Science
AAAS

# Supplementary Materials for

## Structure of the human inner kinetochore bound to a centromeric CENP-A nucleosome

Stanislau Yatskevich, Kyle W. Muir, Dom Bellini, Ziguo Zhang, Jing Yang, Thomas Tischer, Masa Predin, Tom Dendooven, Stephen H. McLaughlin and David Barford

Correspondence to: dbarford@mrc-lmb.cam.ac.uk

**This PDF file includes:**



**Other Supplementary Materials for this manuscript include the following:**

## Materials and Methods

Cloning of the CCAN and CENP-A protein complexes

Genes encoding CENP-O, CENP-P, CENP-Q, CENP-U, CENP-R, CENP-L, CENP-N, CENP-T, CENP-W, CENP-S, CENP-X were synthesized by Thermo Fisher Scientific with codon optimization for *Trichoplusia ni*. Genes for the CENP-OPQUR and CENP-TWSX modules were subsequently cloned separately into a modified Multibac expression system (*42*). TEV cleavable double strep II (DS) tags were added to the C-termini of CENP-U and CENP-P. An N-terminal 6xHis-SNAP tag was added to CENP-T. CENP-L and CENP-N were sub-cloned into a pF1 plasmid with C-terminal 3C-cleavable SNAP-DS tags on CENP-N and an N-terminal 3C-cleavable GST-tag on CENP-L. Genes for CENP-C, CENP-H, CENP-I, CENP-K and CENP-M were cloned from human cDNA. Genes for the CENP-HIKM module were sub-cloned into a pF1 plasmid (*42*) with a C-terminal TEV-cleavable DS tag on CENP-I. The N-terminal fragment of CENP-C (residues 1-544) (CENP-$C^{N}$) was cloned into a modified pFastBac vector with a C-terminal 3C- cleavable SNAP-DS tag.

CENP-$LN^{cm}$ (R46D, R47D, K48E, K155E, K321E mutated in CENP-L), CENP-$LN^{bs}$ (R236, I237, I238, H239, E240 of CENP-N mutated to alanine), CENP-$O^{\Delta 35}$PQUR (CENP-$PQUR^{full\text{-}length}$, CENP-$O^{\Delta 35}$), CENP-$HIKM^{\Delta Head}$ (CENP-$H^{1\text{-}183}$, CENP-$I^{340\text{-}756}$, $CenpM^{full\text{-}length}$, HsCENP-$K^{1\text{-}151}$), CENP-$OPQUR^{Foot}$ (CENP-$OPR^{full\text{-}length}$, CENP-$Q^{133\text{-}268}$, CENP-$U^{295\text{-}418}$), CENP-$OPQUR^{\Delta N}$ (CENP-$OPR^{full\text{-}length}$, CENP-$Q^{79\text{-}268}$, CENP-$U^{250\text{-}418}$) were generated using a HiFi DNA assembly kit (New England Biolabs).

For CENP-$HIK^{Head}$ expression, DNA encoding CENP-I 1-280, CENP-H 204-247, and CENP-K 165-269 were cloned, in that order, as a single polycistron into pRsfDuet using HiFi assembly (New England Biolabs), with a 3C-SNAP-2xStrepII tag in frame with CENP-I, thereby yielding the pRsfDuet CENP-$HIK^{Head}$ expression vector.

The N-terminal fragment of CENP-N (residues 1-212) was cloned into pGex6P-1 in-frame with an N-terminal 3C-cleavable GST tag using HiFi assembly (New England Biolabs, USA).

Human cDNAs comprising CENP-A and H4 were cloned by restriction-ligation into ORF1 and ORF2 of pRsfDuet. A bicistronic cassette encompassing 6xHis3C-tagged H2A and untagged H2B was cloned by restriction-ligation into pAcycDuet.

The coding regions of histone H2A, H2B, H4 and H3 were amplified by PCR from cDNA and cloned into pET28 plasmid. A double Strep-II tag together with a TEV cleavage site was attached to the N-termini of H3 and H2A proteins. For H3 octamer reconstitution, the histones H2A, H2B, H3 and H4 were assembled into a single pET28 plasmid by the USER methodology.

DNA fragment generation and purification

For DNA fragment preparation, two complementary oligos were synthesized by Sigma-Aldrich. Mixing the oligos in 1x PCR reaction mixture, the fragments were produced by single step extension at 68˚C for 1 min. The final products were purified by 1 ml Resource Q anion

exchange chromatography (Cytiva) and stored in a buffer of 2 M NaCl, 10 mM Tris.HCl (pH 7.5), 1 mM EDTA, 2 mM DTT at -20˚C.

The αSat147 sequence used in this study is an X-chromosome α-satellite repeat identified and used previously (*10, 21, 43, 44*). Oligos aSat147F (ATCGAGGAAG TTCATATAAA AGGCAAACGG AAGCATTCTC AGAATATTCT TTGTGATGAT GGAGTTTCAC TCACAGAGCT GAAC) and aSat174R (ATCAAATATC CACCTGCAGA TTCTACCAAA AGTGTATTTG GAAACTGCTC CATCAAAAGG CATGTTCAGC TCTGTGAGTG AAAC) were used for generation of αSat147. Oligs aSat171F(AATCTGCAAG TGGATATTTG GACCGCTTTG AGGCCTTCGT TGGAAACGGG AATATCTTCA CATAAAAACT AAACAGAAGC ATTCTCAGAA ACTTC) and aSat171R (CTACAAAAAG AGTGTTTCAA AACTGCTCTA TCAAAAGGAA TGTTCAACTC TGTGAGTTGA ATGCAATCAT CACAAAGAAG TTTCTGAGAA TGCTTC) were used for AS1 generation. Oligos aSat171BbF (CCGCTTTGAG GCCTTCGTTG GAAACGGGAA TATGTTCACA TAAAAACTAG ACAGAAGCAT TCTCAGAAAC TTCTATGTGA TGTTTGCATT CAACT) and aSat171BbR (TCCAAATGTC CAATTCCAGA TACTACAAAA AGAGTGTTTC AAAACTGCTC TATGAAAAGG AATGTTCAAC TCTATGAGTT GAATGCAAAC ATCACA) were used for AS2 generation. Oligos 171aSatAaF (CCGCTTTGAG GC CTA TGGTGAAAAA GGAA A TATGTTCACA TAAAAACTAG ACAGAAGCAT TCTCAGAAAC TTCTA TGTGA TGTTTGCATT CAACT) and 171aSatAaR (TCCAAATGTC CAATTCCAGA TACTACAAAA AGAGTGTTTC AAAACTGCTC TATGAAAAGG AATGTTCAAC TCTATGAGTT GAATGCAAAC ATCACA) were used for AS2$^{\Delta B\text{-box}}$ generation. Oligos 177a601F (TTGGACCGCT TTGAGGCCTT CGTTGGAAAC GGGAAGAATC CCGGTGCCGA GGCCGCTCAA TTGGTCGTAG ACAGCTCTAG CACCGCTTAA ACGCACGTAC) and 177a601R (ATCGGATGTA TATATCTGAC ACGTGCCTGG AGACTAGGGA GTAATCCCCT TGGCGGTTAA AACGCGGGGG ACAGCGCGTA CGTGCGTTTA AGCGGTG) were used for 177-601 generation. Oligos widom 601F(ATCGAGAATC CCGGTGCCGA GGCCGCTCAA TTGGTCGTAG ACAGCTCTAG CACCGCTTAA ACG CACGTAC GCGCTGTCCC) and widom601R (ATCGGATGTA TATATCTGAC ACGTGCCTGG AGACTAGGGA GTAATCCCCT TGGCGGTTAA AACGCGGGGG ACAGCGCGTA CGTG) were used for 601 generation. Oligos cy5-601F (cy5-ATCGAGAATC CCGGTGCCGA GGCCGCTCAA TTGGTCGTAG ACAGCTCTAG CACCGCTTAA ACG CACGTAC GCGCTGTCCC) and widom601R (ATCGGATGTA TATATCTGAC ACGTGCCTGG AGACTAGGGA GTAATCCCCT TGGCGGTTAA AACGCGGGGG ACAGCGCGTA CGTG) were used for Cy5-601 generation. 53 bp linker DNA was synthesized by IDT and prepared as double-stranded DNA.

DNA sequences used in this study

**AS1**:
AATCTGCAAGTGGATATTTGGACCGCTTTGAGGCCTTCGTTGGAAACGGGAATATCT
TCACATAAAAACTAAACAGAAGCATTCTCAGAAACTTCTTTGTGATGATTGCATTCA
ACTCACAGAGTTGAACATTCCTTTTGATAGAGCAGTTTTGAAACACTCTTTTTGTAG.
**AS2**:
CCGCTTTGAGGCCTTCGTTGGAAACGGGAATATGTTCACATAAAAACTAGACAGAA
GCATTCTCAGAAACTTCTATGTGATGTTTGCATTCAACTCATAGAGTTGAACATTCCT
TTTCATAGAGCAGTTTTGAAACACTCTTTTTGTAGTATCTGGAATTGGACATTTGGA.

**AS2$^{\Delta B\text{-box}}$**:
CCGCTTTGAGGCCTATGGTGAAAAAGGAAATATGTTCACATAAAAACTAG
ACAGAAGCATTCTCAGAAACTTCTATGTGATGTTTGCATTCAACTCATAG
AGTTGAACATTCCTTTTCATAGAGCAGTTTTGAAACACTCTTTTTGTAGT
ATCTGGAATTGGACATTTGGA.
**αSat147**:
ATCGAGGAAGTTCATATAAAAGGCAAACGGAAGCATTCTCAGAATATTCTTTGTGAT
GATGGAGTTTCACTCACAGAGCTGAACATGCCTTTTGATGGAGCAGTTTCCAAATAC
ACTTTTGGTAGAATCTGCAGGTGGATATTTGAT.
**601**:
ATCGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCT
TAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAG
TCTCCAGGCACGTGTCAGATATATACATCCGAT.
**177-601**:
TTGGACCGCTTTGAGGCCTTCGTTGGAAACGGGAAGAATCCCGGTGCCGAGGCCGCT
CAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGC
GTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACA
TCCGAT.
**DNA$^{53}$**:
AATCTGCAAGTGGATATTTGGACCGCTTTGAGGCCTTCGTTGGAAACGGGAAT.

Expression and purification of the CCAN and CENP-A protein complexes

The baculoviruses for expression of all CCAN protein complexes were generated using standard protocols (*42*). All the CCAN sub-complexes were expressed individually in High-5 insect cells. The High-5 insect cell line was not tested for mycoplasma contamination and was not authenticated. Typically, 4L of High-5 cells were infected 2.5% v/v with the P3 cell culture and the cells were harvested 48-72 h after infection by centrifugation.

The cells for CENP-OPQUR, CENP-HIKM, CENP-LN, CENP-C$^{N}$ and their associated mutants were lysed with a sonicator in lysis buffer (50 mM Tris.HCl pH 8.0, 300 mM NaCl, 0.5 mM TCEP and 5% glycerol) supplemented with benzamidine, EDTA-free protease inhibitor tablets and benzonase. Clarified lysate was loaded onto a Strep-Tactin column (Qiagen), immobilized proteins washed with lysis buffer, and the complexes were eluted in a buffer containing 300 mM NaCl, 20 mM HEPES pH 7.8, 1 mM TCEP, 5% glycerol with addition 5 mM desthiobiotin (Sigma). For CENP-HIKM, TEV protease was added to remove the DS tag and 3C protease was added to CENP-LN to remove tags.

For the CENP-OPQUR module, the Strep-Tactin eluate was loaded onto the HiTrap SP HP cation exchange column (Cytiva). The column was washed with CENP-OPQUR wash buffer (20 mM HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP, 5% glycerol) supplemented with 10 mM ATP and 10 mM $MgCl_2$ for five column volumes. The complex was eluted in a single step with five column volumes of CENP-OPQUR elution buffer (20 mM HEPES pH 7.8, 500 mM NaCl, 1 mM TCEP, 5% glycerol), aliquoted and flash frozen in liquid nitrogen.

For CENP-OPQUR$^{Foot}$, CENP-OPQUR$^{\Delta N}$ and CENP-O$^{\Delta 35}$PQUR mutants, the Strep-Tactin eluate was diluted to a final salt concentration of 150 mM NaCl directly and loaded onto the

HiTrap Heparin HP column (Cytiva). The column was washed with CENP-OPQUR wash buffer supplemented with 10 mM ATP and 10 mM $MgCl_2$ for five column volumes. The protein was eluted in 20 mM HEPES pH 7.8, 1 mM TCEP and 5% glycerol with NaCl gradient from 150 mM to 600 mM. The peak fraction was flash frozen in liquid nitrogen.

For CENP-HIKM and CENP-HIKM$^{Head}$ sub-complex, the Strep-Tactin eluate was diluted to a final salt concentration of 150 mM NaCl directly and loaded onto the Resource Q anion exchange column (Merck). The protein was eluted in 20mM HEPES pH 8.0, 1 mM TCEP and 5% glycerol with NaCl gradient from 150mM to 600 mM. The peak fraction was concentrated and further purified using a Superdex 200 column (Cytiva) in gel filtration buffer (20 mM HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP). The eluted CENP-HIKM complex was concentrated and flash frozen in liquid nitrogen.

For CENP-C$^{N}$, the Strep-Tactin eluate was loaded directly onto the Resource Q anion exchange column (Cytiva). The protein was eluted in 20 mM HEPES pH 8.0, 1 mM TCEP and 5% glycerol with NaCl gradient from 300 mM to 600 mM. The peak fraction was concentrated and further purified using Superdex 200 column (Cytiva) in gel filtration buffer (20 mM HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP).

For CENP-LN, CENP-LN$^{cm}$ and CENP-LN$^{bs}$ the Strep-Tactin eluate was diluted to a final salt concentration of 150 mM NaCl directly and loaded onto the HiTrap Heparin HP column (Cytiva). The column was washed with CENP-LN wash buffer (20 mM HEPES pH 7.8, 150 mM NaCl, 1 mM TCEP, 5% glycerol) supplemented with 10 mM ATP and 10 mM $MgCl_2$ for 5 column volumes. The protein was eluted in 20 mM HEPES pH 7.8, 1 mM TCEP and 5% glycerol with NaCl gradient from 150 mM to 600 mM. The peak fraction was concentrated and further purified using Superdex 200 column (Cytiva) in gel filtration buffer (20 mM HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP). The eluted CENP-LN complexes were concentrated and flash frozen in liquid nitrogen.

The CENP-TWSX cell pellet was resuspended in lysis buffer supplemented with 10 mM imidazole, lysed by sonication as described above and loaded onto the cOmplete His-tag purification column (Roche). The complex was eluted in 300 mM imidazole, diluted to final salt concentration of 150 mM directly and loaded onto a HiTrap heparin HP column (Cytiva). The protein was eluted in 20 mM HEPES pH 7.8, 1 mM TCEP and 5% glycerol with a NaCl gradient from 150 mM to 600 mM. The protein was then concentrated and further purified using a Superdex 200 gel filtration column (Cytiva), Peak fractions were concentrated and flash frozen in liquid nitrogen.

To purify CENP-HIK$^{Head}$, BL21 codon plus RIL transformants bearing pRsfDuet HIK$^{Head}$ were grown in ZY media at 37 °C until an OD=600 nm of 0.8 was obtained, at which point expression was induced by addition of 300 μM IPTG, and cells incubated for a further 18 h overnight at 18 °C. Cells were lysed using a homogenizer (Emulsiflex, Avestin) operating at 25 kpsi, and the resulting lysate clarified by centrifugation, followed by application to StrepTactin resin (Qiagen). The resin was washed with lysis buffer, bound proteins eluted in a buffer containing 300 mM NaCl, 20 mM HEPES pH 7.8, 1 mM TCEP, 5% glycerol and the SNAP-2xStrepII tags removed overnight at 10 °C by HRC 3C protease. The CENP-HIK$^{Head}$ complex was further purified using heparin affinity resin (Cytiva) using a linear gradient of buffer HA (20 mM Tris.HCl pH 7.5, 1

mM TCEP) against HB (1 M NaCl, 20 mM Tris.HCl pH 7.5, 1 mM TCEP), and size exclusion chromatography (Superdex S75 16/60, Cytiva; 300 mM NaCl, 20 mM HEPES pH 7.8, 1 mM TCEP). Protein fractions containing the $HIK^{Head}$ were pooled, concentrated (Centricon, Amicon) to 15 mg/ml, flash frozen in liquid nitrogen and stored at -80 °C.

BL21 Star DE3 transformants containing GST-CENP-$N^{NT}$ were cultured in LB at 37 °C until an OD=600 nm of 0.8 was reached. Expression was then induced overnight at 18 °C. Cells were lysed using a homogeniser (Emulsiflex, Avestin) at 25 kpsi. Clarified supernatant was loaded onto an EconoColumn (Biorad) pre-packed with GST Sepharose 4B resin (Cytiva), immobilized proteins washed with a buffer containing 300 mM NaCl, 20 mM HEPES pH 7.8, 1 mM TCEP, 5% glycerol, and eluted by addition of 5 mM reduced glutathione. GST eluate was then purified by heparin affinity chromatography and size exclusion, exactly as described for $HIK^{Head}$.

