CryoEM structure of human CENP-A nucleosome alone and in complex with central domain of human CENP-C

Ahmad Ali-Ahmad[1], Silvija Bilokapić[2], Ingmar B. Schäfer[3], Mario Halić[2*], Nikolina Sekulić[1,4*]

[1]Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, Faculty of Medicine, University of Oslo, 0318 Oslo, Norway.

[2] Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

[3] Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Munich, Germany

[4]Department of Chemistry, University of Oslo, P.O. Box 1033, NO-0315 Blindern, Norway.

*Correspondence should be addressed to nikolina.sekulic@ncmm.uio.no and mario.halic@stjude.org

Keywords: centromere, nucleosome, CENP-A, CENP-C, cryoEM

Abstract:

Centromere is a chromosome locus necessary for genetic stability during cell division. Centromere is defined epigenetically, by nucleosome containing histone H3 variant, CENP-A, upon which constitutive centromere-associated network of proteins (CCAN) is built. CENP-C, a member of CCAN, is considered to be a blueprint of centromere. It interacts with several CCAN components and directly bridges CENP-A nucleosomes to kinetochore in mitosis. In this study, we provide new molecular details on structure of CENP-A nucleosome alone and in complex with CENP-C central region (CENP-C^{CR}), main CENP-A binding module of CENP-C. We confirm high-flexibility of DNA ends as intrinsic features of CENP-A-containing nucleosomes independent on DNA sequence it wraps. We establish that, in vitro, two regions of CENP-C (CENP-C^{CR} and CENP-C^{motif}), both bind exclusively CENP-A nucleosome, albeit CENP-C^{CR} with a higher affinity due to extended hydrophobic area made of CENP-A^{V532} and CENP-A^{V533}. Finally, we identify conformational changes on CENP-A nucleosome that occur upon CENP-C binding. Pronounced DNA unwrapping is facilitated by destabilization of H2A N-terminal tail and N-terminal tail of H4 is stabilized in the upward conformation favored for CENP-A specific H4^{K20} monomethylation. Together, our findings provide foundation for understanding epigenetic basis for centromere formation and functioning.

Introduction:

The centromere is a part of the chromosomal locus that directs accurate segregation of chromosomes during cell division (McKinley and Cheeseman, 2016). Defects in chromosome segregation lead to aneuploidy, a hallmark of cancer (Potapova and Gorbsky, 2017).

DNA sequences underlining human centromeres are AT rich repeats (termed α-satellites), but over the years, DNA sequence is shown to be neither necessary nor sufficient for centromere function. Instead, centromeres are specified epigenetically, by the presence of the histone H3 variant, CENP-A [reviewed in (Sekulic and Black, 2012)]. It is clear that chromatin containing CENP-A nucleosomes must have unique structural properties to organize constitutive centromere associated network (CCAN), so full structural understanding of CENP-A nucleosome alone and in complex with 2 components of CCAN with which it interacts directly and specifically – CENP-C and CENP-N (Carroll et al., 2010) is necessary. Initial clues came from the crystal structure of the CENP-A nucleosome (Tachiwana et al., 2011) that implied octameric histone core, similar to canonical nucleosome, with unwrapped DNA ends. Recently, a cryoEM structure of the human CENP-A nucleosome in complex with human CENP-N has been reported by several groups (Chittori et al., 2018; Pentakota et al., 2017; Tian et al., 2018) revealing high-resolution molecular determinants for the CENP-A/CENP-N interaction. However, high-resolution structure of human CENP-A nucleosome in complex with CENP-C is still missing.

CENP-C is a central node of the CCAN, responsible for interactions from the CENP-A nucleosome on the one end towards the mitosis-specific kinetochore subunits of the Mis12 complex on the other (Musacchio and Desai, 2017). Human CENP-C is a 934 amino acid long disordered protein, depletion of which, leads to cell division defects and chromosome mis-segregation (Milks et al., 2009; Song et al., 2002). Two regions on human CENP-C have been identified as nucleosome binding regions: 1. central region (426-537 a.a), CENP-C^{CR}, that is necessary and sufficient to promote CENP-A nucleosome binding *in vitro* and kinetochore targeting *in vivo* and 2. CENP-C motif (736-758 a.a.), CENP-C^{motif}, that bears conservation across species, but is not sufficient for centromere recruitment in the absence of endogenous CENP-C with the dimerization domain and is dispensable for epigenetic stability of the CENP-A nucleosomes (Guo et al., 2017; Milks et al., 2009; Musacchio and Desai, 2017).

Current molecular understanding of the CENP-A nucleosome/CENP-C interactions are based on the crystal structure of the canonical *D. melanogaster* nucleosomes where the C-

terminal tail of histone H3 is replaced by the mutated C-terminal tail of rat CENP-A, in complex with rat CENP-C motif (Kato et al., 2013).

Here, we report a 3.8 Å cryoEM structure of CENP-A nucleosome that confirms flexibility of DNA ends as an intrinsic property of CENP-A nucleosome, independent of the nature of the underlining DNA sequence and contributed primarily with the N-terminal tail of CENP-A. Further, we find both nucleosome binding domains of CENP-C, CENP-C^{CR} and CENP-C^{motif}, to be specific for CENP-A nucleosomes, however CENP-C^{CR} showing stronger binding. We determined the cryoEM structure of CENP-A nucleosome in complex with CENP-C^{CR} at 3.1 Å resolution and identified CENP-A^{V532} and CENP-A^{V533} as important determinants of extra specificity and affinity to CENP-A/CENP-C interaction. We notice conformational changes on CENP-A upon binding of CENP-C^{CR}. The enhanced DNA unwrapping is facilitated by destabilized H2A C-terminal tail and H4 N-terminal tail in stabilized in the upward conformation.

In summary, our work provides high-resolution view of the CENP-A nucleosome. We establish CENP-A nucleosomes as sole CENP-C binder and we provide molecular understanding for higher specificity of CENP-C^{CR} over CENP-C^{motif}. Finally, our cryoEM structure of human CENP-A nucleosome in complex with human CENP-C^{CR} provides not only the molecular details of the most relevant interaction between CENP-A nucleosome and CENP-C but it also captures conformational changes in the nucleosome taking place upon binding. Together, presented data provide a solid foundation for further studies on organization and interactions within centromeric chromatin.

Results:

1. CENP-A nucleosome has flexible DNA ends independently of the DNA sequence it wraps

Ever since CENP-A has been identified as an epigenetic mark of centromere, the question raised on how does it distinguish itself from the canonical nucleosome. Recent research strongly supports octameric nucleosome, similar to the canonical one (Hasson et al., 2013; Nechemia-Arbely et al., 2017). In the last 10 years, several studies *in vivo* (Hasson et al., 2013; Roulland et al., 2016) and *in vitro* (Conde e Silva et al., 2007; Roulland et al., 2016; Tachiwana et al., 2011; Zhou et al., 2019) have identified flexible DNA ends as a unique feature of CENP-A nucleosome but it stayed unclear to which degree was observed unwrapping contributed by

DNA sequence or crystal packing. To test this, we digested, with MNase, CENP-A nucleosomes assembled on synthetic super-positioning DNA "601" (Lowary and Widom, 1998) and 2 natural α-satellite DNA constructs (Hasson et al., 2013), first with and second without the CENP-B box, a 17bp sequence recognized by CENP-B (Masumoto et al., 1989). Since exact *in vivo/in vitro* nucleosome positioning is not precisely mapped for the sequence that contains CENP-B box, we used the length of the whole α-satellite repeat (171 bp) with CENP-B box at the end. The DNA was digested faster when assembled on CENP-A nucleosomes in comparison to H3 nucleosomes for all three DNA sequences (Figure 1A, Figure S1A). The DNA unwrapping of CENP-A nucleosome was linked to properties of the N-terminal sequence of CENP-A (Panchenko et al., 2011; Roulland et al., 2016). Indeed, when we substitute residues 1-49 of CENP-A with 1-50 of H3, we completely loose the DNA flexibility (Figure S1B). The same is true for the opposite, H3 nucleosome with the N-terminal tails of CENP-A undergoes DNA unwrapping to a similar extent as CENP-A nucleosome (Figure S1B). From here we conclude that flexibility of the DNA ends is an intrinsic property of CENP-A nucleosome, that is regulated by its N-terminal tail and that is independent of DNA sequence.

Next, we used cryoEM to obtain the high-resolution structure of CENP-A nucleosomes on 601 DNA at 3.8 A (Figure 1B, Figure S1C, Figure S2, Table S1). In contrast to the crystal structure, in our cryoEM structure of CENP-A nucleosome we can see the density for the entire 145 bp of DNA. However, the map is less clear and has lower local resolution for the terminal DNA (Figure S2E), indicating local flexibility. The modeled DNA is shifted by 4 Å in comparison to the one in H3 nucleosome (Bilokapic et al., 2018) (Figure 1C). These results are in agreement with our MNase experiments and with recent structure of CENP-A nucleosome aided by antibody (Zhou et al., 2019). Interestingly, despite the low resolution for the DNA, we can clearly model the N-terminus of CENP-A all the way to CENP-A^{R42}, including αN helix (Figure 1D, Figure S1D). αN helix of CENP-A ends at bulky CENP-A^{W47} while H3 continues with one extra turn (Figure 1C). Arginines 42, 43 and 44 are all involved in DNA binding, although the

interactions are slightly different on the two sides of the nucleosome (Figure 1D, Figure S1D). In our structure, we can also clearly see other CENP-A specific parts of the nucleosome (Figure S1D): The C-terminal tail (-LEEGLG), that specifically binds CENP-C and L1 loop containing Arg 80 ("RG-loop"), that specifically binds CENP-N.

2. CENP-C^{CR} competes out CENP-C^{motif} on CENP-A nucleosomes

Initial efforts to identify CCAN proteins that directly and specifically bind CENP-A nucleosome uncovered CENP-C and CENP-N proteins (Carroll et al., 2009, 2010). In those studies, only central region of human CENP-C (CENP-C^{CR}, residues 426-537) was characterized for CENP-A binding (Figure 2A). In 2013, Kato at al. (Kato et al., 2013) reported a crystal structure of canonical H3 nucleosome with grafted C-terminal tail of mutated rat CENP-A in complex with rat CENP-C residues 710-734, a region known as CENP-C motif (CENP-C^{motif}). Analysis of the structure identified the hydrophobic interactions between CENP-A C-terminal tail and CENP-C^{motif} and authors proposed that both CENP-C^{CR} and CENP-C^{motif} bind nucleosomes, with 5-10 times higher affinity for CENP-A then for H3 nucleosomes. To test if CENP-C^{CR} binds the nucleosome with different affinity than CENP-C^{motif}, we made complexes and analyzed their mobility on the native gels. We found that each, CENP-C^{CR} and CENP-C^{motif}, when mixed with CENP-A nucleosomes, make sharp band on the native gel, corresponding to uniform complex (Figure 2B, 2C). Surprisingly though, when mixed with H3 nucleosomes, both CENP-C constructs, result in a smear on the native gel, indicating non-specific binding (Figure 2B, 2C). We conclude that *in vitro* CENP-C^{CR} and CENP-C^{motif} bind only CENP-A nucleosomes specifically.

Further, we have set a competition experiment to determine which CENP-C region has higher affinity for CENP-A nucleosomes. When we assemble human CENP-A/CENP-C^{motif} complex and start adding CENP-C^{CR}, we observe formation of CENP-A/CENP-C^{CR} complex with almost full saturation at 2 x molar ratio (Figure 2D). In contrary, if we start with CENP-A/CENP-C^{CR} complex and add CENP-C^{motif}, we observe only smearing of the gel indicating non-specific

binding of CENP-C^{motif} to CENP-A/CENP-C^{CR} complex (Figure 2E). CENP-C^{motif} is 71 residues shorter than CENP-C^{CR} (40 vs. 111 residues) and it is possible that those extra residues are contributing additional affinity. To test for this, we made truncated CENP-C^{CR} that has only residues predicted to interact with CENP-A C-terminal tail (501-537; CENP-$C^{CR\text{-}short}$) and is comparative in size to CENP-A^{motif}. We confirmed its specificity for CENP-A nucleosomes (Figure S3). Interestingly, CENP-$C^{CR\text{-}short}$, cannot, even non-specifically, bind H3 nucleosome, suggesting that additional 71 residue present in the CENP-C^{CR}, are responsible for non-specific H3 nucleosome binding, likely through interactions with DNA. In difference to CENP-C^{motif}, CENP-$C^{CR\text{-}short}$ can effectively compete out CENP-C^{CR} from pre-assembled CENP-A/CENP-C^{CR} complexes (Figure 2E) demonstrating that residues within nucleosome binding region of CENP-C^{CR} contribute high affinity of binding.

