost identical α-satellite repeats are organized into higher order repeat (HOR) arrays of typically four, seven and eleven repeats [6]. Only a small fraction of α-satellite repeats is associated with the approximately two-hundred CENP-A nucleosomes per centromere [7]. CENP-A nucleosomes are restricted to the hypo-methylated centromere dip region (CDR), the most recently evolved segment of centromeres [6,8,9], and are interspersed with canonical H3 nucleosomes [10]. Assuming at most 147 bp of DNA wraps a histone octamer [11], at least 24 bp of α-satellite repeat sequence forms extranucleosomal linker DNA. Both *in vivo* ChIP-seq mapping and in *vitro* reconstitution of CENP-A nucleosomes on α-satellite sequences indicate that the CENP-A nucleosome is installed precisely towards the 3' end of the repeat, with the centre of the α-satellite repeat palindrome coinciding with the nucleosome dyad axis [12,13]. Human centromeres, except for the Y chromosome centromere, contain the 17 bp B-box element. The majority of B-boxes are arranged on alternating α-satellite repeats [14] positioned such that the B-box on linker DNA partially overlaps a nucleosome entry site [15]. CENP-B engagement by the B-box facilitates unwrapping of the DNA gyre from the CENP-A histone octamer [16], whereas the engagement of the dimeric CENP-B to two separate B-boxes induces looping of centromeric DNA [17].

The centromere-kinetochore assembly of vertebrates comprises large macromolecular complexes at the constriction point of chromosomes visible by electron microscopy in prometaphase cells as plate-like structures ~200 nm in diameter, and attached to 20-30 microtubule filaments [18]. Kinetochores are composed of two main components: the inner kinetochore centromere-binding CCAN complex, and the microtubule-binding KMN complex of the outer kinetochore [19]. The 16-subunit CCAN complex is organized into five modules: CENP-C and CENP-OPQUR, and the DNA-binding modules, CENP-LN, CENP-HIKM and CENP-TWSX. The latter module comprises four histone-fold domain proteins [20]. Cryo-EM structures of human and budding yeast CCAN have elucidated mechanisms of centromere and CENP-A nucleosome recognition [12,21-25]. CCAN encloses extranucleosomal linker DNA within a topologically-enclosed chamber [12,22], a mechanism recently shown to be conserved in the CENP-A-independent holocentric kinetochores of *Lepidoptera* [26]. In vertebrates,

interactions with the CENP-A nucleosome, that confer CCAN specificity for CENP-A rather than H3 nucleosomes, are mediated primarily by the CENP-C subunit through its short linear interaction motifs that bridge the CENP-A histones of CENP-A^{Nuc} with the CCAN modules CENP-LN and CENP-HIKM [12,21,23,27-30]. Human CENP-C incorporates two CENP-A-binding motifs, generating four potential CENP-A-binding sites (two CENP-A nucleosomes) in the CENP-C dimer [31].

In our previous cryo-EM structures of CCAN:CENP-A^{Nuc} and CCAN:DNA complexes, a DNA duplex was observed threading through a 50 Å-long central tunnel running the length of the CCAN molecule [12]. This DNA-binding tunnel is formed by the CENP-LN channel, and CENP-I and CENP-TW subunits. As the DNA exits the tunnel it curves around the CENP-TWSX module, interacting with the histone-fold domains of CENP-TW. The DNA molecules used in the earlier study (the 171 bp α-satellite DNA used to reconstitute the CENP-A^{Nuc} for the CCAN:CENP-A^{Nuc} complex, and the 54 bp DNA used in the CCAN:DNA complex) were of insufficient length to extend beyond CENP-TW, and so in these complexes DNA did not contact CENP-SX of the histone-fold CENP-TWSX module. Previous cell biology and biochemical studies however, had identified CENP-SX as DNA-binding proteins at the centromere [32]. We were therefore missing important details regarding how CCAN interacts with DNA at the centromere, and this information is required for understanding the higher-order architecture of kinetochores arranged on α-satellite repeat arrays. Furthermore, the repeating CCAN:CENP-A nucleosome unit at centromeres is not defined. Thus, the aims of this study are two-fold. First, we aim to obtain a more complete understanding of how CCAN interacts with DNA, and second we seek to provide models for how multiple copies of CCAN and CENP-A nucleosomes assemble onto contiguous α-satellite repeat sequences at human centromeres.

Results

CCAN binds 70 bp of DNA in a loop configuration that wraps the CENP-TWSX module

To understand how centromeric DNA interacts with CCAN, including all of the histone-fold domains (HFDs) of the CENP-TWSX module, we reconstituted CCAN with a longer DNA duplex (native human α-satellite DNA: 171 bp) and used mild cross-linking to stabilize the complex. The cryo-EM structure, resolved to 3.5 Å resolution, showed clearly defined cryo-EM density for ~70 bp of DNA that wraps around CCAN in a loop-like structure (**Fig. 1 and Supplementary Figs 1 and 2a and Supplementary Table 1**). As observed in our previous CCAN:CENP-A^{Nuc} and CCAN:DNA complexes [12], DNA threads through the enclosed CCAN chamber, contacting CENP-TW as it exits the tunnel (**Fig. 1 and Supplementary Fig. 2a, b**).

However, in the new structure, we observed an additional 20 bp of DNA exiting the DNA-binding tunnel, so that in total 40 bp of DNA interacts with the entire CENP-TWSX module. Additionally, a short segment of DNA (~5 bp) extends beyond the CENP-TWSX module and approaches CENP-QU.

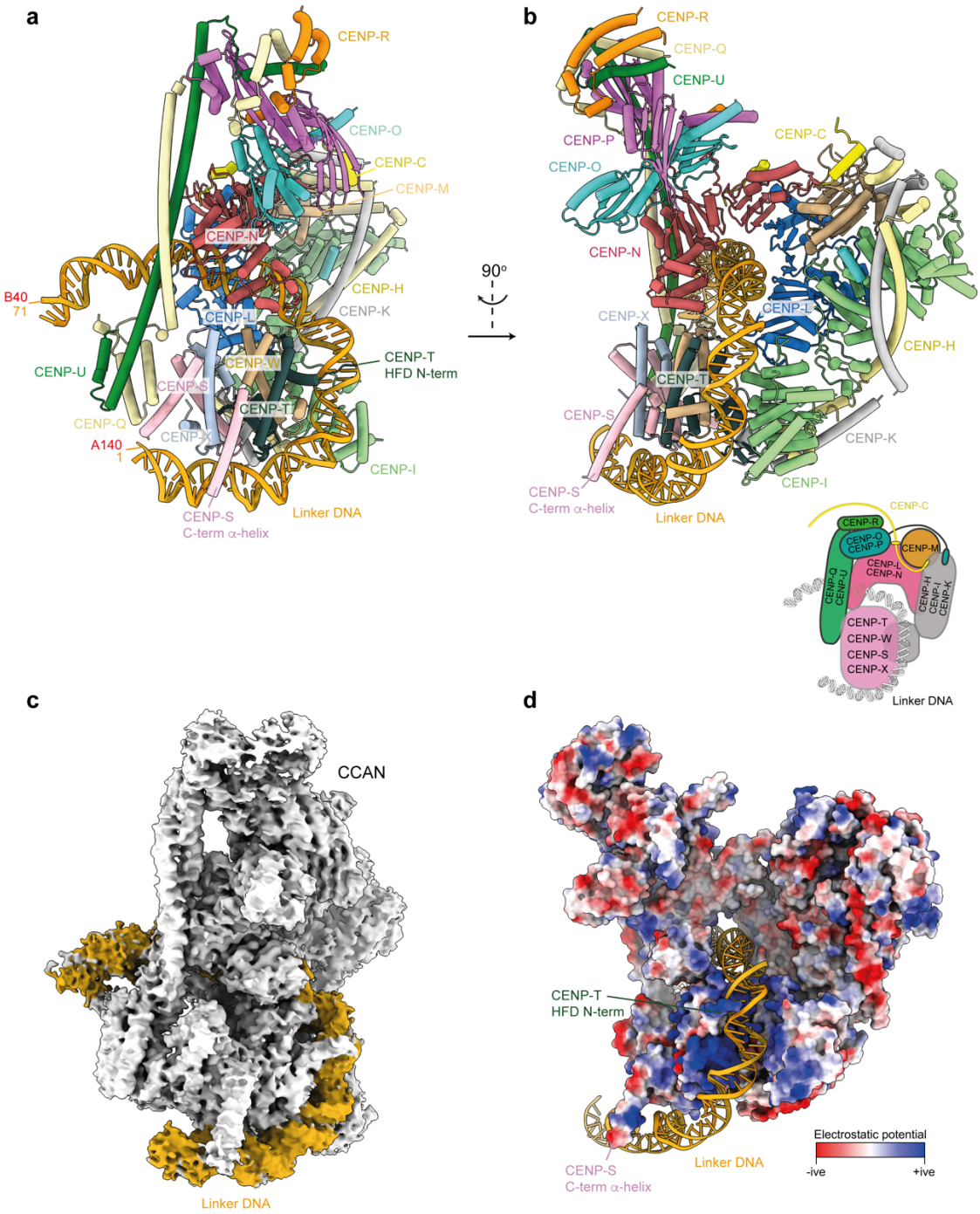


Fig. 1. CCAN binds 70 bp of DNA in a loop configuration. **a** and **b**, Orthogonal views of the CCAN:DNA complex. 41 bp of DNA interacts with the CENP-TWSX module. Insert: Schematic of the CCAN:DNA complex. In panel (**a**), DNA is numbered '1' and '71' at the ordered 5' and 3' ends, and 'A140' and 'B40' to indicate the position of the DNA on a dimeric α-satellite sequence (Fig. 3c). **c**, Cryo-EM map of the CCAN:DNA complex with DNA coloured in gold and CCAN in grey. An

unsharpened version of this map is shown in Supplementary Fig. 1c. **d**, Electrostatic surface of CCAN as viewed in panel B, with DNA shown as a ribbons representation (Supplementary Fig. 2).

The HFDs of the CENP-TWSX module assemble into a heterotetramer that structurally resembles the H3:H4 tetramer of canonical nucleosomes (**Fig. 2a, b**) [11,32]. In the CCAN:DNA complex, DNA wraps around CENP-TWSX comparably to how H3:H4 tetramers wrap DNA in the context of an octameric nucleosome [11]. In both structures DNA has a similar curvature and adopts a left-handed helical geometry. The CENP-TWSX module creates a positively-charged ridge that binds the DNA phosphate backbone (**Fig. 1d**). Strikingly, DNA wraps the H3:H4 and CENP-TWSX tetramers with identical registers such that the DNA major groove projects towards both the dyad axis of the H3:H4 tetramer, and the related pseudo-dyad axis of CENP-TWSX (**Fig. 2a, b**). Consequently, in both structures, at superhelical locations (SHLs) $\pm$0.5, $\pm$1.5, $\pm$2.5, the DNA minor groove projects towards and contacts equivalent residues in the histone-fold domains of the H3:H4 tetramer and CENP-TWSX module. Our structure shows that DNA interacts with conserved basic residues of all four HFD domains of CENP-TWSX (**Fig. 2c, d**), consistent with prior mutagenesis data demonstrating the role of basic CENP-TWSX amino acids in DNA binding *in vitro* and centromere localization *in vivo* [32].

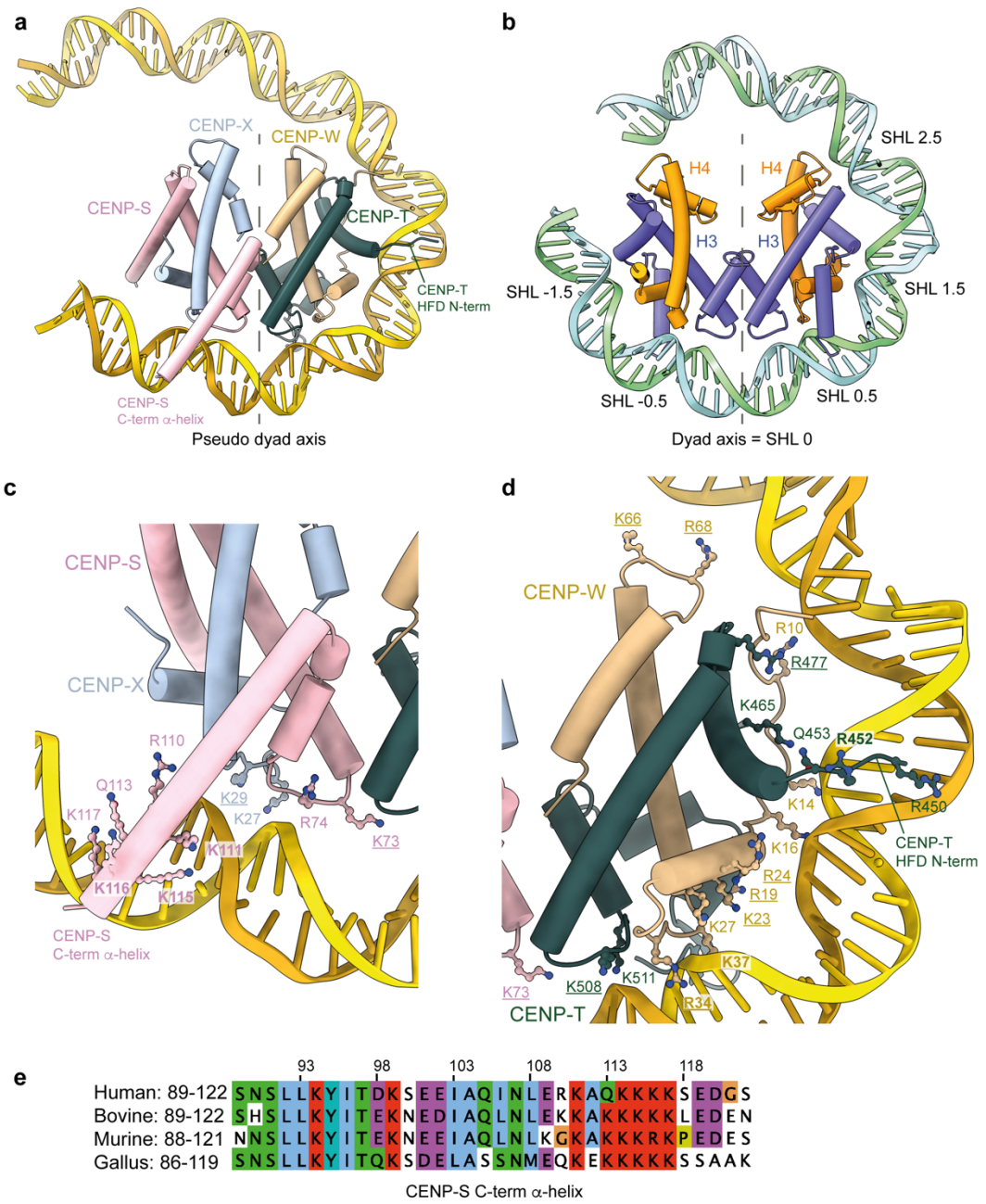
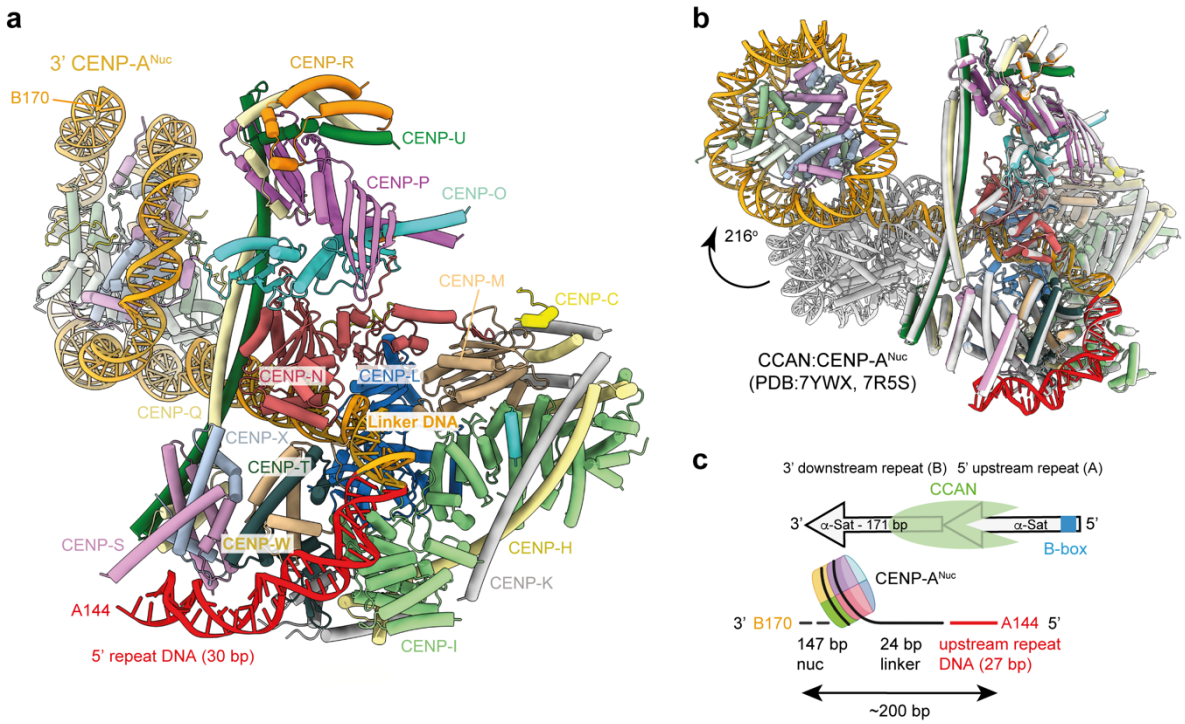


Fig. 2. The CENP-TWSX module binds DNA analogous to the H3:H4 histone tetramer of canonical nucleosomes. a, CENP-TWSX and CCAN-associated 70 bp DNA, 40 bp interacts with the CENP-TWSX module. **b**, H3:H4 tetramer of a nucleosome with almost with 67 bp of DNA. Super helical locations (SHLs) are indicated (PDB:1AOI [11]). The pseudo dyad axes of CENP-TWSX and the H3:H4 tetramer are shown in panels (a) and (b), respectively. **c**, Zoomed-view of panel (a) showing details of basic residues from the HFDs of CENP-S and CENP-X and CENP-S C-terminal α-helix interacting with the DNA phosphate backbone. **d**, Zoomed-view of panel (a) showing details of basic residues from the HFDs of CENP-T and CENP-W, and CENP-T N-terminal extension interacting with the DNA phosphate backbone. In panels (c) and (d) residues previously shown to contribute to CENP-TWSX – DNA binding are underlined [32]. **e**, Multiple sequence alignment of the CENP-S C-terminal α-helix. Alignment created using Jalview [33].

