De novo formation and epigenetic maintenance of centromere chromatin

Citation for published version:

Digital Object Identifier (DOI):
10.1016/j.ceb.2018.12.004

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Current opinion in cell biology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
De novo formation and epigenetic maintenance of centromere chromatin

Junichirou Ohzeki\textsuperscript{1}, Vladimir Larionov\textsuperscript{2}, William C Earnshaw\textsuperscript{3} and Hiroshi Masumoto\textsuperscript{1}

Accurate chromosome segregation is essential for cell proliferation. The centromere is a specialized chromosomal locus, on which the kinetochore structure is formed. The centromere/kinetochore is required for the equal separation of sister chromatids to daughter cells. Here, we review recent findings on centromere-specific chromatin, including its constitutive protein components, its de novo formation and maintenance mechanisms, and our progress in analyses with synthetic human artificial chromosomes (HACs).

Constitutive components and mechanism of maintaining centromere chromatin

In all normal human chromosomes, centromere proteins assemble on a typical centromere-specific sequence of tandem 171-bp repeats, termed alpha-satellite (alphoid) DNA, where they form centromere-specific chromatin (Figure 1a). The first centromere proteins discovered were CENP-A, CENP-B, and CENP-C [4,5].

CENP-A, a histone H3 variant, is a key epigenetic centromere marker. It forms a centromere-specific nucleosome with the canonical histones, H2A, H2B, and H4. A network of constitutive centromere-associated network proteins (termed the CCAN, ICEN, or CENP-A NAC/CAD proteins) is associated with the CENP-A nucleosome [6\textsuperscript{7},7\textsuperscript{8},8\textsuperscript{8},9\textsuperscript{9},10,11,12\textsuperscript{12}] (Figure 1a). These CCAN proteins are localized on centromere chromatin throughout the cell cycle. Among the CCAN proteins, CENP-C and CENP-N bind directly to CENP-A nucleosomes and stabilize their three-dimensional structure [6\textsuperscript{7},7\textsuperscript{8},8\textsuperscript{8},9\textsuperscript{9},10,11,12\textsuperscript{12}] (Figure 1a). In addition, compared to the canonical histone H3 nucleosome, the CENP-A nucleosome has flexibility in DNA binding, in the vicinity of the linker region [13,14]. When amino acid sequences in the αN region of the CENP-A protein, which is involved in the flexibility describing above, were replaced with the corresponding sequences in canonical histone H3, histone H1 became bound and CENP-C assembly with the centromere was lost [15].

CENP-B is the only protein that directly binds to a specific sequence in alphoid DNA—the 17 bp DNA motif known as the CENP-B box. The CENP-B box motif appears most commonly once every 340 bp in alphoid DNA dimers in centromeric alpha satellite higher order repeat domains [5,16]. The CENP-B protein has both a DNA binding domain and a CENP-A interaction domain at the N terminus [17,18]. CENP-B forms homodimers at the C-terminus and interacts with CENP-G via an acidic

**Introduction**

A number of specific proteins assemble on the centromere to form a specialized chromatin structure and maintain its functions. During the mitotic phase, kinetochore components assemble on the centromere chromatin. This centromere chromatin/kinetochore structure is essential for the equal segregation of sister chromatids to each daughter cell by generating accurate attachments to the mitotic spindle that produce mobile forces and determining the separation timing by monitoring the quality of interactions with microtubules. Dysfunctions in centromere chromatin lead to chromosomal instability. Lagging and unsegmented chromosomes form micronuclei, where DNA fragmentation may occur. Re-incorporation of damaged DNA from micronuclei into the nucleus leads to chromosomal DNA rearrangements during subsequent cell cycles [1,2\textsuperscript{2},3]. Chromosomal instability is considered to be one of the major causes of genomic instability and cancer progression. Therefore, mechanisms for assembling, maintaining, and regulating centromere chromatin are crucially important in assuring proper chromosome segregation and avoiding chromosomal/genomic instability.
Proteins involved in centromere/kinetochore organization and maintenance.