In summary, we conclude that CENP-C^{CR} provides major interactions between CENP-A and CENP-C. Our findings are in agreement with studies done in cells that found CENP-C^{CR} to be essential for epigenetic propagation of centromere (Guo et al., 2017).

3. Strong hydrophobic interactions are contributing specificity and high affinity between CENP-A nucleosome and CENP-C^{CR}

Next, we wanted to understand how is CENP-C^{CR} achieving this higher affinity. The crystal structure of fruit fly H3-GIEGGL/rat CENP-C^{motif} has identified 2 types of interactions involved in complex formation: 1. electrostatic interactions of rat CENP-$C^{R717/R719}$ with the acidic patch on H2A/H2B, and 2. hydrophobic interactions between rat CENP-$C^{Y725,\,W726}$ and hydrophobic C-terminal tail of CENP-A. Sequence overlay of rat CENP-C^{motif} with human CENP-C^{motif} and human CENP-C^{CR} reveals existence of analogous residues in human protein but does not explain the higher affinity of the CENP-C^{CR} (Figure 3A). Also, since crystallographic study (Kato et al., 2013) used canonical nucleosome with grafted C-terminal of CENP-A, it failed to capture possible conformational changes that could occur on CENP-A nucleosome upon CENP-C binding. To resolve this, we solved 3.1 Å cryoEM structure of human CENP-A

nucleosome in complex with human CENP-C^{CR} (Figure 3B, Table S2, Figure S4, Figure S5). Our maps show CENP-C bound to both sides of the nucleosome with its CENP-$A^{R521, R522}$ anchoring residues, while the remaining residues in CENP-C^{CR} show fuzzy interactions. In order to visualize larger CENP-C fragment, we have further sorted particles and increased CENP-C occupancy on one of the sides, so that we can trace residues 519 - 536. We observe strong electrostatic interactions between human CENP-$C^{R521,R522}$ and the acidic patch of H2A/H2B, similar to those seen in structure with CENP-C^{motif} (Kato et al., 2013) (Figure 3C, Figure S4). We also observe human CENP-$C^{W530,W531}$ fitting nicely in the hydrophobic cleft formed by the C-terminal tail of CENP-A (Figure 3C, Figure S4B). Further, conformation of C-terminal tail of CENP-A changes slightly upon binding (Figure 3D) and hydrophobic interactions that we observe are extended in comparison to those reported in the crystal structure (Figure 3E). Two bulky tryptophans (CENP-$C^{W530,W531}$) are clamped between CENP- A^{R131} and CENP- $A^{L135, L139}$ while CENP-C^{V532} and CENP-C^{V533} stabilize hydrophobic patch on H4 formed by $H4^{V58}$ with $H4^{V61}$. These interactions explain perturbations observed by NMR in H4 residues on nucleosome upon CENP-C^{CR} binding (Kato et al., 2013). To test if extended hydrophobicity contributed by CENP-$C^{V532,V533}$ plays a role in stronger CENP-C^{CR}/CENP-A nucleosome interactions, we generated CENP-C mutant devoid of residues 530-537, CENP-$C^{CR-\Delta C}$ (Figure 3E) . This mutant fails to make a complex with CENP-A nucleosome (Figure 3G), indicating that the residues following CENP-$C^{W530,W531}$ are important for binding. Point mutations further revealed that replacement of CENP-C^{V533} with aspartic acid is sufficient to eliminate binding to the CENP-A nucleosome (Figure 3G). We also notice that conformation of CENP-A C-terminal tail changes slightly upon binding of CENP-C^{CR} (Figure 3D).

From here, we conclude that human CENP-C^{CR} binds human CENP-A nucleosome through electrostatic interactions between CENP-$C^{R521,R522}$ and acidic patch on the nucleosome and extensive hydrophobic network formed between CENP-$C^{W530, W531, V532, V533}$ and C-terminal tail of CENP-A aided by hydrophobic patch of H4. Our data show that CENP-C^{CR} makes additional

hydrophobic contacts with the CENP-A nucleosome that have not been seen for CENP-C^{motif}. These contacts are required for specific CENP-A nucleosome binding and for the higher affinity of CENP-C^{CR} vs. CENP-C^{motif}.

4. Conformational changes on CENP-A nucleosome upon CENP-C binding

Nucleosome in our experiments has a full length CENP-A, which enables us to assess conformational changes on the nucleosome upon CENP-C binding. It was proposed before that CENP-C binding rigidifies the histone core, making it more histone H3-like, while at the same time it further enhances DNA unwrapping (Falk et al., 2015). Our MNase experiments on CENP-A nucleosome in complex with CENP-C^{CR} or CENP-C^{motif} show the increased DNA digestion in both cases (Figure 4A, Figure S6A). The comparison of the local resolution maps between CENP-A and CENP-A/CENP-C complex (Figure S3E; Figure S5E) indicates more DNA unwrapping in samples with CENP-C bound. To further confirm this we did a careful classification of the entry/exit sites for each of the two samples (Figure S3E; Figure S5E). We find that, although flexible, the DNA at the entry/exit site in CENP-A nucleosome samples only limited space (distance between most extreme DNA conformation is 2 Å) (Figure 4B), while in CENP-A/CENP-C complex the most extreme distance between different subpopulations is 9 Å (Figure 4C). Concomitantly, we observe no density for C-terminal tail of H2A (Figure 4C), indicating disorder in this part of the nucleosome upon CENP-C binding. Strikingly, in cryoEM structure of canonical nucleosome with tightly wrapped DNA, the density for C-terminal tail of H2A is one of the most well resolved parts of the structure (Bilokapic et al., 2018). Removal of the C-terminal tail of H2A leads to decreased stability of nucleosome and alters nucleosome positioning and interactions with H1 (Eickbush et al., 1988; Vogler et al., 2010). Molecular dynamic studies have identified residues $H2A^{K118}$ and $H2A^{K119}$ to interact with DNA most likely securing DNA wrap (Biswas et al., 2011). Indeed, if we remove residues 110-130 in H2A of CENP-A nucleosome (CENP-$A^{\Delta C\text{-tail H2A}}$), we observe increased DNA digestion (Figure 4D, Figure S6B, top), and addition of CENP-C^{CR} to these nucleosomes does not have any additional

effect on the digestion speed (Figure 4D, Figure S6B, bottom). To the contrary, DNA digestion of H3 nucleosome was only mildly affected by removal of the C-terminus of H2A (Figure S6C, top). We next hypothesized that different MNase pattern of CENP-A vs. H3 nucleosome (in the background of ΔC-tail H2A) must be contributed by residues in the N-terminal tail of CENP-A/H3, since this region of the nucleosome is in the close contact with C-terminal H2A. Truly, in ΔC-tail H2A background, MNase digestion of nucleosome with H3 core and CENP-A tail is highly similar to CENP-A nucleosome while nucleosome with CENP-A core and H3 tail is similar to that of the H3 nucleosome (Figure S6C, middle). These results demonstrate contribution of C-terminal tail of H2A to nucleosome DNA wrapping synergistically with N-terminal tail of CENP-A. In the context of CENP-A nucleosome that has shorter αN helix, DNA wrapping is already compromised and removal of C-terminal H2A results in further unwrapping. Consistently, we find that in CENP-A nucleosome with H3 N-tail, binding of CENP-C alone cannot induce DNA unwrapping (Figure S6C, bottom). From here we conclude that N-terminal tail of CENP-A favors DNA unwrapping but C-terminal of H2A is counteracting it. CENP-C binding destabilizes C-terminal of H2A leading to increased DNA unwinding. We find that this alternative, destabilized conformation of H2A C-tail is most likely caused by interaction of bulky hydrophobic CENP-C^{W530,W531} with H2A^{L108}. In addition, CENP-C^{W530,W531} is orienting CENP-A^{R131} to establish salt bridge with H2A^{Q112} (Figure S6D).

Further, we noticed that the N-terminal tail of H4 clearly adopts upwards conformation on both sides of the CENP-A nucleosome, but in the presence of CENP-C this conformation gets more rigidified (Figure 4E). Monomethylation of H4 is specific for CENP-A nucleosome and this modification is necessary for epigenetic establishment of kinetochore (Hori et al., 2014). Recent structural study (Arimura et al., 2019), proposed upwards conformation of H4 tail to be required for establishment and/or maintenance of H4^{K20} monomethylation specific for CENP-A nucleosome. Our structures confirm that H4 tail adopts upward conformation in the

context of CENP-A nucleosome and binding of CENP-C further stabilizes and rigidifies H4 N-terminal tail most likely through interaction with $H4^{V61}$.

In summary, we find that binding of $CENP\text{-}C^{CR}$ to CENP-A nucleosome enhances DNA unwrapping by destabilizing C-terminal tail of H2A while it, at the same time stabilizes histone H4 C-terminal tail in the upward conformation important for centromere specific monomethylation of $H4^{K20}$.

5. Discussion

CENP-A is an epigenetic mark for centromere, a chromosome locus essential for genome integrity. CENP-A is a histone H3 variant with a histone core being 64% identical to H3 and completely divergent tails. Ever since it has been identified as an epigenetic determinant for centromere, the question was raised how does nucleosome, containing CENP-A, distinguishes itself from the rest of the chromatin. Over the years, number of models were proposed, including different histone stoichiometries, presence of non-histone proteins and alternative DNA wrapping [summarized in (Black and Cleveland, 2011)]. Finally, octameric nucleosome with right-handed DNA wrap, much like canonical nucleosome, emerged as a favorite model for human CENP-A nucleosome, based on both *in vitro* and *in vivo* studies (Hasson et al., 2013; Nechemia-Arbely et al., 2017; Sekulic et al., 2010), leaving open the question of what is making CENP-A nucleosomes so unique. Crystal structure of CENP-A nucleosome (Tachiwana et al., 2011), *in vivo* MNase experiments (Hasson et al., 2013) and initial cryoEM studies (Roulland et al., 2016) all pointed towards enhanced DNA unwrapping as CENP-A specific feature where unwrapped DNA results in a different chromatin architecture able to accommodate binding of CCAN components. Indeed, the experiments presented here, together with the structure of CENP-A, confirm high flexibility of the DNA ends as an intrinsic feature of CENP-A nucleosome independent on the DNA sequence, but instead encoded in the N-terminal tail of CENP-A.

Next most intriguing question is how CENP-A nucleosome directly binds 2 CCAN components, CENP-N and CENP-C, and how is their binding changing the nucleosome. Recent work (Chittori et al., 2018; Pentakota et al., 2017; Tian et al., 2018) provides the atomic resolution view at CENP-A/CENP-N interaction, which involves recognition of the solvent-exposed L1 loop on CENP-A and interaction with DNA, but structure of CENP-A nucleosome in complex with CENP-C is still missing. Currently, 2 different parts of CENP-C are proposed to bind CENP-A nucleosome: CENP-C^{CR} and CENP-C^{motif}. CENP-C^{CR} is necessary and sufficient for CENP-A nucleosome binding *in vitro* and centromere targeting and stability *in vivo* while the CENP-C^{motif} can be recruited to kinetochores only in the presence of homo-dimerization domain (Guo et al., 2017; Milks et al., 2009; Musacchio and Desai, 2017). Having two independent CENP-C modules able to bind nucleosome at centromeres with different affinities, led to the proposal (Musacchio and Desai, 2017) of CENP-C directly mediating CENP-A loading during the cell cycle. The model assumes that, CENP-C module with weaker nucleosome binding, CENP-C^{motif}, alternates between binding of H3.3 nucleosome and CENP-A nucleosome while the stronger module, CENP-C^{CR}, stays stably associated with CENP-A nucleosome in all phases of the cell cycle. Here, we show that, indeed, CENP-C^{CR} binds CENP-A nucleosome with higher affinity than CENP-C^{motif}, but neither of the CENP-C modules can make uniform complexes with H3 nucleosome.