Additional protein-DNA contacts stabilize CENP-TWSX – DNA interactions. These involve an extended basic segment immediately N-terminal of the CENP-T HFD that inserts into the

155 DNA minor groove, and a C-terminus α-helix of CENP-S that engages the DNA major groove
156 (**Fig. 2a, c and d**). This CENP-S α-helix becomes ordered on binding DNA, such that a
157 conserved stretch of basic residues stabilize its interactions with the DNA backbone (**Fig. 2c,**
158 **e and Supplementary Fig. 2c, d**), as similarly observed in the crystal structure of a CENP-
159 SX:DNA complex [34]. Our structural data are consistent with previous studies showing that
160 deletion of the C-terminus of CENP-S, including this basic region, abolished CENP-S binding
161 to DNA *in vitro* [32]. Similar observations regarding the interaction of linker DNA with the
162 CENP-TWSX module were obtained by Musacchio and colleagues from a cryo-EM structure
163 of a CCAN:DNA complex (M. Pesenti, I. Vetter and A. Musacchio, unpublished data).
164
165 <u>DNA of CENP-A^{Nuc} reconstituted with longer α-satellite repeat DNA wraps CENP-TWSX</u>
166
167 We next assessed whether CENP-TWSX wraps linker DNA in the context of a CENP-A
168 nucleosome. We assembled CCAN onto a CENP-A^{Nuc} reconstituted with 211 bp DNA that
169 comprised the 171 bp α-satellite repeat and 40 bp from the 5' upstream α-satellite repeat,
170 and determined its structure using cryo-EM to an overall resolution of 4.5 Å (**Supplementary**
171 **Fig. 3**). Both individual CCAN:DNA and CENP-A^{Nuc} modules were resolved in cryo-EM density,
172 showing relative flexibility (**Supplementary Fig 3d**). A total of 200 bp DNA was visible in the
173 cryo-EM map. Similar to our previous CCAN:CENP-A^{Nuc} structure with 171 bp DNA [12], CENP-
174 A^{Nuc} is positioned at the 3' end of the α-satellite repeat (**Fig. 3a, c and Supplementary Fig.**
175 **4a**). The 5' terminus of the CENP-A^{Nuc} DNA gyre is unwrapped by 10 bp (**Fig. 3a and**
176 **Supplementary Fig. 4b**). This unwrapped DNA then leads into extranucleosomal linker DNA
177 that threads through the CCAN DNA-binding tunnel. On exiting the tunnel, DNA wraps
178 around the entire CENP-TWSX module, as seen in the CCAN:DNA complex described above.
179 This DNA comprises linker DNA of the 3' downstream α-satellite repeat together with 30 bp
180 of the 5' upstream α-satellite repeat. Thus, a total of ~65 bp of free DNA 5' of CENP-A^{Nuc}
181 interacts with CCAN (**Fig. 3c**). An unexpected finding was that compared with our previous
182 CCAN:CENP-A^{Nuc} cryo-EM structure [12], CENP-A^{Nuc} adopts a different orientation relative to
183 CCAN. In our new structure, CENP-A^{Nuc} is rotated by 216° and shifted by six bp away from
184 CCAN. A shift of register while maintaining the same DNA phosphate backbone interactions
185 with the CCAN DNA-binding tunnel causes the DNA duplex (and hence CENP-A^{Nuc}) to rotate
186 36° per bp, rotating 216° overall (**Fig. 3b and Supplementary Video 1**). The reasons for the
187 differences in CENP-A^{Nuc} position between the two CCAN:CENP-A^{Nuc} complexes are not
188 entirely clear. A possible explanation is that the longer extranucleosomal DNA linker used in
189 this study optimizes interactions with DNA-binding sites on CCAN, resulting in displacement
190 of CENP-A^{Nuc}. Cryo-EM density for CENP-C is observed at the CENP-C-binding sites on CENP-

191 A^{Nuc}, CENP-LN and CENP-HIKM, however, due to the displaced CENP-A^{Nuc}, there are no
192 direct contacts between CENP-A^{Nuc} and the back-side of CCAN.



193
194 **Fig. 3. Cryo-EM structure of CCAN:CENP-A^{Nuc} showing CENP-TWSX binding DNA. a**, Ribbons
195 representation of the CCAN:CENP-A^{Nuc} complex. DNA coloured in red indicates DNA of the 5'
196 upstream α-satellite repeat. **b**, Superimposition of our new CCAN:CENP-A^{Nuc} complex (with 211 bp
197 DNA) (coloured as in panel A) onto the previous model of CCAN:CENP-A^{Nuc} with a single 171 bp α-
198 satellite repeat DNA (coloured grey) [12]. This shows the difference in position and orientation of the
199 downstream 3' CENP-A^{Nuc}. The two nucleosomes differ in orientation by 216° due to a 6 bp
200 translation of the repeat sequence through the tunnel (**Supplementary Video 1**). **c**, Schematic of the
201 CCAN:CENP-A^{Nuc} on an α-satellite repeat dimer. A total of ~200 bp of DNA is embedded by the
202 complex including 27 bp from the 5' upstream α-satellite repeat. A B-box accessible for CENP-B
203 binding in the upstream repeat is shown in blue. The nucleotide numbering (A144 and B170 in
204 panels a and c) refers to the position of the visible DNA 5' and 3' ends of the complex within the α-
205 satellite sequence. Positions in the α-satellite repeat are numbered 1-171 (5' to 3'), and the
206 upstream and downstream α-satellite repeats are labelled A and B, respectively (Supplementary Fig.
207 4).
208
209 Although the cryo-EM density for the DNA phosphate backbone is well resolved for both the
210 CCAN:DNA and CCAN:CENP-A^{Nuc} complexes, density for individual DNA bases is less well
211 defined such that purine and pyrimidine bases are not distinguished (**Supplementary Figs**
212 **1e, 2a and 3d**). We also do not observe sequence-specific protein-DNA interactions, which
213 might be expected to contribute to a defined position of α-satellite DNA bound to CCAN.
214 This poorly-resolved base-pair density is likely due to the limited resolutions of the
215 CCAN:DNA and CCAN:CENP-A^{Nuc} complexes (3.5 and 4.5 Å, respectively), but would also be
216 consistent with sequence variability at each position of the DNA duplex. However, the well-
217 resolved cryo-EM density for the DNA phosphate backbone indicates that there is a specific

218 mode of DNA recognition by CCAN, clearly illustrated by how the DNA duplex wraps around
219 the CENP-TWSX module with phasing similar to how DNA wraps histone octamers (**Fig. 2a,**
220 **b**).
221
222 In the CCAN:CENP-A^{Nuc} complex structure using the longer DNA described here, we do not
223 observe contacts between CCAN and CENP-A^{Nuc}, in contrast to our previous CCAN:CENP-A^{Nuc}
224 complex [12]. However, the position of CCAN on DNA appears fixed as judged by its defined
225 position on the DNA duplex relative to CENP-A^{Nuc}. This spacing of CCAN and CENP-A^{Nuc} is
226 unlikely caused by the CENP-C subunit. First, the linker connecting the CENP-C binding sites
227 on CENP-A^{Nuc} and CCAN (200 residues) is intrinsically disordered and not visible in the cryo-
228 EM density. Second, CENP-C was also present in the previous structure with shorter DNA [12].
229 Third, in the CCAN:di-CENP-A^{Nuc} complex without CENP-C (discussed below), the position of
230 CCAN relative to the downstream nucleosome is the same as in this new CCAN:CENP-A^{Nuc}
231 complex. In our previous structure [12], the 171 bp of a single α-satellite repeat DNA was too
232 short to both wrap CENP-A^{Nuc} and extend beyond CENP-TW of the CENP-TWSX module. For
233 that complex, to optimise CCAN-DNA interactions, CCAN would be required to shift towards
234 CENP-A^{Nuc}. In the new complex reported here, with 40 bp additional DNA (27 bp of which
235 are ordered), there is sufficient DNA to wrap CENP-A^{Nuc}, pass through the CCAN tunnel and
236 wrap around CENP-TWSX. A fixed spacing between CENP-A^{Nuc} and CCAN positioned on DNA
237 potentially indicates the presence of a CCAN-positioning sequence within α-satellite DNA
238 that might contribute to how the DNA duplex wraps around the histone-fold CENP-TWSX
239 module. This is reminiscent of how CENP-A nucleosomes are positioned precisely on α-
240 satellite DNA, as revealed by *in vitro* reconstitution [12], and the phasing of CENP-A
241 nucleosomes on centromeric chromatin [13], and of DNA nucleosome positioning sequences
242 in general. The DNA-binding subunits of CCAN modules: CENP-LN, CENP-I, CENP-TWSX could
243 all contribute to a defined mode of DNA phosphate backbone recognition.
244
245 <u>Implications for CCAN binding to α-satellite repeat dimers</u>
246
247 Our new CCAN:DNA and CCAN:CENP-A^{Nuc} structures showing that CCAN itself interacts with
248 65 bp of duplex DNA has implications for understanding the higher-order organization of
249 CCAN on repetitive α-satellite repeat CENP-A nucleosome arrays. The 65 bp DNA interacting
250 with CCAN comprises 30 bp from the 5' upstream α-satellite repeat and 35 bp of the 3'
251 downstream α-satellite repeat: 24 bp of extranucleosomal linker DNA and 10 bp of
252 unwrapped CENP-A^{Nuc} DNA (**Fig. 3c**). The 30 bp DNA interacting with CCAN from the 5'
253 upstream α-satellite repeat is DNA of the CENP-A nucleosome gyre when CENP-A
254 nucleosomes are positioned at the 3' end of the α-satellite repeat [12,13]. This suggests that in

255 the context of a CCAN:CENP-A^{Nuc} complex, the upstream α-satellite repeat might be devoid
256 of a nucleosome (either CENP-A or canonical H3). In alternative non-exclusive scenarios, to
257 accommodate CCAN bound between two CENP-A nucleosomes on adjacent α-satellite
258 repeats, a CENP-A nucleosome positioned on the upstream 5' α-satellite repeat might
259 unwrap by 30 bp, or the upstream CENP-A^{Nuc} may force DNA to unwrap from the CENP-
260 TWSX module. A third possibility is that CCAN repositions CENP-A^{Nuc} by causing DNA to slide
261 around the histone octamer. However, the defined position of CENP-A^{Nuc} at the 3' of α-
262 satellite repeats *in vivo* [13] does not support this scenario.
263
264 Substantial evidence indicates that the DNA termini of CENP-A nucleosomes are loosely
265 wrapped and flexible. First, nuclease digestion sensitivity assays showed that CENP-A
266 nucleosomes protect only 100-120 bp DNA, compared to 140-150 bp for canonical H3
267 nucleosomes [13,24,35,36]. Second, FRET data indicated increased flexibility of terminal DNA
268 segments in CENP-A^{Nuc} relative to canonical nucleosomes [16]. Third, both crystal and cryo-EM
269 structures of CENP-A^{Nuc} revealed unwrapping of terminal DNA segments [12,16,24,37-41], in some
270 instances as much as 24-36 bp of flexible DNA at one end [42]. Finally, CENP-B and CENP-C
271 have been implicated in promoting DNA unwrapping from CENP-A^{Nuc} (Refs [16,43,44]).
272
273 <u>CCAN binds a dimeric α-satellite repeat with two CENP-A nucleosomes</u>
274
275 To investigate the feasibility of CCAN binding to linker DNA connecting two adjacent CENP-A
276 nucleosomes, we reconstituted two CENP-A nucleosomes on an α-satellite repeat dimer.
277 The α-satellite repeat dimer was designed to include only linker DNA connecting the two
278 CENP-A nucleosomes: upstream and downstream α-satellite repeats were 147 bp and 171
279 bp, respectively (**Supplementary Fig. 5a**). To stabilize the two nucleosome positions on this
280 324 bp DNA, we replaced 92 bp of native α-satellite repeat DNA at the centre of each CENP-
281 A^{Nuc} site with Widom601 DNA. We attempted to reconstitute CENP-A nucleosomes on a
282 dimeric α-satellite repeat sequence (342 bp) using completely native sequence. Whereas
283 we could generate well-positioned CENP-A nucleosomes on a single α-satellite repeat (as
284 defined by native agarose gels), we found that native sequences exceeding 180-190 bp did
285 not wrap well. Typically, the shifts were incomplete and heterogeneous with substantial
286 amounts of free DNA compared to monomeric α-satellite DNA and Widom-based
287 sequences.
288
289 We observed that CCAN associated with this di-CENP-A^{Nuc} array, as assessed by size
290 exclusion chromatography (SEC) (**Supplementary Fig. 5b, c**). Analysis of the peak SEC
291 fractions using interferometric scattering microscopy (iSCAT) revealed two species: the

major 1,054 kDa complex corresponded to a single CCAN protomer bound to the di-CENP-A^{Nuc} array (CCAN:di-CENP-A^{Nuc}), although with a lower mass than the expected 1,282 kDa. The underestimate of molecular mass using ISCAT may be due to the DNA component which scatters less than the globular proteins used for calibration. A 445 kDa component corresponded to a di-CENP-A^{Nuc} array without CCAN (expected molecular mass of 431 kDa) (**Supplementary Fig. 5d**). The partial dissociation of CCAN from di-CENP-A^{Nuc} is likely due to the low sample concentrations required for single molecule counting in iSCAT.

CCAN and CENP-B binding to the same linker DNA segment is mutually exclusive

Our α-satellite dimer incorporated a B-box within the linker DNA of the downstream α-satellite repeat connecting the two CENP-A nucleosomes (**Supplementary Fig. 6a**). SEC data showed that the CENP-B protein recognized the B-box specifically because CENP-B bound to the di-CENP-A^{Nuc} array with an intact B-box, but not to a di-CENP-A^{Nuc} array with a mutant B-box (**Supplementary Fig. 6b, c**). We next tested whether CCAN and CENP-B bind the di-CENP-A^{Nuc} array simultaneously. We observed that CCAN displaced CENP-B from the array, showing that CCAN and CENP-B engagement of linker DNA is mutually exclusive (**Supplementary Fig. 6d, e**). This is consistent with our model of CCAN:CENP-A^{Nuc}, that with CCAN bound to linker DNA, the B-box is enclosed by the CCAN DNA-binding tunnel and therefore inaccessible to CENP-B.

<u>CCAN binds to linker DNA in the dimeric α-satellite repeat with two CENP-A nucleosomes</u>

To understand the structural basis of how CCAN interacts with linker DNA connecting adjacent CENP-A nucleosomes, we prepared cryo-EM grids with the CCAN:di-CENP-A^{Nuc} complex. Despite cross-linking and the use of additives to reduce sample denaturation at the air-water interface, CCAN:di-CENP-A^{Nuc} particles were not visible on either cryo-electron micrographs or subsequent 2D classes of the cryo-EM dataset. We instead observed a heterogeneous mixture of free DNA, CENP-A^{Nuc} and protein complexes (data not shown), indicative of CCAN:di-CENP-A^{Nuc} denaturation on cryo-EM grids. To stabilize the CCAN:di-CENP-A^{Nuc} complex for cryo-EM analysis, we combined it with the CENP-A nucleosome stabilizing-single chain antibody fragment (ScFv) [40], omitting CENP-C due to its binding site overlapping that of the ScFv (**Supplementary Fig. 7)**, and collected a large cryo-EM dataset on this ScFv-stabilized CCAN:di-CENP-A^{Nuc} complex (**Supplementary Figs 8 and 9 and Supplementary Table 1**).