(a) Centromere DNA and the centromere/kinetochore protein and interaction network. (Left panel) Box shows the key for the component symbols; (middle panel) centromere structure; (right panel) model of centromere/kinetochore interactions. CENP-A and CENP-B assemble on alphoid DNA with H3 nucleosomes, and CCAN proteins assemble on CENP-A nucleosomes. Kinetochore proteins assemble in the mitotic phase to regulate the equal segregation of sister chromosomes through interactions with microtubules. (b) CENP-A replenishing proteins and a model of balancing centromere chromatin and heterochromatin. The M18BP1 complex assembles on alphoid DNA through interactions with CENP-C and CENP-I, and recruits proteins, such as the HJURP-CENP-A-H4 complex and KAT7. RSF1 assembles on alphoid DNA through acetylation and promotes the histone exchange reaction [80]. In this histone exchange, old heterochromatinized nucleosomes are removed, and heterochromatin invasion into the centromere chromatin is prevented. This opens a space for the deposition of a new CENP-A. (c) Model of SUV39H1 overexpression in KAT7-knockout cells. When the mechanism for preventing heterochromatin invasion is lost, due to the lack of KAT7, SUV39H1 overexpression causes CENP-A reduction and the loss of centromere function. (Bottom right) Box shows the key for the component symbols.
region near the C-terminus [19, 20]. CENP-B is required for efficient de novo CENP-A chromatin assembly, when naked alphoid DNA is introduced into cells [21, 22], but it is dispensable in the Y chromosome centromere and in neocentromeres [23], and mice can survive without it [24–26]. CENP-B also enhances heterochromatin formation at ectopic integration sites of alphoid DNA in chromosomal arms that do not assemble other centromere components [22]. Consequently, it is extremely interesting how this functional centromere chromatin switch is regulated.

CENP-C is the key platform protein necessary both for CENP-A chromatin assembly and kinetochore formation. CENP-C interacts with CENP-A, CENP-B, CCAN proteins (CENP-I, CENP-L), a kinetochore protein, MIS12, and a CENP-A-replenishment factor, M18BP1 [27–32]. Two CENP-C assembly pathways on alphoid DNA have been identified [33]. In one, CENP-C directly binds to the CENP-A C-terminus region in CENP-A nucleosomes. In the other pathway, CENP-C is assembled through an interaction with CENP-B. When only one of these pathways is blocked, centromere function is preserved, but when both pathways are blocked, centromere function is lost [33].

Currently, more than one hundred centromere/kinetochore protein components have been revealed. Many are highly conserved in various species, including human and yeast [34, 35*].

In addition to the kinetochore proteins, CENP-A replenishing factors also assemble on the centromere chromatin. Although deposition timing of canonical histone H3.1 is coupled with DNA replication, other H3 variants (i.e. H3.3 and CENP-A) are independent of DNA replication. In human cells, newly synthesized CENP-A deposition (replenishment) occurs in G1 phase [36]. The MIS18 complex (M18BP1, MIS18α, and MIS18β), which is involved in CENP-A replenishment, assembles at the centromere through an interaction with CENP-C from telophase to G1 phase (Figure 1b). The CENP-A chaperone, HJURP, is recruited through an interaction with MIS18α and MIS18β [37–40, 41**]. M18BP1, MIS18α, and MIS18β form a complex with a 2:4:2 stoichiometry, respectively. The timing of MIS18 complex assembly at the centromere is regulated through modulation by protein kinases CDK and PLK1 [42**, 43**, 44–47]. In addition, DNA binding ability of HJURP is regulated by phosphorylation; this DNA binding activity and homodimer formation were reported to be important for CENP-A deposition [41**, 48, 49].

Each new CENP-A molecule that is recruited to the centromere by HJURP is then stabilized as a newly assembled CENP-A nucleosome by chromatin remodeling factors, such as RSF1 [50]. RNA also plays an important role in centromere assembly that is still being actively studied. In human and fly cells, a cis-acting centromere transcript was also shown to stabilize CENP-A nucleosomes [51**, 52**]. In contrast, a trans-acting centromere RNA, ectopically transcribed from an introduced DNA vector, was reported to destabilize chromosome segregation [53]. CENP-C was reported to bind to RNA [54, 55]. The Drosophila melanogaster HJURP homolog, CAL1, interacts with a chromatin remodeling factor complex, FACT, which promotes transcription [56]. In human cells, FACT was reported to interact with the CENP-T/W complex [57]. The newly assembled CENP-A nucleosomes might be stabilized by chromatin remodeling factors, CCAN protein assembly, and/or possibly by centromere DNA transcripts.

**Alphoid DNA is competent in both de novo centromere assembly and HAC formation**

Two types of alphoid DNA arrays have been identified at many human centromeres. The first is based on the higher-order repeat (HOR) units of alphoid DNA. The HOR unit is repeated homogeneously in tandem, to sizes on the order of a mega-base, with more than 97% identity (though the constituent monomers are much more variable). The second type of array is composed of divergent irregular 171-bp monomers. By comparing these two types of alphoid repeats in X chromosome centromeres, evolutionarily ancient transposon sequences were frequently found in the monomeric array, but not in the HOR. Thus, the HOR is considered to be evolutionarily newer than the monomeric alphoid arrays [58]. In addition, all the HORs of human chromosomes contain a CENP-B box, except the alphoid HOR on the Y chromosome.