Further, structural understandings of CENP-C binding are based on crystal structure of chimeric H3 nucleosome with C-terminal tail of CENP-A (Kato et al., 2013). Although, C-terminal tail of CENP-A indeed is the major determinant for CENP-C binding, other parts of CENP-A may also play a role. For example, chimeric CENP-A with N-terminal tail of H3, has reduced recruitment of CENP-C in cells (Roulland et al., 2016). In our high-resolution cryoEM structure of human CENP-A nucleosome in complex with human CENP-C^{CR}, we indeed observe the electrostatic interactions in the acidic patch and hydrophobic interactions with the C-terminal tail of the CENP-A as seen in the crystal structure (Kato et al., 2013). Additionally, we

find that longer hydrophobic stretch on CENP-C^{CR}, formed by CENP-C^{V532} and CENP-C^{V533}, is essential for robustness of CENP-A nucleosome/CENP-C^{CR} interactions, that in agreement with essential role of CENP-C^{CR} in establishing epigenetic stability of CENP-A (Guo et al., 2017). We also find that binding of hydrophobic CENP-C perturbs conformation of C-terminal part of the H2A tail, detaching it from the surface of the nucleosome. H2A C-terminal tail is involved in stabilization of the DNA wrap, so this conformational change in the context of already compromised CENP-A DNA wrap leads to enhanced unwrapping. However, CENP-C induced destabilization of C-terminal tail of H2A does not have an effect in the absence of CENP-A N-terminal tail. It could be that enhanced DNA unwrapping positively regulates recruitment of CENP-C in cell explaining results reported before (Roulland et al., 2016).

In both our structures, CENP-A nucleosome alone and CENP-A/CENP-C^{CR} complex, we see upward conformation of C-terminal tail of H4. At the same time as hydrophobic interactions between CENP-C and H2A^{L108} are destabilizing N-terminal of H2A, we find that hydrophobic interactions between CENP-C and H4^{V60} propagate to stabilize the N-terminal tail of H4. The tail is stabilized in upwards, CENP-A specific conformation, favoring H4^{K20} monomethylation, essential for kinetochore assembly (Arimura et al., 2019; Hori et al., 2014).

Together, our structures provide essential and long-needed high-resolution view of the fundamental building block of centromere, CENP-A nucleosome alone and in complex with CENP-C, protein imprinting the backbone of CCAN. Our biochemical and structural analysis establishes CENP-C as an exclusive and multivalent binder of CENP-A nucleosomes that employs 2 independent modules and homo-dimerization to crosslink sparse CENP-A nucleosomes in chromatin and initiate functional centromere by bringing other CCAN components (Musacchio and Desai, 2017), likely by enhancing DNA unwrapping.

References:

Arimura, Y., Tachiwana, H., Takagi, H., Hori, T., Kimura, H., Fukagawa, T., and Kurumizaka, H. (2019). The CENP-A centromere targeting domain facilitates H4K20 monomethylation in the nucleosome by structural polymorphism. Nat. Commun. *10*, 576.

Bilokapic, S., Strauss, M., and Halic, M. (2018). Histone octamer rearranges to adapt to DNA unwrapping. Nat. Struct. Mol. Biol. *25*, 101.

Biswas, M., Voltz, K., Smith, J.C., and Langowski, J. (2011). Role of Histone Tails in Structural Stability of the Nucleosome. PLOS Comput. Biol. *7*, e1002279.

Black, B.E., and Cleveland, D.W. (2011). Epigenetic Centromere Propagation and the Nature of CENP-A Nucleosomes. Cell *144*, 471–479.

Carroll, C.W., Silva, M.C.C., Godek, K.M., Jansen, L.E.T., and Straight, A.F. (2009). Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. Nat. Cell Biol. *11*, 896–902.

Carroll, C.W., Milks, K.J., and Straight, A.F. (2010). Dual recognition of CENP-A nucleosomes is required for centromere assembly. J. Cell Biol. *189*, 1143–1155.

Chittori, S., Hong, J., Saunders, H., Feng, H., Ghirlando, R., Kelly, A.E., Bai, Y., and Subramaniam, S. (2018). Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N. Science *359*, 339–343.

Conde e Silva, N., Black, B.E., Sivolob, A., Filipski, J., Cleveland, D.W., and Prunell, A. (2007). CENP-A-containing Nucleosomes: Easier Disassembly versus Exclusive Centromeric Localization. J. Mol. Biol. *370*, 555–573.

Eickbush, T.H., Godfrey, J.E., Elia, M.C., and Moudrianakis, E.N. (1988). H2a-specific proteolysis as a unique probe in the analysis of the histone octamer. J. Biol. Chem. *263*, 18972–18978.

Falk, S.J., Guo, L.Y., Sekulic, N., Smoak, E.M., Mani, T., Logsdon, G.A., Gupta, K., Jansen, L.E.T., Duyne, G.D.V., Vinogradov, S.A., et al. (2015). CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. Science *348*, 699–703.

Guo, L.Y., Allu, P.K., Zandarashvili, L., McKinley, K.L., Sekulic, N., Dawicki-McKenna, J.M., Fachinetti, D., Logsdon, G.A., Jamiolkowski, R.M., Cleveland, D.W., et al. (2017). Centromeres are maintained by fastening CENP-A to DNA and directing an arginine anchor-dependent nucleosome transition. Nat. Commun. *8*, 15775.

Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., and Black, B.E. (2013). The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat. Struct. Mol. Biol. *20*, 687–695.

Hori, T., Shang, W.-H., Toyoda, A., Misu, S., Monma, N., Ikeo, K., Molina, O., Vargiu, G., Fujiyama, A., Kimura, H., et al. (2014). Histone H4 Lys 20 Monomethylation of the CENP-A Nucleosome Is Essential for Kinetochore Assembly. Dev. Cell *29*, 740–749.

Kato, H., Jiang, J., Zhou, B.-R., Rozendaal, M., Feng, H., Ghirlando, R., Xiao, T.S., Straight, A.F., and Bai, Y. (2013). A Conserved Mechanism for Centromeric Nucleosome Recognition by Centromere Protein CENP-C. Science *340*, 1110–1113.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning1. J. Mol. Biol. *276*, 19–42.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. (1989). A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. *109*, 1963–1973.

McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nat. Rev. Mol. Cell Biol. *17*, 16–29.

Milks, K.J., Moree, B., and Straight, A.F. (2009). Dissection of CENP-C–directed Centromere and Kinetochore Assembly. Mol. Biol. Cell *20*, 4246–4255.

Musacchio, A., and Desai, A. (2017). A Molecular View of Kinetochore Assembly and Function. Biology *6*, 5.

Nechemia-Arbely, Y., Fachinetti, D., Miga, K.H., Sekulic, N., Soni, G.V., Kim, D.H., Wong, A.K., Lee, A.Y., Nguyen, K., Dekker, C., et al. (2017). Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. J Cell Biol *216*, 607–621.

Panchenko, T., Sorensen, T.C., Woodcock, C.L., Kan, Z., Wood, S., Resch, M.G., Luger, K., Englander, S.W., Hansen, J.C., and Black, B.E. (2011). Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. Proc. Natl. Acad. Sci. *108*, 16588–16593.

Pentakota, S., Zhou, K., Smith, C., Maffini, S., Petrovic, A., Morgan, G.P., Weir, J.R., Vetter, I.R., Musacchio, A., and Luger, K. (2017). Decoding the centromeric nucleosome through CENP-N. ELife *6*.

Potapova, T., and Gorbsky, G.J. (2017). The Consequences of Chromosome Segregation Errors in Mitosis and Meiosis. Biology *6*, 12.

Roulland, Y., Ouararhni, K., Naidenov, M., Ramos, L., Shuaib, M., Syed, S.H., Lone, I.N., Boopathi, R., Fontaine, E., Papai, G., et al. (2016). The Flexible Ends of CENP-A Nucleosome Are Required for Mitotic Fidelity. Mol. Cell *63*, 674–685.

Sekulic, N., and Black, B.E. (2012). Molecular underpinnings of centromere identity and maintenance. Trends Biochem. Sci. *37*, 220–229.

Sekulic, N., Bassett, E.A., Rogers, D.J., and Black, B.E. (2010). The structure of (CENP-A-H4)2 reveals physical features that mark centromeres. Nature *467*, 347–351.

Song, K., Gronemeyer, B., Lu, W., Eugster, E., and Tomkiel, J.E. (2002). Mutational analysis of the central centromere targeting domain of human centromere protein C, (CENP-C). Exp. Cell Res. *275*, 81–91.

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.-Y., et al. (2011). Crystal structure of the human centromeric nucleosome containing CENP-A. Nature *476*, 232–235.

Tian, T., Li, X., Liu, Y., Wang, C., Liu, X., Bi, G., Zhang, X., Yao, X., Zhou, Z.H., and Zang, J. (2018). Molecular basis for CENP-N recognition of CENP-A nucleosome on the human kinetochore. Cell Res. *28*, 374–378.

Vogler, C., Huber, C., Waldmann, T., Ettig, R., Braun, L., Izzo, A., Daujat, S., Chassignet, I., Lopez-Contreras, A.J., Fernandez-Capetillo, O., et al. (2010). Histone H2A C-Terminus Regulates Chromatin Dynamics, Remodeling, and Histone H1 Binding. PLOS Genet. *6*, e1001234.

Zhou, B.-R., Yadav, K.N.S., Borgnia, M., Hong, J., Cao, B., Olins, A.L., Olins, D.E., Bai, Y., and Zhang, P. (2019). Atomic resolution cryo-EM structure of a native-like CENP-A nucleosome aided by an antibody fragment. Nat. Commun. *10*, 2301.

Acknowledgements:

This work was supported by Centre for Molecular Medicine Norway and Norwegian Research Council (N.S. and A.A.). S.B. and M.H. are supported by St. Jude Children's Research Hospital. We would like to thank Elena Conti and the cryo-EM facility at Max Planck Institute for Biochemistry in Martinsried for access to cryo-EM microscopes. Without their support this work would not be possible.

Competing interests:

The authors declare no conflicts of interest.

Authors contributions:

N.S. and M.H. conceived and supervised the project. A.A. carried out mutagenesis, protein expression, purification and assembled nucleosomes and CENP-C complexes and did MNase experiments and analysis. S.B. prepared grids for cryo-EM. S.B., I.S. and M.H. collected cryoEM data. S.B. and M.H. processed cryo-EM data. A.A., N.S., S.B. and M.H. analyzed data and built the structure and A.A. refined the final models. A.A. and N.S. wrote the manuscript and all authors commented on it.

