Structure of the CENP-A nucleosome and its implications for centromeric chromatin architecture

Hiroaki Tachiwana and Hitoshi Kurumizaka*

Laboratory of Structural Biology, Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan

(Received 28 November 2011, accepted 28 December 2011)

Centromeres are dictated by the epigenetic inheritance of the centromere-specific histone H3 variant, CENP-A. The structure of the CENP-A nucleosome has been considered to be the fundamental architecture of the centromeric chromatin. Controversy exists in the literature regarding the CENP-A nucleosome structures, with octasome, hemisome, compact octasome, hexasome, and tetrasome models being reported. Some of these CENP-A nucleosome models may correspond to transient intermediates for the assembly of the mature CENP-A nucleosome; however, their significances are still unclear. Therefore, the structure of the mature CENP-A nucleosome has been eagerly awaited. We reconstituted the human CENP-A nucleosome with its cognate centromeric DNA fragment, and determined its crystal structure. In this review, we describe the structure and the physical properties of the CENP-A nucleosome, and discuss their implications for centromeric chromatin architecture.

Key words: CENP-A, centromere, chromatin, histone, kinetochore, nucleosome

INTRODUCTION

The centromere is the chromosomal location for the assembly of the kinetochore, which provides sites for microtubule attachment in chromosome segregation during cell division (For reviews see Cheeseman and Desai, 2008; Allshire and Karpen, 2008; Santaguida and Musacchio, 2009; Talbert and Henikoff, 2010; Verdaasdonk and Bloom, 2011; Perpelescu and Fukagawa, 2011). The centromere is epigenetically inherited after cell division, and is marked by the centromere-specific protein, CENP-A. CENP-A has been identified as a centromere-specific histone H3 variant (Palmer et al., 1987, 1991; Sullivan et al., 1994), and it is referred to as CENP-A in vertebrates, Cid in flies, and Cse4 in budding yeast. Eight human H3 variants (H3.1, H3.2, H3.3, H3T (H3.4), H3.5, H3.X, H3.Y, and CENP-A) have been identified, and the most distant H3 variant is CENP-A, which shares about 50% amino acid identity with the canonical H3.1, H3.2, or H3.3 (For a review see Hamiche and Shuaib, 2011).

The removal of CENP-A causes chromosome missegregation (Stoler et al., 1995; Buchwitz et al., 1999; Takahashi et al., 2000; Blower and Karpen, 2001; Goshima et al., 2003), indicating that CENP-A functions to establish the active centromere and kinetochore. Actually, the depletion of CENP-A causes the significant dispersion of the fundamental components of the centromere, such as CENP-B and CENP-C, in mouse cells (Howman et al., 2000). These findings suggested that CENP-A is a key architectural component of the centromere-specific chromatin.

The centromeric nucleosome is considered to have a distinct structure, underlying the centromere-specific function, such as kinetochore assembly. Therefore, several centromeric nucleosome models with CENP-A have been proposed. We successfully determined the crystal structure of the human centromeric nucleosome containing histones H2A, H2B, H4, and CENP-A (Tachiwana et al., 2011a). In this review, we describe the structural features of the CENP-A nucleosome, and discuss its functional significance in establishing centromeric chromatin.

The CENP-A nucleosome models

Biochemical studies indicated that CENP-A directly binds to DNA, and forms the centromeric nucleosome with the other histones, H2A, H2B, and H4 (CENP-A nucleosome) (Palmer et al., 1987, 1991; Sullivan et al., 1994; Shelby et al., 1997; Vafa and Sullivan, 1997; Yoda et al., 2000; Tanaka et al., 2004, 2005). To ensure the specific function of centromeres, the CENP-A nucleosome may be structurally distinct from the canonical H3 nucleosome (Luger et al., 1997). Therefore, the CENP-A nucleosome may adopt a very different conformation, as compared to that of the canonical H3 nucleosome (Black et al., 2004, 2007). In this context, several models for the CENP-A nucleosome
have been proposed (For a review see Black and Cleveland, 2011).