2D classification of the selected particles and subsequent 3D reconstruction and classification revealed a highly heterogenous mixture of molecular states (**Supplementary**

329 **Figs 8b-d and 9**). The majority of 2D classes from nearly nine million picked particles
330 comprised either free mono-CENP-A nucleosome or CCAN:DNA with no obvious associated
331 CENP-A nucleosome. *Ab initio* reconstruction and 3D refinement resulted in high resolution
332 maps of CCAN:DNA and a CENP-A mono-nucleosome with bound ScFv (2.9 Å and 2.7 Å,
333 respectively; **Supplementary Fig. 9**). The CCAN reconstruction showed DNA enclosed within
334 the DNA-binding tunnel. We re-extracted CCAN particles in RELION [45] using a larger box and
335 performed further 3D classification (**Supplementary Fig. 9**). This procedure revealed 3D
336 classes with diffuse, poorly-defined EM densities at both 3' and 5' entrances to the CCAN-
337 DNA tunnel that likely represented flexible CENP-A nucleosomes. To determine whether the
338 density at the 3' entrance corresponded to a 3' CENP-A nucleosome, we generated a wide
339 mask covering this density and performed signal-subtraction of density outside the mask.
340 Subsequent 3D reconstruction and classification of the mask-generated recentred particles
341 recovered a CENP-A nucleosome, confirming that the diffuse density observed in 2D classes
342 as CENP-A^{Nuc}. Using the corresponding unmasked particle set, we then calculated a new 3D
343 reconstruction at 9.5 Å resolution that revealed both CCAN and 3' CENP-A^{Nuc}. Consistent
344 with this reconstruction, 2D classes calculated from this particle set showed density features
345 for CCAN and 3' CENP-A^{Nuc} (**Supplementary Fig. 8c**). A similar approach was used to analyse
346 EM density at the 5' entrance to the CCAN-DNA tunnel, resulting in a 3D reconstruction with
347 CCAN and associated 5' CENP-A^{Nuc} at 15 Å resolution (**Supplementary Figs 8d and 9**). In the
348 2D class galleries corresponding to both the 3' CCAN:CENP-A^{Nuc} and 5' CCAN:CENP-A^{Nuc}
349 complexes, we observed diffuse densities at the opposite end of the CCAN DNA-tunnel (i.e.
350 5' and 3', respectively), for example **Supplementary Fig. 9a(i)**. These particles are consistent
351 with CCAN bound to the linker DNA of the di-CENP-A^{Nuc} array. However, the small number
352 of the particles precluded a 3D reconstruction. We therefore generated a composite cryo-
353 EM map of a CCAN:di-CENP-A^{Nuc} complex from the individual 3' CCAN:CENP-A^{Nuc} and 5'
354 CCAN:CENP-A^{Nuc} maps (**Supplementary Fig. 9a(i)** and **Fig. 4b**). One of its projected views
355 matches an experimental 2D class of a CCAN:di-CENP-A^{Nuc} , validating our composite
356 approach (**Supplementary Fig. 9a(i)**).
357
358
359
360

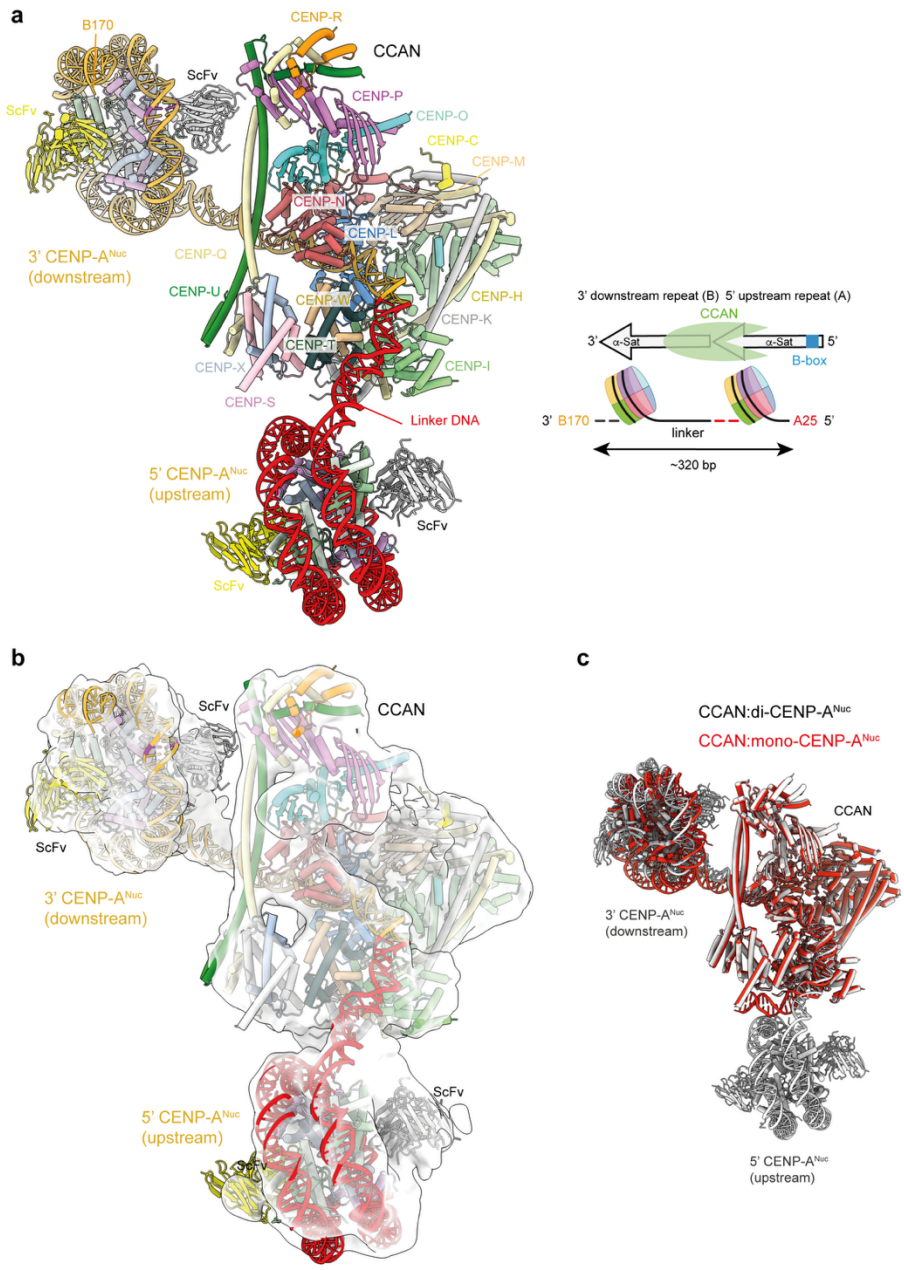


Fig. 4. Structure of the CCAN:di-CENP-A^{Nuc} complex. **a**, Ribbons representation of CCAN:di-CENP-A^{Nuc} complex. Insert: Schematic of the CCAN:di-CENP-A^{Nuc} on a dimer of α-satellite repeats. 20-25 bp of the upstream 5´ CENP-A nucleosome is unwrapped. A total of 300-310 bp of DNA is embedded by the complex. A B-box accessible for CENP-B binding in the upstream repeat is shown in blue. **b**, Cryo-EM map of CCAN:di-CENP-A^{Nuc} complex. **C**. Superimposition of the new CCAN:CENP-A^{Nuc} and CCAN:di-CENP-A^{Nuc} complexes showing the identical position of the 3´ CENP-A^{Nuc}. The nucleotide numbering (A25 and B170 in panel a) is defined in Fig. 3c (Supplementary Fig. 10).

CCAN bound to linker DNA forces DNA unwrapping from CENP-TWSX and 5´ CENP-A^{Nuc}

5´ CCAN:CENP-A^{Nuc} complex: Despite the limited resolution of this map, reliable alignment of CENP-A^{Nuc} into cryo-EM density was facilitated by the ScFv associated with both sides of the nucleosome (**Fig. 4b**). Analysis of this structure indicated that a CENP-A nucleosome

375 positioned 5´ to CCAN is incompatible with DNA wrapping around the CENP-TWSX module
376 (**Fig. 4a, b and Supplementary Fig. 10a)**. Analogous to our previous CCAN:CENP-A^{Nuc} and
377 CCAN:DNA complexes [12], we observed that as DNA exits the CCAN-DNA tunnel, it runs along
378 the CENP-I-CENP-TW channel, although unlike our new CCAN:DNA and CCAN:CENP-A^{Nuc}
379 complexes (this study), does not continue to bend around CENP-SX, but instead projects in a
380 straighter trajectory to join the DNA of the 5´ CENP-A nucleosome. The DNA gyre of the 5´
381 CENP-A nucleosome connected to linker DNA is itself unwrapped by 20-25 bp
382 (**Supplementary Fig. 10c**). Thus, even without DNA wrapping around CENP-SX, the 5' CENP-
383 A nucleosome is forced to unwrap to allow its positioning onto an upstream α-satellite
384 repeat. Our fitting of the 5´ CENP-A^{Nuc} to the cryo-EM density map is consistent with the α-
385 satellite repeat length of 171 bp such that the DNA length from the end of the downstream
386 (3´) α-satellite repeat sequence to the dyad axis of the 5´ CENP-A^{Nuc} is 74 bp (**Fig. 4a**).
387
388 In the 3´ CCAN:CENP-A^{Nuc} complex, the position and orientation of the 3´ CENP-A^{Nuc} in the
389 CCAN:di-CENP-A^{Nuc} complex is the same as our new CCAN:mono-CENP-A^{Nuc} complex with
390 211 bp DNA (this study), differing from that in the previously published CCAN:CENP-A^{Nuc}
391 complex [12] (**Fig. 4a, c and Supplementary Fig. 10a, b**). Thus, the ScFv used for the CCAN:di-
392 CENP-A^{Nuc} structure determination does not affect the orientation of the 3´ CENP-A^{Nuc},
393 consistent with modelling showing that ScFv binding to CENP-A^{Nuc}, when in the mono-CENP-
394 A^{Nuc} position, is possible without a steric clash with CCAN.
395
396 **Discussion**
397
398 In this study we investigated the molecular basis for how human CCAN interacts with CENP-
399 A nucleosomes and DNA of α-satellite repeat arrays, revealing that CCAN engages as much
400 as 70 bp of DNA. As described previously [12], DNA threads through the enclosed tunnel
401 generated by the CENP-LN module, CENP-I and CENP-TW subunits. The longer DNA
402 sequences used in this study showed additional DNA interactions with the CENP-SX subunits
403 of the CENP-TWSX HFD module. The CENP-TWSX subunits interact with DNA analogously to
404 how the H3:H4 heterotetramer, in the context of octameric nucleosomes, wraps its DNA
405 gyre. The specific details of the interactions of DNA with CENP-SX, together with CENP-TW,
406 observed in our structure, are in agreement with biochemical and cell biology data,
407 confirming the proposal that human CCAN incorporates a nucleosome-like particle at the
408 centromere [20,32]. Consistent with this, CENP-TWSX augments the affinity of CCAN for a
409 CENP-A nucleosome three-fold [12]. Thus, this study indicates that CCAN creates a tight
410 attachment to centromeric DNA by both topological entrapment of DNA through the CCAN

411 DNA-binding tunnel (CENP-LN, CENP-I), and by wrapping DNA around a HFD tetramer
412 (CENP-TWSX), mimicking the mechanism by which nucleosomes package DNA.
413
414 The finding that CCAN alone engages 65-70 bp of DNA has implications for the higher-order
415 architecture of CCAN and CENP-A nucleosomes on repetitive α-satellite sequences. The 65
416 bp of extranucleosomal DNA bound by CCAN greatly exceeds the 24 bp linker DNA
417 connecting two adjacent and fully wrapped nucleosomes (147 bp) on tandem α-satellite
418 repeats. In our CCAN:CENP-A^{Nuc} structures, 10 bp of DNA are unwrapped from the 3' CENP-
419 A nucleosome gyre as it enters the CCAN DNA-binding tunnel (Ref. [12] and this study,
420 **Supplementary Fig. 4b**). Thus, to accommodate 65 bp of DNA interacting with CCAN
421 requires ~30 bp DNA of the upstream α-satellite repeat, in addition to the 24 bp of α-
422 satellite extranucleosomal linker DNA and 10 bp of unwrapped DNA from the downstream
423 CENP-A nucleosome (**Fig. 3c**). Since CENP-A nucleosomes are positioned exactly on α-
424 satellite repeats such that the 3' end of the repeat defines the start of the CENP-A
425 nucleosome DNA gyre [12,13], the 30 bp of upstream α-satellite repeat DNA would require the
426 upstream CENP-A nucleosome to unwrap 30 bp of DNA. Such unwrapping is potentially
427 feasible as demonstrated by the extensive levels of DNA unwrapping from CENP-A
428 nucleosomes documented previously [12,16,24,37-42]. To test whether CCAN associates with DNA
429 connecting two adjacent CENP-A nucleosomes, we reconstituted a dimeric α-satellite repeat
430 with CENP-A nucleosomes. SEC and iSCAT data confirmed interactions of CCAN with a
431 dimeric α-satellite repeat with two CENP-A nucleosomes (**Supplementary Fig. 5**). That this
432 interaction required the free linker DNA connecting the two nucleosomes was confirmed by
433 the ability of CCAN to displace CENP-B from its B-box binding site located within the linker
434 DNA (**Supplementary Fig. 6d, e**). Our cryo-EM structure of the CCAN:di-CENP-A^{Nuc} complex
435 revealed CCAN bound to DNA linking two CENP-A nucleosomes (**Fig. 4**). However,
436 accommodating CCAN at this site involved not only 20-25 bp of DNA to be unwrapped from
437 the 5' upstream CENP-A nucleosome, but also dissociation of DNA from CENP-SX. Thus,
438 although the upstream CENP-A nucleosome is unwrapped, the available free DNA duplex is
439 insufficient to bind CENP-SX, presumably because of steric hindrance between the upstream
440 CENP-A nucleosome and CCAN.
441
442 Whether such an arrangement of tandem CENP-A nucleosomes and central CCAN that we
443 observe in our CCAN:di-CENP-A^{Nuc} complex occurs at native human centromeres is unclear.
444 The dissociation of DNA from CENP-SX is inconsistent with the role of conserved basic
445 residues of CENP-SX in binding DNA revealed by our CCAN:DNA and CCAN:mono-CENP-A^{Nuc}
446 structures and from prior *in vitro* binding assays, and the role of CENP-SX in mediating
447 kinetochore assembly in cells [32]. Furthermore, the extensive unwrapping of DNA from a

448 CENP-A nucleosome, although observed for a small population of particles in a cryo-EM
449 study [42], is not consistent with CENP-A nucleosomes defining a 120 bp footprint *in vivo* [13,46],
450 and the size of the nuclease-resistant core observed *in vitro*. Although our cryo-EM
451 structure of the CCAN:di-CENP-A^{Nuc} complex embeds a total of 300-310 bp (**Fig. 4a**), in
452 agreement with the 250-300 bp centromere core-selective footprint identified by single
453 molecule chromatin fibre sequencing (Fiber-seq) [46], we propose two alternative scenarios
454 for the arrangement of CCAN and CENP-A nucleosomes on α-satellite arrays. In one
455 scenario, the kinetochore unit would constitute a monomeric CCAN:CENP-A^{Nuc} complex in
456 which DNA from the upstream α-satellite repeat engages CENP-TWSX such that a total of
457 200-210 bp of DNA is bound by the CCAN:CENP-A^{Nuc} complex (**Fig. 3**). In a second related
458 scenario, a central CENP-A nucleosome would interact with two flanking CCAN protomers in
459 a pseudo-symmetric arrangement of the mono-CCAN:CENP-A^{Nuc} complex (di-CCAN:CENP-
460 A^{Nuc}) (**Fig. 5**). This proposed assembly, with the central CENP-A nucleosome and associated
461 CCAN protomers spanning three α-satellite repeats, embeds 260-270 bp. The 210 bp bound
462 by mono-CCAN:CENP-A^{Nuc} is shorter than the 250-300 bp centromere core-selective
463 footprint [46], however the 260-270 bp bound to the proposed di-CCAN:CENP-A^{Nuc} complexes
464 matches the reported 250-300 bp centromere footprint. We note that Dubocanin et al.,
465 (2025) also identified centromere-associated footprints of 210 bp in size [46]. Although these
466 were assigned as di-nucleosomes, it is also plausible that these methylation-free footprints
467 correspond to mono-CCAN:CENP-A^{Nuc} assemblies. Consistent with this, a recent cryo-ET
468 study of human centromeres found that the majority of putative kinetochore particles were
469 consistent in size with a mono-CCAN complex closely associated with a single CENP-A
470 nucleosome [47]. Additionally, larger globular densities containing multiple nucleosomes and
471 CCAN complexes were also observed, although whether these constitute tandem CCAN and
472 CENP-A^{Nuc} complexes was not defined.
473
474

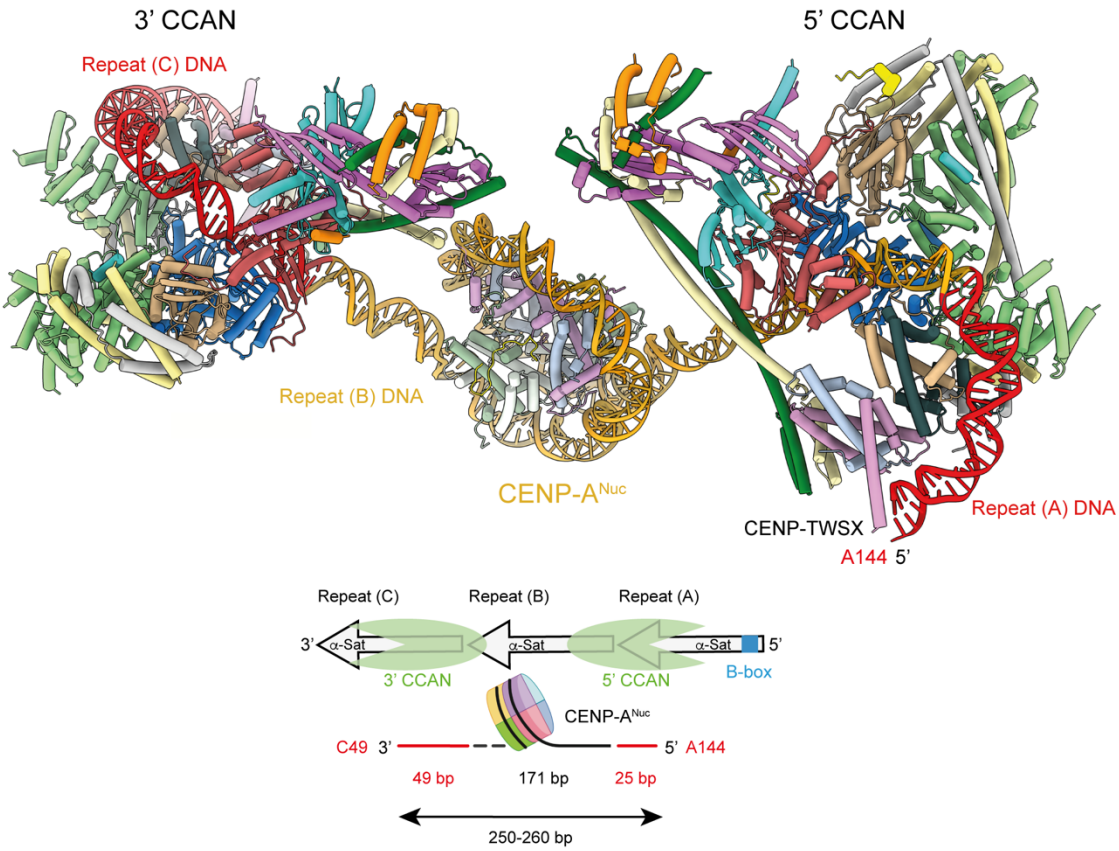


Fig. 5. Model of a di-CCAN:CENP-A^{Nuc} complex. The model was generated from the CCAN:mono-CENP-A^{Nuc} structure by superimposing on CENP-A^{Nuc}. The two CCANs together with a single CENP-A^{Nuc} overlap three α-satellite repeats. In the model, DNA from the central α-satellite sequence is coloured gold. DNA bound to the two CCANs from the 5' and 3' α-satellite repeats (repeats A and C, respectively) is coloured red. Insert: Schematic of the model with DNA coloured red for the 5' and 3' repeats as in the model. A B-box accessible for CENP-B binding in the upstream repeat is shown in blue. B-boxes in the two upstream repeats are occluded by CCAN. 260-270 bp of DNA are bound by the di-CCAN:CENP-A^{Nuc} complex. The nucleotide numbering (A144 and C49) is defined in Fig. 3.

