Structure Basis of CENP-A Nucleosome Recognition of Chicken KNL2

KNL2 による CENP-A ヌクレオソーム認識の分子構造基盤

Honghui Jiang

Graduate School of Frontier Biosciences

Osaka University

Submitted on 30th August 2023
Graduated in September 2023

Abstract

Centromere protein-A (CENP-A), a histone H3 variant, is an important epigenetic mark for centromere specification. For new CENP-A deposition, chicken KINETOCHORE NULL2 (ggKNL2), a component of the licensing factor Mis18 complex, binds directly to CENP-A-containing nucleosome via CENP-C-like motif to recruit the CENP-A chaperone Holliday junction recognition protein (HJURP) to the centromere chromatin and ultimately forming a new CENP-A nucleosome. However, the molecular basis of how the CENP-A nucleosome is recognized by ggKNL2 remains unclear. To address this question, I performed biochemical and structural biological analyses on the CENP-A nucleosome in complex with ggKNL2. To understand the centromere binding mode of KNL2 to the CENP-A nucleosome, the cryo-EM structure of the chicken CENP-A nucleosome in complex with a ggKNL2 fragment containing the CENP-C-like motif was determined. This structure suggests that ggKNL2 distinguishes between CENP-A and histone H3 in the nucleosome using the CENP-C-like motif and its downstream region. Both the C-terminal tail and the RG-loop of the CENP-A nucleosome are simultaneously recognized as CENP-A characteristics by ggKNL2. Based on the structural findings, the mutational analysis was performed using chicken DT40 cells and demonstrated that the interaction of KNL2 with the CENP-A nucleosome is critical for new CENP-A incorporation and cell viability. The CENP-A nucleosome-ggKNL2 interaction is thus essential for KNL2 functions. It is notable that KNL2 shares the CENP-A binding mode with CENP-C, which is a kinetochore component. Furthermore, the structural, biochemical, and cell biology data implied the cell cycle dependent regulation of centromere localization of ggKNL2. ggKNL2 directly binds to the CENP-A nucleosome in interphase, while in mitosis ggKNL2 localizes to centromere depend on CENP-C.

Keywords: CENP-A; CENP-C-like motif; Cryo-EM; KNL2.

Contents

Abstract 2
Chapter 1: General Introduction 4
1.1 Cell cycle 4
1.2 Chromosome & nucleosome 7
1.3 Centromere & CENP-A nucleosome 11
1.3.1 Centromere 12
1.3.2 CENP-A nucleosome 14
1.4 Kinetochore 17
1.4.1 Kinetochore assembles on the centromere 20
1.4.2 CENP-C interact with CENP-A 22
1.5 Mis18 complex & KNL2 24
1.6 CENP-A deposition 26
Chapter 2: KNL2 directly binds to the CENP-A nucleosome via the conserved CENP-C-like motif region 29
2.1 Introduction 29
2.2 Results and discussion 30
2.2.1 KNL2 CC containing fragments stably bind to the CENP-A nucleosome in vitro 32
2.2.2 KNL2 SANTA only fragment shows non-specific binding to the CA nucleosome 35
2.2.3 KNL2$^{456-516}$ fragment does not bind to the CENP-A nucleosome 36
2.2.4 Discussion 36
2.3 Materials and methods 37
Chapter 3: Cryo-EM structural analysis reveals the specific binding mechanism of the CENP-A nucleosome to the CENP-C-like motif of ggKNL2 47
3.1 Results and discussion 47
3.1.1 The conserved arginine residue R527 of KNL2$^{517-560}$ N-terminal region binds to the H2A/H2B acidic patch 53
3.1.2 The conserved ggKNL2 phenylalanine residue F535 and tryptophan residue W536, locating in a 3_{10} helix, recognize the C-terminal region of CENP-A 55
3.1.3 A ggKNL2 C-terminal region Aspartic acid residue D544 directly binds to the CENP-A specific RG-loop 58
3.1.4 Cryo-EM Structural comparison of the CENP-A nucleosome binding between KNL2 and CENP-C 61
3.1.5 Discussion 64
3.2 Materials and methods 66
Chapter 4: A centromeric ggKNL2 localization mechanism in interphase is different from that in mitotic cells 69
4.1 Results and discussion 69
4.2 Materials and methods 75
References 80
Acknowledgement 97
Achievement 99

Chapter 1: General Introduction

1.1 Cell cycle

Millions of eukaryotic species live on Earth[1]. Without cell division, these organisms cannot grow and survive. Unlike binary fission for prokaryotes, eukaryotic cells depend mitosis for cell proliferation. In the 19th century scientists observed the mitotic process in eukaryotes and the term mitosis was first proposed by Walther Flemming[2,3]. Since then, scientists have begun a long search for the principles and mechanisms of the whole cell cycle of mitotic cells.