To purify the CENP-A octamer, fresh BL21 Star DE3 (Invitrogen) transformants harboring pRsfDuet CENP-A/H4 and pAcyc 6xHisH2A/H2B were directly inoculated into LB medium (10 colonies per litre) containing 50 μg/ml kanamycin and 25 μg/ml chloramphenicol. Cells were grown at 37 °C until reaching an OD=600nm of 0.2, at which point the temperature was reduced in two steps (24 °C until an OD=600 nm of 0.4, then 20 °C until an OD=600 nm of 0.6), and expression pursued overnight at 20 °C by induction with 300 μM IPTG.

The harvested CENP-A octamer cell pellet was lysed in a buffer of 2 M NaCl, 40 mM HEPES pH 7.8, 10 mM imidazole, 1 mM TCEP and 5 mM benzamidine with protease inhibitors and benzonase. The octamer was eluted with a gradient of 500 mM imidazole, in lysis buffer. The eluate from Ni-NTA affinity chromatography was diluted into 500 mM NaCl with buffer of 20 mM Tris.HCl pH 7.8, 1 mM TCEP, 5 mM benzamidine. The diluted sample was further loaded onto a heparin column pre-equilibrated with the buffer of 20 mM Tris.HCl pH 7.8, 500 mM NaCl, 1 mM TCEP and 5 mM benzamidine, and the octamer was eluted by gradient to 2M NaCl in the same buffer. The complex was collected and further purified by S200 size exclusion chromatography (Cytiva), concentrated to 3 mg/ml, flash frozen in liquid nitrogen and stored at -80 °C in the buffer of 2 M KCl, 20 mM HEPES pH 7.8, 1 mM TCEP.

The human H3 octamer was prepared by co-expression of H3, H2A, H2B and H4 in B834Rare2 cells. The harvested cell pellet was lysed in a buffer of 50 mM Tris.HCl (pH 8.0), 2 M NaCl, 1 mM EDTA and 1 mM DTT. The H3 octamer was isolated using Strep-Tactin column (Qiagen), and eluted with 4 mM desthiobiotin. The octamer was further purified by S200 size exclusion chromatography (Cytiva), concentrated to 3 mg/ml in a buffer of 10 mM Tris.HCl (pH 7.5), 2 M NaCl, 1 mM EDTA and 2 mM DTT and flash frozen in liquid nitrogen and stored at -80 °C.

CENP-A nucleosome reconstitution

CENP-A, CENP-A mutants or H3 histone octamers were mixed with DNA fragments at 1:0.9 protein:DNA ratio with 7.8 μM final DNA concentration. The histone octamers were then wrapped by gradient dialysis from 2 M KCl to 300 mM KCl buffer with 20 mM HEPES pH 7.5, 1 mM EDTA. The KCl concentration was gradually decreased using a peristaltic pump over 16 h at room temperature. The mixture was then dialyzed for another 2 h against 20 mM HEPES pH 7.5, 300 mM KCl, 1 mM EDTA. The wrapped nucleosomes were then stored at 4 °C (after assessment).

SEC analysis of the CCAN-CENP-A$^{Nuc}$ complexes

To analyse assembly and stability of various CCAN and CCAN-CENP-A$^{Nuc}$ complexes, analytical SEC experiments were performed using Superose 6 3.2/300 column (Cytiva). CCAN complexes were reconstituted at final concentration of 5 μM in CCAN gel filtration buffer (20 mM HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP). The CCAN-CENP-A$^{Nuc}$ complexes were reconstituted at 2:1 ratio (CCAN:CENP-A$^{Nuc}$) with a final CCAN concentration of 5 μM and final nucleosome concentration of 2.5 μM. Chromatograms are plotted as absorbance at 280 nm (a.u., arbitrary unit). The eluted fractions were analyzed by 4-12% SDS-PAGE gels and stained with InstantBlue Coomassie Stain. All reconstitution assays were performed at least in duplicates.

SEC-MALS

SEC-MALS was performed using a Wyatt MALS system. Non-crosslinked and gluteraldehyde crosslinked CCAN$^{\Delta T}$–CENP-A$^{Nuc}$ complexes were injected onto an Superose 6 3.2/300 gel filtration column (GE Healthcare) pre-equilibrated in 20 mM HEPES (pH 7.8), 300 mM NaCl and 1 mM TCEP. The light scattering and protein concentration at each point across the peaks in the chromatograph were used to determine the absolute molecular mass from the intercept of the Debye plot using Zimm's model as implemented in the ASTRA v.5.3.4.20 software (Wyatt Technologies). To determine inter-detector delay volumes, band-broadening constants and detector intensity normalization constants for the instrument, we used aldolase as a standard prior-to sample measurement. Data were plotted with the program Prism v.8.2.0 (GraphPad Software).

Electrophoretic mobility shift assay (EMSA)

For CCAN-DNA EMSAs, 601 DNA was fluorescently labelled with Cy5 fluorophore at the 5' terminus. The reaction was prepared with a constant DNA concentration of 0.5 μM and various concentrations of protein as indicated in the figures. The protein-DNA complexes were incubated in CCAN gel filtration buffer (20 mM HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP) for 30 min. After addition of 5% (v/v) of glycerol, DNA and DNA-protein complexes were resolved by electrophoresis for 30 min at 88V using 0.8% TBE-agarose (w/v). The gel was visualized using Typhoon FLA 9,500 scanner (GE Healthcare) with laser set at 635 nm and Cy5 670BP30 filter. All assays were performed at least in triplicate.

For CCAN-nucleosome EMSAs, nucleosome concentration was kept constant at 2.3 μM and CCAN was added at various concentrations as indicated in the figure legends. The CCAN-nucleosome complexes were incubated in CCAN gel filtration buffer for 30 min. After addition of 5% (v/v) glycerol, nucleosome and CCAN-nucleosome complexes were resolved by gel electrophoresis for 30 min at 88V using 0.8% TBE-agarose (w/v) with added ethidium bromide. The gels were visualized using ChemiDoc XRS+ (Bio-Rad) imaging system. All assays were performed at least in triplicate.

Quantification of the EMSA

Quantification of band intensities was performed using a Fiji scanner. Lanes for each experiment were selected and intensity estimated using a Fiji scanner. The total amount of free DNA or free nucleosome was estimated from the first lane that did not contain any additional protein for each respective repeat. The band intensity for DNA-protein or nucleosome-protein complex was then estimated, subtracting background from the first protein-free lane. To plot the fraction of DNA or nucleosome bound, the signal from the bound fraction was divided by the signal from free DNA or nucleosome fractions, respectively. The data were plotted using Prism 9 (Graphpad). Each position is the mean value from three biological replicates with one standard deviation.

Pull-down Assays

For CENP-$N^N$ binding experiments, CCAN and nucleosome were kept constant at a final concentration of 2.5 μM and 1.25 μM, respectively, in binding buffer A (200 mM NaCl, 20 mM HEPES pH 7.8, 10 mM imidazole pH 8, 1 mM TCEP, 10% glycerol, and 0.05% IGEPAL-CA630), in the combinations indicated in the figure legend, and mixed with 50 μL TALON $Co^{2+}$ resin pre-equilibrated in binding buffer A (Takara Bio, 50% slurry) resulting in a final volume of 100 μL. Where present, CENP-$N^N$ was added to a final concentration of 10 μM. Reactions were incubated with gentle rotation in falcon tubes for 30 mins at 4°C. Thereafter, 40 μL were withdrawn as an input sample, the remaining resin was washed four times with 200 μL binding buffer A, after which eluted via SDS-PAGE loading buffer, and the reactions analyzed by SDS-PAGE.

CENP-LN binding experiments were conducted as above, albeit without titration of CENP-$N^N$, and using binding buffer B (150 mM NaCl, 20 mM HEPES pH 7.8, 10 mM imidazole pH 8, 1 mM TCEP, 10% glycerol, and 0.05% IGEPAL-630). Where indicated, CENP-LN was added at a concentration of 2.5 μM.

All assays were performed at least in duplicate.

Cultivation of human cell lines and synchronization

HEK293 FlpIn-TRex cells (Invitrogen) were cultured in DMEM (Gibco) supplemented with 10% tetracycline-free FBS (PAN Biotech) at 37°C and 5% CO2. For the stable integration of eGFP-CENP-L-wt and eGFP-$CENPL^{cm}$ into the genome, the gene was cloned into the pcDNA5-FRT-TO vector (Invitrogen) with an N-terminal eGFP-tag. HEK293 FlpIn-TREX cells were individually cotransfected with the pCDNA5-FRT-TO-CENP-L plasmids and pOG44, containing the flippase, using the HBS method. Cells were seeded the evening before transfection and the medium was exchanged the next morning. Both plasmids were mixed with 160 mM $CaCl_2$ and 2×HBS buffer (final concentrations: 137 mM NaCl, 5 mM KCl, 0.7 mM $Na_2HPO_4$, 7.5 mM D-glucose and 21 mM HEPES) and added to the cells. The next steps were performed according to the Invitrogen manual. Cells were selected using 100 μg/ml Hygromycin B gold (InvivoGen). All stable cell lines were always kept under selection.

Cells were seeded in a 6-well plate in the presence of thymidine (Sigma) and doxycycline (250pg/mL, Sigma) to induce expression of the eGFP tagged transgenes for 16 h. Subsequently, cells were released by washing three times with pre-warmed medium, 4 h later the Cdk1

inhibitor RO-3306 (Santa Cruz) was added for 18 h. The cells were released again as described above for 90 min, before they were centrifuged onto poly-lysin covered cover-glasses.

Immunofluorescence imaging

Cells on cover-glasses were pre-extracted for 2 min using PHEM + Tx (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 2 mM $MgCl_2$, 2% Triton-X 100) before fixation with ice-cold methanol for 5 min at -20°C. The fixed cells were rehydrated for 10 min in PBS, permeabilized for 10 min in PBS + 0.5% Triton-X 100 and blocked in PBS, 0.5% Triton-X 100 and 2% BSA for 30 min. The following antibodies were used for staining: primary antibodies for 1 h at room temperature: CENP-A (1:500, Abcam ab13939), GFP (1:500, Abcam ab6556), secondary antibodies for 45 min at room temperature: goat anti mouse Alexa-fluor 647 (Invitrogen A21235) and goat anti rabbit Alexa-fluor 488 (Invitrogen A11034). The samples were also stained with 1 µg/mL Hoechst33342 (Sigma) together with the secondary antibodies. The cover-glasses were mounted using Diamond antifade (Thermo Fisher Scientific).

Images were acquired with an SP8 confocal microscope (Leica) equipped either with 405 nm, 488 nm, and 633 nm laser lines. An APO CS2 63x/1.4 oil immersion objective was used. Image resolution was set to 1024x1024 with a 3x zoom-in, bidirectional scanning, 600 Hz scan speed and 2x line average. A Z-spacing of 0.23 µm was always used. Laser power and detector gain were set for each primary antibody individually, but were kept constant between different experiments with the same antibody to ensure reproducibility.

For colocalization, images were prepared using the following workflow in ImageJ: The corresponding acquisition channels (CENP-A and eGFP) were separated and the background was subtracted from each image using the ImageJ background subtraction tool with a rolling ball radius of 5 pixels. A Gaussian blur filter was laid over the images with a sigma radius of 1. Colocalization was subsequently analyzed using the ImageJ plugin JaCoB (*45*) with the built-in object based colocalization method based on distance between geometrical centers of the identified objects. The threshold for each image was set manually to allow clear identification of kinetochores based on the CENP-A signal.

Immunoblotting of cell lysates

Cells were seeded with a density of 0.5x106 cells/mL in a 24-well plate. Tetracycline was added and incubated for 20 h. Cells were collected by pipetting them up and down in the plate and were subsequently washed once with PBS. The cell pellet was taken up in NuPAGE LDS loading buffer (Invitrogen) and heated at 95°C for 5 min before loading onto a NuPAGE Bis-Tris 4-12% gel (Invitrogen). Immunoblotting was performed using a Mini Trans-Blot cell (BioRad) and Tris-Glycine buffer onto a nitrocellulose membrane (Amersham). Primary antibodies were incubated in 5% milk/PBS over night at 4°C and secondary antibodies in 5% Milk/PBS for 1 h at 22°C. Antibodies used were: CENP-L (Fisher PA5-114281, 1:250), GFP (Abcam ab6556, 1:1000), actin-HRP (Santa Cruz sc-47778-HRP, 1:2000), secondary anti-rabbit (Thermofisher 31462, 1:10.000).

Crystallization, crystallography data collection and structure solution

Diffraction quality crystals of CENP-HIK$^{Head}$ (12 mg/ml) and CENP-OPQUR$^{Foot}$ (15 mg/ml) were obtained using the sitting drop vapor diffusion method by mixing equal volumes (200 nL) of protein and 10% PEG 8K, 100 mM imidazole pH 8.0 at 277 K or 15% PEG 2Kmme, 40 mM sodium formate, 200 mM Bis-Tris propane pH 6.9 at 18 °C, respectively. Crystals of CENP_HIK$^{Head}$ grew after 30-60 days. CENP-OPQUR$^{Foot}$ was treated with subtilisin in a ratio of 1:600 of enzyme per CENP-OPQUR$^{Foot}$ complex for 30 mins at room temperature and quenched with 1 mM phenylmethylsulfonyl (PMSF) before setting up the crystallization plate. CENP-HIK$^{Head}$ and CENP-OPQUR crystals were cryo-protected with glycerol [25% (v/v)] prior to flash freezing in liquid nitrogen. Diffraction data were collected on Diamond Light Source beamline I24 at 100 K with wavelength of 0.9795 Å, and processed with IMosflm (*46*), which allowed for precise prediction of low-resolution spots by adjusting the mosaic block size parameter from 100 to 2. Anisotropy cut-offs of diffraction limits were applied using the server STARANISO (Global Phasing). The CENP-HIK$^{Head}$ and CENP-OPQUR$^{Foot}$ structures were solved by molecular replacement with Phaser (*47*) using a Phyre2-generated model (*48*) and part of the cryo-EM model from this study, respectively. Interactive building was performed with COOT (*49*), refinement with REFMAC5 (*50*) and PHENIX (*51*), and validation with MolProbity (*52*). There are 0 and 0.3% Ramachandran outliers in the models of CENP-OPQUR and CENP-HIK$^{Head}$, respectively.

Reconstitution of the AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$ complex for cryo-EM

CENP-C$^{N}$ was added directly to the reconstituted AS2-CENPA$^{Nuc}$ at 2:1 molar ratio and final nucleosome concentration of 5 μM. The sample was applied on grids directly as described below.

Reconstitution of the CCAN$^{\Delta T}$ complex for cryo-EM

The CCAN$^{\Delta CT}$ complex was reconstituted by mixing purified CENP-OPQUR, CENP-HIKM and CENP-LN at 1:1:1 ratio at a final concentration of 5 μM. The CCAN complex was incubated for at least half an hour at room temperature. To prepare the complex for structural studies, we performed on-column glutaraldehyde (GA) cross-linking. 500 μl of 0.5% of GA was injected onto the Superose 6 Increase 10/300 column (Cytiva) with the flow rate set to 0.25 ml/min. The GA was injected for 7 ml, after which the CCAN sample was applied and eluted under an isocratic flow of 0.25 ml/min in CCAN gel filtration buffer (20 mM HEPES pH 7.8, 300 mM NaCl, 1 mM TCEP). The reaction was quenched with Tris.HCl buffer pH 8.0 at a final concentration of 50 mM straight after protein complex elution. The peak fraction of the CCAN sample was taken, concentrated and re-injected onto a Superose 6 3.2/300 column (Cytiva) coupled to a MicroÄkta (Cytiva). The resulting peak for the CCAN complex was then concentrated to 1 mg/ml and used for structural studies. For non-cross-linked CCAN$^{\Delta T}$ complex, purified CENP-OPQUR, CENP-HIKM, CENP-LN and CENP-C$^{N}$ at 1:1:1:1 ratio were mixed together at a final concentration of 5 μM and applied on grids directly. For non-cross-linked CCAN$^{\Delta C}$ -DNA complex, the components were mixed in 1:1:1:1:1 ratio (CENP-OPQUR: CENP-LN: CENP-HIKM:CENP-TWSX:DNA) at a final concentration of 5 μM and applied on grids directly.

Reconstitution of the CCAN$^{\Delta T}$ -CENP-A$^{Nuc}$ complex for cryo-EM

The cross-linked CCAN$^{\Delta T}$ -CENP-A$^{Nuc}$ complex as well as CCAN-CENP-A$^{Nuc}$ complex was prepared identically to the cross-linked isolated CCAN$^{\Delta CT}$ complex but with addition of CENP-A nucleosome and either CENP-C$^{N}$ protein alone or with CENP-TWSX. The protein ratios used were 1:1:1:1:2 (CENP-OPQUR: CENP-LN: CENP-HIKM:CENP-A$^{Nuc}$:CENP-C$^{N}$) or 1:1:1:1:1:2 (CENP-OPQUR: CENP-LN: CENP-HIKM:CENP-TWSX:CENP-A$^{Nuc}$:CENP-C$^{N}$), to a final nominal concentration of 5 μM.

Cryo-EM grid preparation and data acquisition.