Our models are consistent with the published numbers of CENP-A, CENP-C and CENP-T proteins at centromeres, and the region of the centromere occupied by CENP-A nucleosomes. Within canonical centromeres, the two-hundred CENP-A nucleosomes [7] are restricted to a region of ~100-500 kilobases [6,13] that defines kinetochore assembly. Since 100-500 kilobase comprises 600-3,000 α-satellite repeats, this indicates one CENP-A nucleosome per 3-15 α-satellite repeats. The question of how many CENP-A nucleosomes a CENP-C dimer binds at the centromere is unclear. Since each human CENP-C subunit contains two-CENP-A binding motifs, in principle a CENP-C dimer could bind two CENP-A nucleosomes and two CCAN protomers, consistent with *in vitro* data [31]. However, for human CENP-C, only one CENP-A binding motif is essential [13,48-50], and chicken, *Xenopus* and *S. cerevisiae* CENP-C all contain one CENP-A binding motif. For the human kinetochore, possible models are for one CENP-C dimer for either one or two CENP-A nucleosomes, and one or two CCAN protomers. Suzuki et al reported 215 copies of CENP-C (108 CENP-C

498 dimers) and 72 copies of CENP-T [51] per centromere, data roughly consistent with the
499 number of CENP-A nucleosomes. The three models of CCAN:CENP-A^{Nuc} assembly discussed
500 here: (i) 1xCCAN:1xCENP-A^{Nuc}:1xCENP-C dimer (**Fig. 3**), (ii) 1xCCAN:2xCENP-A^{Nuc}:1xCENP-C
501 dimer (**Fig. 4**), and (iii) 2xCCAN:1xCENP-A^{Nuc}:1xCENP-C dimer (**Fig. 5**) are all feasible
502 candidates given the published copies of CENP-A, CENP-C and CENP-T proteins per
503 centromere. Models (i) and (ii) would occupy two α-satellite repeats, whereas model (iii)
504 would occupy three repeats. Assuming 200 CENP-A nucleosomes per centromere and
505 contiguous tandem repeats of the CCAN:CENP-A^{Nuc} assembly, the number of α-satellite
506 repeats required is 400 (models (i) and (ii) or 800 (model (iii)), consistent with the 600-3,000
507 α-satellite repeats of the ~100-500 kilobase CENP-A region of the centromere [6,13]. Our
508 models are thus consistent with the published numbers of CENP-A, CENP-C and CENP-T
509 proteins, and the region of the centromere occupied by CENP-A nucleosomes.
510
511 In a model of a CENP-C dimer binding to a CCAN:di-CENP-A^{Nuc} complex, a possible scenario is
512 for the pair of C-terminal CENP-A binding sites (CENP-C motif) to interact with either the up-
513 stream or down-stream CENP-A nucleosome and the pair of N-terminal CENP-A binding sites
514 (central region) to interact with the remaining CENP-A nucleosome. As there is one CCAN
515 protomer in this complex, only one CENP-C subunit of the dimer would bind CCAN. Chicken
516 CENP-C dimers are observed to form higher-order oligomers based on the crystal packing of
517 chicken CENP-C cupin dimers, supported by biochemical and cell biology data [52]. The same
518 study also presented evidence for oligomerization of human CENP-C cupin domains [52], as
519 had been reported previously [53], although *in vitro* oligomers of cupin dimers were not
520 observed [31]. Formation of higher-order oligomers of CENP-C might be expected to
521 oligomerize the CCAN:CENP-A^{Nuc} assemblies. However, we are unable to reliably predict the
522 exact architecture of this higher-order assembly. Hara et al., (2023) [52] proposed that CENP-C
523 oligomerization would lead to increased CENP-C concentration that might stabilize CCAN
524 assembly.
525
526 Our findings that CCAN displaces CENP-B from B-box linker DNA indicated that the binding
527 of CCAN and CENP-B to a single α-satellite repeat-CENP-A^{Nuc} module is mutually exclusive
528 (**Supplementary Fig. 6**). B-boxes are generally present on alternating α-satellite repeats [6].
529 For all three models of CCAN:CENP-A^{Nuc} complexes discussed above (**Figs 3-5**), a B-box in the
530 upstream α-satellite repeat is accessible for CENP-B binding, although only in models (i) and
531 (ii) would CENP-B bind to B-box elements on alternat α-satellite repeats.
532
533 In summary this work, combined with recent 2D and 3D mapping of CCAN, CENP-A
534 nucleosomes and DNA-binding proteins at native centromeres [46,47], suggest potentially

535 heterogenous structures of CCAN and CENP-A nucleosomes assembled on α-satellite repeat
536 arrays including mono-CCAN:CENP-A^{Nuc}, di-CCAN:CENP-A^{Nuc} and CCAN:di-CENP-A^{Nuc}
537 complexes.
538
539 **Supplementary Video 1.** Video illustrates the difference in position of the 3' CENP-A
540 nucleosome between our previous CCAN:CENP-A^{Nuc} complex with 171 bp DNA (grey
541 cartoon) [12] and the new structure of CCAN:CENP-A^{Nuc} with 211 bp DNA (ASW6) (colour-
542 coded according to subunits as in Fig. 1). Relative to the previous CCAN:CENP-A^{Nuc} structure,
543 in the new model, the DNA register of the α-satellite repeat is shifted 6 bp (translation of 20
544 Å) away from CCAN (towards the 3' end). Because the DNA phosphate backbone contacts
545 with CCAN are preserved, this results in a 216° rotation of CENP-A^{Nuc}.
546
547 **Methods**
548
549 DNA cloning
550
551 Protein complexes for CCAN and CENP-A histone octamer were prepared as previously
552 described [12], except CENP-Q was modified to CENP-Q^{55-268}, CENP-U was modified to CENP-
553 $U^{235-418}$, CENP-I was changed to isoform 1 from previously isoform 2, and CENP-B full length
554 was cloned into pOPINS plasmid, with a double strep II-sumo-TEV tag at their N-termini. The
555 DNA fragment ASW14 was designed as two repeats from α-satellite DNA of chromosome 2
556 with 91 base pairs centred on their dyad positions replaced with Widom601 sequence
557 aligned at the dyad base. A B-box sequence is in the linker region. Six ASW14 fragments
558 flanked by GATATC (EcoRV recognition site) were assembled into the pMK-RQ plasmid. The
559 plasmid containing ASW14 fragments was produced by using Plasmid Plus Giga Kit
560 (QIAGEN). The DNA fragment was released by EcoRV-HF (NEB) at 37 °C for 16 h. The digested
561 sample was diluted five-fold with PBS and loaded onto a pre-equilibrated (PBS) 1 mL
562 Resource Q Column (GE 17-1177-01). The column was washed with 10 mL PBS (buffer A),
563 followed by a wash with 20 mL of 50% of buffer B (PBS with 850 mM NaCl). The DNA
564 fragment was eluted using a gradient to 70% of buffer B at 1 mL/min over 120 min. The
565 target fragment eluted around 60% buffer B. The fractions containing DNA were checked on
566 an agarose gel, and DNA was precipitated using 1:1 of isopropanol at 20 °C for 30 min at
567 15,000g. The DNA pellet was dissolved in the buffer needed for its following application.
568
569 The sequence of the generated ASW14 fragment is:
570
571 1 ATCAGGAAAT ATCTTCACAT AAAAACTAGA CAATTGGTCG TAGACAGCTC TAGCACCGCT
572 61 TAAACGCACG TACGCG*C*TGT CCCCCGCGTT TTAACCGCCA AGGGGATTAC TCCCTAGTCT
573 121 CCTTGTAGTA TCTGCAAGTG GACATTCGGA GCACTTTGTG GC**CTTCGTTG GAAACGGGA**A

```
181     TATGTTCACA TAAAAACTAG ACAATTGGTC GTAGACAGCT CTAGCACCGC TTAAACGCAC
241     GTACGCGCTG TCCCCCGCGT TTTAACCGCC AAGGGGATTA CTCCCTAGTC TCCTTGTAGT
301     ATCTGGAATT GGACATTTGG AGAT
```

The Widom601 sequence underlined. Dyad base of the two CENP-A^{Nuc} indicated in italics. ASW14 with mutant B-box: The B-box of CTTCGTTGGAAACGGGA in bold of ASW14 was mutated to CAAAATTGGTAATTTTA in the B-box mutant α-satellite repeat dimer.

The 212 bp DNA fragment for the CCAN:CENP-A^{Nuc} complex (ASW6) has an additional 38 bp 5' of an α-satellite repeat sequence (AS2) with a central 92 bp Widom601 sequence. ASW6 sequence is (B-box element in bold):

```
1       CATACTCTTT TTGTAGTATC TGCAAGTGGA CATTCGGAGC ACTTTGTGGC CTTCGTTGGA
61      AACGGGAATA TGTTCACATA AAAACTAGAC AATTGGTCGT AGACAGCTCT AGCACCGCTT
121     AAACGCACGT ACGCGCTGTC CCCCGCGTTT TAACCGCCAA GGGGATTACT CCCTAGTCTC
181     CTTGTAGTAT CTGGAATTGG ACATTTGGAG AT
```

AS2 (α-satellite repeat sequence from chromosome 2), also defined in Yatskevich et al., [12]:

```
1       CCGCTTTGAG GCCTTCGTTG GAAACGGGAA TATGTTCACA TAAAAACTAG ACAGAAGCAT
61      TCTCAGAAAC TTCTATGTGA TGTTTGCATT CAACTCATAG AGTTGAACAT TCCTTTTCAT
121     AGAGCAGTTT TGAAACACTC TTTTTGTAGT ATCTGGAATT GGACATTTGG A
```

α-satellite repeat sequence as defined in Hasson et al., (Supplementary Figure 6a) [13].

```
1                                                               ACAGAAGCAT
11      TCTCAGAAAC TTCTATGTGA TGTTTGCATT CAACTCATAG AGTTGAACAT TCCTTTTCAT
81      AGAGCAGTTT TGAAACACTC TTTTTGTAGT ATCTGGAATT GGACATTTGG ACCGCTTTGA
141     GGCCTTCGTT GGAAACGGGA ATATGTTCAC ATAAAAACTA G
```

We define the start and end of the α-satellite repeat based on the positioning of the CENP-A nucleosome on the α-satellite repeat sequence from our 2.4 Å resolution cryo-EM structure [12]. Specifically, we numbered the 3' end of the repeat (nucleotide position #171) as the entry site of the CENP-A^{Nuc} DNA gyre. This then positions the CENP-A^{Nuc} dyad axis as nucleotide #98, and the 5' entry into the DNA gyre as #25. Interestingly, the dyad axis of CENP-A^{Nuc} at nucleotide #98 maps to the centre of the 19 bp palindrome of human α-satellite repeats (TTCAACTCA**T**AGAGTTGAA), described in the 1978 papers first reporting primate α-satellite sequences [54,55]. Our definition of the α-satellite repeat is 50 bp 5' to that defined by Hasson et al., (2013) [13].

CCAN:DNA complex assembly

CCAN:DNA complexes were assembled in CCAN buffer (300 mM NaCl, 20 mM Hepes pH 7.8, 0.5 mM TCEP) as follows. 6 µM CCAN was incubated at 22 °C for 30 mins before addition of an equal volume of 6 µM 171 bp (AS2 α-satellite repeat [12]) DNA (AS2 described in [12]), and incubated for a further 30 mins at at 22 °C, resulting in a final CCAN:DNA concentration of 3 µM and a final volume of 100 µL. The CCAN:DNA complex was then centrifuged at 10 °C in a

621 table-top centrifuge for 10 min. To stabilize complexes for cryo-EM analysis, on-column
622 crosslinking was performed.
623
624 On-column crosslinking of CCAN:DNA
625
626 An Agilent 500 column coupled to a MicroAkta was pre-equilibrated in CCAN buffer. 100 µL
627 2.5 mM $BS(PEG)_5$ (ThermoFisherScientific) was injected into the system for 1 mL at a flow
628 rate of 0.07 µL/min. The CCAN:DNA complex was then injected into the column and eluted
629 at the same flow rate into a 96-well plate in 50 µL fractions. Each plate well was
630 supplemented with 5 µL 1 M Tris.HCl buffer pH 7.4 to quench the crosslinker during elution.
631 Fractions containing CCAN:DNA were pooled, buffer exchanged into 150 mM NaCl, 20 mM
632 Hepes pH 7.8, 0.5 mM TCEP and concentrated to 1.5 mg/mL (Amicon 100 kDa, 0.5 mL).
633 Following concentration, CCAN:DNA complexes were again centrifuged at 10 °C in a table-
634 top centrifuge for 10 min.
635
636 CENP-A mono-nucleosome (ASW6) and CENP-A di-nucleosome (ASW14) reconstitution
637
638 The CENP-A histone octamer was prepared as described previously [12]. Either ASW6 or
639 ASW14 DNA were desalted into buffer B (20 mM HEPES pH 7.5, 2 M KCl). DNA at 8 µM was
640 mixed with the CENP-A histone octamer at molar ratios of 1.1:1 protein:DNA for mono-
641 CENP-A^{Nuc} and 2.1:1 protein:DNA for di-CENP-A^{Nuc}. The histone octamers were then wrapped
642 with DNA using the gradient dialysis method starting from 2 M KCl (in 20 mM HEPES pH 7.5,
643 1 mM EDTA) to 300 mM KCl over 16 h at 22 °C. The mixture was then dialysed against 20
644 mM HEPES pH 7.5, 300 mM KCl, 1 mM EDTA for a further 5 h. The wrapped mono-CENP-A
645 and di-CENP-A nucleosomes were assessed by agarose gel and stored at 4 °C.
646
647 Assembly of the CCAN:CENP-A^{Nuc} (ASW6 DNA) complex
648
649 CENP-A^{Nuc} and CCAN components (CENP-C^N, CENP-LN, CENP-HIKM, CENP-TWSX, and CENP-
650 $OPQUR^{\Delta N}$) were mixed at a 1:1 molar ratio at 5 µM in PBS buffer and TCEP. The sample was
651 injected into an Agilent Bio SEC-5 1000A 4.6x300 column in CCAN gel filtration buffer (20
652 mM HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP). The peak fractions were analysed using a 4-
653 12% SDS-PAGE gel, stained with Instant Blue Coomassie Stain. To prepare the complex for
654 structural analysis, on-column cross-linking using $BS(PEG)_5$ (ThermoFisherScientific) was
655 performed. 100 µl of 2.5 mM $BS(PEG)_5$ was injected into an Agilent Bio SEC-5 1000A 4.6x300
656 column (equilibrated in PBS buffer) at a flow rate set of 0.07 µL/min, then 0.2 mL buffer,
657 after which the CCAN:CENP-A^{Nuc} sample in (100 µl) was applied and eluted in PBS buffer. The
658 sample was eluted into a 96-well plate in 50 µL fractions. Each plate well was supplemented
659 with 5 µL 1 M Tris.HCl buffer pH 7.4 to quench the crosslinker during elution. The peak

660 fractions of the CCAN:CENP-A^{Nuc} samples were pooled and concentrated to 1.5 mg/mL for
661 cryo-EM grid preparation.
662
663 Assembly of the CCAN:di-CENP-A^{Nuc} (ASW14 DNA) complex
664
665 For assembly of the CCAN:di-CENP-A^{Nuc} (ASW14 DNA) complex, we used the same
666 procedure as for the CCAN:mono-CENP-A^{Nuc} (ASW6 DNA) complex except we used 20 mM
667 HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP buffer in the cross-linking final gel filtration step.
668
669 Assembly of the CCAN$^{\Delta CENP-C}$:di-CENP-A^{Nuc} (ScFv) (ASW14 DNA) complex
670
671 *1. Cross-linking di-CENP-A^{Nuc} and ScFv (di-CENP-A^{Nuc}:ScFv)*
672 Di-CENP-A^{Nuc} and ScFv were combined at a 1:2.3 molar ratio and dialysed into a buffer of 20
673 mM HEPES pH 7.5, 50 mM NaCl. Cross-linking of the di-CENP-A^{Nuc}:ScFv complex was
674 performed using a 24 mL Superose 6 column by injecting 500 μL of 2.5 mM BS(PEG)$_5$ at a
675 flow rate of 0.25 mL.min^{-1} for 7 mL before loading the CENP-A^{Nuc}:ScFv complex onto the
676 column at the same flow rate. The cross-linking was assessed using SDS-PAGE gel. The cross-
677 linked fractions were pooled and concentrated to 10 μM.
678
679 *2. Assembly of cross-linked di-CENP-A^{Nuc}:ScFv with CCAN*
680 All CCAN components (omitting CENP-C^{N}) were without DTT. The components were at a
681 1:1:1:1:1 molar ratio (CENP-OPQUR:CENP-LN:CENP-HIKM:CENP-TWSX:di-CENP-A^{Nuc}:ScFv)
682 with CCAN components at a concentration of 5 μM. The resultant complex was cross-linked
683 on an Agilent Bio SEC-5 1000A 4.6x300 column. Collected fractions were concentrated to
684 1.5-2 mg/mL for cryo-EM grid preparation.
685
686 CCAN displaces CENP-B from di-CENP-A^{Nuc}-B-box
687
688 *Di-CENP-A^{Nuc}-CENP-B binding test*. Di-CENP-A^{Nuc} and CENP-B were mixed at a 1:2 ratio at 5
689 μM final concentration and injected into a Superose 6 Increase 10/300 column (Cytiva).
690 Appropriate fractions were collected and concentrated.
691
692 *Di-CENP-A^{Nuc}-CENP-B-CCAN binding test*. An equal molar ratio of CCAN$^{\Delta CENP-C}$ components
693 (CENP-LN, CENP-HIKM, CENP-TWSX, and CENP-OPQUR$^{\Delta N}$) was added to the di-CENP-
694 A^{Nuc}:CENP-B complex at a final concentration of 5 μM of di-CENP-A^{Nuc} and 10 μM CENP-B
695 and then loaded onto an Agilent Bio SEC-5 1000A 4.6x300 column. The eluted fractions
696 were visualized using a 4-12% SDS-PAGE gel, stained with InstantBlue Coomassie Stain. In

control experiments to assess the elution volume of CENP-B alone, and the di-CENP-A^{Nuc}:CENP-B complex, the individual samples were run on the same Agilent Bio SEC-5 1000A 4.6x300 column as the di-CENP-A^{Nuc}:CENP-B:CCAN sample (**Supplementary Fig. 6d-i**).