Precise and orderly mitosis is essential for eukaryotic cell proliferation. Normally, through mitosis, eukaryotic cells equally divide chromosomes replicated exponentially in the nucleus into two daughter cells. This process is accompanied by cytoplasmic division, which eventually produces two daughter cells with the same genetic material as the parent cell and with intact cell composition and structure.

A complete cell cycle is the process that a cell goes through from the completion of one division to the end of the next division. Basically, one cell cycle is roughly divided into interphase and mitotic (M) phase.

Interphase is divided into three main phases: G1, S, G2. Cells in G1 phase have just divided from the parent cell and are smaller in size. During this period, cells need to synthesize some proteins and RNA to promote cell growth and, at the same time, prepare for the next phase. The next is S phase. Chromatin is structural loose to provide enough space for RNA and DNA polymerase, so the chromatin DNA is replicated and linked to each other as sister chromatids through centromeres. Also, during replication, the two sister chromatids are brought close to each other by the action of cohesins[4]. After DNA replication, the cells perform centriole replication and continue to grow in G2 phase, providing adequate preparation for the M phase that follows. In total, interphase is transition for the M phase to prepare sufficient cellular constitutions for daughter cells.

M phase includes mitosis and cytokinesis. under the cell cycle control, many of the M

phase events are regulated with the existence of mitotic kinases. In terms of mitosis, it contains four phases: prophase, metaphase, anaphase and telophase. At the beginning, the cell should be ready for accurate segregation. Nuclear envelope is broke down, nucleolus is disappeared and the chromatin is spirally condensed into chromosomes which defines the beginning of prometaphase. Besides, two centrosomes are separated and then send out spindle fiber to connect to the centromere of chromosome via kinetochore. With the force from the mitotic spindle, chromosomes are aligned at cell equator, which is called 'Metaphase plate' - a central plane of the cell, in metaphase. Normally, the chromosomes oscillate along the spindle axis instead of standing still in plate. The end of metaphase occurs when the sister chromatids separate. This is an irreversible process which also means that mitosis has officially entered the anaphase. During this process, cohesins are disconnected by separase, two chromatids are separated from the centromere, and the spindle fiber pulls the two chromatids in opposite directions toward the opposite poles[5]. At this point, the number of chromosomes doubles. In anaphase, usually, the two chromosome sets keep approaching towards their respective cell poles while the pole to pole distance get increasing[6]. The activity of many kinases that control the cell cycle is altered during chromosome segregation, such as the gradual loss of CDK1 (Cyclin dependent kinase) activity, in preparation for returning to interphase. In contrast to prophase, during telophase the spindle fibers gradually disappear, nuclear envelop and nucleoli re-form, and the chromosomes gradually decondense into chromatin and are confined to the nucleus[7]. In parallel with telophase, the cytoplasm is divided equally into two daughter cells, called cytokinesis.

The above is an overview of the entire mitotic cell cycle and is somewhat general in nature (Fig 1)[6]. But whether interphase or M-phase, the mode and the specific molecular mechanisms of division throughout the cell cycle are often different in different species. The information of diversity about cell division can help to better understand key regulations of the whole cell cycle and also the evolution information of eukaryotic cell mitosis. The related research is certainly full of significance. While exploring the mysteries of nature, it brings more information and inspiration to the treatment of diseases related to cell division, such as chromosomal abnormalities and even cancer.