1) Octasome model. Based on biochemical evidence obtained with in vitro reconstituted nucleosomes containing CENP-A, the octasome model has been proposed. In this model, two CENP-A molecules form a histone octamer with two each of histones H2A, H2B, and H4, and the histone octamer wraps the DNA in a left-handed orientation, as in the canonical H3 nucleosomes (Shelby et al., 1997; Yoda et al., 2000; Tanaka et al., 2005; Camahort et al., 2009). The crystal structure of the human CENP-A nucleosome is mostly consistent with the octasome model (Fig. 1) (Tachiwana et al., 2011a). Therefore, the octasome is a strong candidate for the mature CENP-A nucleosome structure.

2) Hemisome model. Based on atomic force microscopy (AFM) observations of Drosophila Cid-containing nucleosomes (Dalal et al., 2007), a centromeric nucleosome that is half the height of the canonical H3 nucleosomes has been proposed. The hemisome model is further supported by a biochemical study, in which the histones within the Drosophila Cid nucleosomes failed to crosslink into an octameric form, unlike the histones within the canonical H3 nucleosomes (Dalal et al., 2007). An in vitro reconstitution assay with Drosophila Cid and an in vivo supercoiling assay with Saccharomyces cerevisiae Cse4 revealed that the nucleosomes containing Cid or Cse4 induced positive supercoiling of the DNA (Furuyama and Henikoff, 2009). These results suggested that the centromeric nucleosomes containing Cid or Cse4 may right-handedly wrap DNA. However, in vitro reconstitution experiments with Cse4 supported the left-handed wrapping of DNA (Camahort et al., 2009; Kingston et al., 2011; Dechassa et al., 2011).

3) Compact octasome model. Based on the crystal structure of the DNA-free CENP-A-H4 tetramer, the CENP-A nucleosome with a compact histone octamer has been proposed. In this compact octasome model, the interface between two CENP-A molecules is rotated, as compared to the canonical H3-H3 interface, and the overall structure of the CENP-A nucleosome thus becomes more compact than that of the canonical H3 nucleosome (Sekulic et al., 2010).

4) Hexasome model. Based on chromatin immunoprecipitation (ChIP) experiments, it has been reported that the amounts of histones H2A and H2B are markedly diminished in the centromeric DNA from budding yeast. Biochemical experiments suggested that Scm3, a Cse4 chaperone, may be incorporated into the Cse4-containing nucleosomes, instead of the H2A-H2B dimers (Mizuguchi et al., 2007). The Scm3-Cse4-H4 nucleosome may specifically form on the centromere-specific DNA sequences, but not on other chromosome regions (Xiao et al., 2011).

5) Tetrasome model. Based on the ChIP experiments, H2B, H2A, and Htz1 (an ortholog of the vertebrate H2AZ variant) reportedly interact only weakly with centromeric DNA sequences in budding yeast (Mizuguchi et al., 2007). Similarly, in fission yeast, H2B interacts weakly with centromere DNA sequences (Williams et al., 2009). These observations suggested that the CENP-A nucleosome may be deficient in H2A and H2B.

Some of these CENP-A nucleosome models may represent transient, intermediate forms for mature CENP-A nucleosome assembly during the cell cycle. The mature CENP-A nucleosome must be disrupted by the replication machinery in S phase. In fact, an in vivo fluorescence pulse-chase study revealed that the CENP-A molecules bound to centromeres are reasssembled onto each daughter DNA strand after replication (Jansen et al., 2007). However, the assembly of newly synthesized CENP-A does not occur in S phase. The CENP-A is incorporated into centromeres during early G1 phase (Jansen et al., 2007; Schuh et al., 2007), and the mature CENP-A nucleosome may thus be established during G1 phase. Therefore, the various putative intermediates of the CENP-A nucleosomes may exist until the mature CENP-A nucleosome is established. The CENP-A nucleosome models may be reconciled, if the hemisome, compact octasome, hexasome, and tetrasome are representatives of each step in the CENP-A nucleosome maturation pathway (Fig. 2).