Mass photometry

To assess the oligomeric state of CCAN:di-CENP-A^{Nuc} and $CCAN^{\Delta CENP-C}$:di-CENP-A^{Nuc}, 18 µL of sample at approximately 100 nM was applied to wells in silicone gaskets on clean coverslips (Refeyn, UK) on a OneMP instrument (Refeyn). Movies were imaged for 60 s using the AcquireMP software (Refeyn) and the results were processed using the DiscoverMP software. The data were analysed by using automatic Gaussian fitting. The contrast intensities of landing events due to interference scattering were converted to masses using a calibration performed with a standard set of protein oligomers (P1, Refeyn) under similar buffer conditions. Data events were exported and plotted in Prism (GraphPad).

Cryo-EM grid preparation and data collection

CCAN:DNA sample vitrification

Immediately prior to plunge-freezing, CCAN:DNA samples were supplemented with n-octyl-β-D-glucopyranoside (Glycon-biochem) to a final concentration of 0.05%. Quantifoil 300 mesh copper R1.2/1.3 grids (Quantifoil Micro Tools) were rendered hydrophilic by glow discharge in a Edwards S150B glow discharger for 1 min 15 s, setting 6, 30-35 mA, 1.2 kV, 0.2 mBar (0.15 Torr). Subsequently, 3 µL sample was applied to the grids, which were then mounted in a Vitrobot IV (ThermoFisherScientific), blotted with Whatman Filter paper 1 (0-0.5 s waiting time, 2 s blotting time, -7 blotting force), and plunge frozen in liquid ethane maintained at a temperature of -170 °C.

CCAN:DNA data acquisition

10,628 micrograph movies were collected on a Titan Krios (ThermoFisherScientific) electron microscope, equipped with a K3 direct electron detector (Gatan) and energy filter (Gatan) with a slit width of 20 eV, operating at 300 keV with a flux of 25 e^-/pixel/s, at a nominal magnification of 105k, yielding a pixel size of 0.725 Å/pixel and a total dose of 40e/$Å^2$. Automated data acquisition was performed using EPU (ThermoFisherScientific) with aberration-free image shift enabled, and a defocus range of -0.8 to -2.2 µm in 0.2 µm increments.

CCAN:CENP-A^{Nuc} complex sample vitrification and cryo-EM data acquisition

To prepare the CCAN:CENP-A^{Nuc} complex for cryo-EM, 3.5 µL of sample at a concentration of 2 mg/mL was applied to glow-discharged Quantifoil R2/1 300 mesh copper grids (Quantifoil

735 Micro Tools). Grids were blotted for 2 s at 4°C , 100% humidity, and a blotting force of -10,
736 then plunge-frozen in liquid ethane using a Vitrobot Mark IV (ThermoFisherScientific). Data
737 were collected on a Titan Krios microscope (ThermoFisherScientific) operated at 300 kV,
738 equipped with a K3 direct electron detector (Gatan) and an energy filter (Gatan) with an
739 energy slit of 20 eV. Images were acquired at a nominal magnification of 81,000×,
740 corresponding to a pixel size of 1.072 Å. Movies were recorded at a dose rate of 15 e^-/pixel/s
741 with an exposure time of 3 s over 40 frames. Automated data collection was performed
742 using EPU (ThermoFisherScientific) with aberration-free image shift (AFIS) enabled. The
743 defocus range was set from -1.0 to -2.0 µm in 0.2 µm increments. A total of 50,781 movie
744 frames were collected.
745
746 *CCAN:di-CENP-A^{Nuc} complex sample vitrification and cryo-EM data acquisition*
747 To prepare the CCAN:di-CENP-A^{Nuc} complex for cryo-EM, 3.5 µL of sample at a concentration
748 of 2 mg.mL^{-1} was applied to glow-discharged Quantifoil R1.2/1.3 300 mesh copper grids
749 (Quantifoil Micro Tools). Grids were blotted for 2 s at 4°C , 100% humidity, and a blotting
750 force of -7, then plunge-frozen in liquid ethane using a Vitrobot Mark IV
751 (ThermoFisherScientific). Data were collected on a Titan Krios microscope
752 (ThermoFisherScientific) operated at 300 kV, equipped with a K3 direct electron detector
753 (Gatan) and an energy filter (Gatan) with an energy slit of 20 eV. Images were acquired at a
754 nominal magnification of 81,000×, corresponding to a pixel size of 0.928 Å. Movies were
755 recorded at a dose rate of 25 e^-/pixel/s with an exposure time of 1.3 s over 82 frames.
756 Automated data collection was performed using EPU (ThermoFisherScientific) with
757 aberration-free image shift (AFIS) enabled. The defocus range was set from -1.0 to -2.0 µm in
758 0.2 µm increments. A total of 28,703 movie frames were collected.
759
760 <u>Cryo-EM data processing</u>
761
762 *CCAN:DNA complex*
763 Micrograph movies were imported into RELION [45] and motion-corrected using RELION's own
764 implementation of MotionCor2 [56], followed by contrast transfer function estimation using
765 the RELION wrapper for CTFFIND4 [57]. Particle picking was performed using the crYOLO
766 general model and particles were extracted in RELION with a box size of 150 pixels. Particles
767 were subsequently exported to cryoSPARC and subjected to 2D classification. Particles
768 exhibiting clear secondary structure features were used to generate an *ab initio*
769 reconstruction. The initial particle set was then subjected to three rounds of heterogeneous
770 refinement against one true volume (the CCAN:DNA *ab initio* reconstruction) and three
771 noise decoy volumes arising from *ab initio* reconstruction of particles with no obvious

772 proteo-typic features. Particles accumulating in the CCAN:DNA class were selected for each
773 subsequent round of refinement until few or no particles reclassified into the noise decoy
774 classes. The cleaned particle set was then subjected to homogeneous refinement, followed
775 by 3D variability analysis. Particles exhibiting the most continuous DNA and CCAN subunit
776 density were re-extracted without binning in RELION and subjected to particle polishing and
777 CTF refinement. Finally, polished and CTF-corrected particles were re-imported into
778 cryoSPARC where a final round of homogeneous refinement was performed, yielding a 3D
779 reconstruction of CCAN:DNA to an overall resolution of 3.54 Å. The resulting map was
780 filtered according to local resolution, as estimated in cryoSPARC, and then used for structural
781 model building (in COOT) [58] and real-space refinement in PHENIX [59].
782
783 *CCAN:CENP-A^{Nuc} complex*
784 For the CCAN:CENP-A^{Nuc} complex, movie frames were motion-corrected, followed by
785 estimation of Contrast transfer function (CTF) parameters using the patch-based method in
786 cryoSPARC [60]. Particle picking was performed with crYOLO [61] using the general model, and
787 particles were extracted with a box size of 400 pixels. Reference-free 2D classification was
788 then conducted to clean the particle dataset. Classes exhibiting clear secondary structure
789 features were selected for *ab-initio* reconstruction and heterogeneous refinement in
790 cryoSPARC. One class with 1,525,520 particles showing nucleosome features was selected
791 for non-uniform refinement, yielding a reconstruction of the CENP-A^{Nuc}:CENP-C^{N} at 2.4 Å
792 resolution. Another dominant class containing 3,815,026 particles was import into RELION
793 for further 3D classification. Three classes displaying CCAN feature were selected for particle
794 subtraction using a mask covering 3' CENP-A^{Nuc} region. After an additional round of 2D and
795 3D classification on the recentered particle dataset, 325,201 well-aligned nucleosome
796 particles were reverted to their original, un-subtracted particles for further 3D classification.
797 This yielded a subset of 39,036 particles corresponding to the CCAN:CENP-A^{Nuc} complex,
798 which was refined to 4.5 Å resolution using Refine3D followed by post-processing.
799
800 *CCAN:di-CENP-A^{Nuc} complex*
801 For the CCAN:di-CENP-A^{Nuc} complex, all movie frames were motion-corrected, followed by
802 estimation of Contrast transfer function (CTF) parameters using the patch-based method in
803 cryoSPARC. Particle picking was performed with crYOLO using the general model, and the
804 particles were extracted with a box size of 400 pixels. Reference-free 2D classification was
805 then conducted to clean the particle dataset, and classes exhibiting clear secondary
806 structure features were selected for *ab-initio* reconstruction and heterogeneous refinement
807 in cryoSPARC. Three well-populated classes with clear structural features were further
808 selected for non-uniform refinement, CTF refinement and local refinement, resulting three

high-resolution maps: CENP-A^{Nuc}:ScFv (1:1), CENP-A^{Nuc}:ScFv (1:2) and CCAN:DNA, resolved at 2.7 Å, 2.7 Å and 2.9 Å, respectively.

To dissect the CENP-A^{Nuc} association with CCAN, the CCAN:DNA data was exported to RELION [45] for 3D classification, which revealed diffuse density at both the 3' and 5' end of the DNA. To recover CENP-A^{Nuc} at each side, two masks were generated and to subtract the density outside either the 3' or 5' CENP-A^{Nuc} regions. After re-centering the particle on the 3' or 5' CENP-A^{Nuc}, nucleosome density was resolved on both sides, as confirmed by 2D classification and 3D classification showing dominant classes with clear nucleosome views. These dominant classes were reverted to their original un-subtracted particles for further 3D classification, yielding two subsets: 9,661 particles of CCAN:3' CENP-A^{Nuc} and 5,396 particles of CCAN:5' CENP-A^{Nuc}. These were refined to 9.5 Å and 15.5 Å respectively. Subsequent multibody refinement and post-process modestly improved the map quality of both the CCAN:DNA and CENP-A^{Nuc} modules. Finally, the maps of CCAN with 3' and 5' CENP-A^{Nuc}, were combined to generate a composite map of the CCAN di-CENP-A^{Nuc} complex.

<u>Cryo-EM model building and refinement</u>

For the CCAN:DNA structure, the PDB entry 7R5S was used as the initial model. The model was fitted into the cryo-EM map of CCAN:DNA, and the DNA wrapping around CENP-TWSX was guided by a predicted CCAN:DNA model generated using AlphaFold3, followed by manual correction in COOT and real-space refinement in PHENIX. For the CCAN:CENP-A^{Nuc} structure, the newly refined CCAN:DNA complex model was docked into the corresponding map, followed by docking of a 3' CENP-A^{Nuc} using PDB entry 7YWX. The models were then combined in COOT and real-space refined in PHENIX. For the CCAN:di-CENP-A^{Nuc} model, a composite map combining CCAN:3' CENP-A^{Nuc} and CCAN:5' CENP-A^{Nuc} was used. The CCAN:DNA model was first fitted into the map, followed by docking of CENP-A^{Nuc} models (PDB: 6E0P) to the 3' and 5' nucleosomes, respectively. All models were then combined in COOT, and the linker DNA region was manually built and refined. Additional manual corrections were also performed in COOT. Figures were generated using ChimeraX [62].

PDB and cryo-EM maps have been deposited with RCSB and EMDB, respectively. Accession numbers are listed in **Supplementary Table 1**. Source data are provided with this paper. Correspondence and requests for materials should be addressed to David Barford.

Data availability

PDB coordinates generated in this study have been deposited with RCSB under accession codes 9TAW [https://www.rcsb.org/structure/9TAW], 9TAX [[https://www.rcsb.org/structure/9TAX] and 9TAY [https://www.rcsb.org/structure/9TAY]. Cryo-EM maps generated in this study have been deposited with EMDB under accession codes EMD-55755 []
EMD-55756 []
EMD-55757 []
EMD-55758 []
EMD-55759 []
Previously published protein coordinates used in study: 7R5S [https://www.rcsb.org/structure/7R5S], 7YWX [https://www.rcsb.org/structure/7YWX], 6E0P [https://www.rcsb.org/structure/6E0P] and 1AOI [https://www.rcsb.org/structure/1AOI]. Accession numbers are also listed in **Supplementary Table 1**. Source data are provided with this paper. Correspondence and requests for materials should be addressed to David Barford.

Acknowledgements
We are grateful to the LMB EM Facility and the UK's national Electron Bioimaging Centre (eBIC) (under proposal BI37660 funded by the Wellcome Trust, MRC and BBSRC) for help with the EM data collection; J. Grimmett, T. Darling and I. Clayson for scientific computing; K. Turton for help with insect cell expression. We thank Andrea Musacchio, Marion Pesenti and Ingrid Vetter for sharing unpublished experimental observations. For the purpose of open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising.

Funding UKRI/Medical Research Council MC_UP_1201/6 (D.B.)
Cancer Research UK C576/A25675 (D.B.).

Author Contributions The project was conceived by D.B. Experiments were performed by C.Y, K.W.M., J.Y., Z.Z., D.B. The paper was written by D.B. with input from all other authors.

Competing Interests Authors declare that they have no competing interests.

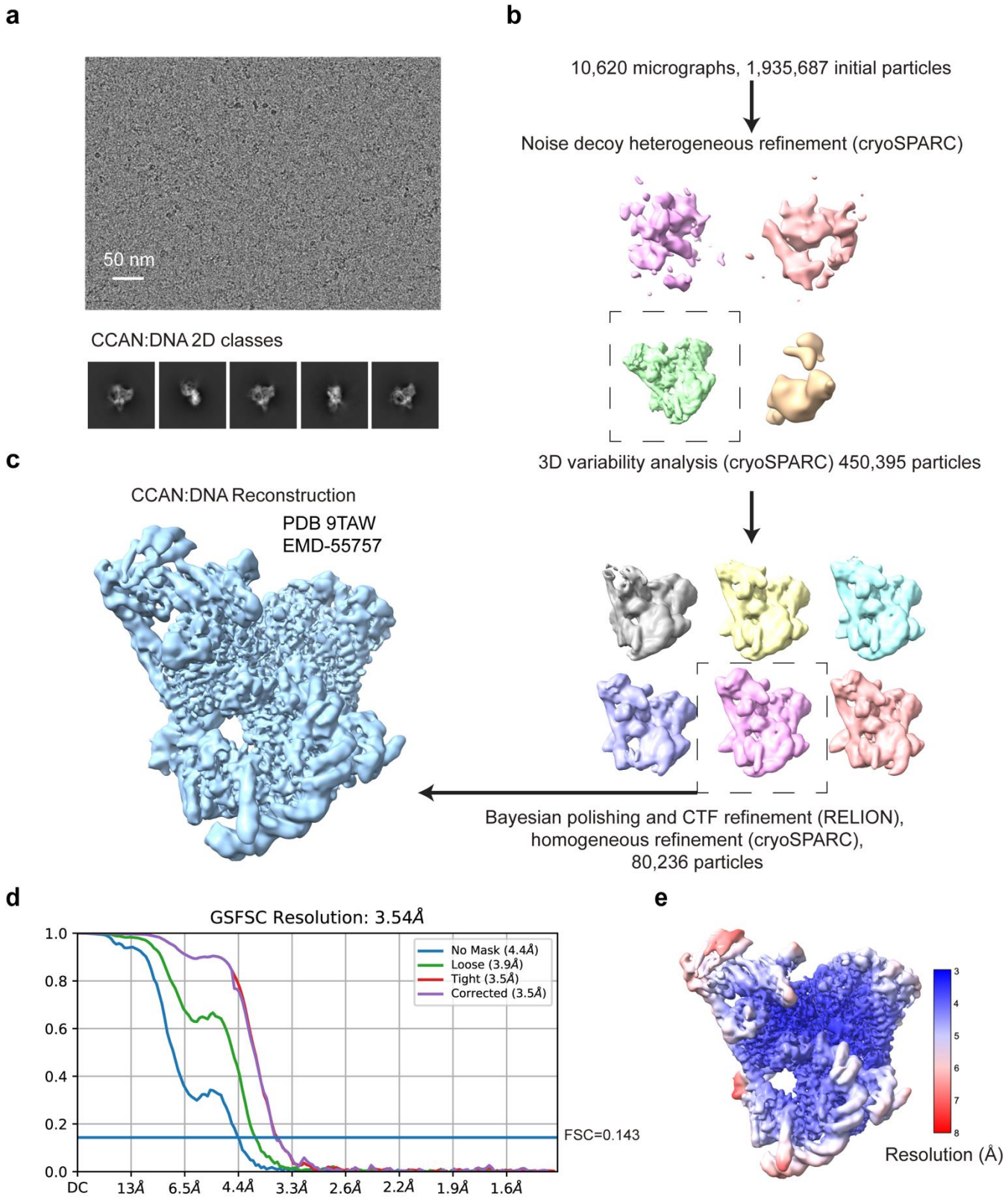
References

1. Ariyoshi, M. & Fukagawa, T. An updated view of the kinetochore architecture. *Trends Genet* **39**, 941–953 (2023).
2. McAinsh, A.D. & Marston, A.L. The Four Causes: The Functional Architecture of Centromeres and Kinetochores. *Annu Rev Genet* **56**, 279–314 (2022).
3. Musacchio, A. & Desai, A. A Molecular View of Kinetochore Assembly and Function. *Biology (Basel)* **6**(2017).