3 μl of prepared complexes at a concentration of approximately 1 mg/ml was applied to Quantifoil 300 mesh copper R1.2/1.3 grids (Quantifoil Micro Tools), glow discharged with Edwards S150B glow discharger for 1 min 15 sec, setting 6, 30-35 mA, 1.2 kV, 0.2 mBar (0.15 Torr). The grids were then flash frozen in liquid ethane using a Thermo Fisher Scientific Vitrobot IV (0-0.5 s waiting time, 2 s blotting time, -7 blotting force) using Whatman filter paper 1. Cryo-EM images were collected on Thermo Fisher Scientific Titan Krios microscope operating at 300 keV using a Gatan K3 camera. A magnification of 105 k was used for the apo-CCAN$^{\Delta CT}$, and CCAN-CENP-A$^{Nuc}$-AS1 samples, yielding pixel sizes of 0.86 Å/pixel and 0.831 Å/pixel, respectively. Additional data were collected for CCAN-CENP-A$^{Nuc}$-AS1 at a magnification of 81 k, yielding nominal pixel sizes of 0.93 Å. The images were recorded at a dose rate of 25 $e^{-}$/px/s with 2 s exposure and 60 frames. For AS2-CENPA$^{Nuc}$-CENP-C$^{N}$, CCAN$^{\Delta C}$ -DNA complex and CCAN-CENP-A$^{Nuc}$ complex, a magnification of 105 k was used yielding pixel sizes of 0.853 Å/pixel. The images were recorded at a dose rate of 16 $e^{-}$/px/s with 2.1 s exposure and 40 frames. The Thermo Fisher Scientific automated data-collection program EPU was used for data collection, with AFIS. Defocus values ranged from -1.6 to -3.0 μm, at an interval of 0.2 μm.

Cryo-EM data processing

For the CCAN$^{\Delta CT}$ -DNA, AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$, CCAN$^{\Delta C}$-DNA, AS1-CENP-A$^{Nuc}$-CENP-C, and CCAN$^{\Delta T}$-CENP-A$^{Nuc}$-AS2 reconstructions, micrograph movie frames were aligned with MotionCor2 (*53*), and contrast-transfer function (CTF) estimation was performed by CTFFIND (*54*), as integrated into RELION3.1 (*55*). Particle picking was performed with SPHIRE-crYOLO (*56*). For the apo-CCAN and CCAN$^{\Delta T}$-CENP-A$^{Nuc}$-AS1 reconstructions, the same processing steps were instead carried out with Warp (*57*). Default parameters were used for data processing unless stated otherwise.

For apo-CCAN, reference-free 2D classification in RELION yielded classes with clear secondary structure which were then exported to cryoSPARC (*58*) for *ab initio* model generation, using a model number of 2. The resulting map was then used in 3D classification in RELION (*55*). Particle classes bearing the most homogeneous density were then subjected to 3D refinement, followed by iterative CTF parameter refinement and Bayesian polishing.

All masks created in RELION were generated with by extending corresponding binary maps by 6 pixels and applying a 10-pixel soft edge. Masks used in cryoSPARC processing were generated automatically by the program itself. Prior to 3D classification or refinement operations in RELION and cryoSPARC, input maps were low-pass filtered to a resolution of 40 Å.

Consensus refinement thus yielded a volume with an overall resolution of 3.2 Å. Multibody refinement was then applied using partially overlapping masks around the CENP-OPQURN, and CENP-HIKMLN lobes, yielding resolutions of 3.9 Å, and 3.2 Å respectively. To further improve the continuity of the CENP-OPQURN density, particle subtraction was applied to remove the CENP-HIKML signal, and particles were exported to cryoSPARC, where they were further heterogeneously classified against a true and two noisy decoy volumes. Resulting particles were then reimported into RELION (*55*) and refined, yielding a 3.6 Å volume. To derive volumes with more continuous CENP-HIK$^{Head}$ density, the CENP-OPQURN signal was subtracted, and resulting particles classified in 3D without alignment, using a T value of 8. Particle classes corresponding to the raised and lowered states of CENP-HIK$^{Head}$ were then refined, producing reconstructions with resolutions of 4.87 Å and 4.11 Å. Reconstruction of the full raised and lowered CCAN volumes was achieved by reverting the subtraction of refined particle sets, and a final round of consensus 3D refinement.

For CCAN$^{\Delta T}$ -CENP-A$^{Nuc}$, rigorous 2D classification, in RELION, was essential to identify complexes containing all constituents. Classes with clear features corresponding to CCAN-DNA and CENP-A$^{Nuc}$ were exported to cryoSPARC (*58*) for *ab initio* model generation, as for apo-CCAN. The particle set arising from 2D classification was then classified iteratively in 3D against the resulting models, in RELION. 3D consensus refinement of the best-defined classes produced an initial reconstruction with an overall resolution of 10 Å. Signal subtraction of the CENP-A$^{Nuc}$ component of these particles, followed by 3D classification with alignment, and 3D consensus refinement, yielded a reconstruction of the CCAN-DNA subdomain with an overall resolution of 8.7 Å, enabling assignment of the register of DNA emanating from the CENP-A nucleosome. To further improve the resolution of the CCAN-CENP-A$^{Nuc}$ map, particles originating from two data collection sessions were scaled to the same pixel size (*59*), 1.662 Å, and exported to cryoSPARC. Two rounds of further heterogeneous refinement against two noise decoys and a single true map were performed, followed by homogeneous and non-uniform refinement (*60*), yielding a CCAN$^{\Delta T}$ - CENP-A$^{Nuc}$ volume with an overall resolution of 8.72 Å. All resolution values were estimated according to the Gold-Standard Fourier Shell Correlation (GS-FSC) = 0.143 criterion. Volume post-processing and local resolution estimation were performed in RELION. Identical procedures were follwoed for the CCAN -CENP-A$^{Nuc}$, resulting in a density volume with an overall resolution of 12 Å.

For CCAN$^{\Delta T}$-DNA and CENP-A$^{Nuc}$-CENP-C$^{N}$ complexes, data were initially subjected to 2D classification in cryoSPARC into 100 classes using default parameters. Classes corresponding to CCAN and CENP-A nucleosomes were picked and *ab initio* models generated (using a model number of 2). *Ab initio* models were further refined individually to obtain better-defined cryo-EM densities. Heterogeneous refinement in cryoSPARC with the refined CENP-A$^{Nuc}$-CENP-C$^{N}$ volume, CCAN$^{\Delta T}$ -DNA volume and four decoy noise volumes was performed against the entire set of automatically picked particles using default parameters. Particles corresponding to CENP-A$^{Nuc}$-CENP-C$^{N}$ and CCAN$^{\Delta T}$ -DNA were selected and kept separate in the subsequent data processing workflow. Another round of heterogeneous refinement was performed using either CENP-A$^{Nuc}$-CENP-C$^{N}$ volume or CCAN$^{\Delta T}$ -DNA volume and five dummy noise volumes for each particle set using default parameters. For CENP-A$^{Nuc}$-CENP-C$^{N}$, particles were then exported back to RELION 3.1 and two successive rounds of CTF refinement and Bayesian polishing were performed, and a final, polished particle dataset was used a final round of 3D refinement in RELION, yielding a final resolution of 2.7 Å (GS-FSC). For CCAN$^{\Delta T}$ -DNA, the

particle set was further cleaned by 2D classification using 50 classes and mask of 280 Å, and the best 2D class averages with well-resolved secondary structure were used for two successive rounds of non-uniform refinement in cryoSPARC using default settings, yielding a final resolution of 4.5 Å.

For CCAN$^{\Delta C}$-DNA, 2D classification was performed in cryoSPARC using 100 classes and mask of 280 Å. The best classes corresponding to CCAN$^{\Delta C}$-DNA were selected and initial model generated *ab initio*, with model number of 2. *Ab initio* models were further refined individually to obtain better-defined cryo-EM densities in cryoSPARC using default settings. Two rounds of heterogeneous refinement were performed in cryoSPARC with refined CCAN$^{\Delta C}$-DNA volume and five decoy noise classes against the entire dataset of automatically picked particles using default settings. The best particles were subsequently refined in RELION 3.1 with two successive rounds of CTF refinement and Bayesian polishing. The final, polished dataset was used in a final round of 3D refinement in RELION to obtain a resolution of 2.83 Å.

For AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$ 2D classification was performed in cryoSPARC using 100 classes and mask of 150 Å. The best classes were selected and initial model generated ab initio, with model number of 2. *Ab initio* models were further refined individually to obtain better-defined cryo-EM densities. One round of heterogeneous refinement were performed in cryoSPARC with refined AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$ volume and five decoy noise classes against the entire dataset of automatically picked particles using default settings. Subsequently. 2D classification with 100 classes and mask of 150 Å was used to select the best resolved AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$ classes. The best particles were refined using homogeneous and non-uniform refinement in cryoSPARC to obtain a final resolution of 2.44 Å.

CryoSPARC implementation of 3D variability analysis (3DVA) was used to understand conformational heterogeneity of the AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$ complex. 3 modes were analysed with 250k particles as initial input and resolution limit of 7 Å. The results were visualized in Chimera by splitting the first mode of 3DVA into 20 clusters (Supplementary movie 1).

Cryo-EM model building and refinement

Apo-CCAN$^{\Delta CT}$:

Homology models for CENP-H, CEHP-I, CENP-K and CENP-L were generated by one-to-one threading in Phyre2 (*48*), based on *S. cerevisiae* CCAN (*17*) (PDB 6QLE). Crystal structures were available for CENP-N$^{NT}$ (*7*) (PDB 6EQT) and CENP-M (*18*) (PDB 4P0T). These models were docked into the corresponding map density by rigid body fitting in Chimera and COOT. CENP-O, CENP-P, CENP-Q, CENP-U, CENP-R were built *de novo*. Models encompassing the entire CCAN$^{\Delta CT}$ were manually rebuilt, extended, and corrected in COOT (*49*), followed by real-space refinement in PHENIX (*51*) against a composite map. Default parameters for structure refinements in Phenix were used including secondary structure and Ramachandran constraints. The crystal structure of CENP-OPQUR aided the building of the CENP-OPQUR module in the cryo-EM maps. We further compared and modified our models according to AlphaFold2 (*26, 61*) predictions, which were released in the course of this study. The AlphaFold2 predictions particularly aided the building of the C-terminal regions of CENP-Q and CENP-U and CENP-R.

CCAN-DNA:

Models for CCAN$^{\Delta CT}$ and ideal B-form DNA were rigid-body fitted into the corresponding volumes in Chimera, followed by manual correction in COOT (*49*) and real-space refinement in PHENIX (*51*) using default parameters.

CENP-A$^{Nuc}$-CENP-C:

PDB ID 6MUO model (*10*) was used as the initial model. The model was fitted into the cryo-EM map in ChimeraX (*62*), manually corrected in COOT (*49*) and refined in PHENIX (*51*). The DNA sequence was manually re-build in COOT (*49*).

CCAN-CENP-A$^{Nuc}$:

Higher resolution density maps of CENP-A$^{Nuc}$, isolated CCAN$^{\Delta T}$-DNA, and the CCAN$^{\Delta T}$-DNA volume resulting from focused refinement were fitted into the 10 Å consensus volume for the complete CCAN-CENP-A$^{Nuc}$ complex, and resampled on a common origin. Models for CCAN$^{\Delta CT}$, and CENP-A$^{Nuc}$-CENP-C were rigid-body fitted into the corresponding map density regions. The phase of the DNA gyre emerging from the nucleosome could be readily discerned, and was modelled through manual building and real-space correction in COOT (*49*), invoking chain self-restraints to prevent distortion of proper geometry at low resolution.

AlphaFold2 Predictions

AlphaFold2 (*26*) predictions were run using versions of the program installed locally and on ColabFold with the MMseqs2 MSA option. For the CENP-C$^{PEST}$ – CENP-LN prediction, input sequences were CENP-C (residues 201-400), CENP-N (221 to C-term), CENP-L (all residues); and for CENP-C$^{PEST}$ – CENP-HIKM, input sequences were: CENP-C (residues 101-350), CENP-H (residues 1-190), CENP-I (residues 341 to C-term), CENP-K (residues 1-160), CENP-M (all residues). For CENP-OPQUR prediction, input sequences were: CENP-O and CENP-P (all residues), CENP-Q (residues 193 to C-term), CENP-U (residues 351 to C-term), CENP-R (all residues). All models were subsequently refined against experimental densities. An AlphaFold2 prediction of CENP-QU$^{Foot}$ generated a four α-helix bundle with high confidence that fitted perfectly into a 3D class of CCAN$^{\Delta CT}$ featuring corresponding map density, which we subsequently experimentally validated with a structure of full CCAN bound to DNA.

Modeling the CCAN-CENP-A$^{Nuc}$-α-satellite array

A heptad CCAN-CENP-A$^{Nuc}$ array was built based on the 171 bp CCAN-CENP-A$^{Nuc}$-AS2 model as a repeating unit, with modeling also guided by the CCAN$^{\Delta C}$-DNA complex structure. Connecting an upstream 5' CCAN-CENP-A$^{Nuc}$ repeat onto the downstream CCAN-CENP-A$^{Nuc}$ repeat required that 22 bp at the 3' end of the upstream CENP-A$^{Nuc}$ was unwrapped. This is because ~42 bp of linker DNA, emerging from the downstream CENP-A$^{Nuc}$, threads through the CENP-LN:CENP-HIKM:CENP-TWSX compartment, exceeding the 24 bp of linker DNA for a 171 bp α-satellite repeat, assuming 147 bp of wrapped nucleosomal DNA. Unwrapping of the 3' end of the upstream CENP-A$^{Nuc}$ is consistent with the results of the CENP-A$^{Nuc}$-AS2 structure.

Unwrapping also alleviates a steric clash with adjacent CCAN-CENP-A$^{Nuc}$ repeats. The same approach was applied to consecutive repeats of the array to generate the heptad array.

Mass spectrometry

Purified proteins were prepared for mass spectrometric analysis by in solution enzymatic digestion, including reduction (10 mM DTT) and alkylation (50 mM Iodoacetamide). Protein samples were digested with trypsin overnight (16 h) and the resulting peptides were analyzed by nanoflow liquid chromatography-tandem mass spectrometry (LC-MS/MS) using an Ultimate U3000 HPLC system (ThermoFischer Scientific) to deliver a flow of approximately 300 nl/min. Desalting of tryptic peptides was performed at a flow rate of 10 µl/min using a C18 Acclaim PepMap100, 100 µm x 20 mm trapping column (ThermoFischer Scientific, 164564-CMD). This was followed by analytical separation using a C18 Acclaim PepMap100 column (Thermo Scientific, 2 µm particle size, 75 µm x 250 mm, 164536) and elution with a 37 min gradient of acetonitrile (3.2% to 32%) and 0.1% formic acid. The analytical column outlet was interfaced via a 10 µm ID SilicaTip emitter (New Objective) to the nanoflow electrospray ionization source, with a hybrid quadrupole orbitrap mass spectrometer (Thermo Scientific, Q-Exactive Orbitrap). LC–MS/MS data were then processed using MaxQuant (v1.6.10.43) with searches carried out by the in-built Andromeda search algorithm against a Uniprot *Homo sapiens* database (v20211004, 78,139 entries) with oxidation (Met) and phosphorylation (Ser, Thr and Tyr) enabled as variable modifications. All results were analyzed as a target-decoy false discovery rate (FDR) of 1% for both peptide-spectrum matches and protein identifications.

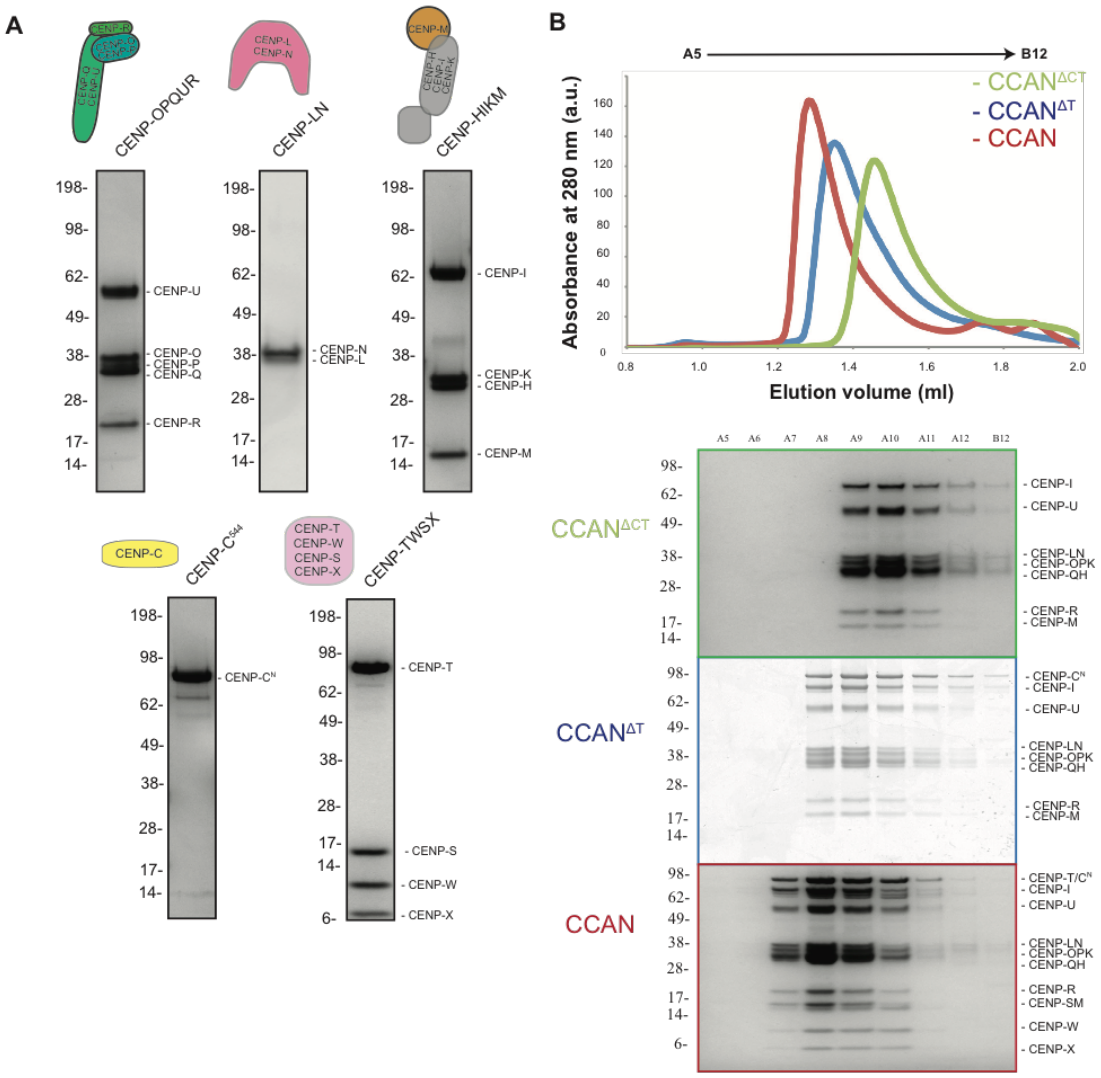


**Fig. S1. Reconstitution of the human CCAN complexes.**
(**A**) Coomassie-blue-stained SDS-PAGE gels of the purified CCAN modules used in this study as well as their cartoon schematic. (**B**) Upper panel: Size-exclusion chromatography profiles (Superose 6 Increase 3.2/300 (Cytiva)) of the reconstituted CCAN, CCAN$^{\Delta T}$ and CCAN$^{\Delta CT}$ complexes showing single high molecular weight peaks for each complex. Lower panel: Coomassie-blue-stained SDS-PAGE gels of the reconstituted CCAN complexes showing that peak fractions contain all of the CCAN modules used in the reconstitution.