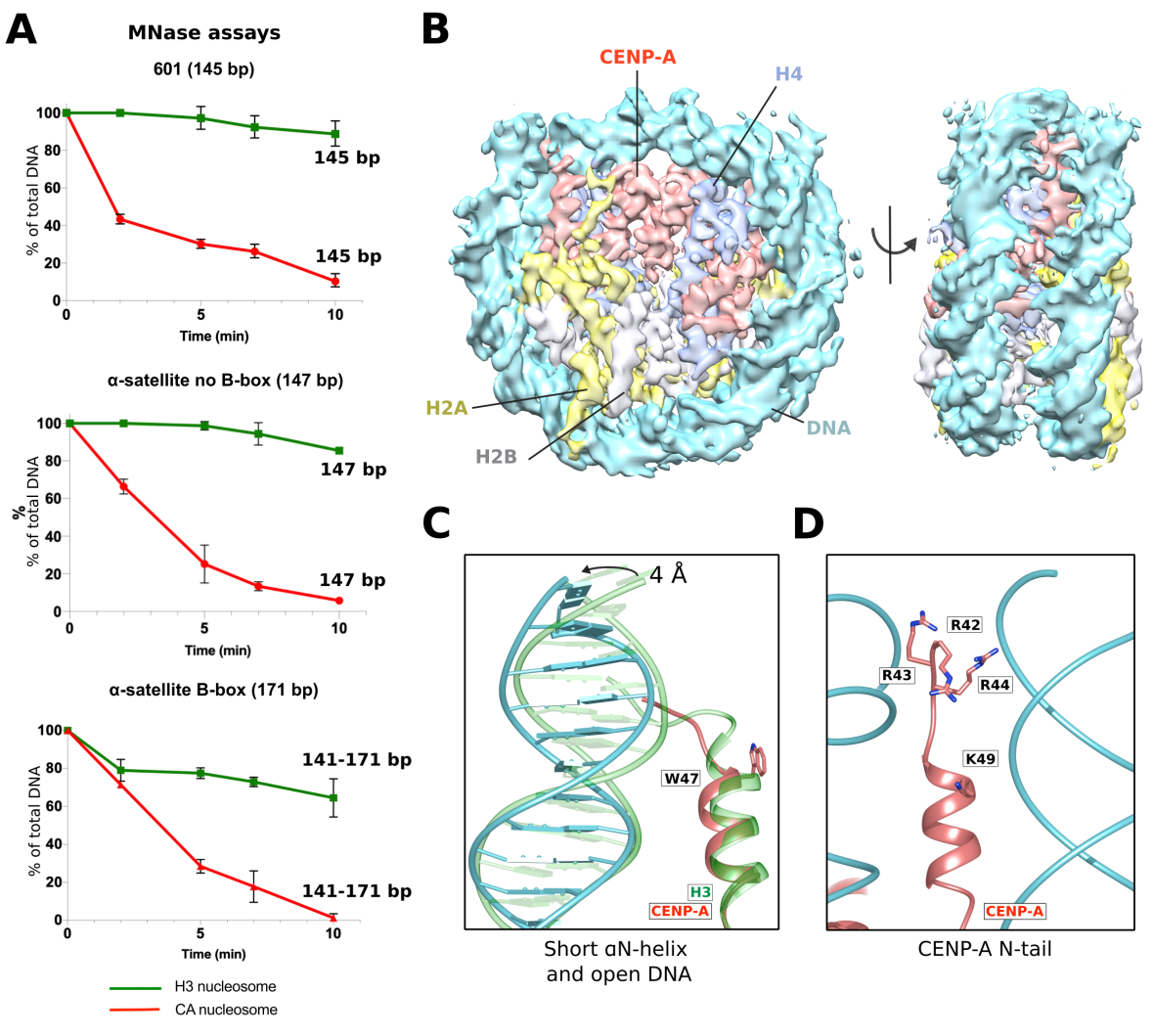


Figure 1: ***CENP-A nucleosome has flexible DNA ends independently of the wrapped DNA sequence.*** ***A.*** *Graphs showing the relative abundance of undigested DNA as a function of time during digestion with micrococcal nuclease (MNase) for 3 types of DNA wrapped around CENP-A nucleosome (red) and H3 nucleosome (green). For α-satellite DNA with initial size of 171 bp, range of sizes (141-171) corresponding to DNA lengths above NCP (nucleosome core particle) is presented. Standard error bar is shown for each time point based on 3 independent experiments. Virtual gels from Bioanaylzer are in Figure S1A.* ***B.*** *CryoEM density map of the human CENP-A nucleosome, colour-coded for histones and DNA.* ***C.*** *Overlay of N-terminal tail of CENP-A (red) and H3 (green; PDB 6ESF) illustrating shorter αN helix of CENP-A (obstructed by the presence of bulky CENP-A^{W47}) and DNA moved by 4 Å.* ***D.*** *N-terminal tail of CENP-A (red) makes contacts with the DNA (cyan) at the entry/exit sites.*

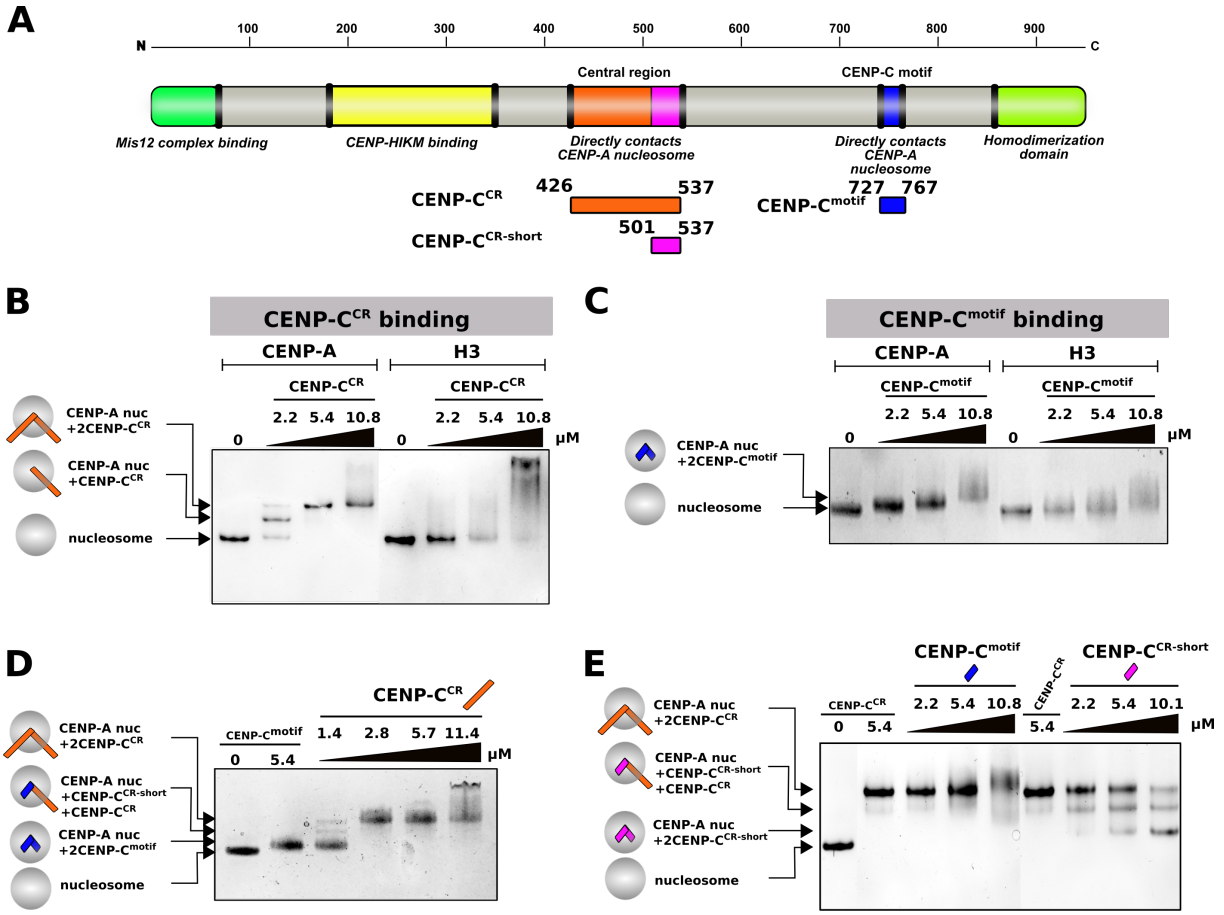


Figure 2: ***Both CENP-C^{CR} and CENP-C^{motif}, bind specifically CENP-A nucleosome and CENP-C^{CR} easily competes out CENP-C^{motif} bound to CENP-A nucleosome. A.*** *Schematic diagram of full-length CENP-C protein, indicating parts involved in interactions with other proteins or homodimerization. Constructs used in this study are depicted below the diagram.* ***B.*** *Native PAGE gel stained with EtBr showing complexes formed between CENP-A or H3 nucleosome and CENP-C^{CR}. Lane 1. CENP-A nucleosome, Lanes 3-5. Increasing amounts of CENP-C^{CR} are added to CENP-A nucleosome. Generation of sharp band with slower mobility indicates formation of a specific CENP-A/CENP-C^{CR} complex. Lane 5. H3 nucleosome. Lanes 6-8. Increasing amounts of CENP-C^{CR} are added to H3 nucleosome. Smear on the gel indicates formation of non-specific CENP-A/CENP-C^{CR} complexes.* ***C.*** *Same experiment as in B using CENP-C^{motif}.* ***D.*** *Native gel showing CENP-C^{CR} competing out CENP-C^{motif} bound to CENP-A nucleosome. Lane 1. CENP-A nucleosome, Lane 2. CENP-A/CENP-C^{motif} complex. Lane 3-6. Increasing amounts of CENP-C^{CR} are added to pre-formed CENP-A/CENP-C^{motif} complex. Formation of slower migrating bands indicates that longer CENP-C^{CR} is replacing shorter CENP-C^{motif} bound to CENP-A nucleosome.* ***E.*** *Native gel showing inability of CENP-C^{motif} to compete out CENP-C^{CR} bound to CENP-A nucleosome. Lane 1. CENP-A nucleosome. Lane 2. and 6. CENP-C^{motif}/CENP-A nucleosome complex. Lanes 3-5. Increasing amounts of CENP-C^{motif} are added to pre-formed CENP-A/CENP-C^{CR} complex. Formation of smear at high amounts of CENP-C^{motif} added, indicates that CENP-C^{motif}, at high concentrations, non-specifically binds CENP-A/CENP-C^{CR} complex rather than replacing bound CENP-C^{CR}. Lane 7-9. Increasing amounts of CENP-C$^{CR\text{-}short}$ are added to pre-formed CENP-A/CENP-C^{CR} complex. Formation of bands with higher mobility indicates that smaller CENP-C$^{CR\text{-}short}$ is effectively replacing bigger CENP-C^{CR} bound to CENP-A nucleosome. For B-E, 2.4 μM nucleosomes are used in all experiments.*