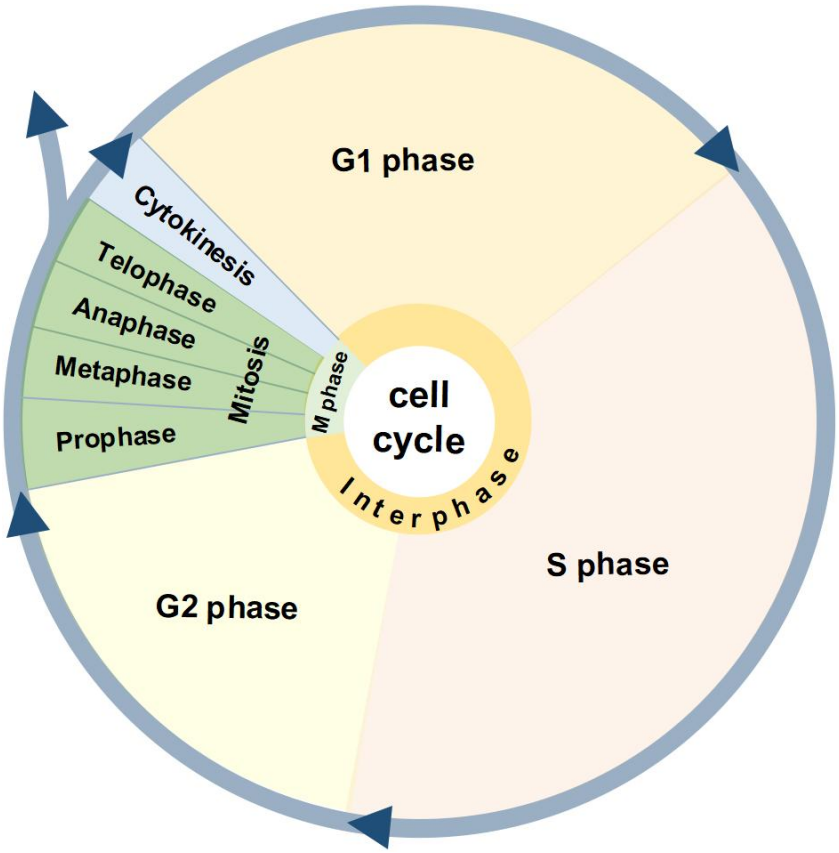


Figure 1. The cellular growth and division cycle.

A Cartoon of the main segments of the cell cycle. During interphase (G1,S,G2), the cell accomplishes sufficient biosynthesis to become two. In mitosis (M), cell parts are reorganized so the mitotic spindle can achieve the equipartition of the chromosomes and centrosomes, leaving the distribution of more numerous components, such as ribosomes, to the laws relating to large numbers and the process of cytokinesis, in which the cell itself divides into two daughters. In tissues, cells can continue further rounds of division or can exit the cell cycle.

Figure adapted from McIntosh. 2016[6].

1.2 Chromosome & nucleosome

The term ‘chromatin’ was first proposed by W. Flemming and ‘chromosome’ was create by Heinrich Wilhelm Waldeyer in the late 19th century to describe its tendency to be dyed dark by particular dyes[8,9]. In the early 20th century, Morgan and his collaborators first introduced the important concept of chromosomes contain genes[10]. Chromosomes composed mainly of deoxyribonucleic acid (DNA) and histones, and are highly condensed from filamentous chromatin (Fig 2)[11]. In fact, chromatin and chromosome, the main carriers of genetic information, are the same substance with the same composition, but with different forms during interphase and metaphase.

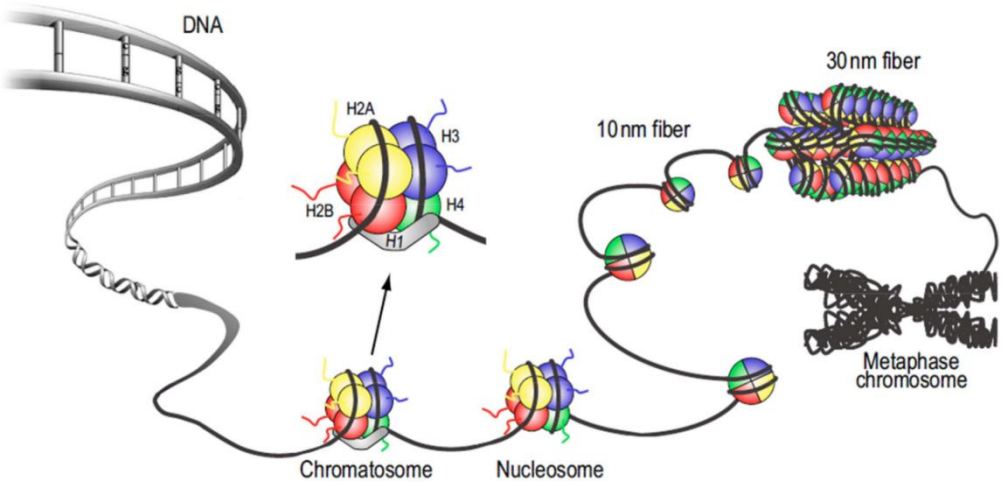


Figure 2. Overview of Chromatin: composition and levels of organization.

The fundamental unit of chromatin is the nucleosome, consisting of DNA, wrapped around an octamer of two copies of each of the four core histones H2A (yellow), H2B (red), H3 (blue) and H4 (green). For further compaction, one molecule of the linker histone H1 (gray) can be bound, resulting in a chromatosome. Nucleosomes with linker DNA constitutes the 10 nm fibers, which can be further compacted by stacking of nucleosomes, leading to the 30 nm fiber. Little is known about higher-order chromatin structures finally leading to a condensation of the genetic material to the level of metaphase chromosomes.

Figure adapted from Shuaib. 2012[11].

In 1977, Arne Leth Bak described the higher-order structure of human mitotic chromosomes[12]. Although chromosome and chromatin have different high-level structures, they both show the classical 'beads on a string' structure[13–16]. String stands for DNA strand as the linker, while beads means nucleosome. The electron microscopic and biochemical evidence of this repeat unit were shown by P. Oudet, M. Gross-Bellard, and P. Chambonthat, and name it 'nucleosome'[17].