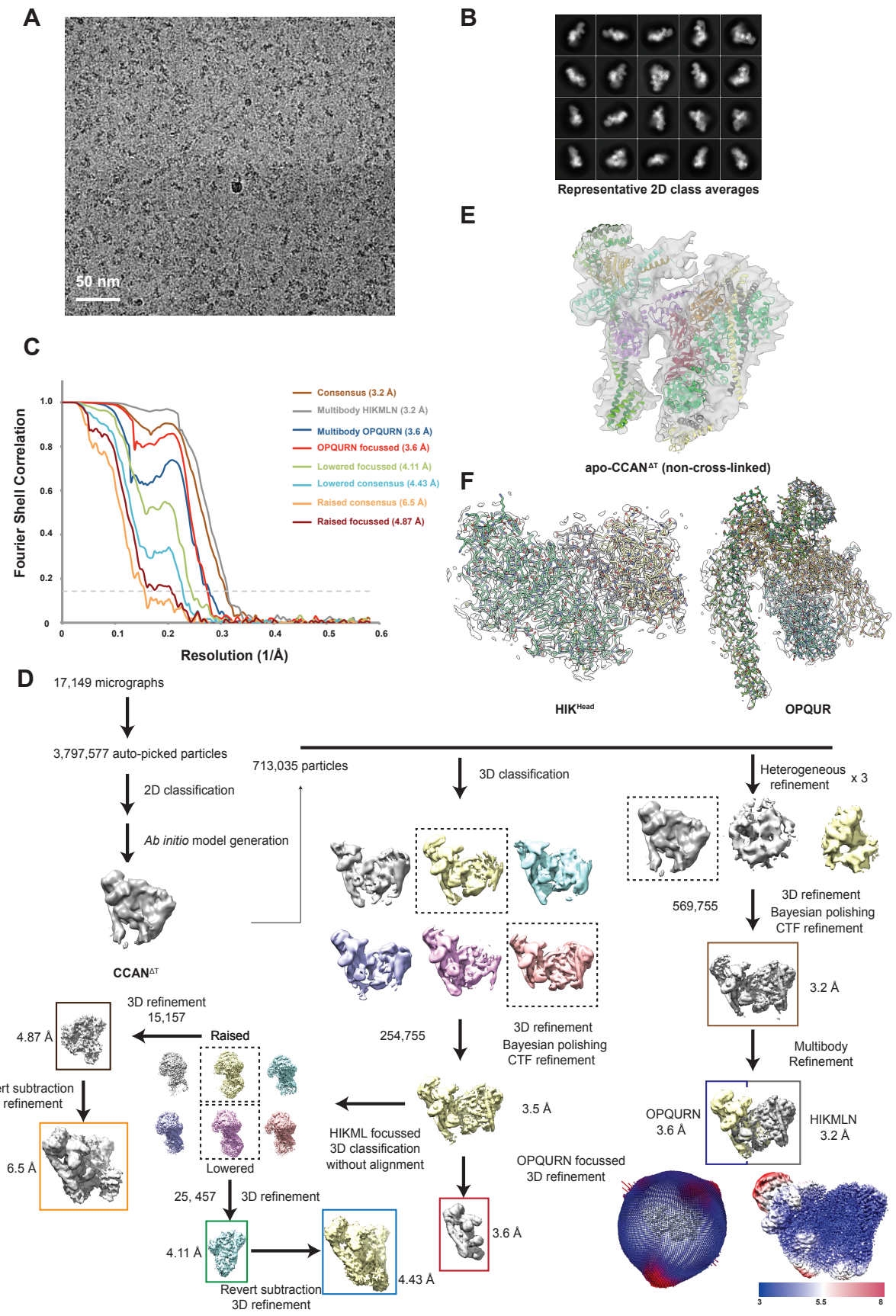
A
50 nm
B
Representative 2D class averages
E
apo-CCAN$^{\Delta T}$ (non-cross-linked)
C
Fourier Shell Correlation
Resolution (1/Å)
Consensus (3.2 Å)
Multibody HIKMLN (3.2 Å)
Multibody OPQURN (3.6 Å)
OPQURN focussed (3.6 Å)
Lowered focussed (4.11 Å)
Lowered consensus (4.43 Å)
Raised consensus (6.5 Å)
Raised focussed (4.87 Å)
F
HIK$^{Head}$
OPQUR
D
17,149 micrographs
3,797,577 auto-picked particles
2D classification
*Ab initio* model generation
CCAN$^{\Delta T}$
713,035 particles
3D classification
Heterogeneous refinement
x 3
569,755
3D refinement
Bayesian polishing
CTF refinement
3.2 Å
Multibody Refinement
OPQURN
3.6 Å
HIKMLN
3.2 Å
254,755
3D refinement
Bayesian polishing
CTF refinement
3.5 Å
HIKML focussed
3D classification
without alignment
OPQURN focussed
3D refinement
3.6 Å
3D refinement
15,157
Raised
4.87 Å
Revert subtraction
3D refinement
6.5 Å
Lowered
25, 457
3D refinement
4.11 Å
Revert subtraction
3D refinement
4.43 Å
3
5.5
8

**Fig. S2. Cryo-EM data of the human apo-CCAN complexes**.
(**A**) A typical cryo-electron micrograph of apo-CCAN$^{\Delta CT}$, representative of 17,149 micrographs. (**B**) Galleries of 2D class averages of apo-CCAN$^{\Delta CT}$, representative of 100 2D classes. (**C**) FSC curves shown for the cryo-EM reconstructions of apo-CCAN$^{\Delta CT}$. FSC curves are defined in panel F. (**D**) Workflow for cryo-EM data processing as well as an angular distribution plot and local resolution map of apo-CCAN. (**E**) Cryo-EM reconstruction of uncross-linked apo-CCAN$^{\Delta T}$ determined at 8 Å resolution, with fitted atomic model of CCAN$^{\Delta CT}$. (**F**) Crystal structures of CENP-HIK$^{Head}$ (left) and CENP-OPQUR (right) with associated 2Fo-Fc electron density map.

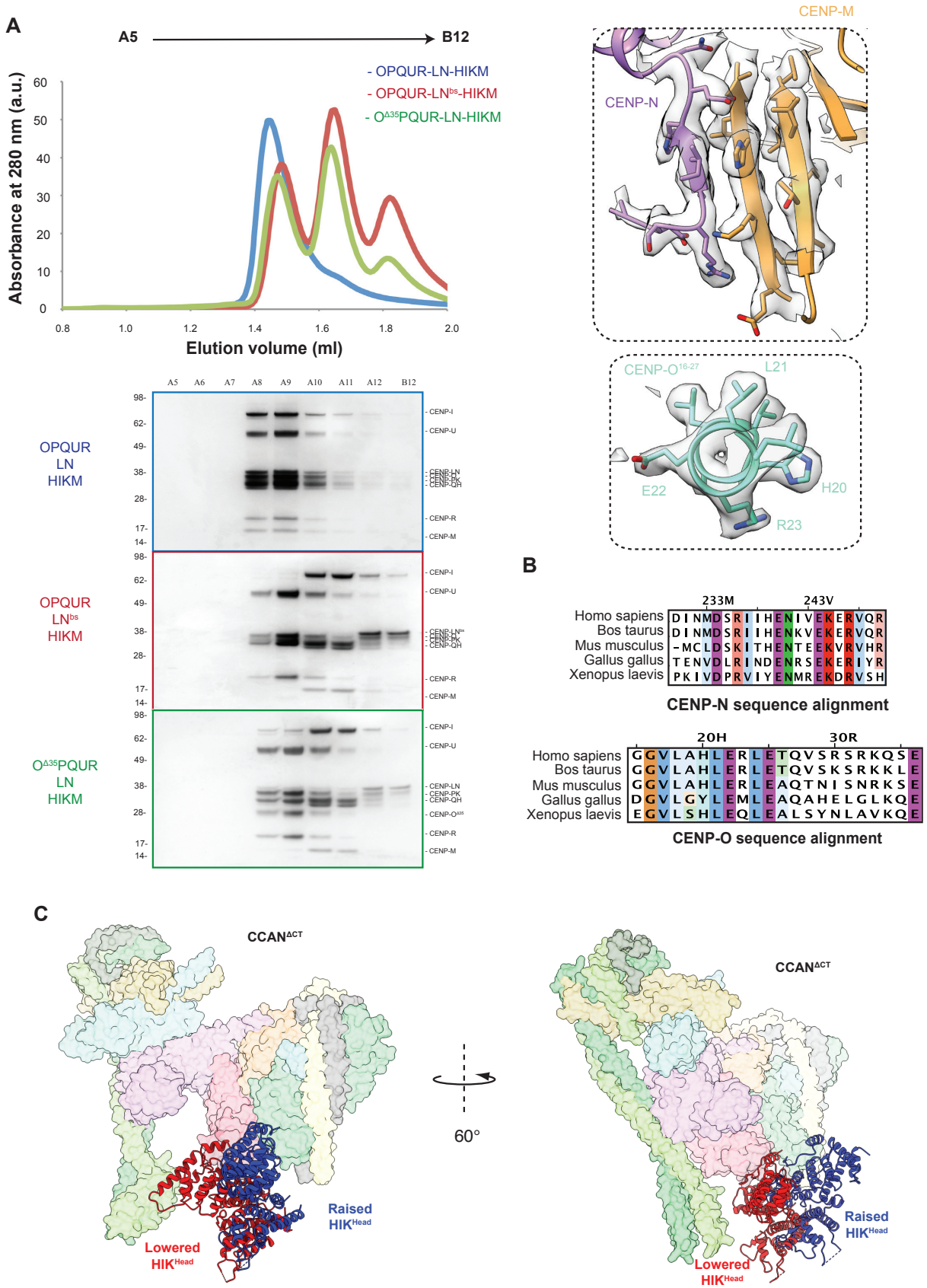
A
A5
B12
- OPQUR-LN-HIKM
- OPQUR-LN^bs-HIKM
- O^Δ35PQUR-LN-HIKM
Absorbance at 280 nm (a.u.)
Elution volume (ml)
OPQUR LN HIKM
OPQUR LN^bs HIKM
O^Δ35PQUR LN HIKM
CENP-I
CENP-U
CENP-LN
CENP-PK
CENP-QH
CENP-R
CENP-M
CENP-O^Δ35
CENP-M
CENP-N
CENP-O^16-27
L21
E22
H20
R23
B
233M
243V
Homo sapiens
Bos taurus
Mus musculus
Gallus gallus
Xenopus laevis
CENP-N sequence alignment
20H
30R
CENP-O sequence alignment
C
CCAN^ΔCT
Raised HIK^Head
Lowered HIK^Head
60°

**Fig. S3. Testing the assembly principles of the CCAN$^{\Delta CT}$.**
(**A**) Comparative SEC profiles (Superose 6 increase 3.2/300 (Cytiva)) of either fully assembled or mutated CCAN$^{\Delta CT}$, composed of CENP-OPQUR, CENP-LN and CENP-HIKM modules. The two mutants (i) CENP-N β-strand mutant (CENP-N$^{bs}$) where 236-240 amino acids were replaced with alanines, and (ii) CENP-O mutant where the N-terminal 35 amino acids were deleted (CENP-O $^{\Delta 35}$ mutant) show defects in apo-CCAN$^{\Delta CT}$ assembly, suggested by the disassembly of the apo-CCAN$^{\Delta CT}$ into three stable modules. Lower panel: Coomassie-blue-stained SDS-PAGE gels for each experiment confirmed the composition of each peak. Right panels: cryoEM density for the interaction between CENP-M and CENP-N as well as cryoEM density for the CENP-O N-terminal α-helix inserted into the CENP-HIK module. (**B**) Multiple sequence alignments (MUSCLE) of the β-strand of CENP-N, and N-terminus of CENP-O shows a high degree of sequence conservation in these regions of CCAN. (**C**) Two conformational states (raised and lowered) of the CENP-HIK$^{Head}$ sub-module.

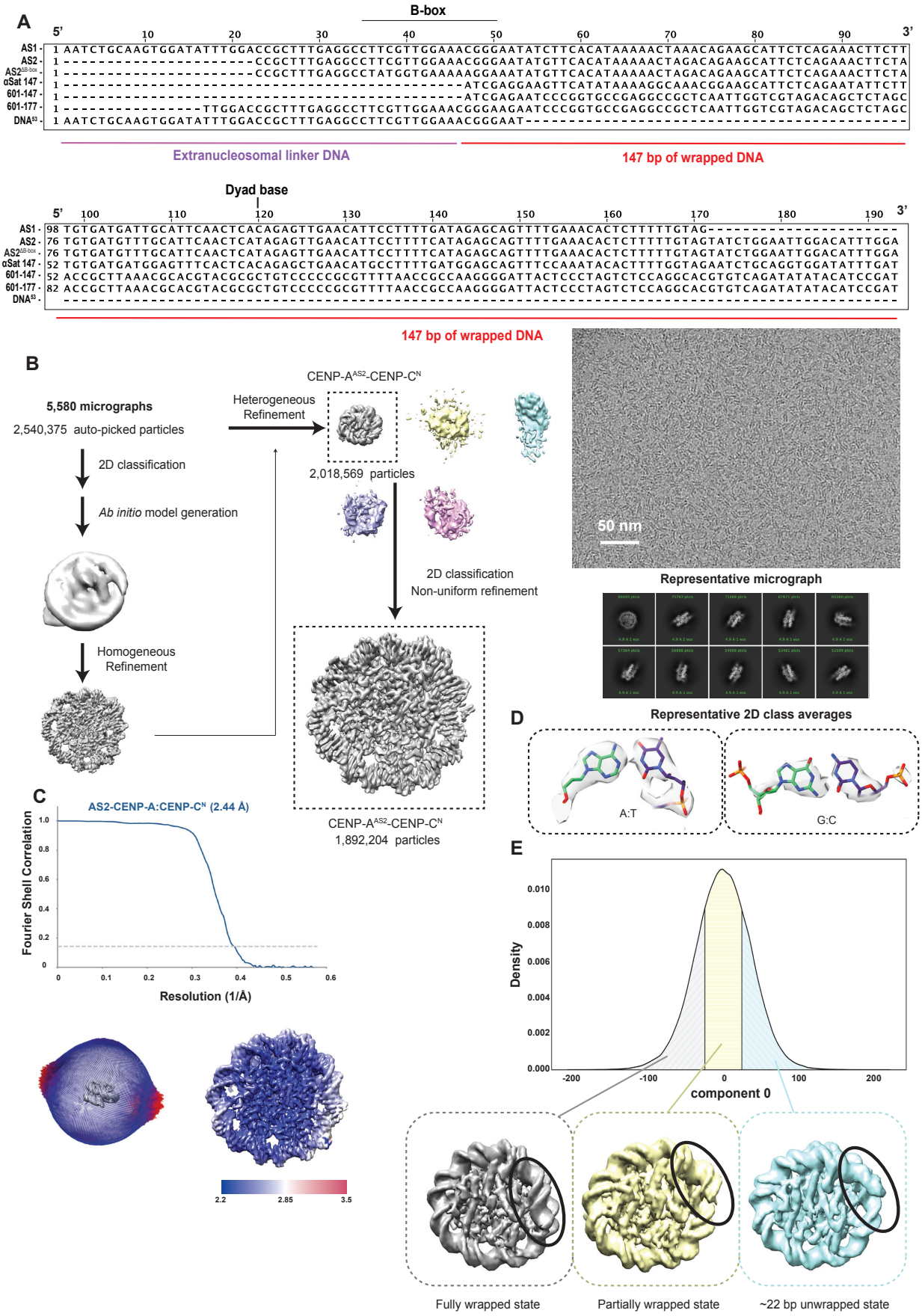
A
B-box
5'
3'
AS1
AS2
AS2ΔB-box
αSat 147
601-147
601-177
DNA53
Extranucleosomal linker DNA
147 bp of wrapped DNA
Dyad base
147 bp of wrapped DNA
B
5,580 micrographs
2,540,375 auto-picked particles
2D classification
Ab initio model generation
Homogeneous Refinement
CENP-AAS2-CENP-CN
Heterogeneous Refinement
2,018,569 particles
2D classification
Non-uniform refinement
CENP-AAS2-CENP-CN
1,892,204 particles
50 nm
Representative micrograph
Representative 2D class averages
C
AS2-CENP-A:CENP-CN (2.44 Å)
Fourier Shell Correlation
Resolution (1/Å)
D
A:T
G:C
E
Density
component 0
Fully wrapped state
Partially wrapped state
~22 bp unwrapped state

**Fig. S4. Reconstitution of the CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ complex.**
(**A**) MUSCLE sequence alignment of all the nucleosome reconstitution DNA constructs that were used in this study to reconstitute CCAN-CENP-A$^{Nuc}$ complexes. The 5' and 3' ends of the DNA referred to in the text are indicated. Numbering scheme is based on the α-satellite repeat defined by {Tanaka, 2005 #302}. The αSat-147 sequence used in this study is X-chromosome α-satellite identified and used previously {Hasson, 2013 #223}. The dyad axis is positioned on a local palindrome. (**B**) Workflow for cryo-EM structure determination of the AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$ complex; right panels: cryo-electron micrograph and 2D class averages of the AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$ complex. (**C**) FSC curve, angular distribution plot and local resolution map for the AS2-CENP-A$^{Nuc}$-CENP-C$^{N}$ complex. (**D**) Representative density for the nucleotide base pairs wrapped around the CENP-A nucleosome. (**E**) The relative probabilities of observing the nucleosome in wrapped (grey, left), intermediate (yellow, middle) and unwrapped (cyan, right) states by kernel density estimation of the eigenvalues for the first eigenvector in a 3D variability analysis experiment in cryoSPARC.