When a HOR with several ten-kb sized components was introduced into human HT1080 cells, de novo centromere assembly occurred, and a human artificial chromosome (HAC) was formed via amplification of the introduced DNA (Figure 2a) [59, 60]. Strikingly, no de novo centromere assembly or HAC formation occurred with alphoid DNA that lacked a CENP-B box (i.e. the Y alphoid HOR, the monomeric alphoid DNA of chromosome 21, and a synthetic alphoid DNA with a mutation in all CENP-B boxes). Thus, CENP-B protein and the CENP-B box are indispensable for efficient de novo centromere assembly and HAC formation [21, 22, 59–62]. It appears that the relevant function of CENP-B is to assemble CENP-A de novo on input alphoid DNAs. Newly assembled CENP-A is structurally stabilized by an interaction with the CENP-B N-terminal domain (Figure 2b) [17]. The assembly of CCAN proteins and/or alphoid RNA might
also contribute to the stabilization of the newly assembled CENP-A nucleosome.

Compared to the evolutionarily conserved centromere proteins, centromere DNAs evolved rapidly and diverged among species. Some strains of mice have chromosomes with expanded centromere repeats (minor satellite DNA) that contain CENP-B boxes, but no other sequences in common with primate alpha satellite DNA. These arrays produce stronger centromeres that assemble more CENP-A than observed in other mouse strains. Chromosomes with stronger centromeres were reported to be more likely to distribute to the egg-proximal side during female meiosis by the inter-strain crosses, and thus, they have a better chance of succeeding in the egg (that is, they will pass to the next generation, rather than being eliminated in the polar body) [63**]. Acquisition of a CENP-B box on centromeric satellite DNA could confer an advantage to chromosome survival in female meiosis, as described in the centromere-drive/meiotic-drive theory [64]. This hypothesis could possibly explain why only Y centromere DNAs have no CENP-B
box, both in human and mouse genomes, because the Y chromosome is not subject to female meiosis. In recent years, centromere DNAs with and without CENP-B boxes have been found in new-world monkey species [65,66]. It is an interesting puzzle how the CENP-B box was acquired through evolution.

It should be noted that, although de novo centromere assembly could occur when a CENP-B box was combined with alpheid DNA, no centromere assembly occurred when a GC-rich repeat derived from a pBR plasmid fragment was combined with a CENP-B box [21]. Thus, in addition to the CENP-B box, alpheid DNA might have other advantageous properties for de novo centromere assembly. These could include AT-rich sequences and/or the ability to form non-B-form DNA upon CENP-B binding [67**].

The discovery that HACs could be generated from synthetic alpheid DNA has advanced the functional analysis of centromeres. Moreover, HACs have been applied as chromosomal vectors [68]. For example, synthetic HACs are useful for manipulating the epigenetic landscape of chromatin. These HACs contain synthetic alpheid DNA monomers with a tetracycline operator (tetO) sequence alternating with monomers bearing CENP-B boxes (tetO-alpheid DNA). The synthetic arrays can be targeted in vivo with tetracycline repressor (tetR) proteins fused to any protein or protein domain of interest. Binding to the tetO site, thus, tethers the chimeric protein to the tetO-alpheid DNA within the centromere. This approach enables removal or addition of histone modifications as well as the recruitment of other proteins that interact with the chimeric protein (Figure 3) [69–71]. The functional consequences are readily assayed by scoring for HAC loss and levels of critical proteins such as CENP-A and CENP-C. HACs are also useful as markers of chromosomal instability and as vectors for gene therapy [72,73].

**Chromatin assembly balance determines centromere assembly on alpheid DNA**

In addition to CENP-A chromatin, heterochromatin is also assembled on alpheid DNA, when it carries histone H3 methylated on the K9 residue (H3K9me3). The H3K9me3 modification serves as a binding site for HP1 assembly. HP1 is reported to recruit cohesin and Aurora B in the chromosomal passenger complex (CPC). Cohesin maintains sister chromosome cohesion and the CPC improves the accuracy of chromosome segregation [74,75]. Heterochromatin can also spread to adjacent chromatin regions [76,77]. When heterochromatin invades the centromere, CENP-A chromatin and kinetochore functions are no longer maintained, and eventually, they are inactivated [69,78]. Interestingly, levels of H3K9me3 modification on alpheid DNA differ, depending on cell type; for example, the activity is high in HeLa cells and low in HT1080 cells. In HeLa cells, high H3K9me3 modification activity on alpheid DNA inhibits de novo centromere assembly on introduced alpheid DNA. Conversely, in HT1080 cells, low H3K9me3 modification activity permits de novo centromere assembly on introduced alpheid DNA [79]. Therefore, for de novo assembly of CENP-A chromatin, the epigenetic chromatin modification status against heterochromatinization is important. Established centromere chromatin might possess some type of barrier mechanism to protect against heterochromatin invasion.