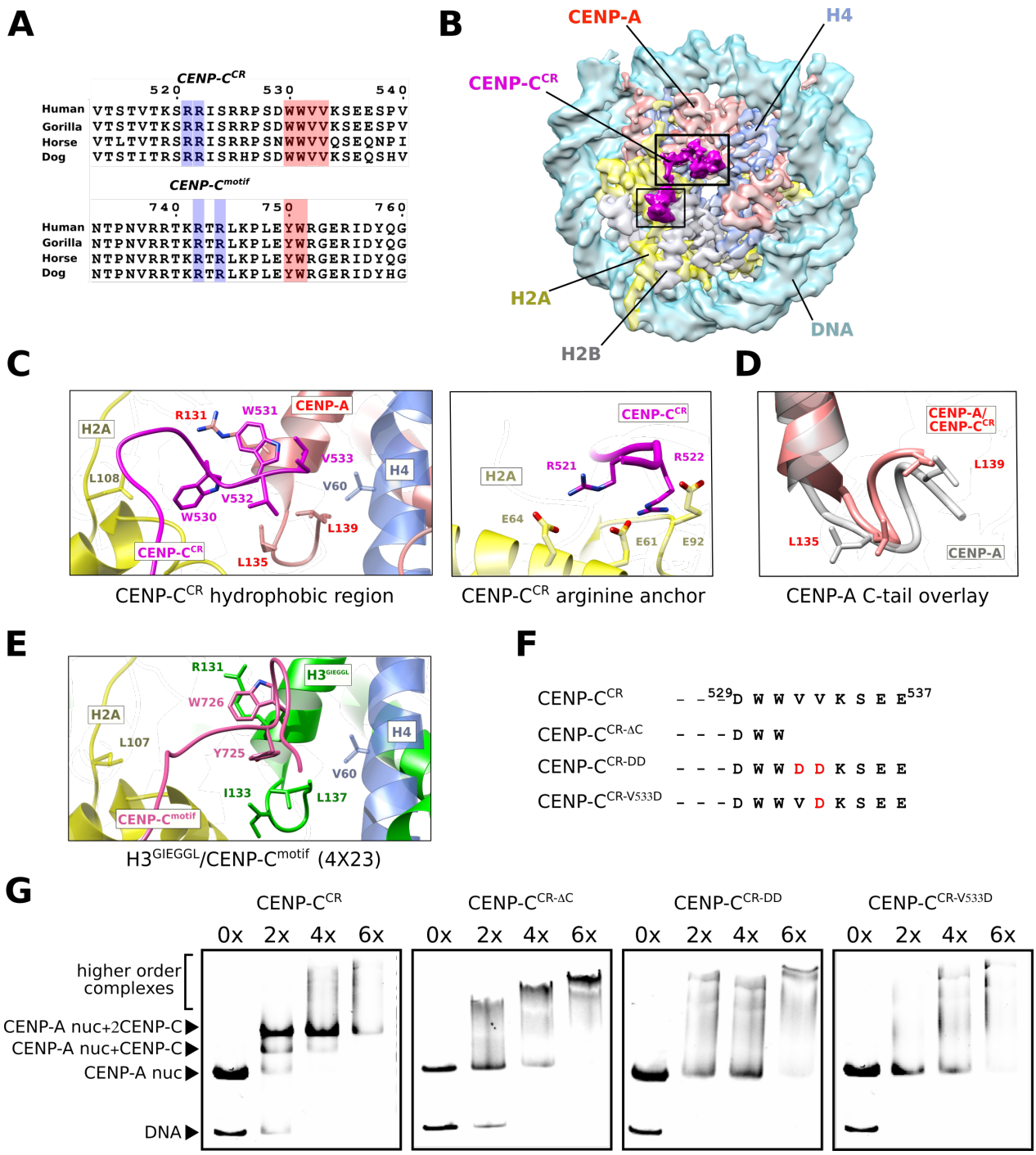


Figure 3. Strong hydrophobic interactions are contributing specificity and high affinity between CENP-A nucleosome and CENP-C^CR. ***A.*** *Sequence alignment of CENP-C^CR and CENP-C^motif regions from different mammals. Conserved residues involved in CENP-A binding are highlighted in blue (electrostatic interactions) and pink (hydrophobic interactions).* ***B.*** *CryoEM density map of the human CENP-A nucleosome, color-coded for histones, in complex with CENP-C^CR (magenta).* ***C.*** *Ribbon diagram showing interactions of CENP-C^CR hydrophobic region (left) and CENP-C^CR arginine anchor (left) with CENP-A nucleosome.* ***D.*** *Overlay of CENP-A C-terminal tail before (gray) and after (red) binding of CENP-C^CR.* ***E.*** *Interactions between H3-GIEGGL and rat CENP-C^motif as observed in PDB 4X23.* ***F.*** *Schematic diagram of mutated CENP-C^CR sequences used to test importance of CENP-C^V532 and CENP-C^V533 for generation of CENP-A/CENP-C^CR complexes.* ***G.*** *Native gels showing impaired ability of mutated CENP-C^CR to form complexes with CENP-A nucleosome. Molar ratio of CENP-C^CR/CENP-A nucleosomes is shown above each lane.*

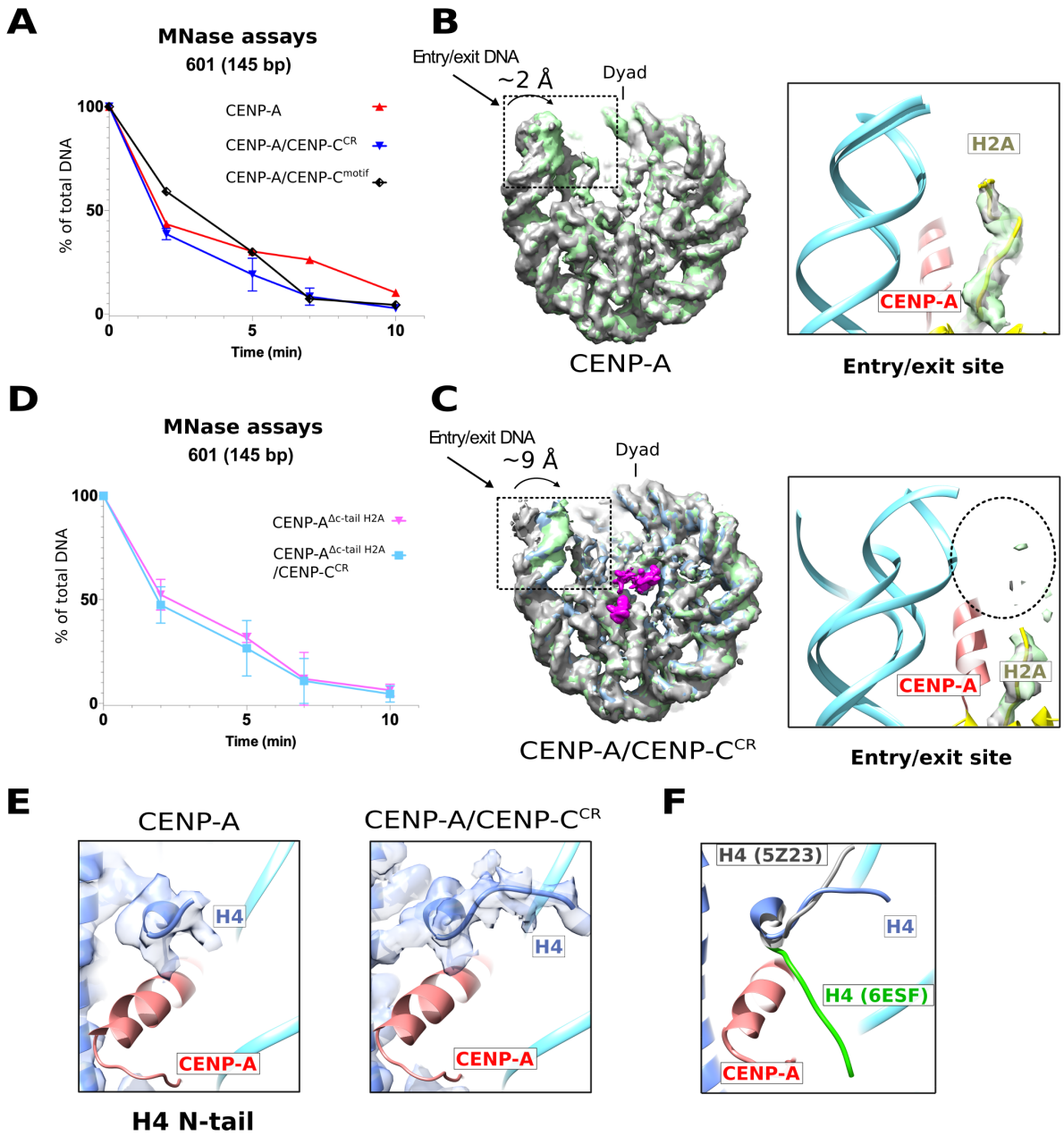


Figure 4. Induced conformational changes of CENP-A nucleosome upon CENP-C^{CR} binding. ***A.*** *Graph showing the relative abundance of undigested DNA (145 bp) as a function of time during digestion with micrococcal nuclease (MN-ase) for CENP-A nucleosome (red), CENP-A/CENP-C^{CR} complex (blue) and CENP-A/CENP-C^{motif} complex (black).* ***B.*** *Overlay of three cryoEM maps of CENP-A nucleosome obtained by sorting on the DNA entry/exit site (Figure S2). Distance between most open (grey) and most closed (green) map is 2 Å. Entry/exit site is boxed and corresponding model in shown on the right. Note that density for H2A C-tail is well defined.* ***C.*** *Same type of data as in A. for CENP-A nucleosome assembled with tailless H2A (H2A, 1-109) alone (pink) and in complex with CENP-C^{CR} (light blue).* ***D.*** *Overlay of three cryoEM maps of CENP-A/CENP-C^{CR} complex obtained by sorting on the DNA entry/exit site (Figure S5). Distance between most open (grey) and most closed (green) map is 9 Å. Entry/exit site is boxed and corresponding model is shown on the right. Note the absence of clearly defined density for H2A C-tail.* ***E.*** *Structure of H4 N-terminal tail with corresponding density in the CENP-A nucleosome (right) and CENP-A/CENP-C^{CR} complex (left).* ***F.*** *Overlay of H4 N-tail from CENP-A/CENP-C^{CR} complex (blue) with the H4 N-tail from H3 nucleosome (green; PDB 6ESF) and H3^{CATD} (grey; PDB 5Z23). For A and C. Standard error bar is shown for each time point based on 3 independent experiments and virtual gels are in Figure S6.*

Supplementary materials for

CryoEM structure of human CENP-A nucleosome alone and in complex with central domain of human CENP-C

Ahmad Ali Ahmad[1], Silvija Bilokapić[2], Ingmar B. Schäfer[3], Mario Halić[2]*, Nikolina Sekulić[1,4]*

[1]Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, Faculty of Medicine, University of Oslo, 0318 Oslo, Norway.

[2]Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

[3]Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Munich, Germany

[4]Department of Chemistry, University of Oslo, P.O. Box 1033, NO-0315 Blindern, Norway.

*Correspondence should be addressed to nikolina.sekulic@ncmm.uio.no and mario.halic@stjude.org

Materials and methods:

Protein purification. Human histones, CENP-A and the CENP-C central domain, were expressed and purified as previously described in (Sekulic and Black, 2016). Briefly, the CENP-A/H4 hetero-tetramer was expressed from a bicistronic plasmid in *E. coli* pLysS under soluble conditions and purified on a hydroxyapatite column followed by cation exchange chromatography. H3, H4, H2A and H2B were expressed in inclusion bodies and purified under denaturing conditions using a Sephacryl size exclusion column followed by cation exchange chromatography. H2A/H2B and H3/H4 histones were subsequently co-refolded to form hetero-dimers and hetero-tetramers, respectively, and purified with size exclusion chromatography.

GST-tagged recombinant human CENP-C central region (a.a. 426–537) (CENP-C^{CR}), CENP-C motif (a.a. 727-767) (CENP-C^{motif}) and short CENP-C central region (a.a. 501-537) (CENP-$C^{CR-short}$) were expressed and affinity-purified on a glutathione column. GST was subsequently cleaved overnight by PreScission protease and separated from CENP-C using cation exchange and size exclusion chromatography.

PCR site-directed mutagenesis was performed to generate CENP-C$^{V532D/V533D}$, CENP-C^{V533D}, CENP-C$^{426-531}$ and CENP-C$^{426-533}$. All mutants were expressed and purified as described for the central CENP-C region.

The H3 histone with a CENP-A N-tail$^{1-49}$ (H3$^{CENP-A\ (N-tail)}$), CENP-A histone with a H3 N-tail$^{1-50}$ (CENP-A$^{H3(N-tail)}$) (figure S1B) and a H2A histone lacking the C-terminal residues 110-130 were cloned using In-Fusion® HD cloning strategy (Takara Bio).

145 bp 601 super-positioning DNA and 147 bp α-satellite DNA were purified as described in (Dyer et al., 2003). Briefly, XL10 cells transformed with pUC57 plasmids containing 6 x 147 bp α-satellite DNA and 8 x 145 bp 601 super-positioning DNA (gift from Ben Black UPenn) were cultivated, DNA was extracted with phenol/chloroform, digested using EcoRV restriction enzyme and further purified using anion-exchange chromatography. 171 bp α-satellite DNA with a CENP-B box was amplified from plasmid using PCR and further purified using anion-exchange chromatography.

601 (145 bp):
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGT
ACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATA
TATACATCGAT

α-satellite no B-box (147 bp):
ATCAAATATCCACCTGCAGATTCTACCAAAAGTGTATTTGGAAACTGCTCCATCAAAAGGCATGTTC
AGCTCTGTGAGTGAAACTCCATCATCACAAAGAATATTCTGAGAATGCTTCCGTTTGCCTTTTATAT
GAACTTCCTCGAT

α-satellite **B-box** (171 bp):
GGAGGA**TTTCGTTGGAAACGGGA**TCAACTTCCCATAACTGAACGGAAGCAAACTCAGAACATTCT
TTGTGATGTTTGTATTCAACTCACAGAGTTGAACCTTCCTTTGATAGTTCAGGTTTGCAACACCCTT
GTAGTAGAATCTGCAAGTGTATATTTTGACCACTTTGGA

Assembly of nucleosomes and nucleosome complexes. CENP-A and H3 nucleosomes were assembled using 601 (145 bp), α-satellite no B-box (147 bp) or α-satellite with B-box (171 bp) DNA. H2A/H2B hetero-dimers, (CENP-A/H4)$_2$ hetero-tetramers and DNA were mixed with a molar ratio of 2:1:1 at high salt concentration (2M NaCl). A gradient dialysis to low salt was performed overnight with a flow rate of 1.5 mL/min using a two-channel peristaltic pump as described in (Luger et al., 1999). Assembled nucleosomes were then uniquely positioned on the DNA by a thermal shift for 2h at 55 °C. CENP-A nucleosome and CENP-C^{CR} were complexed by adding 2.2 moles of CENP-C^{CR} to each mole of CENP-A nucleosome. The complex quality was checked on a 5% native PAGE gel.