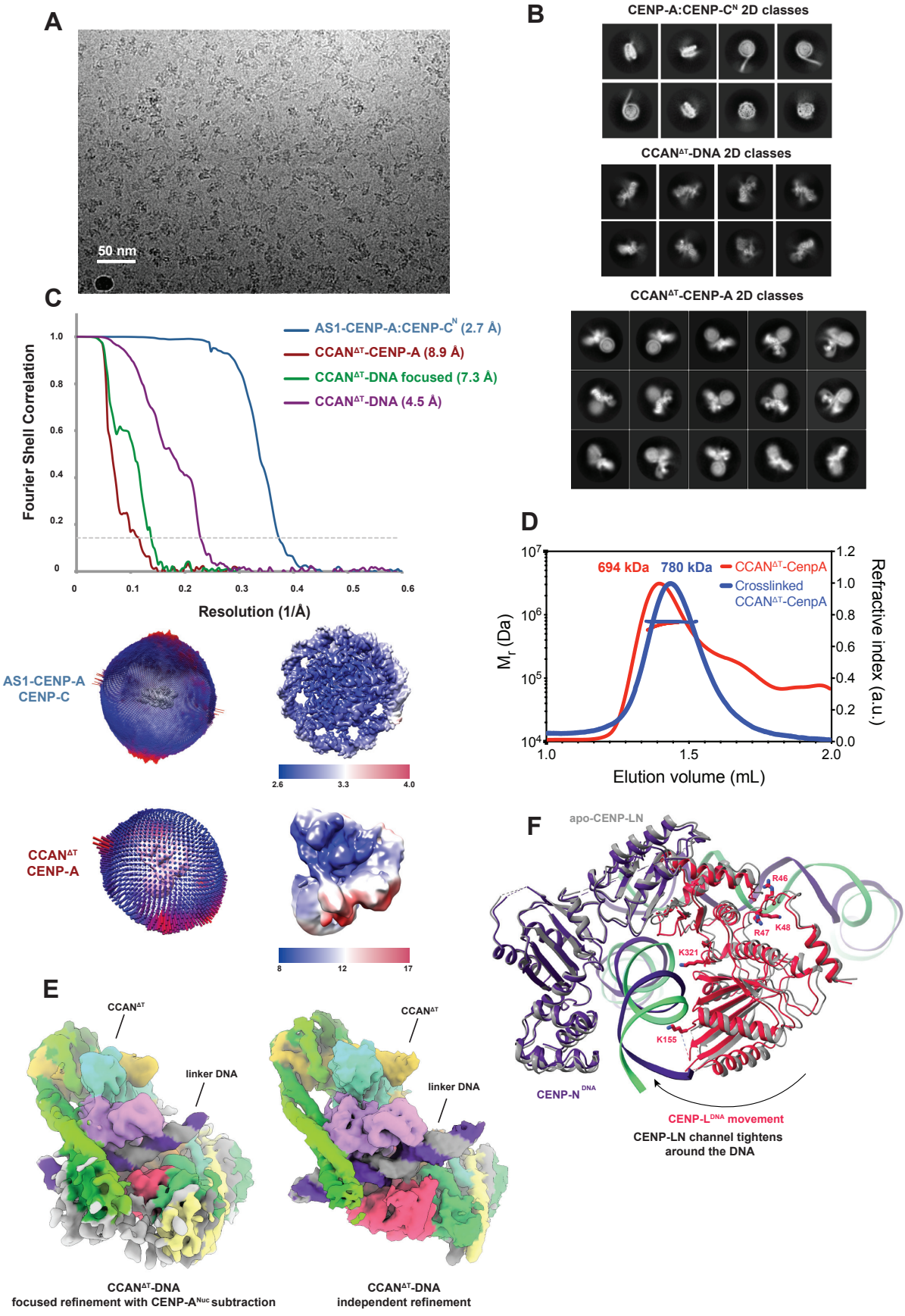
A
50 nm
B
CENP-A:CENP-C^N 2D classes
CCAN^ΔT-DNA 2D classes
CCAN^ΔT-CENP-A 2D classes
C
AS1-CENP-A:CENP-C^N (2.7 Å)
CCAN^ΔT-CENP-A (8.9 Å)
CCAN^ΔT-DNA focused (7.3 Å)
CCAN^ΔT-DNA (4.5 Å)
Fourier Shell Correlation
Resolution (1/Å)
AS1-CENP-A
CENP-C
CCAN^ΔT
CENP-A
D
694 kDa
780 kDa
CCAN^ΔT-CenpA
Crosslinked CCAN^ΔT-CenpA
M_r (Da)
Refractive index (a.u.)
Elution volume (mL)
E
CCAN^ΔT
linker DNA
CCAN^ΔT-DNA
focused refinement with CENP-A^Nuc subtraction
CCAN^ΔT-DNA
independent refinement
F
apo-CENP-LN
R46
K48
R47
K321
K155
CENP-N^DNA
CENP-L^DNA movement
CENP-LN channel tightens
around the DNA

**Fig. S5. Cryo-EM data of the human CCAN$^{\Delta T}$ complexes with CENP-A$^{Nuc}$, CENP-C$^{N}$ and DNA (AS1 DNA sequence).**

(**A**) A typical cryo-electron micrograph of CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ AS1 complex, representative of 38,345 micrographs. (**B**) Galleries of 2D class averages of CENP-A$^{Nuc}$-CENP-C$^{N}$ (top panel), CCAN$^{\Delta C}$-DNA (middle panel) and CCAN$^{\Delta CT}$-CENP-A$^{Nuc}$ (bottom panel), representative of 100 2D classes. (**C**) FSC curves, angular distribution plots and local resolution maps for the cryo-EM reconstructions of various complexes used in this study. FSC curves are defined in **Figure S6**. (**D**) Representative SEC-MALS data for non-cross-linked as well as cross-linked CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ complex. The predicted mass of the CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ complex is 710 kDa, the predicted mass of CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ with two copies of CENP-C$^{N}$ is 796 kDa. (**E**) Cryo-EM maps comparing CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ and CCAN$^{\Delta CT}$-DNA showing the similar mode of linker DNA interaction in the two complexes. (**F**) Superimposition of the CENP-LN module in apo-CCAN (grey) onto the CCAN-DNA complex (indigo and red) shows that the CENP-LN channel contracts when bound to DNA. Residues of CENP-L that contact DNA within the CENP-LN channel (K155, R321) and are in proximity to the DNA of the CENP-A$^{Nuc}$ gyre (R46, R47, K48) and mutated for the CCAN charge mutant (CCANcm) are indicated.

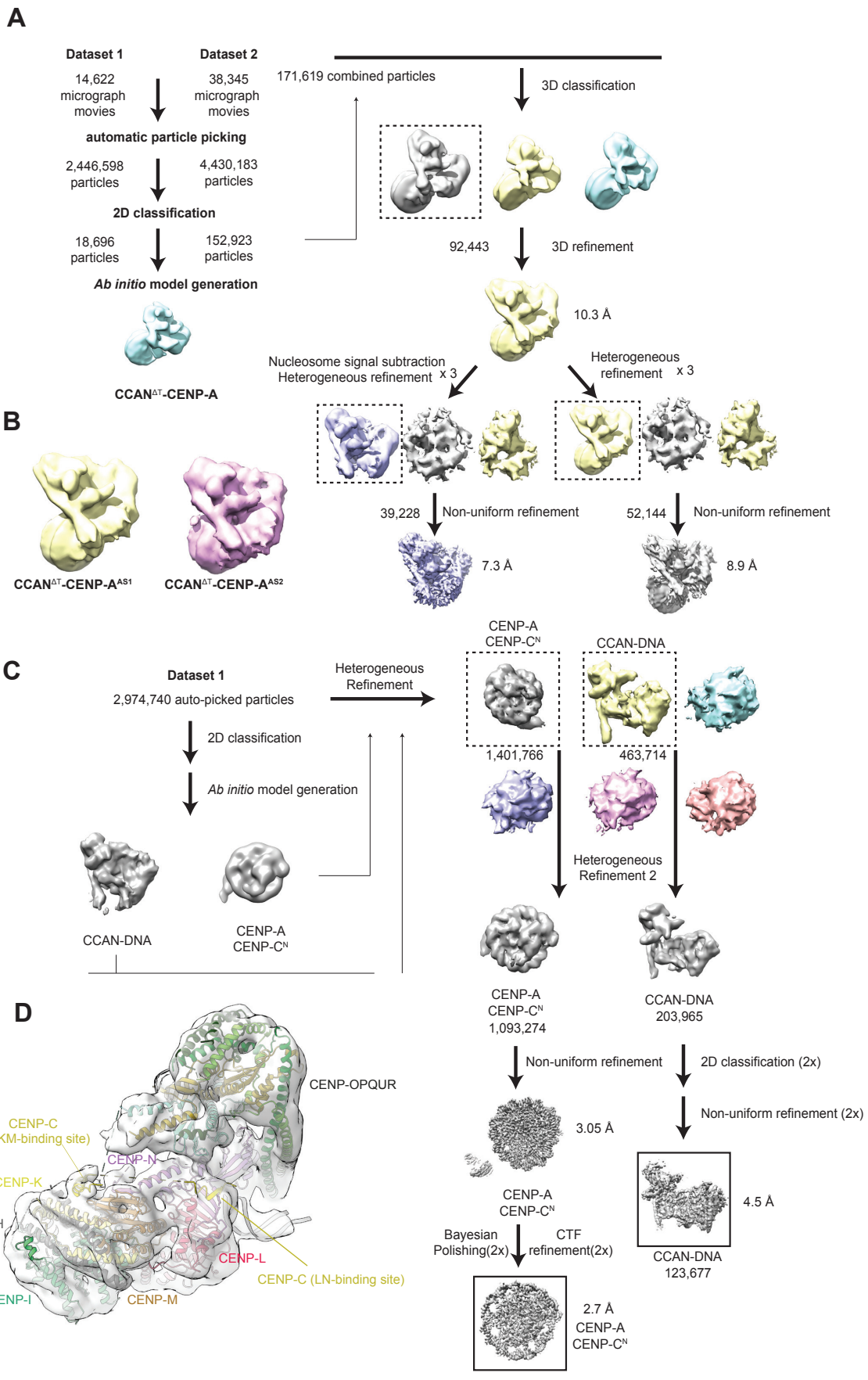
A
Dataset 1
14,622 micrograph movies
Dataset 2
38,345 micrograph movies
automatic particle picking
2,446,598 particles
4,430,183 particles
2D classification
18,696 particles
152,923 particles
Ab initio model generation
CCANΔT-CENP-A
171,619 combined particles
3D classification
92,443
3D refinement
10.3 Å
Nucleosome signal subtraction
Heterogeneous refinement x 3
Heterogeneous refinement x 3
B
CCANΔT-CENP-AAS1
CCANΔT-CENP-AAS2
39,228
Non-uniform refinement
7.3 Å
52,144
Non-uniform refinement
8.9 Å
C
Dataset 1
2,974,740 auto-picked particles
2D classification
Ab initio model generation
CCAN-DNA
CENP-A CENP-CN
Heterogeneous Refinement
CENP-A CENP-CN
CCAN-DNA
1,401,766
463,714
Heterogeneous Refinement 2
CENP-A CENP-CN
1,093,274
CCAN-DNA
203,965
Non-uniform refinement
3.05 Å
CENP-A CENP-CN
2D classification (2x)
Non-uniform refinement (2x)
4.5 Å
CCAN-DNA
123,677
Bayesian Polishing(2x)
CTF refinement(2x)
2.7 Å
CENP-A CENP-CN
D
CENP-OPQUR
CENP-C (HIKM-binding site)
CENP-N
CENP-K
CENP-H
CENP-L
CENP-C (LN-binding site)
CENP-I
CENP-M

**Fig. S6. Workflow for cryo-EM reconstructions of CENP-A$^{Nuc}$-CENP-C$^{N}$, CCAN$^{\Delta T}$-DNA and CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ complexes.**

(**A**) Workflow for cryo-EM reconstruction of CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ for the AS1 DNA sequence (CCAN$^{\Delta T}$-CENP-A$^{AS1}$). An identical work-flow was applied for CCAN$^{\Delta T}$-CENP-A$^{AS2}$. (**B**) The structures of CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ reconstituted with the two α-satellite repeat DNA sequences AS1 and AS2 (defined in **fig. S4A**) are essentially identical. For the CCAN$^{\Delta T}$-CENP-A$^{AS2}$ complex, the dataset comprised 33,097 micrographs and 53,292 particles were used in the final reconstruction. (**C**) Workflow for cryo-EM reconstruction of CCAN$^{\Delta T}$-DNA and CENP-A$^{Nuc}$-CENP-C$^{N}$. (**D**) AlphaFold2 generated models of CENP-C$^{PEST}$ interactions with the CENP-LN and CENP-HIKM modules readily fitted into the unassigned EM density in the CCAN$^{\Delta T}$-DNA complex.