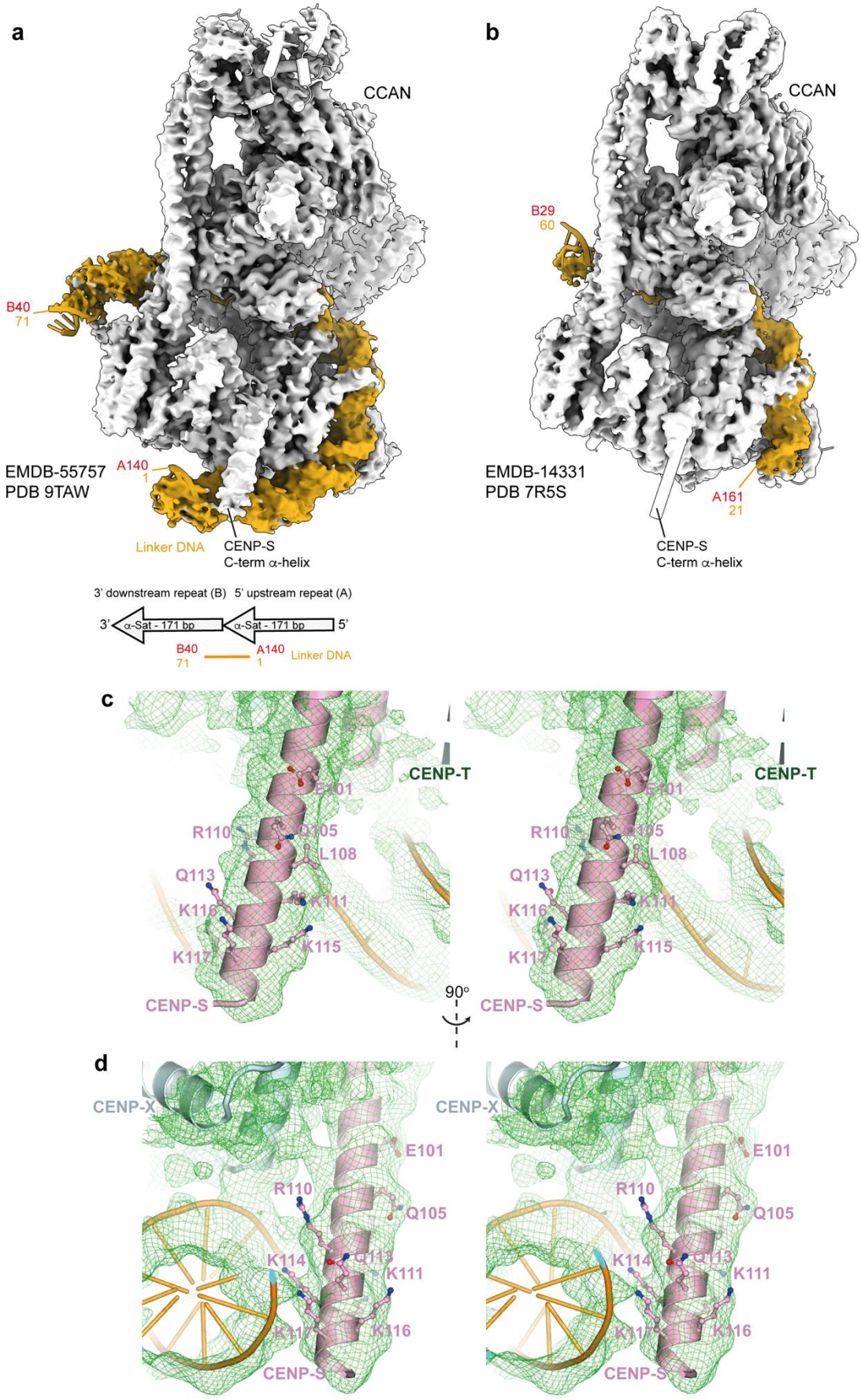
887 4. McAinsh, A.D. & Kops, G. Principles and dynamics of spindle assembly checkpoint
888 signalling. *Nat Rev Mol Cell Biol* **24**, 543–559 (2023).
889 5. Fischer, E.S. Kinetochore-catalyzed MCC formation: A structural perspective. *IUBMB*
890 *Life* **75**, 289–310 (2023).
891 6. Altemose, N. et al. Complete genomic and epigenetic maps of human centromeres.
892 *Science* **376**, eabl4178 (2022).
893 7. Bodor, D.L. et al. The quantitative architecture of centromeric chromatin. *Elife* **3**,
894 e02137 (2014).
895 8. Logsdon, G.A. et al. The structure, function and evolution of a complete human
896 chromosome 8. *Nature* **593**, 101–107 (2021).
897 9. Gershman, A. et al. Epigenetic patterns in a complete human genome. *Science* **376**,
898 eabj5089 (2022).
899 10. Blower, M.D., Sullivan, B.A. & Karpen, G.H. Conserved organization of centromeric
900 chromatin in flies and humans. *Dev Cell* **2**, 319–30 (2002).
901 11. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal
902 structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–60
903 (1997).
904 12. Yatskevich, S. et al. Structure of the human inner kinetochore bound to a centromeric
905 CENP-A nucleosome. *Science* **376**, 844–852 (2022).
906 13. Hasson, D. et al. The octamer is the major form of CENP-A nucleosomes at human
907 centromeres. *Nat Struct Mol Biol* **20**, 687–95 (2013).
908 14. Corda, L. & Giunta, S. Chromosome-specific centromeric patterns define the centeny
909 map of the human genome. *Science* **389**, eads3484 (2025).
910 15. Fujita, R. et al. Stable complex formation of CENP-B with the CENP-A nucleosome.
911 *Nucleic Acids Res* **43**, 4909–22 (2015).
912 16. Nagpal, H. et al. CENP-A and CENP-B collaborate to create an open centromeric
913 chromatin state. *Nat Commun* **14**, 8227 (2023).
914 17. Chardon, F. et al. CENP-B-mediated DNA loops regulate activity and stability of
915 human centromeres. *Mol Cell* **82**, 1751–1767 e8 (2022).
916 18. Magidson, V. et al. Adaptive changes in the kinetochore architecture facilitate proper
917 spindle assembly. *Nat Cell Biol* **17**, 1134–44 (2015).
918 19. Yatskevich, S., Barford, D. & Muir, K.W. Conserved and divergent mechanisms of inner
919 kinetochore assembly onto centromeric chromatin. *Curr Opin Struct Biol* **81**, 102638
920 (2023).
921 20. Hori, T. et al. CCAN makes multiple contacts with centromeric DNA to provide distinct
922 pathways to the outer kinetochore. *Cell* **135**, 1039–52 (2008).
923 21. Pesenti, M.E. et al. Structure of the human inner kinetochore CCAN complex and its
924 significance for human centromere organization. *Mol Cell* **82**, 2113–2131 e8 (2022).
925 22. Dendooven, T. et al. Cryo-EM structure of the complete inner kinetochore of the
926 budding yeast point centromere. *Sci Adv* **9**, eadg7480 (2023).
927 23. Tian, T. et al. Structural insights into human CCAN complex assembled onto DNA. *Cell*
928 *Discov* **8**, 90 (2022).
929 24. Yan, K. et al. Structure of the inner kinetochore CCAN complex assembled onto a
930 centromeric nucleosome. *Nature* **574**, 278–282 (2019).
931 25. Hinshaw, S.M. & Harrison, S.C. The structure of the Ctf19c/CCAN from budding yeast.
932 *Elife* **8**(2019).

26. Yu, C.W.H. et al. Architecture and Function of Holocentric CENP-A-Independent Kinetochores. *BioRxiv* (2025).
27. Kato, H. et al. A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340**, 1110–3 (2013).
28. Klare, K. et al. CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. *J Cell Biol* **210**, 11–22 (2015).
29. McKinley, K.L. et al. The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. *Mol Cell* **60**, 886–98 (2015).
30. Pentakota, S. et al. Decoding the centromeric nucleosome through CENP-N. *Elife* **6**(2017).
31. Walstein, K. et al. Assembly principles and stoichiometry of a complete human kinetochore module. *Sci Adv* **7**(2021).
32. Nishino, T. et al. CENP-T-W-S-X forms a unique centromeric chromatin structure with a histone-like fold. *Cell* **148**, 487–501 (2012).
33. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–91 (2009).
34. Zhao, Q. et al. The MHF complex senses branched DNA by binding a pair of crossover DNA duplexes. *Nat Commun* **5**, 2987 (2014).
35. Dechassa, M.L. et al. Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat Commun* **2**, 313 (2011).
36. Kingston, I.J., Yung, J.S. & Singleton, M.R. Biophysical characterization of the centromere-specific nucleosome from budding yeast. *J Biol Chem* **286**, 4021–6 (2011).
37. Migl, D. et al. Cryoelectron Microscopy Structure of a Yeast Centromeric Nucleosome at 2.7 A Resolution. *Structure* **28**, 363–370 e3 (2020).
38. Tachiwana, H. et al. Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **476**, 232–5 (2011).
39. Ariyoshi, M. et al. Cryo-EM structure of the CENP-A nucleosome in complex with phosphorylated CENP-C. *EMBO J* **40**, e105671 (2021).
40. Zhou, B.R. et al. Atomic resolution cryo-EM structure of a native-like CENP-A nucleosome aided by an antibody fragment. *Nat Commun* **10**, 2301 (2019).
41. Roulland, Y. et al. The Flexible Ends of CENP-A Nucleosome Are Required for Mitotic Fidelity. *Mol Cell* **63**, 674–685 (2016).
42. Boopathi, R. et al. Phase-plate cryo-EM structure of the Widom 601 CENP-A nucleosome core particle reveals differential flexibility of the DNA ends. *Nucleic Acids Res* **48**, 5735–5748 (2020).
43. Falk, S.J. et al. Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science* **348**, 699–703 (2015).
44. Ali-Ahmad, A., Bilokapic, S., Schafer, I.B., Halic, M. & Sekulic, N. CENP-C unwraps the human CENP-A nucleosome through the H2A C-terminal tail. *EMBO Rep* **20**, e48913 (2019).
45. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S.H.W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* **478**, 4169–4185 (2021).

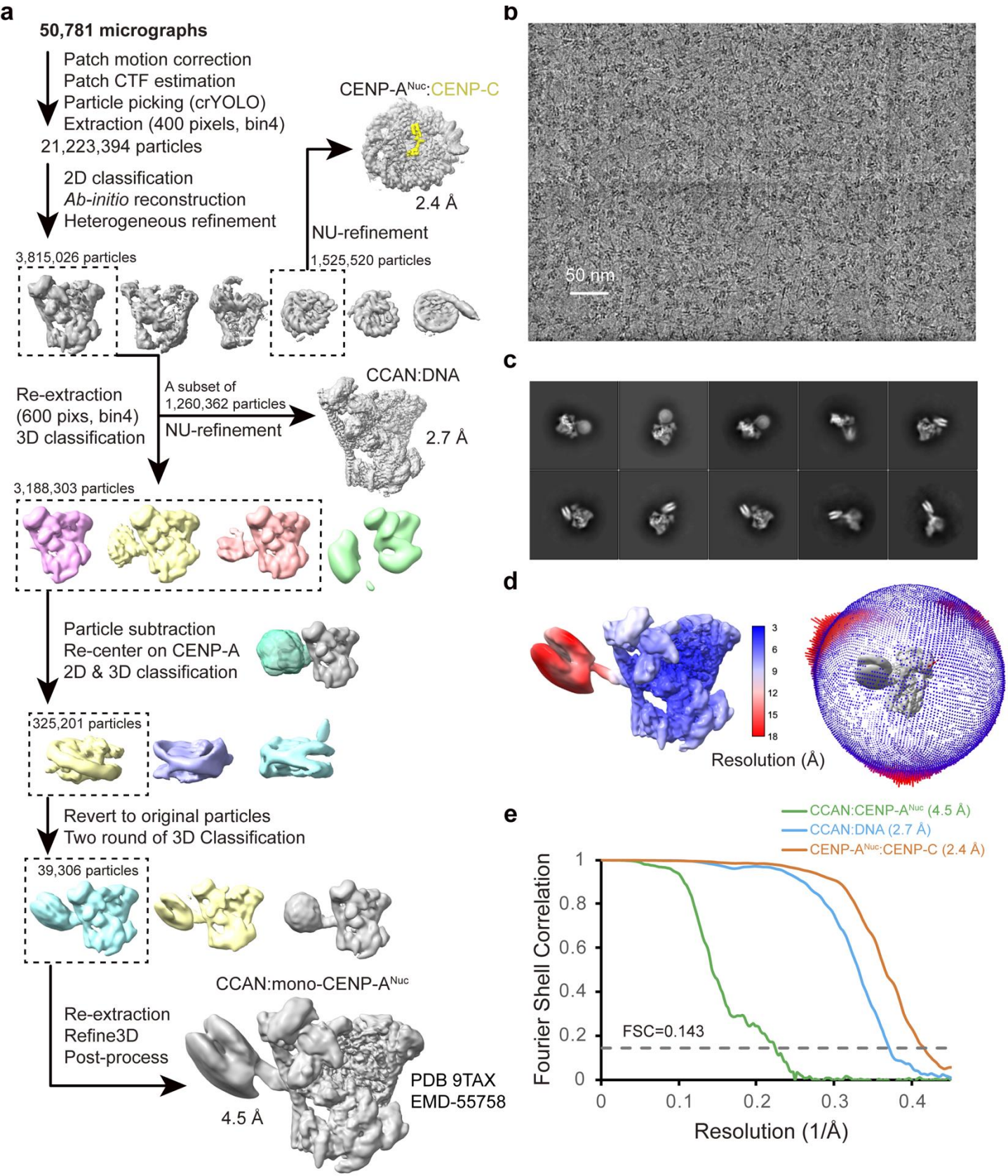
46. Dubocanin, D. et al. Conservation of dichromatin organization along regional centromeres. *Cell Genom* **5**, 100819 (2025).
47. Kixmoeller, K., Tarasovetc, E.V., Mer, E., Chang, Y.W. & Black, B.E. Centromeric chromatin clearings demarcate the site of kinetochore formation. *Cell* **188**, 1280–1296 e19 (2025).
48. Watanabe, R. et al. CDK1-mediated CENP-C phosphorylation modulates CENP-A binding and mitotic kinetochore localization. *J Cell Biol* **218**, 4042–4062 (2019).
49. Guo, L.Y. et al. Centromeres are maintained by fastening CENP-A to DNA and directing an arginine anchor-dependent nucleosome transition. *Nat Commun* **8**, 15775 (2017).
50. Pesenti, M.E. et al. Reconstitution of a 26-Subunit Human Kinetochore Reveals Cooperative Microtubule Binding by CENP-OPQUR and NDC80. *Mol Cell* **71**, 923–939 e10 (2018).
51. Suzuki, A., Badger, B.L. & Salmon, E.D. A quantitative description of Ndc80 complex linkage to human kinetochores. *Nat Commun* **6**, 8161 (2015).
52. Hara, M. et al. Centromere/kinetochore is assembled through CENP-C oligomerization. *Mol Cell* **83**, 2188–2205 e13 (2023).
53. Trazzi, S. et al. The C-terminal domain of CENP-C displays multiple and critical functions for mammalian centromere formation. *PLoS One* **4**, e5832 (2009).
54. Rosenberg, H., Singer, M. & Rosenberg, M. Highly reiterated sequences of SIMIANSIMIANSIMIANSIMIANSIMIAN. *Science* **200**, 394–402 (1978).
55. Manuelidis, L. & Wu, J.C. Homology between human and simian repeated DNA. *Nature* **276**, 92–4 (1978).
56. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).
57. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216–21 (2015).
58. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
59. Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352–67 (2012).
60. Punjani, A., Zhang, H. & Fleet, D.J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214–1221 (2020).
61. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2**, 218 (2019).
62. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14–25 (2018).