Preparation of the CENP-A nucleosome crystal

To obtain structural information about the mature CENP-A nucleosome, we reconstituted the CENP-A nucleosome with recombinant human histones H2A, H2B, H4, and CENP-A. A gel filtration analysis revealed that the purified histones H2A, H2B, H4, and CENP-A formed a complex containing each of these histones in 1:1:1:1 stoichiometry, in the absence of DNA (Tachiwana et al., unpublished). The H2A-H2B-H4-CENP-A complex efficiently formed nucleosomes by the salt dialysis method. The human α-satellite sequence (sat4; Tanaka et al., 2004) was selected as a template DNA for the CENP-A nucleosome reconstitution. This α-satellite sequence positions the CENP-A nucleosome, in which the 17 base pair DNA sequence (CENP-B box) that interacts with the centromere-specific DNA-binding protein, CENP-B, is located at the edge of the nucleosomal DNA (Tanaka et al., 2005).

For crystallization of the CENP-A nucleosome, we first designed the 71-mer sat4 fragment, containing one-half of the sat4 α-satellite sequence with the CENP-B box at the
This DNA fragment contained an extra four base overhang, 5’ TGCA 3’, at the distal edge from the CENP-B box. The 146 base pair DNA fragment for the CENP-A nucleosome assembly was then prepared by the self-ligation of this 71-mer with the 5’ TGCA 3’ overhang. However, the CENP-A nucleosome reconstituted with the 146 base pair DNA fragment only generated microcrystals. Therefore, we next prepared a 147 base pair DNA fragment, in which the 71-mer sat4 fragment containing an extra 5-base overhang, 5’ GTAAC 3’, at the distal edge from the CENP-B box was self-ligated. In this 147 base pair α-satellite DNA derivative, the CENP-B box sequence was located at the edges, and an A:A mismatch was introduced at the center of the 147 base pair DNA fragment. The crystal structure of the CENP-A nucleosome was then determined with this 147 base pair DNA (Fig. 1).

**The CENP-A nucleosome is an octasome** In the crystal structure, two CENP-A molecules form an octamer comprising two each of histones H2A, H2B, and H4, and the DNA is wrapped around the CENP-A octamer in the left-handed orientation (Fig. 1), as in the canonical human H3 nucleosomes (Tsunaka et al., 2005; Tachiwana et al., 2010, 2011b; Iwasaki et al., 2011). This is consistent with the octasome model for the CENP-A nucleosome. Biochemical studies also revealed the left-handed DNA wrapping of the CENP-A nucleosome (Tachiwana et al., 2011a). Human CENP-A introduced left-handed (negative) supercoils into the DNA, when the CENP-A nucleosomes were formed on closed circular plasmid DNA by a salt-dialysis method (Conde e Silva et al., 2007; Sekulic et al., 2010; Tachiwana et al., 2011a).

The same results were obtained, when the CENP-A nucleosomes were assembled in the presence of the human histone chaperones, NAP1 and somatic nuclear autoantigenic sperm protein (sNASP), under physiological salt conditions (Tachiwana et al., 2011a). Therefore, the major form of the CENP-A nucleosome has the left-handed orientation of the DNA wrapped around the histone octamer, containing two each of histones H2A, H2B, H4, and CENP-A.