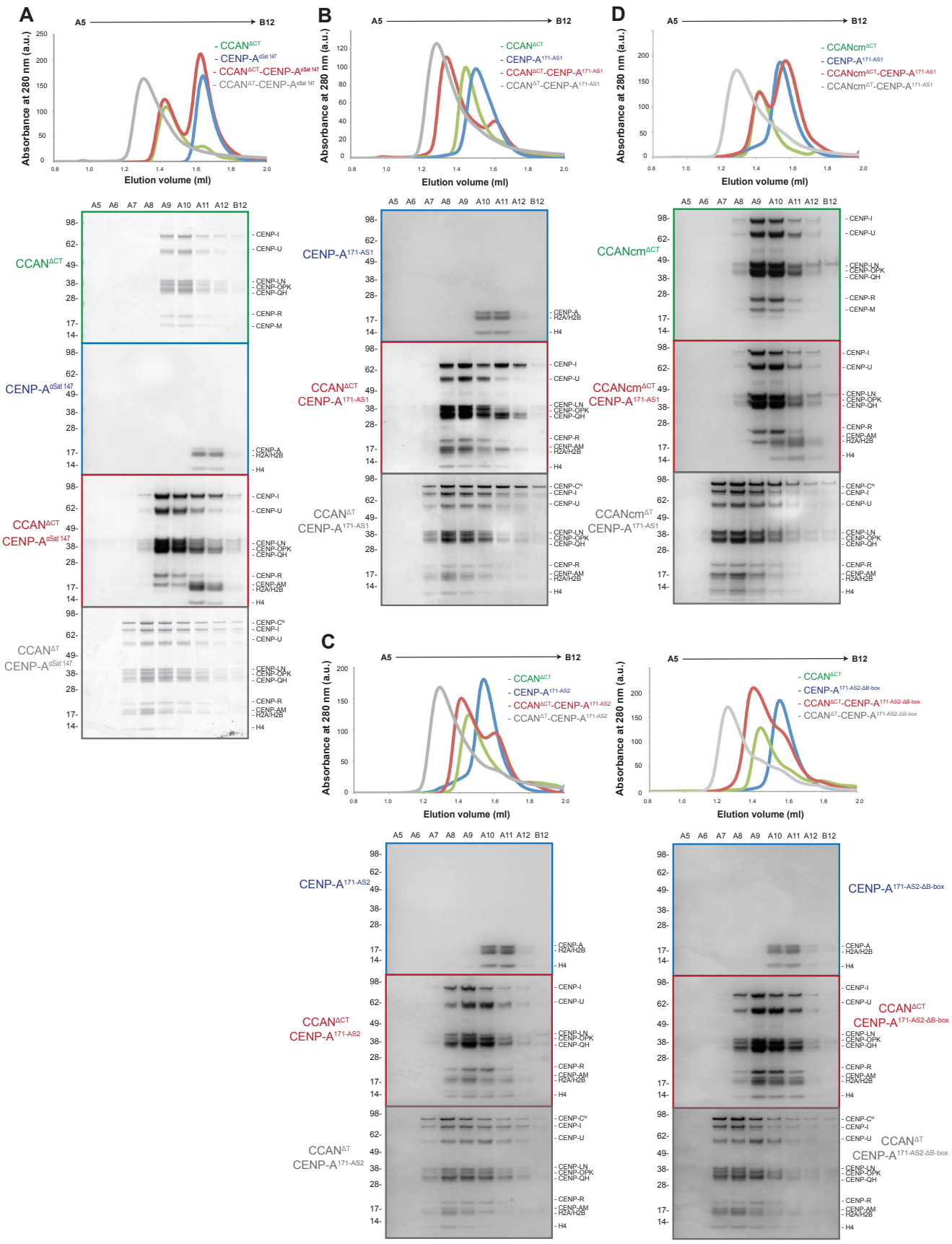
A
B
C
D
A5
B12
Absorbance at 280 nm (a.u.)
Elution volume (ml)
- CCAN^ΔCT
- CENP-A^αSat 147
- CCAN^ΔCT-CENP-A^αSat 147
- CCAN^ΔT-CENP-A^αSat 147
- CCAN^ΔCT
- CENP-A^171-AS1
- CCAN^ΔCT-CENP-A^171-AS1
- CCAN^ΔT-CENP-A^171-AS1
- CCANcm^ΔCT
- CENP-A^171-AS1
- CCANcm^ΔCT-CENP-A^171-AS1
- CCANcm^ΔT-CENP-A^171-AS1
- CCAN^ΔCT
- CENP-A^171-AS2
- CCAN^ΔCT-CENP-A^171-AS2
- CCAN^ΔT-CENP-A^171-AS2
- CCAN^ΔCT
- CENP-A^171-AS2-ΔB-box
- CCAN^ΔCT-CENP-A^171-AS2-ΔB-box
- CCAN^ΔT-CENP-A^171-AS2-ΔB-box
CCAN^ΔCT
CENP-A^αSat 147
CCAN^ΔT
CENP-A^αSat 147
CENP-A^171-AS1
CCAN^ΔCT
CENP-A^171-AS1
CCAN^ΔT
CENP-A^171-AS1
CCANcm^ΔCT
CCANcm^ΔCT
CENP-A^171-AS1
CCANcm^ΔT
CENP-A^171-AS1
CENP-A^171-AS2
CCAN^ΔCT
CENP-A^171-AS2
CCAN^ΔT
CENP-A^171-AS2
CENP-A^171-AS2-ΔB-box
CCAN^ΔCT
CENP-A^171-AS2-ΔB-box
CCAN^ΔT
CENP-A^171-AS2-ΔB-box

**Fig. S7. Reconstitution of the CCAN$^{\Delta CT}$ and CCAN$^{\Delta T}$ with CENP-A$^{Nuc}$ wrapped using different DNA substrates.**

(**A-D**) Comparative SEC profiles (Superose 6 increase 3.2/300 (Cytiva)) of either CCAN$^{\Delta CT}$ or CCAN$^{\Delta T}$ with CENP-A$^{Nuc}$ and their corresponding Coomassie-blue-stained SDS-PAGE gels. All reconstitutions were performed at least in two independent duplicates. (**A**) CCAN$^{\Delta CT}$ does not bind to CENP-A$^{Nuc}$ reconstituted with αSat147 sequence as seen by elution of two separate complexes (CCAN$^{\Delta CT}$-CENP-A$^{\alpha Sat\text{-}147}$). Addition of CENP-C$^{N}$ allowed formation of the CCAN$^{\Delta T}$-CENP-A$^{Nuc}$ complex with the αSat147 DNA sequence (CCAN$^{\Delta T}$-CENP-A$^{\alpha Sat\text{-}147}$). (**B**) CCAN$^{\Delta CT}$ readily bound to CENP-A$^{Nuc}$ reconstituted with AS1 DNA sequence (αSat171 based). The binding was further augmented by addition of CENP-C$^{N}$. (**C, D)** CCAN$^{\Delta CT}$ bound to CENP-A$^{Nuc}$ reconstituted with the AS2 DNA sequence (αSat171 based) and AS2 DNA sequence which lacks the B-box (AS2$^{\Delta B\text{-}box}$). The binding was further strengthened by addition of CENP-C$^{N}$. (**D**) CCAN$^{\Delta CT}$ charge mutant (CCANcm$^{\Delta CT}$), in which the positive patch on the CENP-L backside proximal to the nucleosomal DNA gyre (R46, R47, K48) as well as two positively-charged residues lining the central DNA-binding groove (K155, K321), were mutated to invert the charge, assembled into CCANcm$^{\Delta CT}$ yet failed to bind to the CENP-A$^{Nuc}$ reconstituted with the AS1 DNA sequence. Addition of CENP-C restored binding.

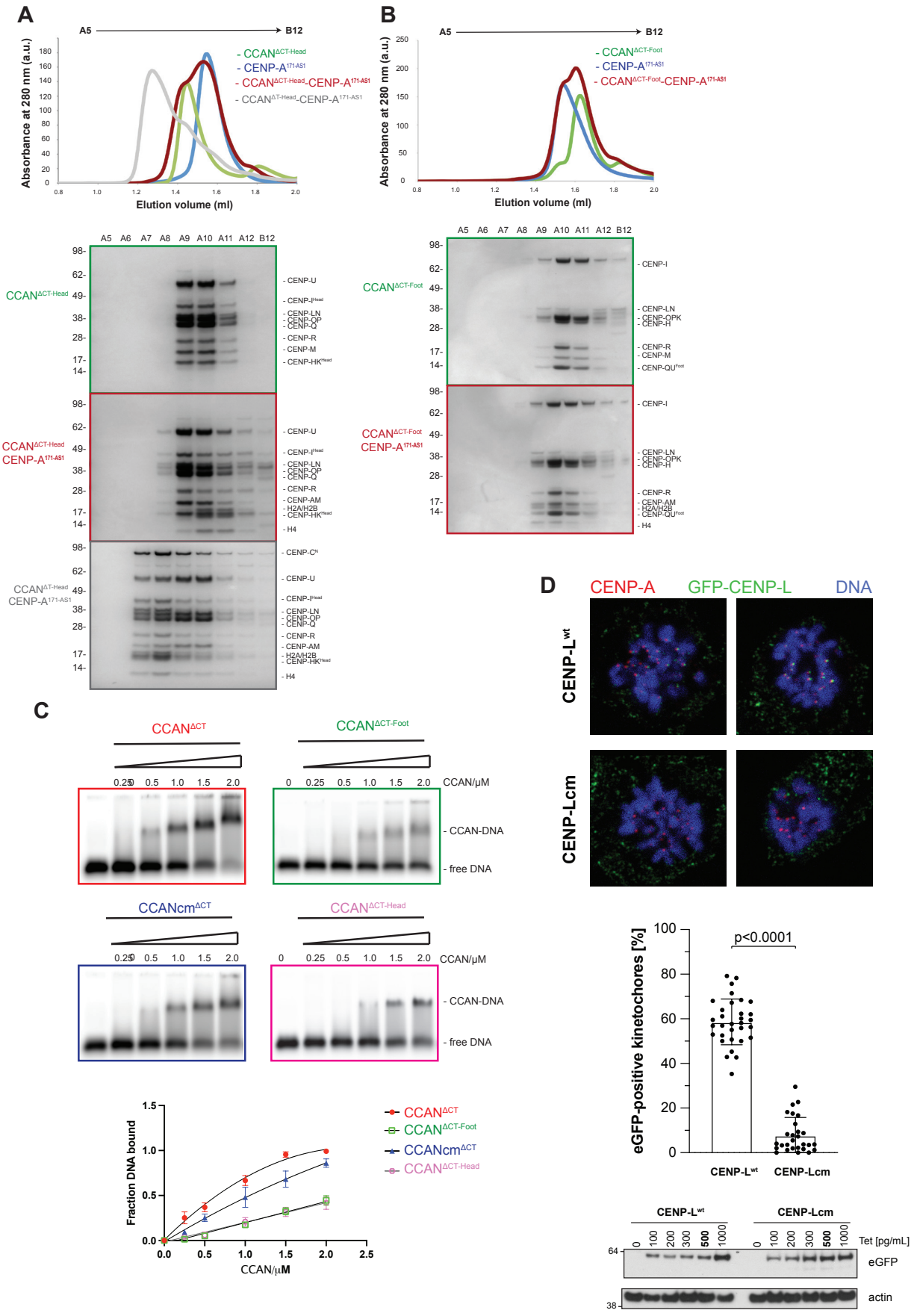
A
A5
B12
Absorbance at 280 nm (a.u.)
Elution volume (ml)
- CCAN^ΔCT-Head
- CENP-A^171-AS1
- CCAN^ΔCT-Head-CENP-A^171-AS1
- CCAN^ΔT-Head-CENP-A^171-AS1
B
- CCAN^ΔCT-Foot
- CENP-A^171-AS1
- CCAN^ΔCT-Foot-CENP-A^171-AS1
A5 A6 A7 A8 A9 A10 A11 A12 B12
CCAN^ΔCT-Head
CCAN^ΔCT-Head CENP-A^171-AS1
CCAN^ΔT-Head CENP-A^171-AS1
CCAN^ΔCT-Foot
CCAN^ΔCT-Foot CENP-A^171-AS1
- CENP-U
- CENP-I^Head
- CENP-LN
- CENP-OP
- CENP-Q
- CENP-R
- CENP-M
- CENP-HK^Head
- CENP-AM
- H2A/H2B
- H4
- CENP-C^N
- CENP-I
- CENP-OPK
- CENP-H
- CENP-QU^Foot
C
CCAN^ΔCT
CCAN^ΔCT-Foot
CCANcm^ΔCT
CCAN^ΔCT-Head
CCAN/μM
- CCAN-DNA
- free DNA
Fraction DNA bound
CCAN/μM
D
CENP-A
GFP-CENP-L
DNA
CENP-L^wt
CENP-Lcm
eGFP-positive kinetochores [%]
p<0.0001
Tet [pg/mL]
eGFP
actin

**Fig. S8. Multiple domains of the CCAN are necessary for stable binding to the CENP-A$^{Nuc}$.**
(**A, B**) Comparative SEC profiles (Superose 6 increase 3.2/300 (Cytiva)) of CCAN$^{\Delta CT}$ mutants where either HIK$^{Head}$ group or OPQUR$^{Foot}$ were deleted and their corresponding Coomassie-blue-stained SDS-PAGE gels below. Deletion of either HIK$^{Head}$ group or OPQUR$^{Foot}$ abolished binding of CCAN$^{\Delta CT}$ to CENP-A$^{Nuc}$. All reconstitutions were done at least in two independent duplicates. (**C**) EMSA assays between either wild-type and mutant CCAN$^{\Delta CT}$ and free 601 Widom DNA (147 bp) labelled with Cy5 dye. The fraction of bound DNA was quantified across three independent replicates as described in the Methods and plotted as a mean value with one standard deviation. Deleting CENP-QU$^{Foot}$ also resulted in defects in CCAN assembly. (**D**) Top panel: mitotic HEK293 cells expressing either eGFP-CENP-Lwt or eGFP-CENP-Lcm were fixed and stained using the indicated antibodies. Exemplary single Z-stacks from two individual cells each are shown to illustrate co-localization of the eGFP tagged constructs with kinetochores marked by CENP-A. The co-localization of the eGFP tagged constructs with kinetochores was analysed. N = 32 cells for CENP-L$^{wt}$ and 28 cells for CENP-Lcm. P-value was calculated using a Mann-Whitney U-test. Bottom panel: Whole cell lysate was prepared from HEK293 cells expressing either eGFP-CENP-L$^{wt}$ or eGFP-CENP-Lcm. Immunoblotting was performed against the indicated proteins. A tetracycline (Tet) concentration of 500pg/mL was used for all experiments with these cell lines.

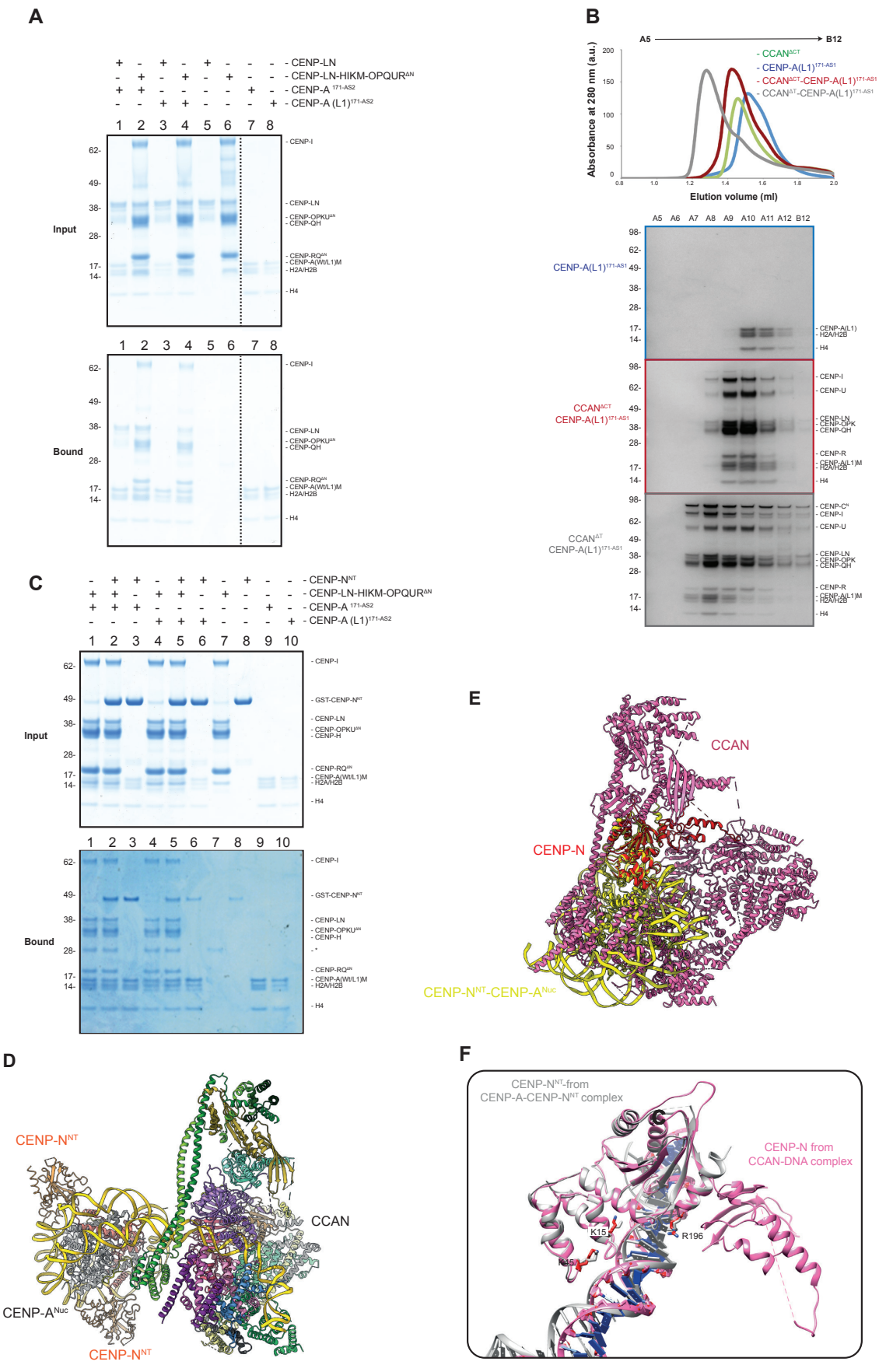
A
B
C
D
E
F
Input
Bound
Elution volume (ml)
Absorbance at 280 nm (a.u.)
CCAN
CENP-N
CENP-A^Nuc
CENP-N^NT
CENP-N^NT-CENP-A^Nuc
CENP-N^NT-from CENP-A-CENP-N^NT complex
CENP-N from CCAN-DNA complex
K15
R196