Binding experiments: 2,4 μM of CENP-A and H3 nucleosomes assembled on 601 (145 bp) DNA were mixed with different amounts of CENPC^{CR}, CENP-C^{motif} or CENP-C$^{CR\text{-}short}$ and incubated for 1 hour at 4°C. Complexes formation was checked on a 5% native PAGE gel.

Competition experiments. 2,4 μM of CENP-A/CENP-C^{motif} complex was mixed with different amounts of CENP-C^{CR} and the competition between the two domains was checked using 5% native PAGE gel. 2,4 μM of CENP-A/CENP-C^{CR} complex was mixed with different amounts of CENP-C^{motif} and CENP-C$^{CR\text{-}short}$. The competition was then followed using 5% native PAGE gel.

Micrococcal nuclease digestion. 2 μg of nucleosomes were incubated with 2 Kunitz units of Micrococcal nuclease (NEB) in buffer containing 10 mM Tris HCl pH 7.5, 3 mM $CaCl_2$ and 1 mM DTT at room temperature. Reactions were quenched at different time points (2, 5, 7, 10 and 20 minutes) by the addition of 250 μL of PB buffer (Qiagen QIAquick PCR Purification Kit) supplemented with 10 mM of EGTA. DNA from each sample was purified with the QIAquick PCR purification kit, and the extent of DNA digestion was quantified by 2100 Bioanalyzer (Agilent). All experiments were done in triplicates.

CryoEM grid preparation and data collection and processing. CENP-A nucleosome and CENP-A/CENP-C^{CR} complex were prepared as described above. 3 μl of the sample (1-1.2 mg/ml) was applied to freshly glow-discharged Quantifoil R2/1 holey carbon grid. After 3 s blotting time, grids were plunge-frozen in the liquid ethane using FEI Vitrobot automatic plunge freezer. Humidity in the chamber was kept at 95%.

Electron micrographs were recorded on FEI Titan Krios at 300 kV with a Gatan Summit K2 electron detector (~4700 micrographs) (Cryo-EM facility at MPI for Biochemistry Martinsried, Germany). The image pixel size was 0.65 Å per pixel on the object scale. Data were collected in a defocus range of 7 000 Å – 30 000 Å with a total exposure of 100 e/Å^2. 50 frames were collected and aligned with the Unblur software package using a dose filter (Grant and Grigorieff, 2015).

Several thousand particles were manually picked and carefully cleaned in Relion to remove inconsistent particles. The resulting useful particles were then used for semi-automatic and automatic particle picking in Relion. The contrast transfer function parameters were determined using CTFFIND4 (Rohou and Grigorieff, 2015). The 2D class averages were generated with the Relion software package (Scheres, 2012). Inconsistent class averages were

removed from further data analysis. The 3D refinements and classifications were subsequently done in Relion. All final refinements were done in Relion using the auto refine option. The initial reference was filtered to 60 Å in Relion. C1 symmetry was applied during refinements for all classes, except for the highest resolution 2.9 Å class, for which we used C2 symmetry. Particles were split into 2 datasets and refined independently and the resolution was determined using the 0.143 cut-off (Relion auto refine option). Local resolution was determined with Resmap. All maps were filtered to local resolution using Relion with a B-factor determined by Relion.

Model building. The model was built in Coot (Emsley et al., 2010) and refined using Phenix real_space_refine (Adams et al., 2010). Figures are prepared with Chimera (Pettersen et al., 2004).

Data availability. Data will be deposited in the EMDB and PDB data banks upon publication.

References:

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2003). Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. In Methods in Enzymology, (Academic Press), pp. 23–44.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Grant, T., and Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. ELife *4*, e06980.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. In Methods in Enzymology, (Academic Press), pp. 3–19.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

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

Sekulic, N., and Black, B.E. (2016). Preparation of Recombinant Centromeric Nucleosomes and Formation of Complexes with Nonhistone Centromere Proteins. Methods Enzymol. *573*, 67–96.

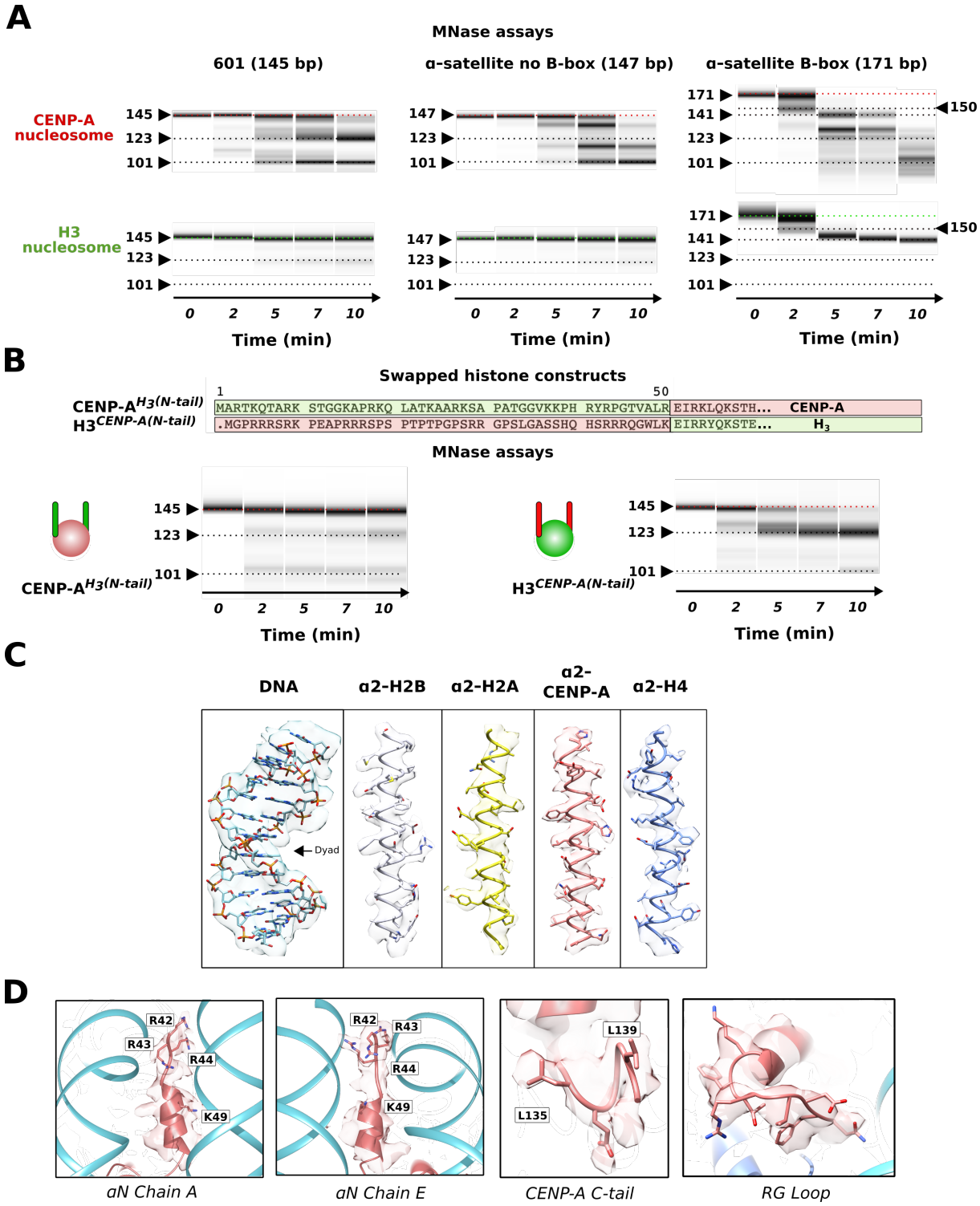


Figure S1. CENP-A nucleosome structural features. A. *Virtual gels from Bioanalyzer showing DNA digestion for CENP-A and H3 nucleosomes on 3 different DNA templates.* ***B.*** *(Top) Sequence overlay of N-terminus of H3 and CENP-A indicating swapped sequences used in CENP-A$^{H3(N\text{-}tail)}$ and H3$^{CENP\text{-}A(N\text{-}tail)}$ constructs. (Bottom) Virtual gels of MNase digestion for CENP-A$^{H3(N\text{-}tail)}$ and H3$^{CENP\text{-}A(N\text{-}tail)}$ nucleosomes assembled on 601 DNA.* ***C.*** *Representative cryoEM map of CENP-A nucleosome illustrating quality of map and model fitting.* ***D.*** *EM density of CENP-A specific features on the nucleosome: αN helix (2 different sides), C-terminal tail and RG-loop.*

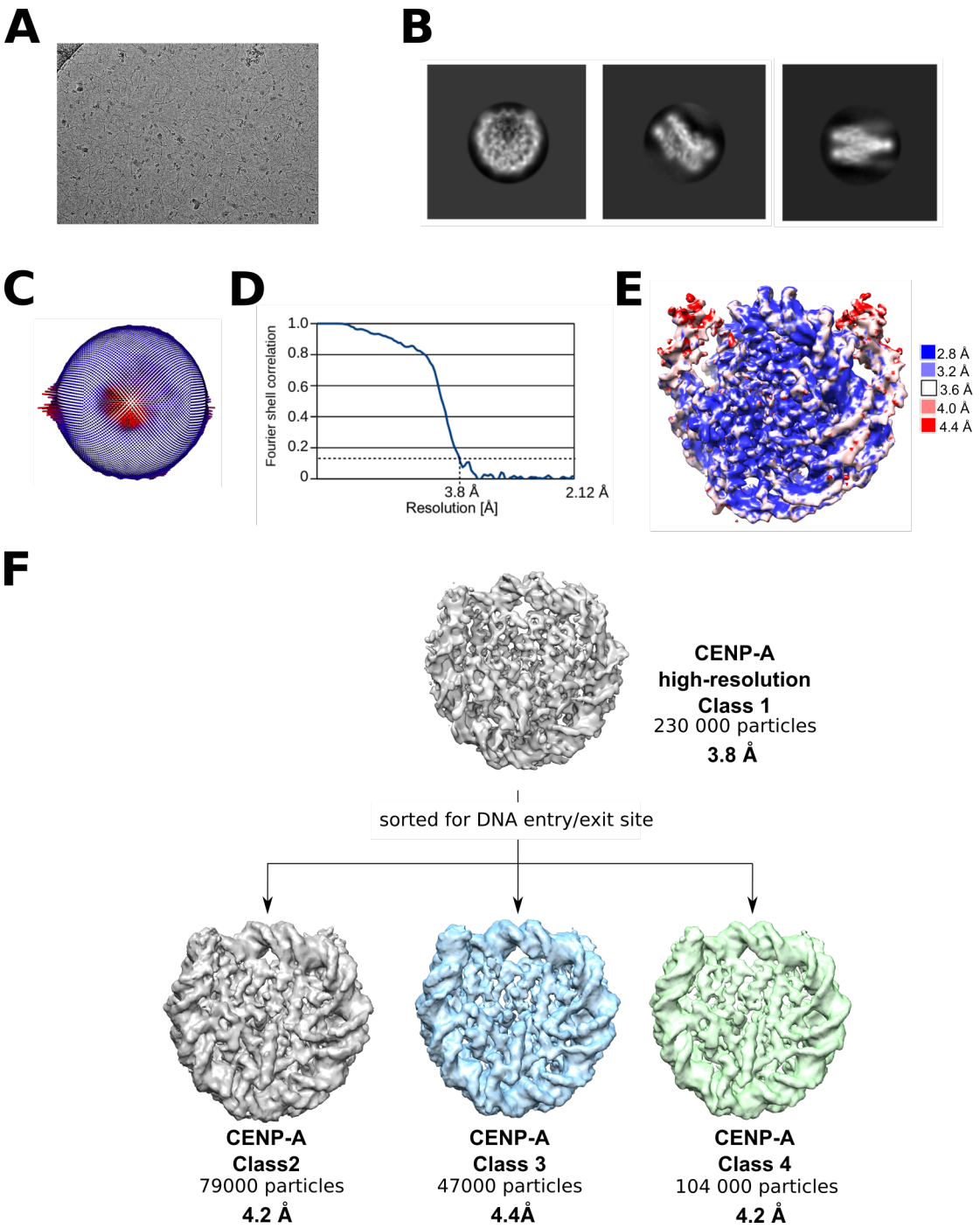


Figure S2. CryoEM analysis of CENP-A nucleosome. A. *Representative cryo-EM raw micrograph.* ***B.*** *Subset of selected 2D class averages.* ***C.*** *Euler angle distribution of particles used in the final 3D reconstruction.* ***D****. Fourier shell correlation (FSC) curves of the final density map (CENP-A high resolution).* ***E.*** *Local resolution of the final 3D density map.* ***F.*** *Particles used for high-resolution CENP-A map were further classified for DNA entry/exit site in order to highlight differences at this part of the nucleosome. Gray map (Class 2) has loosest DNA wrap and green map (Class 4) has tightest DNA wrap. Blue map presents particles that where in-between two extreme conformations.*