**The CENP-A nucleosome is not compacted** A previous structural study with the DNA-free CENP-A-H4 tetramer suggested that the CENP-A nucleosome may have a compact structure, as compared to that of the canonical H3 nucleosomes (Sekulic et al., 2010). In fact, the interface between the two CENP-A molecules was
rotated, as compared to the H3-H3 interface in nucleosomes, and therefore, the CENP-A-H4 tetramer formed a more compact structure in the DNA-free form (Sekulic et al., 2010). However, this compacted CENP-A-H4 tetramer structure is not observed in the CENP-A nucleosome. In the nucleosome, the CENP-A-H4 tetramer structure is quite similar to the H3-H4 tetramer structure (Tachiwana et al., 2011a). A small angle X-ray scattering (SAXS) analysis also revealed that the CENP-A nucleosome is not compacted in solution (Tachiwana et al., 2011a). Consistent SAXS results were reported with budding yeast Cse4-containing nucleosomes (Dechassa et al., 2011). Therefore, the CENP-A nucleosome structure may be conserved in the budding yeast centromeric nucleosome containing Cse4. The compact CENP-A-H4 tetramer may be a preassembled form, or the compact CENP-A octasome may exist as a transient intermediate in the formation of the mature CENP-A nucleosome (Fig. 2) (Panchenko et al., 2011).

**Flexible nature of the DNA within the CENP-A nucleosome** The crystal structure of the CENP-A nucleosome revealed an overall similarity to that of the canonical H3 nucleosome. However, the DNA structure was significantly different between the CENP-A and H3 nucleosomes. In the H3 nucleosomes, 145–147 base pairs of DNA were visible in the crystal structures (Fig. 1) (Luger et al., 1997; Davey et al., 2002; Richmond and Davey, 2003; Makde et al., 2010). In contrast, in the CENP-A nucleosome, only 121 base pairs were visible, and 13 base pairs of both DNA edges were completely disordered (Tachiwana et al., 2011a). The disordered DNA regions in the CENP-A nucleosome contain the CENP-B box, where the centromeric DNA binding protein, CENP-B, specifically binds (Earnshaw et al., 1987; Masumoto et al., 1989, 1993; Yoda et al., 1992; Tanaka et al., 2001). Biochemical studies suggested that the DNA segments at both edges of the CENP-A nucleosome were more susceptible to exonuclease degradation than those of the H3 nucleosome (Conde e Silva et al., 2007; Sekulic et al., 2010; Tachiwana et al., 2011a; Panchenko et al., 2011). Interestingly, the nucleosomes containing budding yeast Cse4 also exhibited similar susceptibility of the DNA edges to an exonuclease attack (Kingston et al., 2011; Dechassa et al., 2011).

The flexible nature of the DNA edges is also supported by SAXS analyses. In this method, the apparent diameter of the nucleosome can be estimated by the $D_{\text{max}}$ value of the SAXS data. Interestingly, the $D_{\text{max}}$ value of the CENP-A nucleosome (165 Å) was obviously larger than that of the H3.1 nucleosome (147 Å) (Tachiwana et al., 2011a). Since the size of the histone octamer containing CENP-A in the crystal structure was quite similar to that of the canonical histone octamers containing H3, the larger $D_{\text{max}}$ value of the CENP-A nucleosome may be caused by the flexible DNA segments. To confirm this possibility, we constructed a CENP-A nucleosome model containing DNA segments that were detached from the histone surface at both edges of the nucleosome (Tachiwana et al., 2011a). For this purpose, the CENP-B box DNA structure complexed with the CENP-B DNA-binding domain was used (Tanaka et al., 2001). This CENP-A nucleosome model generated SAXS parameters that were quite consistent with the actual SAXS data obtained with the CENP-A nucleosome (Tachiwana et al., 2011a). These results strongly suggested that the DNA edges are flexible in the CENP-A nucleosome.

The flexible DNA edges were also reported in the budding yeast Cse4 nucleosomes. Consistent with the human CENP-A nucleosome, the SAXS data revealed that the $D_{\text{max}}$ value of the Cse4 nucleosome was apparently larger than that of the H3 nucleosome, because of the flexible nature of the DNA segments at the edges of the Cse4 nucleosome (Dechassa et al., 2011). Therefore, the flexibility of the DNA in the CENP-A nucleosome may be a common property of the centromeric chromatin in eukaryotes. Black and colleagues showed that the poly-nucleosome containing CENP-A adopted a more compact, higher order folded structure than that of the H3 poly-nucleosome, although the DNA at the edges of the CENP-A mononucleosome was flexible in the poly-nucleosome (Panchenko et al., 2011). These flexible DNA regions may be important to form the centromeric chromatin architecture, and/or may function to serve as binding sites for the centromeric DNA-binding proteins, such as CENP-B (Earnshaw et al., 1987; Masumoto et al., 1989; Yoda et al., 1992), CENP-C (Saitoh et al., 1992; Sugimoto et al., 1994; Yang et al., 1996), and CENP-N (Carroll et al., 2009).