Indeed, heterochromatinization was blocked by acetylation of tetO-alpheid chromatin by a tethered HAT (histone acetyltransferase) in the synthetic HAC system [79]. Recent studies have identified endogenous acetylating enzymes involved in the CENP-A chromatin assembly process. KAT7 is an acetylating enzyme active in the G1 phase; it is recruited onto alpheid DNA through an interaction with M188P1 [80]. Centromeres in KAT7-knockout cells were sensitive to overexpression of the H3K9 tri-methylating enzyme, SUV39H1, which forms pericentromere heterochromatin. Specifically, the overexpression of SUV39H1 in KAT7-knockout cells resulted in high frequencies of chromosome loss and micronucleus formation (Figure 1c). Histone H3 acetylation, induced by the tethering of KAT7 onto tetO-alpheid DNA, recruits a remodeling factor, RSF1, and enhances histone turnover. This reduces pre-existing H3K9me3 nucleosomes and promotes assembly of ectopically expressed histone H3 variants (i.e. H3.3 and CENP-A). Thus, KAT7 can prevent heterochromatin from spreading into the centromere by promoting a histone exchange reaction through RSF1 (Figure 1b). This histone exchange reaction may also be involved in the CENP-A deposition pathway. KAT7 also induces H3K14 acetylation (H3K14ac) [81,82]. H3K14ac was sufficient for evicting histone through a histone chaperone, NAP1, in vitro [83]. Thus, KAT7 induction of H3K14ac might also be involved in the histone exchange process on alpheid DNA. HAT1 is another acetyltransferase involved in CENP-A assembly. Chicken HAT1 induced the acetylation of histones H4K5 and H4K12. This recruited a nucleosomal complex composed of HJUFP, CENP-A, and histone H4 to the centromere [84]. *Drosophila melanogaster* HAT1 is also involved in the new CENP-A assembly pathway [85].

It is becoming clear that transcriptionally active or neutral modifications, such as H3K9ac, H3K4me2, H3K36me2, and H4K20me1 promote centromere assembly [80,86–90]. In addition, the modification of CENP-A through α-amino trimethylation is also involved in centromere assembly [91,92]. Transcriptional activity also affects centromere maintenance. Strong transcription causes loss of centromere activity, but moderate levels of transcription are compatible with centromere activity [93]. According to the evidence that
Schematic diagrams show evaluations of CENP-A assembly activity in the tethering assay. 

(a) (top) Structure of a synthetic tetO-alphoid DNA dimer-repeat unit. One monomer of the dimer has a CENP-B box; the other monomer has the tetO operator sequence, which binds the tet repressor protein (tetR). Dimers are ligated in tandem to form a 60 kb plasmid, and (middle) introduced into cells. The cells generate either (bottom) a tetO-HAC or a tetO site ectopically integrated into the host chromosome. 

(b) A schematic diagram shows three assays where the protein of interest is fused to the tetR sequence; upon tetR binding, the fused protein is tethered to the tetO-alphoid site (top) in the HAC or (middle and bottom) to the ectopic integration site in the chromosome. Assays were
cis-transcripts stabilize the CENP-A nucleosome, it is reasonable to postulate that an unknown mechanism might exist for initiating transcription within centromere DNA. Further studies are necessary to elucidate a mechanism that can explain how histone modifications, chromatin proteins, and transcription events are involved in centromere assembly and maintenance.