**Fig. S9. Assembly of CCAN on the CENP-A nucleosome is not dependent on the L1-loop.**
(**A**) Pulldown assays of CENP-LN and CENP-LN-HIKM-OPQUR$^{\Delta N}$ against CENP-A$^{Nuc}$ wild-type or L1 loop mutant, in which the L1-loop RG motif was mutated to AA, labeled CENP-A (L1), at 150 mM NaCl. A dashed line indicates the boundary of cropped gel images. While wild-type CENP-A$^{Nuc}$ retains CENP-LN (lanes 1), binding is entirely abolished by the L1-loop mutation (lane 2). In contrast, CENP-LN-HIKM-OPQUR$^{\Delta N}$ binds equally well to either nucleosome, confirming that CCAN assembly onto the nucleosome does not require the L1-loop: lanes 2 (wild-type CENP-A), and 4 (L1-loop mutant CENP-A). A gently truncated version of CENP-QU N-termini (CENP-OPQUR$^{\Delta N}$, see Methods) had to be used in these assays as full-length CENP-QU is insoluble under this salt condition. (**B**) Comparative SEC profiles of CCAN$^{\Delta CT}$ and CCAN$^{\Delta C}$ reconstituted with a CENP-A$^{Nuc}$ (L1). CENP-A(L1)$^{Nuc}$ was reconstituted with the AS1 DNA sequence. SEC as well as Coomassie-blue-stained SDS-PAGE gels indicated that both CCAN$^{\Delta CT}$ and CCAN$^{\Delta T}$ formed stable complexes with CENP-A(L1)$^{Nuc}$. (**C**) Pulldown assays of GST-CENP-N$^{NT}$ and CENP-LN-HIKM-OPQUR$^{\Delta N}$ against wild-type and L1 loop mutant variants of CENP-A$^{Nuc}$, at 200 mM NaCl. CENP-N$^{NT}$ is dependent on the L1 loop for nucleosome binding: lanes 3 (wild-type CENP-A), 6 (L1-loop mutant CENP-A), and lanes 8 (GST-CENP-N resin binding control). Addition of a four-fold molar excess of GST-CENP-N$^{NT}$ over CENP-LN-HIKM-OPQUR$^{\Delta N}$ does not impede CCAN retention by CENP-A nucleosomes, however binding of the GST-CENP-N$^{NT}$ but not the CENP-LN-HIKM-OPQUR$^{\Delta N}$ is still dependent on the L1-loop: lanes 1/2 (binding of CENP-LN-HIKM-OPQUR$^{\Delta N}$ to wild-type CENP-A$^{Nuc}$ in the absence and presence of GST-CENP-N$^{NT}$, respectively), and lanes 4/5 (same experiment, albeit with L1-loop mutant CENP-A$^{Nuc}$). A contaminant band is indicated with an asterisk. (**D**) Modelling shows that CCAN and CENP-N$^{NT}$ can simultaneously bind to a CENP-A$^{Nuc}$-171 bp module with CCAN interaction with extranucleosomal DNA and CENP-N$^{NT}$ interactions with the CENP-A$^{Nuc}$ L1 loop. (**E**) Superimposing CENP-N$^{NT}$ of the CENP-N$^{NT}$-CENP-A$^{Nuc}$ complex {Pentakota, 2017 #125;Chittori, 2018 #126;Tian, 2018 #317;Allu, 2019 #263}, onto CENP-N$^{NT}$ of CCAN in the context of CCAN-CENP-A$^{Nuc}$ (this study) shows that CENP-L, CENP-HIKM and CENP-QU would clash extensively with CENP-A$^{Nuc}$ of the CENP-N$^{NT}$-CENP-A$^{Nuc}$. **(F**) Comparative figure showing that the same basic residues of CENP-N in either free CENP-N$^{NT}$ in the CENP-N$^{NT}$-CENP-A$^{Nuc}$ complex, or in the context of CCAN in the CCAN-CENP-A$^{Nuc}$ complex (this study) contact the DNA gyre of CENP-A$^{Nuc}$, although at different SHLs.

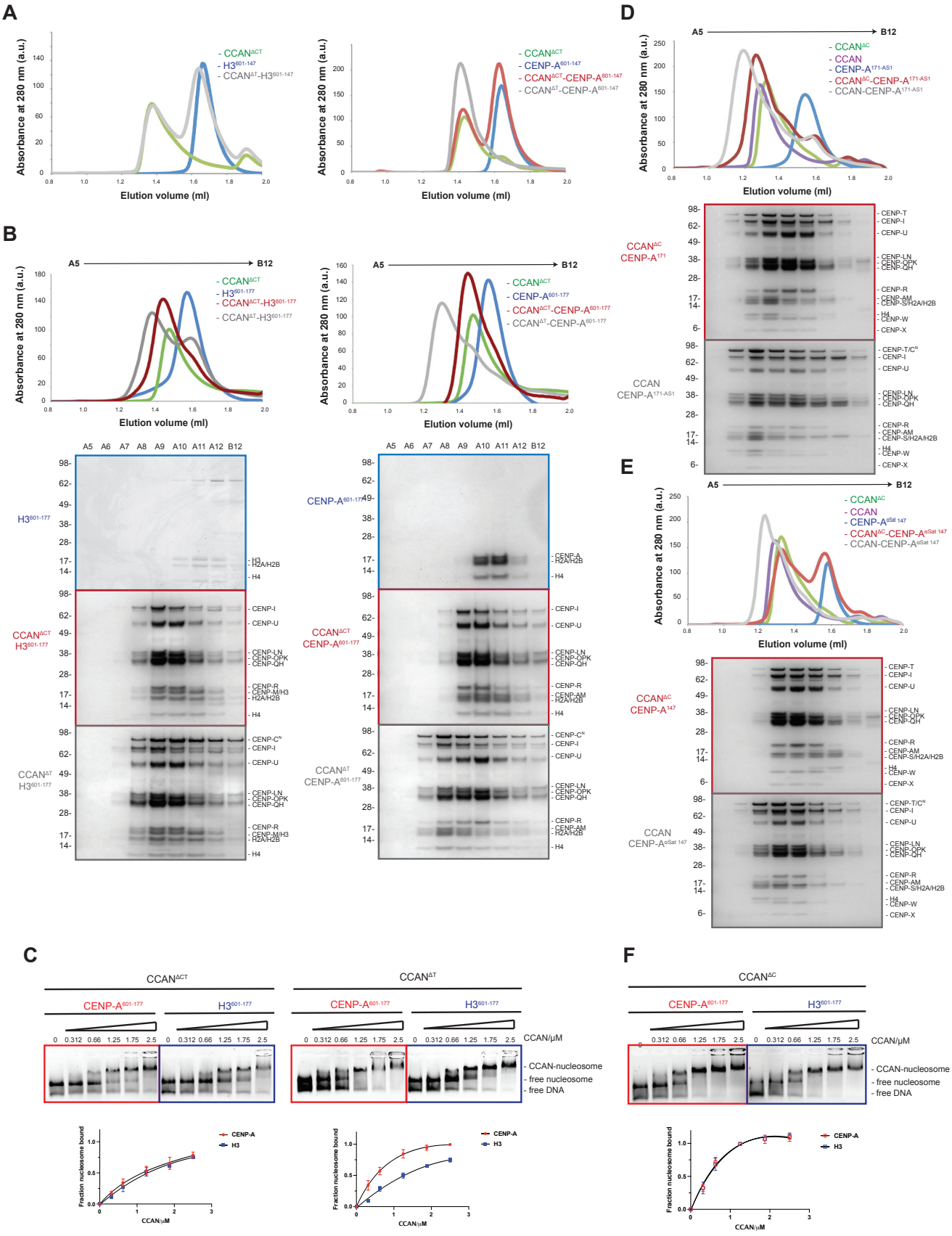
A
Absorbance at 280 nm (a.u.)
Elution volume (ml)
- CCAN^ΔCT
- H3^601-147
- CCAN^ΔT-H3^601-147
- CENP-A^601-147
- CCAN^ΔCT-CENP-A^601-147
- CCAN^ΔT-CENP-A^601-147
B
A5
B12
- H3^601-177
- CCAN^ΔCT-H3^601-177
- CCAN^ΔT-H3^601-177
- CCAN^ΔCT-CENP-A^601-177
- CCAN^ΔT-CENP-A^601-177
A5 A6 A7 A8 A9 A10 A11 A12 B12
H3^601-177
CCAN^ΔCT H3^601-177
CCAN^ΔT H3^601-177
CENP-A^601-177
CCAN^ΔCT CENP-A^601-177
CCAN^ΔT CENP-A^601-177
- H3
- H2A/H2B
- H4
- CENP-I
- CENP-U
- CENP-LN
- CENP-OPK
- CENP-QH
- CENP-R
- CENP-M/H3
- CENP-AM
- CENP-C^N
- CENP-A
D
- CCAN^ΔC
- CCAN
- CENP-A^171-AS1
- CCAN^ΔC-CENP-A^171-AS1
- CCAN-CENP-A^171-AS1
CCAN^ΔC CENP-A^171
CCAN CENP-A^171-AS1
- CENP-T
- CENP-W
- CENP-X
- CENP-T/C^N
E
- CENP-A^αSat 147
- CCAN^ΔC-CENP-A^αSat 147
- CCAN-CENP-A^αSat 147
CCAN^ΔC CENP-A^147
CCAN CENP-A^αSat 147
C
CCAN^ΔCT
CCAN^ΔT
CENP-A^601-177
H3^601-177
0 0.312 0.66 1.25 1.75 2.5
CCAN/μM
- CCAN-nucleosome
- free nucleosome
- free DNA
Fraction nucleosome bound
CCAN/μM
F
CCAN^ΔC

**Fig. S10. The selectivity between CCAN and CENP-A$^{Nuc}$ is conferred by CENP-C$^{N}$ and reconstitution of the complete CCAN-CENP-A$^{Nuc}$ complex.**
(**A**) Comparative SEC profiles (Superose 6 increase 3.2/300 (Cytiva)) of CCAN binding to either H3$^{Nuc}$ or CENP-A$^{Nuc}$ reconstituted with 601 Widom DNA (147 bp). While no binding was observed between CCAN$^{\Delta T}$ and H3$^{Nuc}$, CCAN$^{\Delta T}$ bound to CENP-A$^{Nuc}$ only in the presence of CENP-C$^{N}$. (**B**) Comparative SEC profiles of CCAN binding to either H3$^{Nuc}$ or CENP-A$^{Nuc}$ reconstituted with 177-601 Widom DNA (177 bp) and their corresponding Coomassie-blue-stained SDS-PAGE gels. CCAN$^{\Delta T}$ bound equally well to both H3$^{Nuc}$ and CENP-A$^{Nuc}$ but only binding to CENP-A$^{Nuc}$ was enhanced in the presence of CENP-C$^{N}$. (**C**) EMSA assays to assess nucleosome binding affinity between either CCAN$^{\Delta CT}$ or CCAN$^{\Delta T}$ and either H3$^{Nuc}$ or CENP-A$^{Nuc}$ reconstituted with 177-601 Widom DNA (177 bp). The nucleosome was visualized using ethidium bromide. The fraction of bound nucleosome was quantified across three independent replicates as described in the Methods and plotted as mean value with one standard deviation. (**D**) Reconstitution of the CCAN-CENP-A$^{Nuc}$ complex reconstituted with AS1 DNA sequence (αSat171 based) as assessed by SEC and Coomassie-blue-stained SDS-PAGE gels. CCAN$^{\Delta C}$ readily bound to CENP-A$^{Nuc}$, and further shift is observed upon addition of CENP-C$^{N}$. (**E**) CCAN$^{\Delta C}$ did not detectably bind to CENP-A$^{Nuc}$ reconstituted with αSat147 DNA sequence. The binding was restored in the presence of CENP-C$^{N}$. All reconstitutions were performed in at least two independent duplicates.

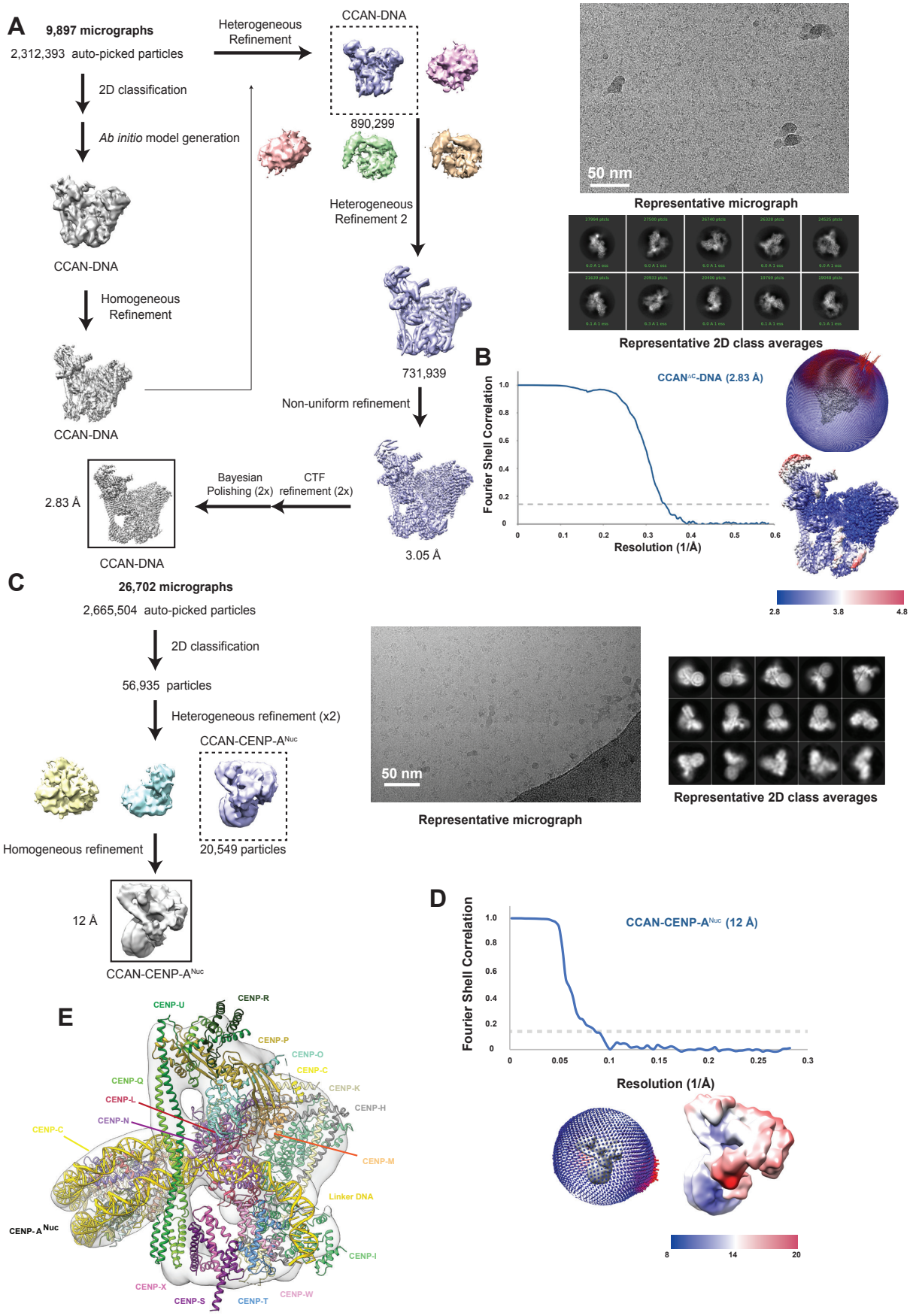
A
9,897 micrographs
2,312,393 auto-picked particles
Heterogeneous Refinement
CCAN-DNA
2D classification
Ab initio model generation
890,299
CCAN-DNA
Heterogeneous Refinement 2
Homogeneous Refinement
731,939
CCAN-DNA
Non-uniform refinement
2.83 Å
Bayesian Polishing (2x)
CTF refinement (2x)
3.05 Å
CCAN-DNA
50 nm
Representative micrograph
Representative 2D class averages
B
CCANΔC-DNA (2.83 Å)
Fourier Shell Correlation
Resolution (1/Å)
2.8
3.8
4.8
C
26,702 micrographs
2,665,504 auto-picked particles
2D classification
56,935 particles
Heterogeneous refinement (x2)
CCAN-CENP-ANuc
20,549 particles
Homogeneous refinement
12 Å
CCAN-CENP-ANuc
50 nm
Representative micrograph
Representative 2D class averages
D
CCAN-CENP-ANuc (12 Å)
Fourier Shell Correlation
Resolution (1/Å)
8
14
20
E
CENP-U
CENP-R
CENP-P
CENP-O
CENP-C
CENP-Q
CENP-K
CENP-L
CENP-H
CENP-N
CENP-C
CENP-M
Linker DNA
CENP-A Nuc
CENP-I
CENP-X
CENP-W
CENP-S
CENP-T

**Fig. S11. Workflow for cryo-EM reconstruction of CCAN$^{\Delta C}$-DNA.**
(**A**) Work-flow for reconstruction of CCAN$^{\Delta C}$-DNA for which 9,897 micrographs were used. Right panel: a typical electron micrograph and representative 2D classes of CCAN$^{\Delta C}$-DNA complex. (**B**) FSC curve, angular distribution plot and local resolution map for the CCAN$^{\Delta C}$-DNA complex. (**C**) Cryo-EM data processing workflow for the CCAN-CENP-A$^{Nuc}$ AS2 complex. Right panel: a typical electron micrograph and representative 2D classes of CCAN-CENP-A$^{Nuc}$ AS2 complex. (**D**) FSC curve, angular distribution plot and local resolution map for the CCAN-CENP-A$^{Nuc}$ AS2 complex. (**E**) Complete atomic model of the CCAN-CENP-A$^{Nuc}$ complex fitted into the low-resolution cryo-EM reconstruction of CCAN-CENP-A$^{Nuc}$ complex determined using AS2 α -satellite repeat.