**Possible binding sites for transacting factors on the nucleosomal CENP-A** In the CENP-A nucleosome structure, three regions, the N-terminal region (residues 1–45), the loop 1 region (residues 78–84), and the C-terminal region (residues 135–140), of CENP-A are exposed to the solvent, and are accessible (Fig. 3A) (Tachiwana et al., 2012). The CENP-A N-terminal region is disordered, and may flutter around the CENP-A nucleosome core (Tachiwana et al., 2011a). The CENP-A loop 1 region is two amino acids longer than the H3 loop 1 region, and protrudes from the disk-like surface of the CENP-A nucleosome (Sekulic et al., 2010; Tachiwana et al., 2011a). The C-terminal six residues are also disordered in the CENP-A nucleosome (Tachiwana et al., 2011a). These regions may be potential binding sites for the CENP-A binding proteins in the nucleosome.

The CENP-A N-terminal region is not required for centromere targeting (Shelby et al., 1997; Chen et al., 2000; Vermaak et al., 2002), but it is essential for cell survival (Morey et al., 2004; Ravi and Chan, 2010). In the plant
Arabidopsis thaliana, a chimeric CENP-A, in which the N-terminal tail of CENP-A is replaced by the corresponding region of the canonical histone H3, rescues mitosis, but renders the plants sterile (Ravi and Chan, 2010). Therefore, the CENP-A N-terminal region may have distinct functions in mitosis and meiosis. It is interesting to postulate the existence of factors that specifically interact with the CENP-A N-terminal region in mitosis and meiosis. Structural and physicochemical studies revealed that this region may be responsible for the DNA flexibility at the edges of the CENP-A nucleosome (Tachiwana et al., 2011a; Panchenko et al., 2011).

The CENP-A loop 1 region is involved in the CENP-A targeting domain (CATD), which was mapped as the CENP-A region required for its centromere targeting (Black et al., 2004, 2007). The functional importance of loop 1 was suggested by the swapping experiment with the fly CENP-A ortholog, Cid (Vermaak et al., 2002). The CATD domain contains the binding site for the CENP-A-specific chaperone, HJURP (Foltz et al., 2009; Dunleavy et al., 2009; Shuaib et al., 2010; Bernad et al., 2011; Barnhart et al., 2011; Moree et al., 2011). In both budding yeast and fission yeast, Scm3 was identified as a functional homolog of human HJURP, which promotes Cse4 assembly at centromeres (Stoler et al., 2007; Camahort et al., 2007; Mizuguchi et al., 2007; Williams et al., 2009; Fidoux et al., 2009; Sanchez-Pulido et al., 2009). Structural studies revealed that part of the CATD region functions as the binding site for human HJURP and yeast Scm3, but both chaperones may not directly interact with the loop 1 region of CENP-A and Cse4 (Fig. 3B) (Zhou et al., 2011; Hu et al., 2011; Cho and Harrison, 2011). Consistently, point and deletion mutants of the human CENP-A loop 1 were proficient in centromere targeting (Tachiwana et al., 2011a). Structural studies also revealed that HJURP and Scm3 bind to the CENP-A (or Cse4)-H4 dimer, and their binding may be incompatible with the CENP-A-H4 tetramer formation and the DNA binding of CENP-A-H4 (Fig. 3B) (Zhou et al., 2011; Hu et al., 2011; Cho and Harrison, 2011). These facts indicated that, although HJURP and Scm3 may function to recruit CENP-A in centromeres, they may not bind to the nucleosomal CENP-A and Cse4, which directly bind to DNA as the CENP-A-H4 tetramer and Cse4-H4 tetramer forms. Interestingly, the CENP-A loop 1 mutants were not stably retained in centromeric chromatin, as compared to wild type CENP-A, although their centromere targeting activities were proficient (Tachiwana et al., 2011a). This may be explained by postulating the existence of a loop 1 binding factor that is required for the stable maintenance of the mature CENP-A nucleosome at centromeres.