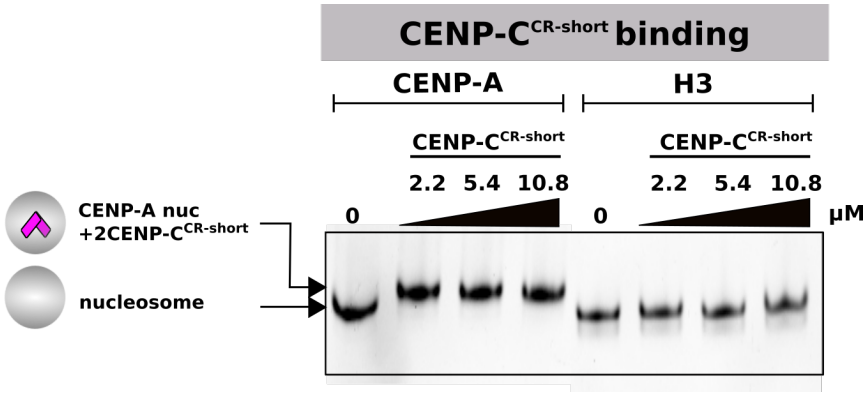


Figure S3. CENP-C$^{CR\text{-}short}$ does not bind H3 nucleosome. *Native PAGE gel stained with EtBr probing binding of CENP-C$^{CR\text{-}short}$ to CENP-A and H3 nucleosome. Lane 1. CENP-A nucleosome. Lanes 3-5. Increasing amounts of CENP-C$^{CR\text{-}short}$ are added to CENP-A nucleosome. Generation of sharp band with slower mobility indicates formation of a specific complex. Lane 5. H3 nucleosome. Lanes 6-8. Increasing amounts of CENP-C^{CR} are added to H3 nucleosome. Band corresponding to H3 nucleosome is not changing mobility, indicating absence of CENP-C$^{CR\text{-}short}$/H3 nucleosome complex formation. Interestingly, absence of residues 426-500 in CENP-C$^{CR\text{-}short}$ vs. CENP-C^{CR} prevents non-specific binding of H3 nucleosome (Figure 2B). 2.4 μM nucleosomes are used in all conditions.*

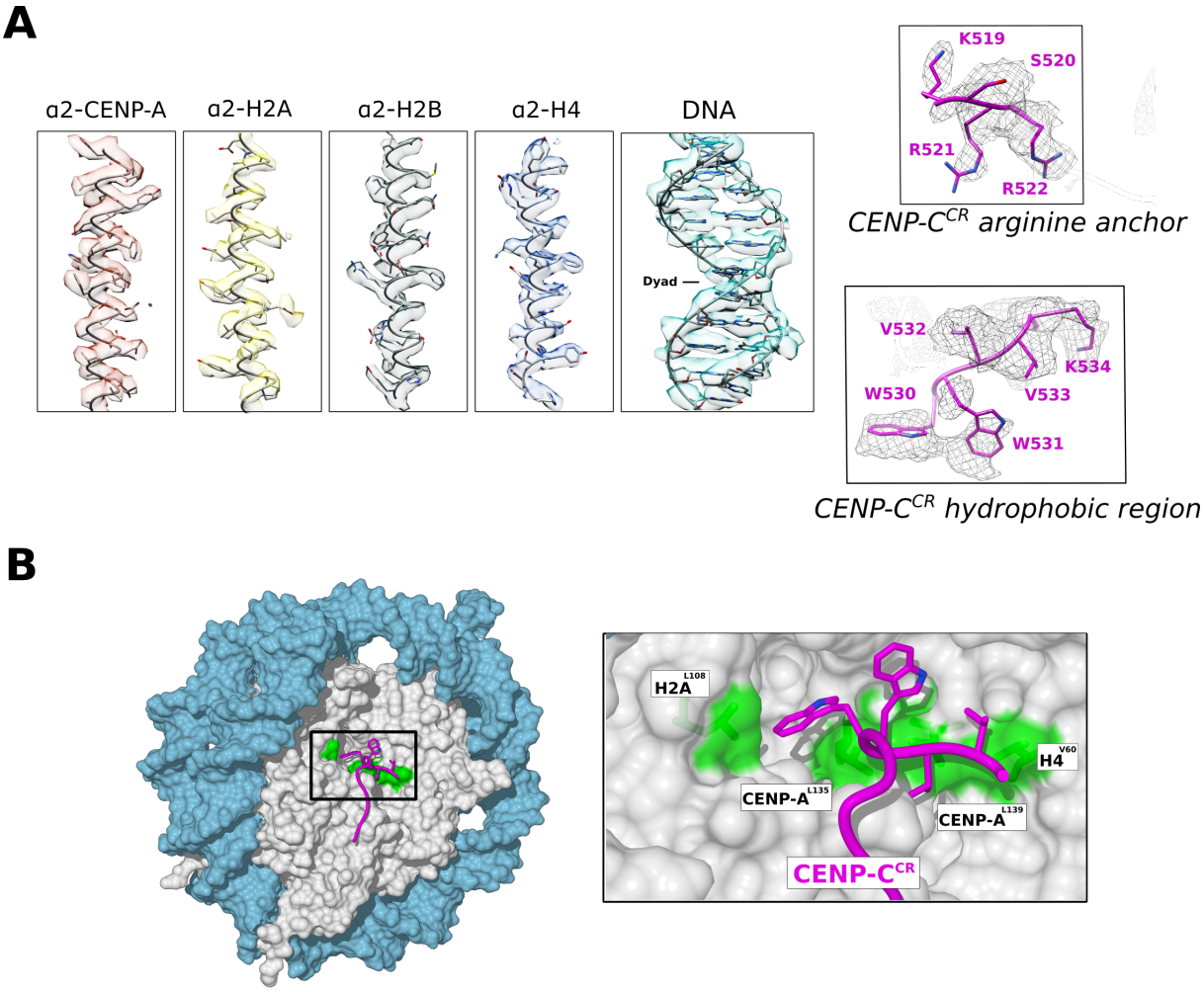


Figure S4. CENP-A nucleosome/CENP-C^{CR} complex structure. A. *Representative cryoEM densities showing fitted model for DNA and each of the histones (left), arginine core and hydrophobic regions of CENP-C^{CR} (right).* ***B.*** *Space filling model of nucleosome (histone core – grey, DNA - blue), showing a hydrophobic groove (green) on the nucleosome formed by H2A^{L108}, CENP-A^{L135}, CENP-A^{L139} and H4^{V60}. CENP-C^{CR} is shown as purple coil with hydrophobic sidechains in stick representation.*

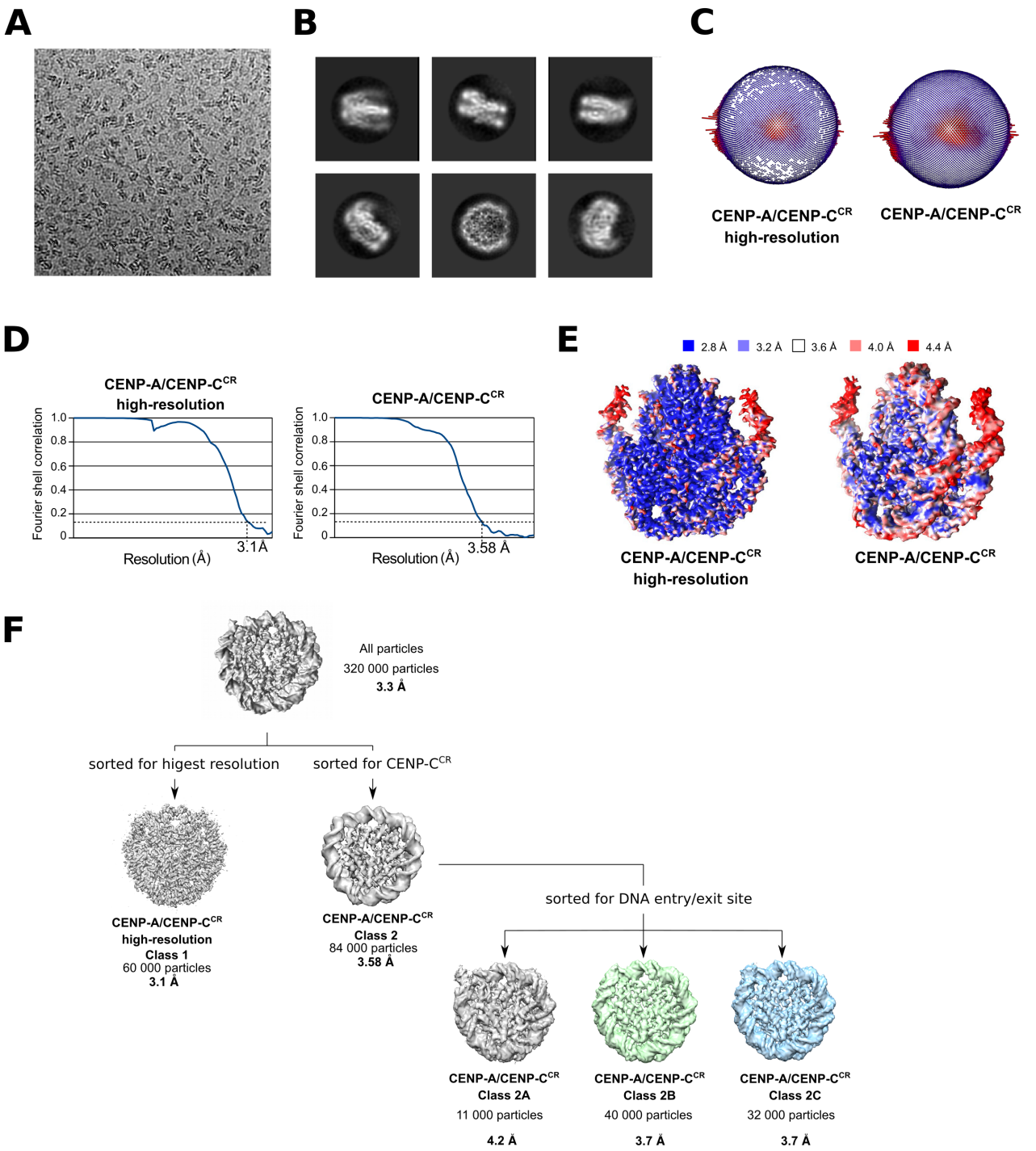
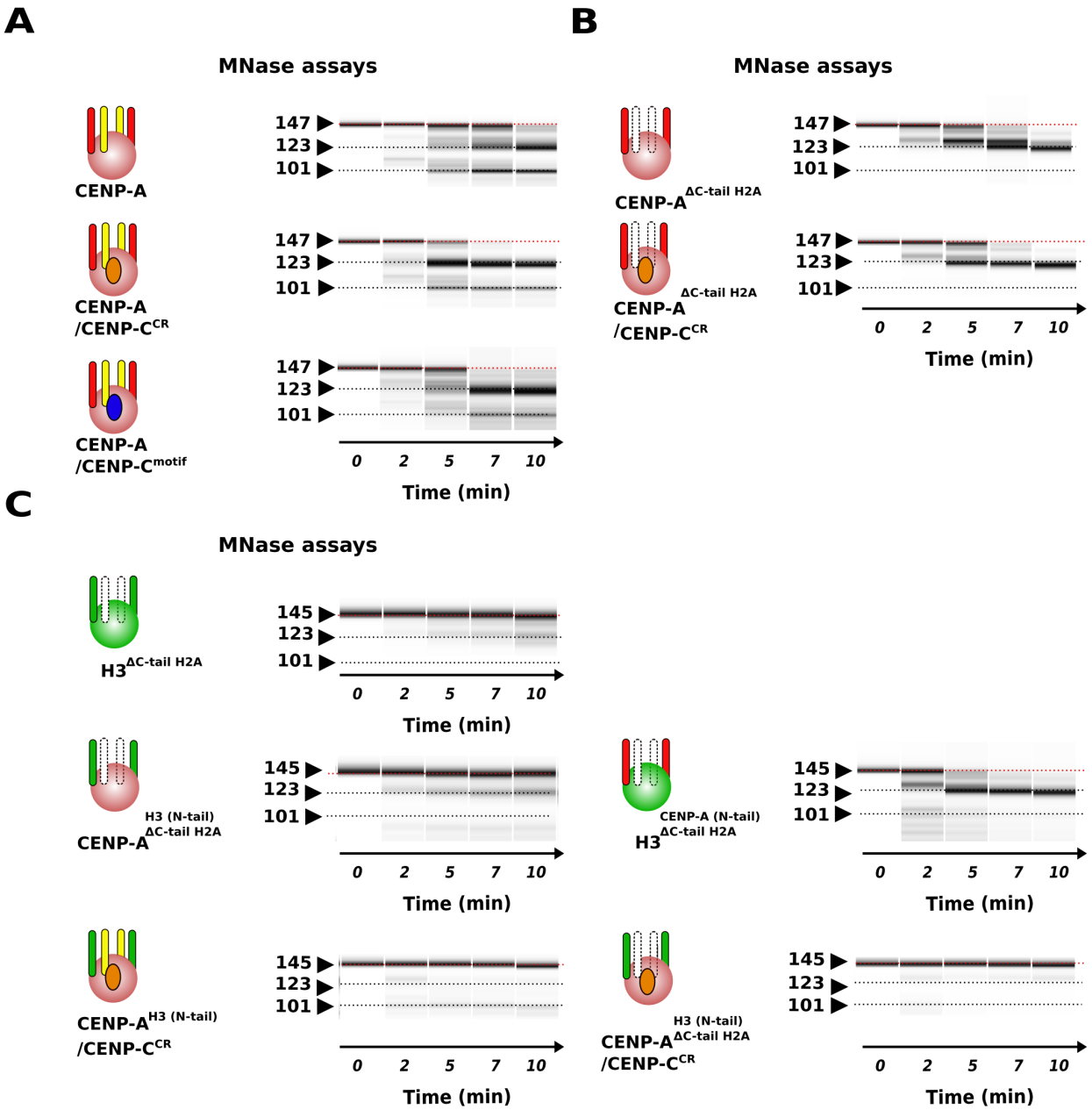