**Chromatin factors involved in CENP-A assembly and maintenance**

Two HeLa cell lines were generated to evaluate how chromatin factors and histone modifiers affected CENP-A assembly on alphoid DNA. One cell line contained the HAC, and the other contained tetO-alphoid DNA integrated into a host chromosome (Figure 3a) [70]. In the first assay (Figure 3b), the tetR was fused with several chromatin proteins of interest. Upon tetR binding to the tetO site on the HAC (tet-O-HAC) [66], each protein of interest was tethered to that site. Then, the effect of each protein on levels of endogenous CENP-A at the tet-O-HAC centromere was quantified. In other assays, tetR-fused proteins were tethered to the ectopic integration site of the tetO-alphoid DNA. Each construct was then assessed for whether it induced the assembly of either de novo endogenous CENP-A or exogenously over expressed CENP-A (Figure 3b). The results revealed that CENP-C, CENP-I, and HJURP contributed to the assembly of endogenous CENP-A on the ectopic tetO-alphoid DNA site. In contrast, other MIS18 complex proteins (M18BP1, MIS18a, MIS18b), which assemble to the centromere downstream of CENP-C, were not sufficient to assemble endogenous CENP-A on the ectopic tetO-alphoid integration site (Figure 3c). These results suggest that the presence of CENP-C and CENP-I might regulate the activity of MIS18 or the HJURP complex in de novo endogenous CENP-A deposition, even on a non-centromeric tetO-alphoid site. Alternatively, HJURPimerization and binding through the tetR might mimic or bypass the regulation by CENP-C and CENP-I. Taken together, these findings indicate that CENP-C and CENP-I might promote HJURPimerization and binding to DNA. However, it remains unclear whether CENP-C and CENP-I directly regulate this process, or whether some other proteins (or modifications) mediate the process.

In these experiments, most tethered proteins that promote open chromatin, such as acetyltransferase subunits or the RSF1 remodeling factor, could promote the assembly of exogenously overexpressed CENP-A or a placeholder histone H3.3 on the ectopic tetO-alphoid DNA integration site (Figure 3b). A potential mechanism could be that these chromatin modifiers and the remodeling factor promote histone exchange or turnover on the tetO-alphoid DNA site. This would create openings for the deposition of a new histone H3.3 or CENP-A. Interestingly, some of these factors (M18BP1, SSRP1, HAT1) selectively assembled CENP-A rather than histone H3.1 or H3.3 at the ectopic tetO-alphoid DNA integration site [70]. Indeed, in chicken and Drosophila melanogaster cells, FACT (SSRP, SPT16) and HAT1 are normally involved in CENP-A assembly [56,84,85]. These findings suggested that open chromatin modifications and/or remodeling mechanisms are very likely involved in the CENP-A deposition process.

In the same experiments, tethered heterochromatin factors caused reductions in CENP-A levels on HAC centromeres (Figure 3b) [70]. When HPI was tethered to the tet-O-HAC, a hierarchical centromere protein reduction occurred on the HAC centromere, in the following order: CENP-H, CENP-C, then CENP-A [94]. In addition, at chicken centromeres, CENP-T and CENP-H levels fell when H4K20me1 was removed [87]. It is interesting that heterochromatin spreading destabilized CCAN protein assembly, which consequently, destabilized CENP-A.

**Closing remarks**

By now, the list of structural components that assemble on centromere chromatin is largely complete. In recent years, the functions of these core components have been studied by a wide range of methods, including in vitro reconstitution, crystal structures, and other analyses. Parallel studies of natural and synthetic human artificial chromosomes (HAGs) have focused on characterizing the chromatin environment that is essential for kinetochore assembly. In particular, the tethering of various chromatin modifiers to the HAC and ectopic integration sites has been applied to investigate de novo centromere assembly and maintenance in vivo. This approach has led to new insights that have suggested that centromere chromatin assembly and maintenance are controlled positively and negatively on repetitive DNA through chromatin modifications and remodeling activities. Thus, the tethering approach is a powerful method for evaluating how chromatin modifications and remodeling can affect
centromere organization and establishment. Together these studies are beginning to reveal how the centromere is organized dynamically on repetitive DNA.

Conflict of interest statement
Nothing declared.

Acknowledgements
This work was supported by MEXT KAKENHI [grant numbers 16H04747, 16H04141, and 18H04271] and the Kazusa DNA Research Institute Foundation (to H. Masumoto). Work in the W.C. Earnshaw laboratory was funded by the Wellcome Trust, of which W.C. Earnshaw is a Principal Research Fellow [grant number 073915]. Work in the V. Larionov laboratory was supported by an Intramural Research Program of the National Institutes of Health, the National Cancer Institute, and the Center for Cancer Research.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


of M18BP1 on a Mis18 hexamer is necessary for CENP-A loading. *eLife* 2017, 6:e20352. These reference provides composition of human M18BP1 complex contains hetero hexamer of MIS18α and MIS18β.


Showed an experimental evidence how centromeric DNA has been selected for evolution.


Showed DNA sequence features favorable for centromere formation of non-B-form DNA which is enriched at centromeres, and suggests CENP-B contribution in the induced non-B form DNA formation.