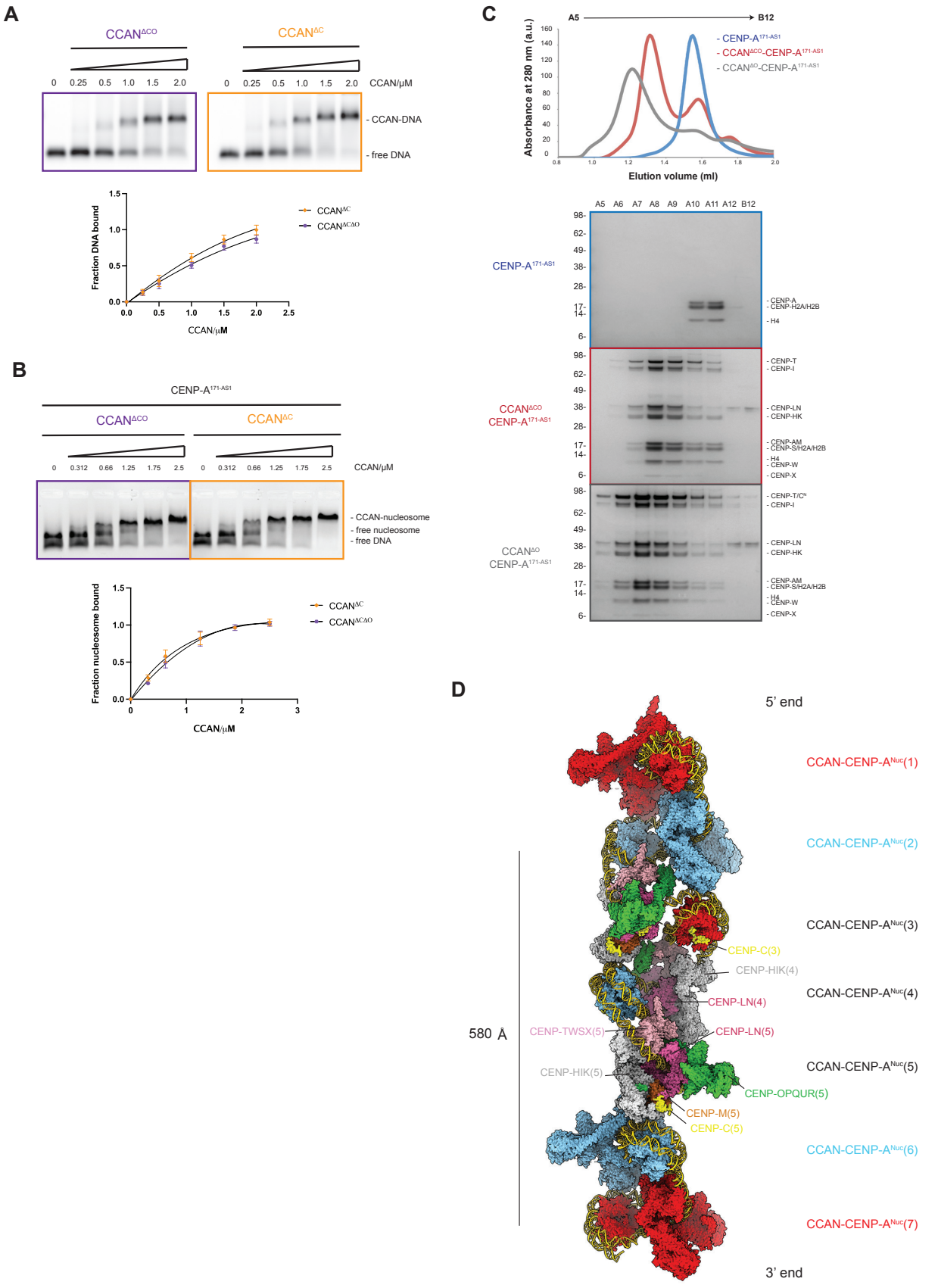
A
CCAN^ΔCO
CCAN^ΔC
0 0.25 0.5 1.0 1.5 2.0
CCAN/μM
- CCAN-DNA
- free DNA
Fraction DNA bound
CCAN^ΔC
CCAN^ΔCΔO
CCAN/μM
B
CENP-A^171-AS1
CCAN^ΔCO
CCAN^ΔC
0 0.312 0.66 1.25 1.75 2.5
CCAN/μM
- CCAN-nucleosome
- free nucleosome
- free DNA
Fraction nucleosome bound
CCAN^ΔC
CCAN^ΔCΔO
CCAN/μM
C
A5
B12
Absorbance at 280 nm (a.u.)
- CENP-A^171-AS1
- CCAN^ΔCO-CENP-A^171-AS1
- CCAN^ΔO-CENP-A^171-AS1
Elution volume (ml)
A5 A6 A7 A8 A9 A10 A11 A12 B12
CENP-A^171-AS1
- CENP-A
- CENP-H2A/H2B
- H4
CCAN^ΔCO
CENP-A^171-AS1
- CENP-T
- CENP-I
- CENP-LN
- CENP-HK
- CENP-AM
- CENP-S/H2A/H2B
- H4
- CENP-W
- CENP-X
CCAN^ΔO
CENP-A^171-AS1
- CENP-T/C^N
- CENP-I
- CENP-LN
- CENP-HK
- CENP-AM
- CENP-S/H2A/H2B
- H4
- CENP-W
- CENP-X
D
5' end
CCAN-CENP-A^Nuc(1)
CCAN-CENP-A^Nuc(2)
CCAN-CENP-A^Nuc(3)
CENP-C(3)
CENP-HIK(4)
CCAN-CENP-A^Nuc(4)
CENP-LN(4)
CENP-TWSX(5)
CENP-LN(5)
580 Å
CENP-HIK(5)
CCAN-CENP-A^Nuc(5)
CENP-OPQUR(5)
CENP-M(5)
CENP-C(5)
CCAN-CENP-A^Nuc(6)
CCAN-CENP-A^Nuc(7)
3' end

**Fig. S12. Biochemical analysis of the CCAN$^{\Delta CO}$ and CCAN$^{\Delta C}$ complexes**.
(**A**) EMSA assays between CCAN$^{\Delta CO}$ and CCAN$^{\Delta C}$ and free 601 Widom DNA (147 bp) labelled with Cy5 dye. The fraction of bound DNA was quantified across three independent replicates as described in the Methods and plotted as a mean value with one standard deviation. (**B**) EMSA assays to assess nucleosome binding affinity between between CCAN$^{\Delta CO}$ and CCAN$^{\Delta C}$ using CENP-A$^{Nuc}$ reconstituted with AS1-171 α-satellite (171 bp) DNA. The nucleosome was visualized using ethidium bromide. The fraction of bound nucleosome was quantified across three independent replicates as described in the Methods and plotted as mean value with one standard deviation. (**C**) Comparative SEC profiles (Superose 6 increase 3.2/300 (Cytiva)) of CCAN$^{\Delta CO}$ and CCAN$^{\Delta O}$ binding to CENP-A$^{Nuc}$ reconstituted with AS1-171 DNA (171 bp) and their corresponding Coomassie-blue-stained SDS-PAGE gels. CCAN$^{\Delta CO}$ bound to CENP-A$^{Nuc}$ and this binding was further augmented when the CCAN$^{\Delta O}$ complex was used. (**D**) Model of a heptad CCAN-CENP-A$^{Nuc}$ array. The array forms a left-handed helix with a pitch of 580 Å, comprising five repeats. The model shows extensive interactions between CCAN and CENP-A$^{Nuc}$ of neighboring repeats, including contacts between CENP-TWSX and the upstream CENP-A$^{Nuc}$ DNA gyre. Color coding: all subunits of CCAN-CENP-A$^{Nuc}$ of repeats 1 and 7, and 2 and 6 are in red and cyan, respectively to show repeat structure. CCAN modules of repeats 3-5 are colored according to Fig. 5B, with CENP-A$^{Nuc}$ histones in red, cyan, red, respectively. This model assumes contiguous α-satellite repeats are occupied by CENP-A$^{Nuc}$, and that each CENP-A$^{Nuc}$ is associated with a single CCAN assembly. However, CENP-B binding to the B-box of the extranucleosomal DNA segment of the CENP-A$^{Nuc}$-α-satellite repeat would sterically block CCAN assembly on that repeat. Furthermore, canonical H3 nucleosomes far outnumber CENP-A$^{Nuc}$, and it is possible many α-satellite repeats are unoccupied *(33)*.

## Table S1. Cryo-EM data collection, refinement and validation statistics.

| **Complex** | **apoCCAN$^{\Delta CT}$**<br>PDB 7PKN,<br>EMD-13473 | **CENP-A$^{AS1\text{-}Nuc}$ CENP-C$^{N}$**<br>PDB 7PII,<br>EMD-13437 | **CCAN$^{\Delta T}$ DNA**<br>PDB 7R5V,<br>EMD-14341 | **CCAN$^{\Delta T}$ CENP-A$^{AS1\text{-}Nuc}$**<br>PDB 7YYH,<br>EMD-14375 | **CCAN$^{\Delta C}$ DNA**<br>PDB 7R5S,<br>EMD-14336 | **CENP-A$^{AS2\text{-}Nuc}$ CENP-C$^{N}$**<br>PDB 7R5R,<br>EMD-14334 | **CCAN CENP-A$^{AS2\text{-}Nuc}$**<br>PDB 7YWX,<br>EMD-14351 |
|---|---|---|---|---|---|---|---|
| **Data collection and processing** | | | | Dataset1, Dataset2 | | | |
| Microscope | TFS Titan Krios | TFS Titan Krios | TFS Titan Krios | TFS Titan Krios | TFS Titan Krios | TFS Titan Krios | TFS Titan Krios |
| Voltage (keV) | 300 | 300 | 300 | 300 300 | 300 | 300 | 300 |
| Camera | Gatan K3 | Gatan K3 | Gatan K3 | Gatan K3 | Gatan K3 | Gatan K3 | Gatan K3 |
| Magnification | 105,000 | 105,000 | 105,000 | 105,000 85,000 | 105,000 | 105,000 | 105,000 |
| Pixel size at detector (Å/pixel) | 0.86 | 0.831 | 0.831 | 0.831 0.93 | 0.853 | 0.853 | 0.853 |
| Total electron exposure ($e^{-}$/Å$^2$) | 50 | 50 | 50 | 50 40 | 40 | 40 | 40 |
| Exposure rate ($e^{-}$/px/s) | 25 | 25 | 25 | 25 16 | 16 | 16 | 16 |
| Number of frames | 60 | 60 | 60 | 60 40 | 40 | 40 | 40 |
| Defocus range (μm) | 1.5-3.0 | 1.5-3.0 | 1.5-3.0 | 1.5-3.0 1.2-2.6 | 1.5-3 | 1.5-3 | 1.2-2.6 |
| Automation software | EPU | EPU | EPU | EPU EPU | EPU | EPU | EPU |
| Energy filter slit width (eV) | 20 | 20 | 20 | 20 20 | 20 | 20 | 20 |
| Micrographs collected (no.) | 17,149 | 14,622 | 14,622 | 14,622 38,345 | 9,897 | 5,580 | 26,702 |
| Total extracted particles (no.) | 3,797,577 | 2,974,740 | 2,974,740 | 6,876,781 (combined) | 2,312,393 | 2,540,375 | 2,665,504 |
| **For each reconstruction:** | | | | | | | |
| Final particles (no.) | 569,755 | 1,093,274 | 123,677 | 52,144 (combined) | 731,939 | 1,892,204 | 20,549 |
| Point-group | C1 | C1 | C1 | C1 | C1 | C1 | C1 |
| Resolution (global , Å) | 3.2 | 2.7 | 4.49 | 8.9 | 2.82 | 2.44 | 12 |
| FSC 0.5 (unmasked/masked) | 4.11/3.8 | 3.47/2.95 | 7.88/5.75 | 16.17/12 | 3.87/3.2 | 3.17/2.68 | 20.2/18.8 |
| FSC 0.143 (unmasked/masked) | 3.6/3.2 | 3.1/2.7 | 5.8/4.49 | 14/8.9 | 3.32/2.82 | 2.99/2.44 | 16.5/12 |
| Map resolution range (local, Å) | 3.2-20 | 2.7-20 | 4.5-20 | 7.3-20 | 2.83-20 | 2.44-20 | 8-20 |
| Map sharpening *B* factor (Å$^2$) | -50 | - 76.6 | - 222.4 | -20 | -30 | -50 | -100 |
| Map sharpening methods | RELION3.1 | RELION3.1 | RELION3.1 | RELION3.1 | RELION3.1 | RELION3.1 | RELION3.1 |
| **Model Composition** | | | | | | | |
| Protein (residues) | 2184 | 779 | 2184 | 2963 | 3126 | 779 | 3940 |
| DNA (nucleotides) | 0 | 245 | 67 | 305 | 82 | 266 | 352 |
| **Model refinement** | | | | | | | |
| Refinement software | PHENIX | PHENIX | PHENIX | PHENIX | PHENIX | PHENIX | PHENIX |
| - real or reciprocal space | Real space | Real space | Real space | Real space | Real space | Real space | Real space |
| -resolution cutoff | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| Model-Map scores | | | | | | | |
| -CCvolume/mask | 0.73/0.75 | 0.63/0.65 | 0.72/0.73 | 0.6/0.62 | 0.78/0.78 | 0.64/0.66 | 0.59/0.61 |
| *B* factors (Å$^2$) | | | | | | | |
| Protein residues | 125.85 | 185.34 | 173.05 | 182.65 | 83.32 | 185.39 | 118.86 |
| DNA | | 205.96 | 507.76 | 268.55 | 245.18 | 187.14 | 171.24 |
| R.m.s. deviations from ideal values | | | | | | | |
| Bond lengths (Å) (#>4σ) | 0.002 (0) | 0.010 (2) | 0.004 (6) | 0.007 (6) | 0.002 (0) | 0.003 (1) | 0.009 (18) |
| Bond angles (°) (#>4σ) | 0.494 (1) | 0.922 (31) | 0.939 (19) | 0.933 (48) | 0.465 (3) | 0.524 (20) | 1.025 (97) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Validation** | | | | | | | |
| MolProbity score | 1.30 | 1.72 | 1.78 | 1.75 | 1.45 | 1.39 | 1.63 |
| CaBLAM outliers (%) | 1.19 | 1.22 | 1.29 | 1.33 | 1.03 | 1.35 | 1.30 |
| Clashscore | 5.47 | 8.23 | 13.48 | 11.29 | 5.94 | 5.41 | 9.62 |
| Poor rotamers (%) | 0.00 | 0.31 | 0.21 | 0.23 | 0.82 | 0.00 | 0.32 |
| C-beta deviations (%) | 0 | 0 | 0.04 | 0.03 | 0 | 0 | 0.05 |
| EM ringer score | 1.56 | 3.69 | *n.d. | *n.d. | 2.17 | 3.6 | *n.d. |
| Ramachandran plot Favored (%) Outliers (%) | 97.970.00 | 96.050.00 | 97.270.00 | 96.910.03 | 97.320.07 | 97.50.00 | 97.350.18 |

*Not determined (n.d.)

**Table S2. Data collection and refinement statistics (molecular replacement)**

| | CENP-HIK$^{Head}$ (PDB:7PB4) | CENP-OPQUR (PDB:7PB8) |
|---|---|---|
| **Data collection** | | |
| Space group | $P2_12_12_1$ | C2 |
| Cell parameters | | |
| *a*, *b*, *c* (Å) | 44.6, 56.1, 176.7 | 119.6, 51.3, 110.0 |
| α, β, γ (°) | 90.0, 90.0, 90.0 | 90.0, 117.5, 90.0 |
| Resolution (Å) | 53.5-2.5 (2.6-2.5)* | 46.3-2.7 (3.9-3.7) |
| $R_{sym}$ or $R_{merge}$ | 0.13 (0.7) | 0.084 (0.35) |
| $I / \sigma I$ | 8.5 (2.5) | 7.6 (1.4) |
| Completeness (%) | 99.5 (88.8) | 77.5 (73.8) |
| Redundancy | 3.7 (4.9) | 2.5 (2.3) |
| **Refinement** | | |
| Resolution (Å) | 53.5-2.5 (2.6-2.5) | 46.3-3.7 (3.8-3.7) |
| Reflections (N) | 11771 | 6104 |
| $R_{work}$ / $R_{free}$ | 0.21/0.29 | 0.26/0.34 |
| Atoms | | |
| Protein (N) | 2868 | 5497 |
| Ligand/ion (N) | 0 | 0 |
| Water (N) | 0 | 0 |
| *B*-factors | | |
| Protein | 42.4 | 111.0 |
| Ligand/ion | N/A | N/A |
| Water | N/A | N/A |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.009 | 0.005 |
| Bond angles (°) | 1.186 | 0.92 |

Both experiments were performed on single crystals.
*Values in parentheses are for highest-resolution shell.

**Movie S1.**

3D variability analysis demonstrates a continuous conformational heterogeneity of the DNA termini of the CENP-A$^{Nuc}$.

**Movie S2.**

Overview of the cryo-EM reconstruction and molecular model of the CCAN$^{\Delta C}$-DNA complex.