Figure S5. CryoEM analysis of CENP-A/CENP-C^{CR} complex. A. *Representative cryoEM raw micrograph.* ***B.*** *Subset of selected 2D class averages.* ***C.*** *Euler angle distribution of particles used in the final 3D reconstruction for CENP-A/CENP-C^{CR} high-resolution and CENP-A/CENP-C^{CR} complex enriched for CENP-C^{CR}.* ***D.*** *Fourier shell correlation (FSC) curves of the final density map for CENP-A/CENP-C^{CR} high-resolution and CENP-A/CENP-C^{CR} enriched for CENP-C^{CR}.* ***E.*** *Local resolution of the final 3D density maps.* ***F.*** *First particles were sorted for high-resolution and this map was used for initial model building. Map enriched in CENP-C^{CR} was generated to increase map quality around CENP-C^{CR}. Particles used for the later map were further classified for DNA entry/exit site in order to highlight extend of DNA unwrapping. Gray map (Class 2) has loosest DNA wrap and green map (Class 4) has tightest DNA wrap. Blue map presents particles that where in-between two extreme conformations.*


A
MNase assays
CENP-A
CENP-A /CENP-C^CR
CENP-A /CENP-C^motif
147
123
101
0 2 5 7 10
Time (min)
B
MNase assays
CENP-A^ΔC-tail H2A
CENP-A ΔC-tail H2A /CENP-C^CR
147
123
101
0 2 5 7 10
Time (min)
C
MNase assays
H3^ΔC-tail H2A
CENP-A^H3 (N-tail) ΔC-tail H2A
CENP-A^H3 (N-tail) /CENP-C^CR
H3^CENP-A (N-tail) ΔC-tail H2A
CENP-A^H3 (N-tail) ΔC-tail H2A /CENP-C^CR
145
123
101
0 2 5 7 10
Time (min)

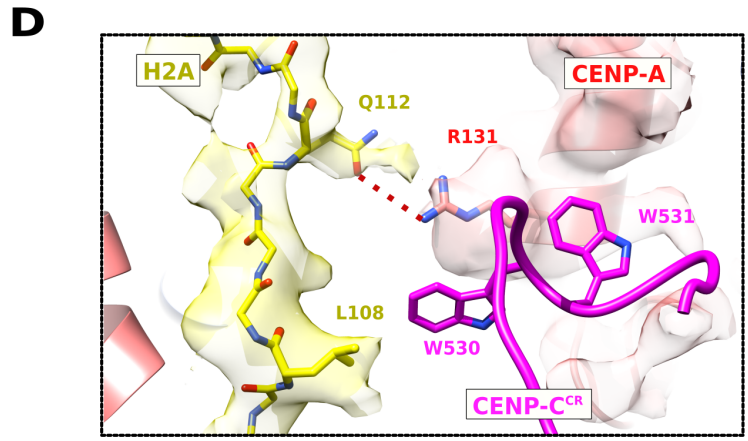

D
H2A
CENP-A
Q112
R131
W531
L108
W530
CENP-C^CR

Figure S6. Conformational changes on CENP-A nucleosome upon CENP-C^{CR} binding. A-C. *Virtual gels for MNase digestion.* ***A.*** *MNase digestion for CENP-A nucleosome alone and in complex with CENP-C^{CR} or CENP-C^{motif} (also in Figure 4A)* ***B.*** *MNase digestion for CENP-A nucleosome assembled with H2A lacking 110-130 residues alone (top) or in complex with CENP-C^{CR} (bottom) showing similar magnitude of digestion.* ***C.*** *(Top) MNase digestion of H3$^{\Delta C\text{-tail H2A}}$ indicating that removal of H2A$^{110\text{-}130}$ does not have an effect on DNA digestion speed in the context of H3 nucleosome. (Middle) MNase digestion of CENP-A$^{H3(N\text{-tail}),\ \Delta C\text{-tail H2A}}$ and H3$^{CENP\text{-}A(N\text{-tail}),\ \Delta C\text{-tail H2A}}$ indicating that removal of H2A$^{110\text{-}130}$ does not have an effect on DNA digestion speed in the context of CENP-A$^{H3(N\text{-tail})}$ nucleosome, but digestion is slightly increased in the context of H3$^{CENP\text{-}A(N\text{-tail})}$. (Bottom) MNase digestion for CENP-A$^{H3(N\text{-tail})}$ is unaffected with CENP-C^{CR} binding independently of the presence of H2A$^{110\text{-}130}$.* ***D.*** *Interactions between H2A C-terminal tail and CENP-C^{CR}.*

Table S1: Cryo-EM data collection, refinement and validation statistics for CENP-A nucleosome

	Class 1 EMD-xxx PDB ID	Class 2 EMD-xxx PDB ID	Class 3 EMD-xxx PDB ID	Class 4 EMD-xxx PDB ID
Data collection and processing				
Magnification				
Voltage (kV)	200	200	200	200
Electron exposure (e–/Å^2)	80	80	80	80
Defocus range (μm)	-0.7 – -3.0	-0.7– -3.0	- 0.7 – -3.0	- 0.7 – -3.0
Pixel size (Å)	1.06	1.06	1.06	1.06
Symmetry imposed	C1	C1	C1	C1
Initial particle images (no.)	~ 230 000	~ 230 000	~ 230 000	~ 230 000
Final particle images (no.)	170 000	79 000	47 000	104 000
Map resolution (Å) FSC threshold	3.8	4.2	4.4	4.2
Map resolution range (Å)	4-6.0	4-6.0	4-6.0	4-6.0
Refinement				
Initial model used	6E0C			
Model resolution (Å) FSC threshold	3.8			
Model resolution range (Å)	235-3.8			
Map sharpening *B* factor (Å^2)	-100			
Model composition				
Nonhydrogen atoms	12010			
Protein residues	765			
Ligands	0			
R.m.s. deviations				
Bond lengths (Å)	0.009			
Bond angles (°)	1.025			
Validation				
MolProbity score.	1.22			
Clashscore	4.4			
Poor rotamers (%)	0			
Ramachandran plot				
Favored (%)	98.66			
Allowed (%)	1.34			
Disallowed (%)	0			

Table S2: Cryo-EM data collection, refinement and validation statistics for CENP-A nucleosome/ CENP-C

	Class 1 EMD-xxx PDB ID	Class 2 EMD-xxx PDB ID	Class 2A EMD-xxx PDB ID	Class 2B EMD-xxx PDB ID	Class 2C EMD-xxx
Data collection and processing					
Magnification					
Voltage (kV)	300	300	300	300	300
Electron exposure (e–/Å^2)	100	100	100	100	100
Defocus range (μm)	-0.7 – -3.0	-0.7– -3.0	- 0.7 – -3.0	-0.7 – -3.0	-0.7 – -3.0
Pixel size (Å)	0.65 (1.3)	0.65 (1.3)	0.65 (1.3)	0.65 (1.3)	0.65 (1.3)
Symmetry imposed	C1	C1	C1	C1	C1
Initial particle images (no.)	~ 320 000	~ 320 000	~ 84 000	~ 84 000	~ 84 000
Final particle images (no.)	60 000	84 000	11 000	40 000	32 000
Map resolution (Å) FSC threshold	2.9	3.5	4.2	3.7	3.7
Map resolution range (Å)	2.6-5.0	3.3 – 5.0	4.0 – 5.0	3.5-5.0	3.5-5.0
Refinement					
Initial model used	6BUZ	Class 1	Class 1	Class 1	
Model resolution (Å) FSC threshold	2.9	3.5	4.2	3.7	
Model resolution range (Å)	235-2.9	235-2.9	235-3.6	235-3.8	
Map sharpening *B* factor (Å^2)	-100	-100	-100	-100	
Model composition					
Nonhydrogen atoms	11697	12091	11791	11744	
Protein residues	728	771	736	741	
Ligands	0	0	0	0	
R.m.s. deviations					
Bond lengths (Å)	0.01	0.008	0.008	0.007	
Bond angles (°)	0.953	0.957	0.889	0.871	
Validation					
MolProbity score.	1.11	1.12	1.3	1.47	
Clashscore	3.19	3.3	5.24	5.44	
Poor rotamers (%)	0	0	0	0	
Ramachandran plot					
Favored (%)	98.88	98.8	97.92	96.97	
Allowed (%)	1.12	1.2	2.08	3.03	
Disallowed (%)	0	0	0	0	