The N-terminal, loop 1, and C-terminal regions of CENP-A could contain binding sites for factors, such as RSF (Perpelescu et al., 2009), MgcRacGAP (Lagana et al., 2010), CENP-C (Guse et al., 2011), and CENP-N (Carroll et al., 2009), which may be required for the maturation and/or maintenance of the CENP-A nucleosome. CENP-C and CENP-N are members of the constitutive centromere-associated network (CCAN) proteins, which are required for the formation of the mitotic kinetochore (Minoshima et al., 2005; Obuse et al., 2004; Izuta et al., 2006; Okada et al., 2006; Foltz et al., 2006; McClelland et al., 2007; Cheeseman et al., 2008; Hori et al., 2008; Amano et al., 2009). The interactions of these essential factors with the CENP-A nucleosome may be the key steps for the assembly of centromeric chromatin, which is required for mitotic kinetochore formation.

Fig. 3. Locations of the N-terminal region (residues 1–45), the loop 1 region (residues 78–84), and the C-terminal region (residues 135–140) of CENP-A. The N-terminal, loop 1, and C-terminal regions are shown in the CENP-A nucleosome (A) and in the HJURP-CENP-A-H4 complex structure (PDB ID: 3R45) (B). The disordered CENP-A N-terminal region of nucleosome (A) is represented by a dashed black line, and the truncated N-terminal region of CENP-A in the HJURP-CENP-A-H4 complex (B) is represented by a dashed black line with a cylinder (αN helix).
It has been reported that CENP-C, an essential component of the functional centromere, directly binds to the C-terminal six amino acid residues of CENP-A (Guse et al., 2011). This region is completely disordered in the CENP-A nucleosome, and is expected to be located at the accessible disk-like surface of the CENP-A nucleosome (Tachiwana et al., 2011a). Therefore, CENP-C may bind to the C-terminal region of CENP-A within the nucleosome. CENP-C also possesses DNA binding activity (Sugimoto et al., 1994; Yang et al., 1996). Further studies will be required to understand the functional significance of the CENP-A and DNA binding activities of CENP-C within the CENP-A nucleosome.

Concluding remarks The crystal structure of the CENP-A nucleosome has been solved. The structure revealed that the CENP-A nucleosome adopts an octasome-like structure, containing two each of histones H2A, H2B, H4, and CENP-A, with the DNA left-handedly wrapped around the octamer. The global structure of the octamer containing CENP-A is similar to that of the histone octamer containing canonical H3, but the DNA structure in the CENP-A nucleosome is very different from that in the H3 nucleosome. In addition, the CENP-A N-terminal, loop 1, and C-terminal regions, which are specific for the CENP-A nucleosome, may be important for forming and maintaining the proper centromeric chromatin architecture. The CENP-A nucleosome structure presented here may be the mature CENP-A nucleosome, and it provides a basic framework for centromeric chromatin. However, various possible transient intermediates, such as the tetrasome, hexasome, hexasome, and compact octasome, are also important targets for structural biology to understand how the CENP-A nucleosome is assembled during G1 phase.

This work was supported in part by Grants-in-Aid from the Japanese Society for the Promotion of Science (JSPS), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. H. K. was also supported by the Waseda Research Institute for Science and Engineering.

REFERENCES


Dunleavy, E. M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-
The CENP-A nucleosome architecture


Obuse, C., Yang, H., Nozaki, N., Goto, S., Okazaki, T., and Yoda, K. (2004) Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. Genes Cells 9, 105–120.