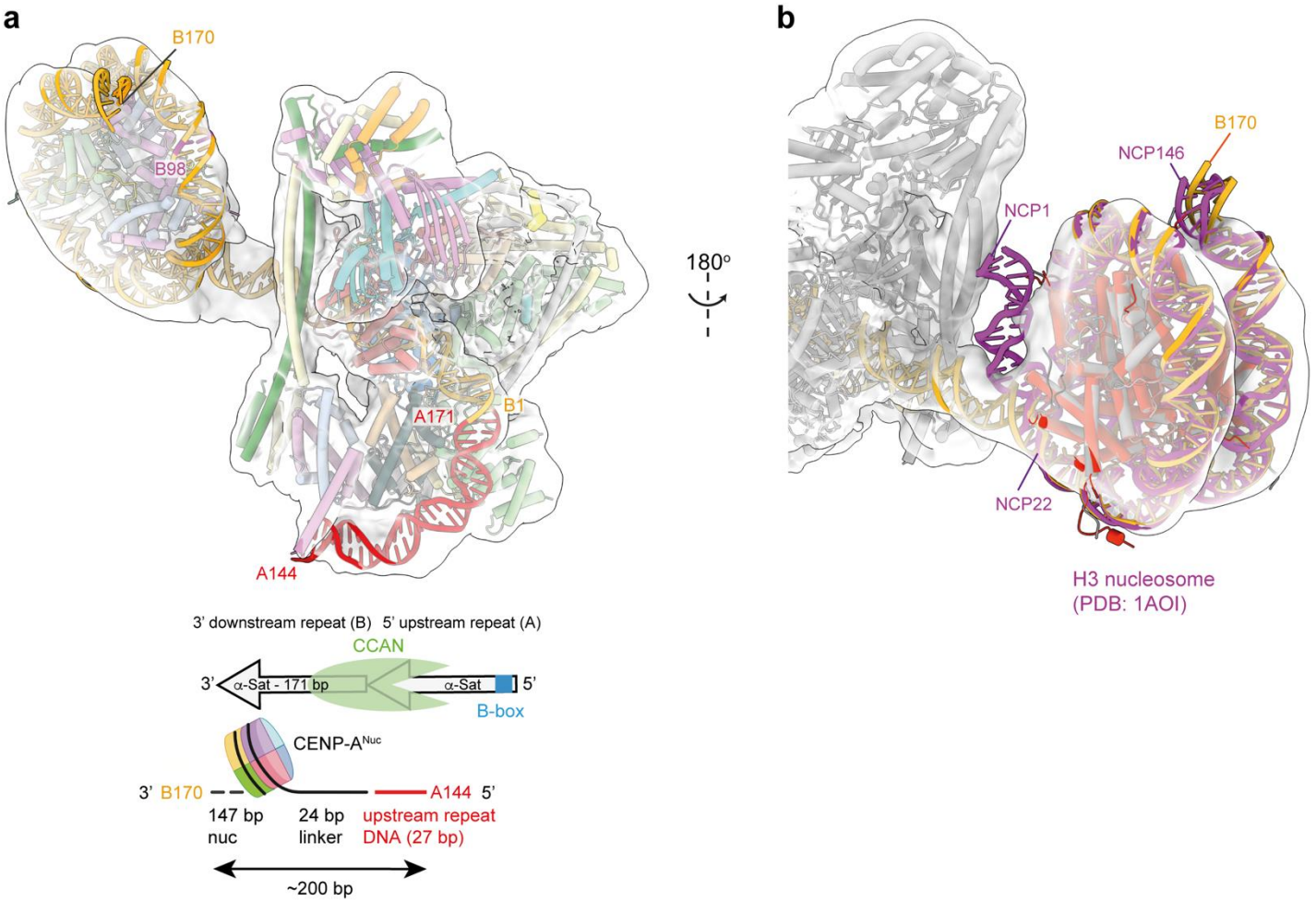
Supplementary Fig. 1. Cryo-EM data and processing workflow for the CCAN:DNA complex. **a**, Representative cryo-electron micrograph of 10,628 collected and 2D class averages for CCAN:DNA. **b**, Cryo-EM processing workflow. **c**, 3D reconstruction. A sharpened map is shown in Fig. 1c. **d**, FSC curves. **e**, Cryo-EM reconstruction colour-coded according to local resolution.



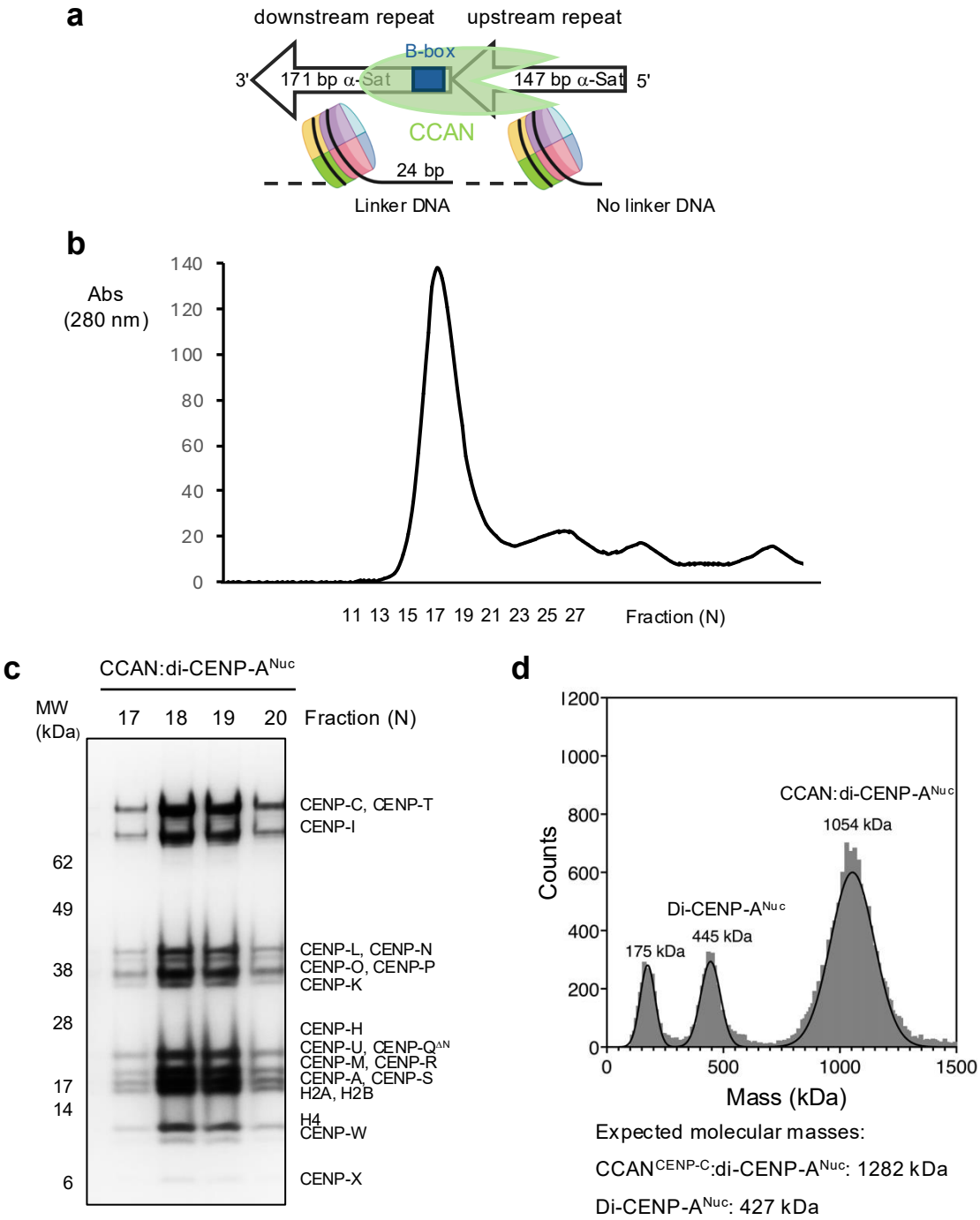
Supplementary Fig. 2. Cryo-EM density fits for DNA and CENP-S α-helix of the CCAN:DNA complex. **a**, Cryo-EM map of CCAN:DNA complex with CCAN density in grey and DNA density in orange. 70 bp of DNA are resolved in density. A140 and B40 refer to the positions on a dimeric α-satellite repeat sequence shown below and as defined in Methods and Fig. 3. **b**, Cryo-EM map of previous CCAN:DNA complex using 54 bp DNA which has 40 bp ordered [12]. Base pair numbering matches (**a**). **c** and **d**, Orthogonal stereo views of cryo-EM density map for the ordered C-terminal α-helix of CENP-S. Residues shown in Fig. 2c are indicated. View in (**c**) as in Fig. 2c.



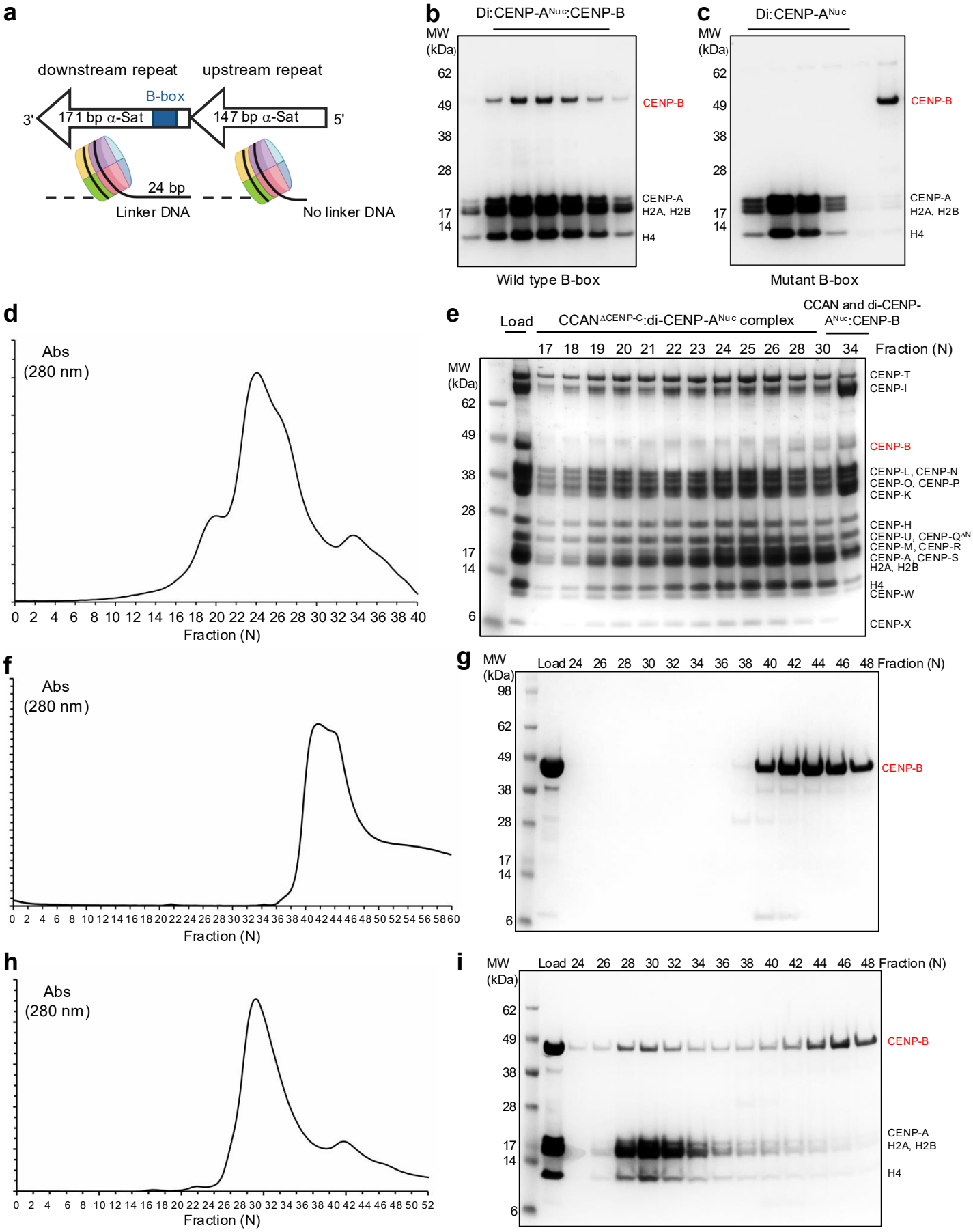
Supplementary Fig. 3. Cryo-EM images and processing workflow for CCAN:mono-CENP-A^{Nuc} complex. **a**, Cryo-EM processing workflow. **b**, Representative cryo-electron micrograph of 50,781 collected. **c**, 2D class average gallery. **d**, Cryo-EM reconstruction colour-coded according to local resolution. **e**, FSC curves. The CCAN:DNA coordinates and map were deposited with codes PDB ID 28OP and EMD-56683. The CENP-A^{Nuc}:CENP-C coordinates and map were not deposited in PDB/RCSB with this study.



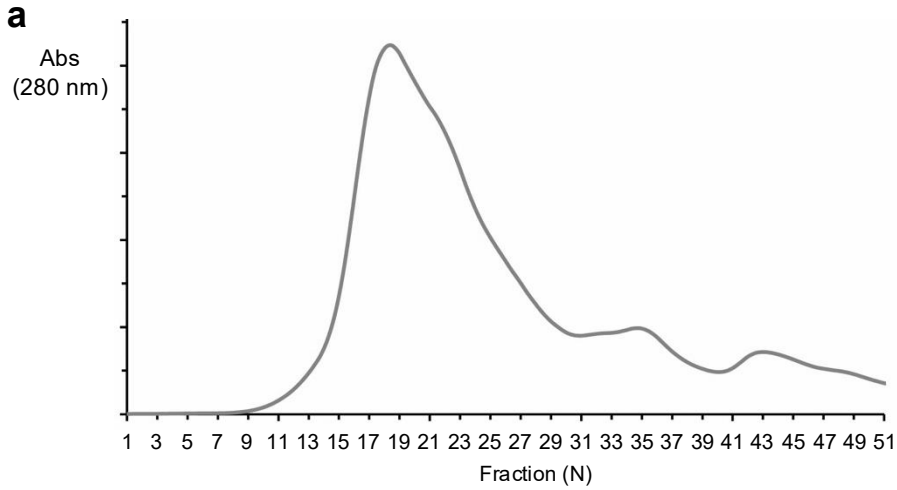
Supplementary Fig. 4. Cryo-EM density fits for DNA of the CCAN:CENP-A^{Nuc} complex. a, Cryo-EM map of CCAN:CENP-A^{Nuc} complex with CCAN density in grey and DNA density in red (5' upstream repeat (A)) and orange (3' downstream repeat (B)). A144 and B144 refer to the positions on a dimeric α-satellite repeat sequence shown in the schematic below and as defined in Methods. B98 indicates the dyad axis of CENP-ANuc. **b**, Superimposition of a canonical H3 nucleosome (NCP) (DNA in magenta, histones in red) onto the CENP-A nucleosome (DNA in orange, histones in grey). This illustrates that ~20 bp of CENP-A^{Nuc} is unwrapped as the DNA gyre enters the CCAN DNA-binding tunnel.

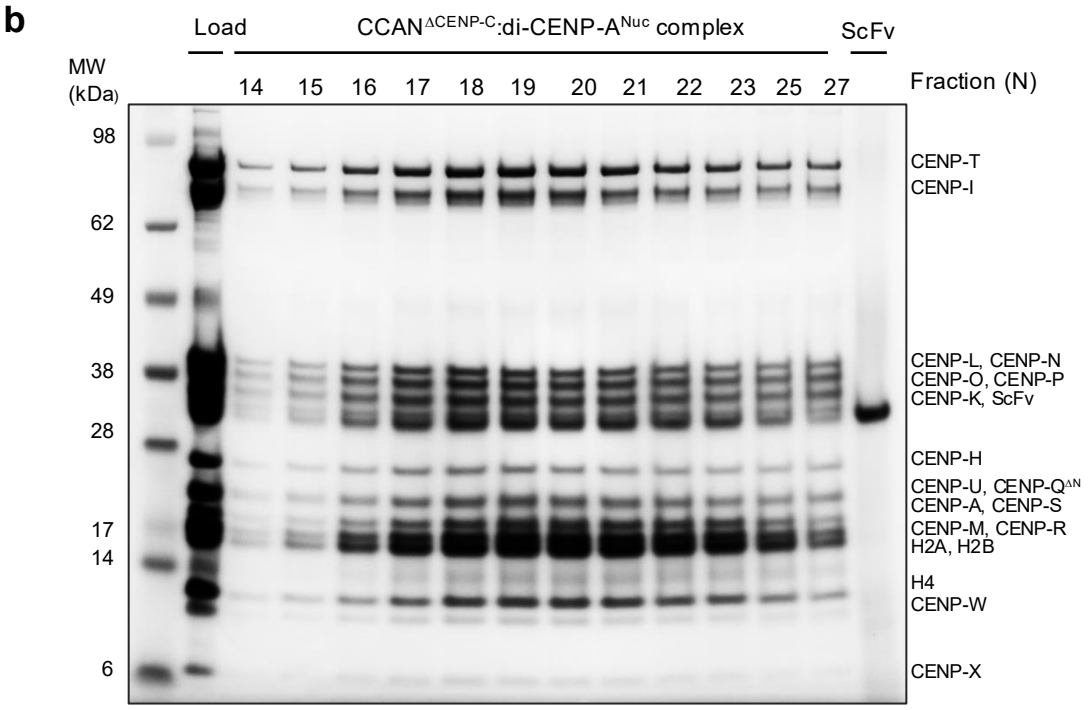


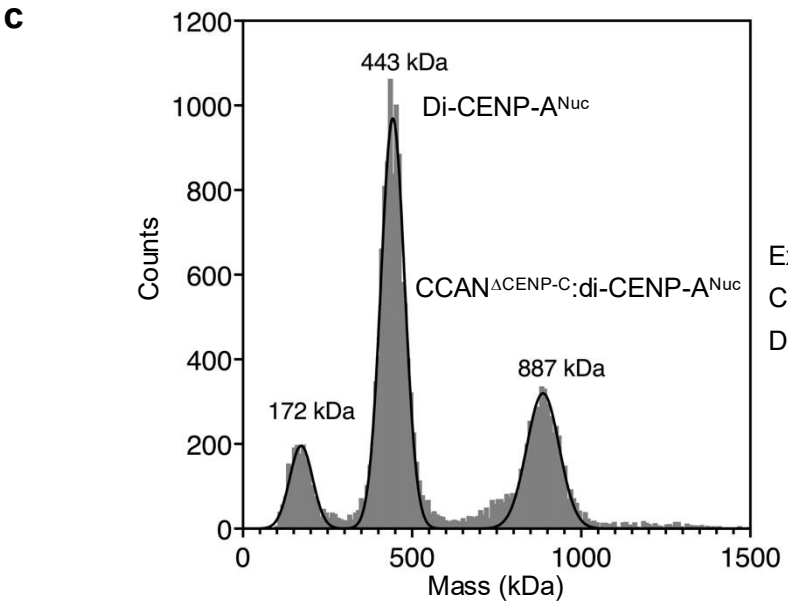
Supplementary Fig. 5. Reconstitution of CCAN:di-CENP-A^{Nuc} complex. a, Schematic of CCAN:di-CENP-A^{Nuc} complex showing α-satellite repeat dimer used in this study. The 3' (downstream) repeat is 171 bp whereas the 5' upstream repeat is 147 bp, i.e. without linker DNA. This design ensures only one CCAN-binding linker DNA per α-satellite repeat dimer. **b** and **c**, Size exclusion chromatogram (b) and associated SDS PAGE gel (c) of purified CCAN:di-CENP-A^{Nuc} complex. **d**, iSCAT data for CCAN:di-CENP-A^{Nuc} complex showing a mixture of di-CENP-A^{Nuc} alone (445 kDa) and the CCAN:di-CENP-A^{Nuc} complex (1,054 kDa).



Supplementary Fig. 6. CCAN association with linker DNA displaces CENP-B bound to linker DNA B-box. a, Schematic of the CCAN:di-CENP-A^{Nuc} complex showing α-satellite repeat dimer used in this study with B-box in blue. SDS PAGE gels of: **b**, di-CENP-A^{Nuc}:CENP-B complex (with wild type B-box) and **c**, showing CENP-B does not bind to di-CENP-A^{Nuc} with a mutated B-box. **d**, and **e**, Size exclusion chromatogram (d) and associated SDS PAGE gel (e) showing that CCAN binding to di-CENP-A^{Nuc} displaces CENP-B in a load sample of di-CENP-A^{Nuc}, CCAN and CENP-B (run on Agilent SEC-5 1000A 4.6x300 column). **f**, and **g**, Control SEC (f) and associated SDS PAGE gel (g) of CENP-B protein run on the same Agilent column. **h**, and **i**, Size exclusion chromatogram (h) and associated SDS PAGE gel (i) of di-CENP-A^{Nuc}:CENP-B complex on the same Agilent column. CENP-B in lanes 28-34 in panel (e) is from di-CENP-A^{Nuc}:CENP-B complex (compare with panel (i)).





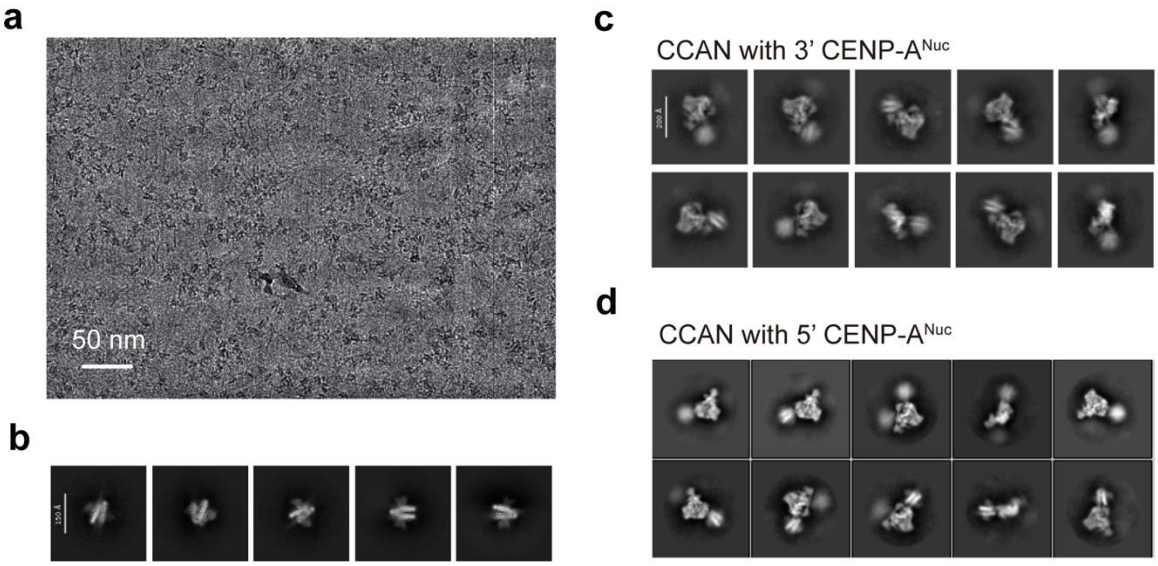


Expected molecular masses:

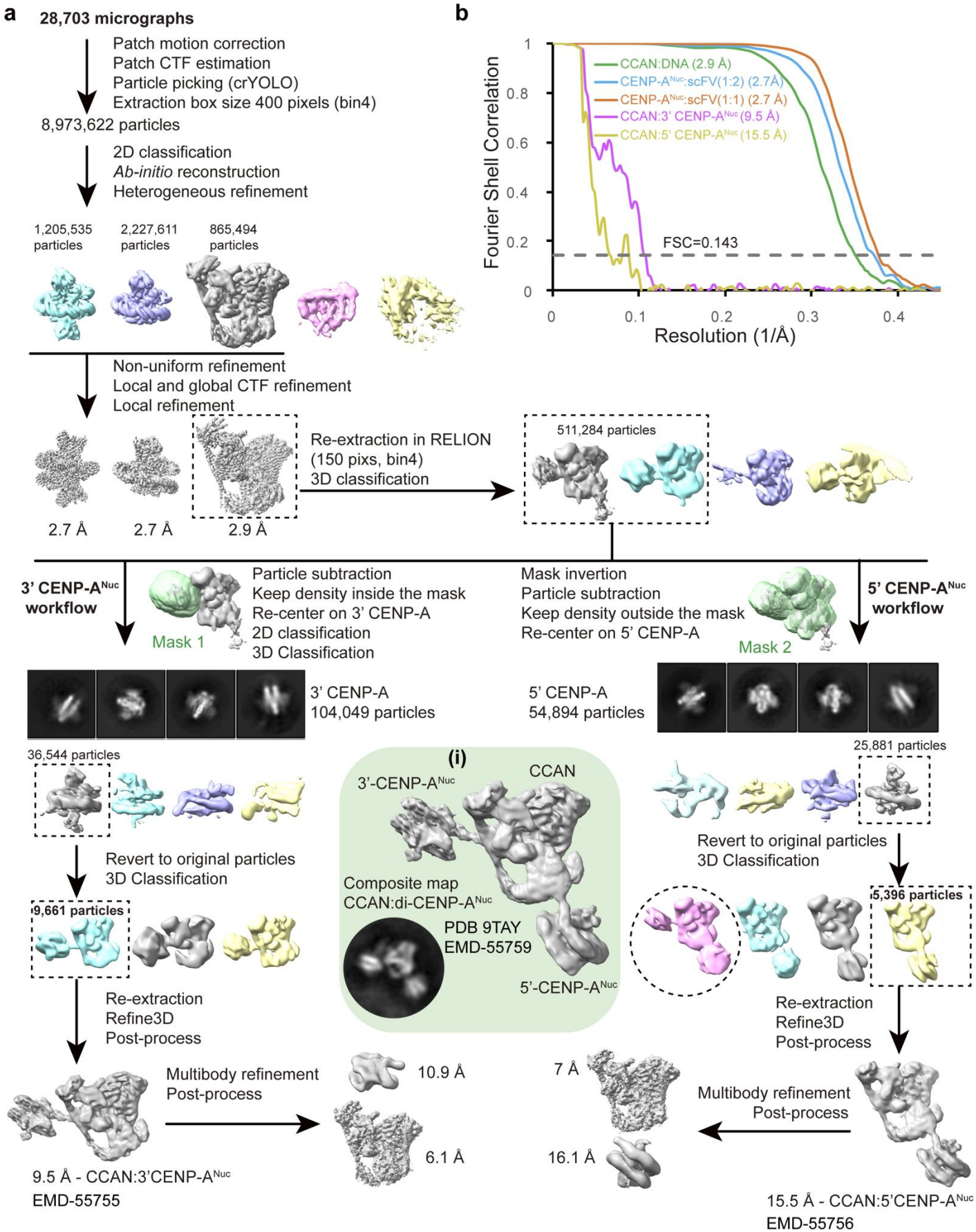
CCAN$^{\Delta CENP-C}$:di-CENP-A^{Nuc}:ScFv 1004 kDa

Di-CENP-A^{Nuc}: 427 kDa

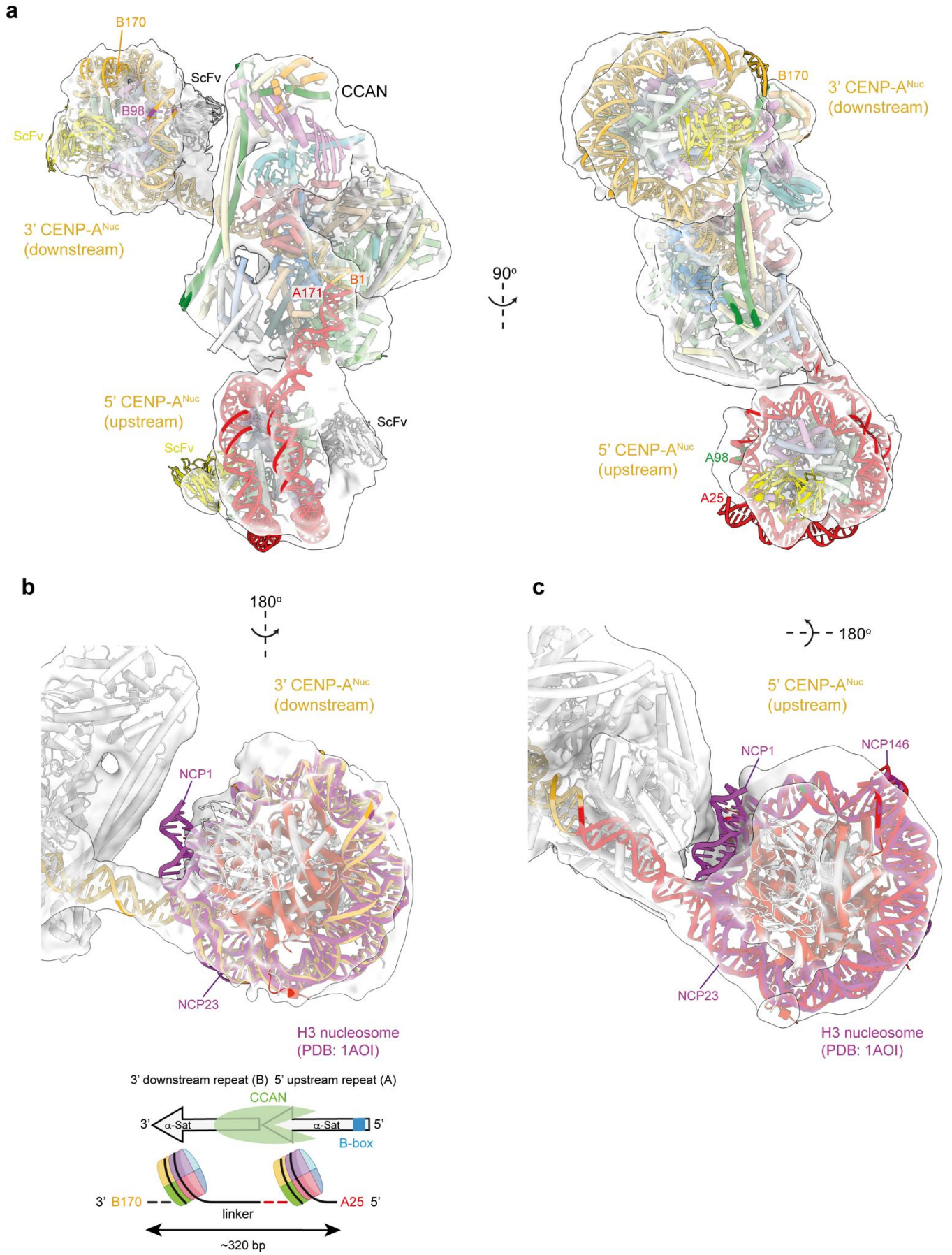
Supplementary Fig. 7. Reconstitution of CCAN$^{DCENP-C}$:di-CENP-A^{Nuc} complex for cryo-EM. a and **b**, Size exclusion chromatogram (A) and associated SDS PAGE gel (B) of purified CCAN$^{\Delta CENP-C}$:di-CENP-A^{Nuc}:ScFv complex. **c**, iSCAT data for CCAN$^{\Delta CENP-C}$:di-CENP-A^{Nuc}:ScFv complex.



Supplementary Fig. 8. Cryo-EM data for the CCAN:di-CENP-A^{Nuc} complex.
a, Representative cryo-electron micrograph of 28,703 collected. **b**, 2D class averages of CENP-A^{Nuc}. **c**, 2D class averages of CCAN with 3' CENP-A^{Nuc}. **d**, 2D class averages of CCAN with 5' CENP-A^{Nuc}.



Supplementary Fig. 9. Cryo-EM workflow for CCAN:di-CENP-A^{Nuc} complex reconstruction. a, Cryo-EM processing workflow with insert **(i)** showing the composite map and 2D class average. **b**, FSC curves. Cryo-EM maps and coordinates for the CCAN:DNA, CENP-A^{Nuc}:ScFv (1:2) and CENP-A^{Nuc}:ScFv (1:1) complexes were not deposited with this study.



Supplementary Fig. 10. Cryo-EM density fits for DNA of the CCAN:di-CENP-A^{Nuc} complex. a, Cryo-EM map of CCAN:di-CENP-A^{Nuc} complex with CCAN density in grey and DNA density in red (5' upstream repeat (A)) and orange (3' downstream repeat (B)). A144 and B144 refer to the positions on a dimeric α-satellite repeat sequence shown in the schematic below and as defined in Methods. B98 indicates the dyad axis of CENP-ANuc. **b**, Superimposition of a canonical H3 nucleosome (NCP) (DNA in magenta, histones in red) onto the CENP-A nucleosome (DNA in orange, histones in grey). This illustrates that ~20 bp of CENP-A^{Nuc} is unwrapped as the DNA gyre enters the CCAN DNA-binding tunnel.

Extended Data Table 1. Cryo-EM data collection, refinement, and validation statistics

Complex	CCAN:DNA PDB 9TAW EMD-55757 (Supplementary 1c and sharpened map Fig. 1c)	CCAN:mono-CENP-A^{Nuc} PDB 9TAX EMD-55758	CCAN:di-CENP-A^{Nuc} PDB 9TAY EMD-55755, EMD55756, EMD-55759 (composite map), EMD-56612 (consensus map) (Supplementary Fig. 9)	
Data collection and Processing				
Microscope	Titan Krios	Titan Krios	Titan Krios	Titan Krios
Voltage (keV)	300	300	300	300
Camera	K3	K3	K3	K3
Magnification	105,000	81,000	81,000	81,000
Pixel size at detector (Å/pixel)	0.725	1.072	0.928	0.928
Total electron exposure (e^{-}/Å^2)	40	40	40	40
Exposure rate (e-/pixel/sec)	25	15	18	18
Number of frames	40	40	82	82
Defocus range (μm)	0.8-2.2	1-2	1.0-2.0	1.0-2.0
Automation software	EPU	EPU	EPU	EPU
Energy filter slit width (eV)	20	20	20	20
Micrographs collected (no.)	10,628	50,781	28,703	28,703
Total extracted particles (no.)	1,963,587	21,223,394	8,973,622	8,973,622
For each reconstruction:			CCAN:3'-CENP-A^{Nuc}	CCAN:5'-CENP-A^{Nuc}
Final particles (no.)	80,236	39,306	9,661	5,396
Point-group	C1	C1	C1	C1
Resolution (global, Å)	3.54	4.5	9.5	15.5
Resolution range (local, Å)	3.54-3.20	4.5-4.18	9.5-9.20	15.5-15.20
Map sharpening *B* factor (Å^2)	-73	-20	-50	-50
Map sharpening methods	cryoSPARC	RELION5	RELION5	RELION5
Model composition				
Protein (residues)	3165	3962	5565	
RNA/DNA (nucleotides)	162	398	636	
Refinement				
Refinement package	PHENIX 1.20.1	PHENIX 1.20.1	PHENIX 1.21	
real or reciprocal space	real	real	real	
Initial model used (Alphafold2)	7R5S/AlphaFold3	7YWX	7R5S/6E0P	
Model-Map scores				
CC Volume/Mask	0.74/0.77	0.80/0.80	0.46/0.55	
B factors (Å^2)				
Protein residues	113.26	358.74	488.77	
RNA/DNA	278.79	854.25	832.12	
R.m.s. deviations from ideal values				
Bond lengths (Å)	0.004	0.004	0.008	
Bond angles (°)	0.740	0.652	0.674	
Validation				
MolProbity score	1.52	1.73	1.82	
CaBLAM outliers	1.22	1.19	1.87	
Clashscore	6.70	14.00	13.12	
Poor rotamers (%)	0	0	0	
C-beta deviations	0	0	0	
EMRinger score	1.26	0.47	*n.d.	
Ramachandran plot				
Favored (%)	97.16	97.63	96.85	
Allowed (%)	2.78	2.3	3.12	
Outliers (%)	0.06	0.08	0.04	

*Not determined (n.d.)