So what is nucleosome? As the first level of chromosomal high-level structures, the nucleosome of the repeat unit is composed of DNA and proteins forming the nucleosome core particle (NCP). The shape of a nucleosome resembles a flattened cylinder - a DNA strand is coiled around the periphery of the histone octamer (Fig 3)[18,19].

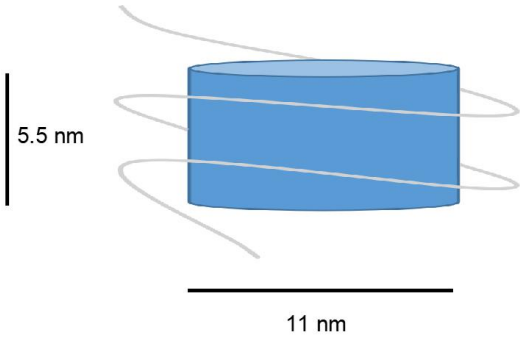


Figure 3. The nucleosome is the fundamental repeating subunit of chromatin — the first level of higher-order packaging of chromosomal DNA by histones. Each nucleosome particle consists of ~200 base pairs of DNA (the actual repeat length varies among different eukaryotic species) wrapped around a histone protein core, leading to an approximate sixfold-length compaction of the DNA. The histone core consists of an octamer of pairs of four histones (H2A, H2B, H3 and H4) related by a single dyad axis. The 'core particle', derived by nuclease digestion of chromatin, is a metastable product with 146 base pairs wrapped around the histone octamer ~90 base pairs per turn. A 'linker' region of DNA between core particles is more susceptible to nuclease degradation than the core particle DNA and is associated with histone H1. The core particle is shaped like a squat cylinder, with a diameter of ~11 nm and a height of ~5.5 nm and histones (blue)). The four histones of the octamer associate by their highly α-helical globular regions.

Figure adapted from Olins et al., 2003[19].

For each nucleosome unit, the average length of DNA molecules is about 200 bp, of which the length around the octamer is about 146 bp[20]. The canonical H3 nucleosome octamer consists a pair of four histones, namely H2A, H2B, H3 and H4 (Fig 4)[11,20]. In addition, the fifth histone, H1, is mainly associated with linker DNA[19]. Under physiological conditions, histones are basic proteins with positive charge, which help to bind to negatively charged DNA. In addition, a variety of non-histone proteins are also present around the nucleosome that are involved in the maintenance of chromosome structure or the execution of chromosome-related functions.

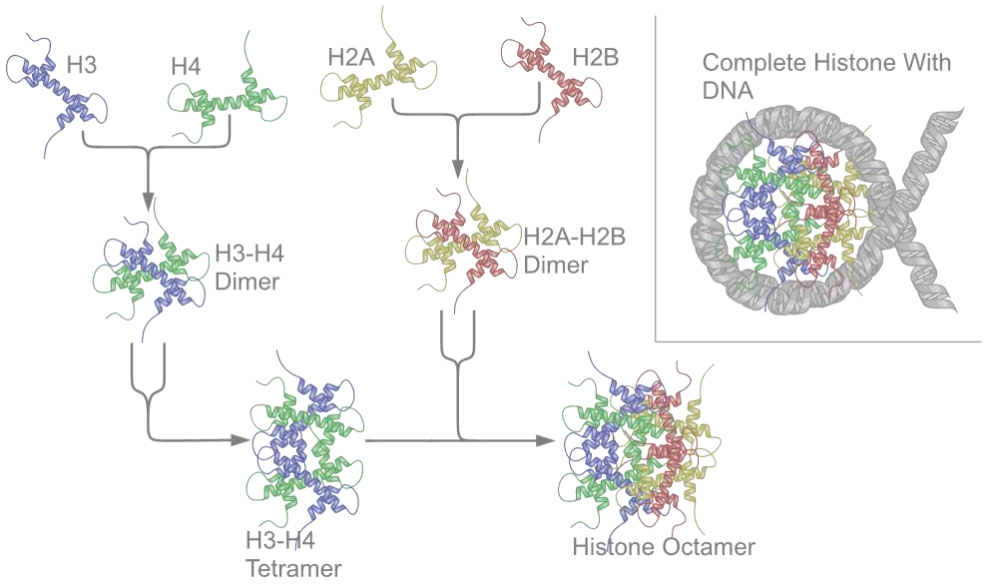


Figure 4. Schematic illustration of histones assembly into Nucleosome.

Two molecules of each of the four core histone proteins (H2A, H2B, H3 and H4) form the histone octamer via formation of one tetramer of H3 and H4 and two dimers of H2A and H2B. These entities are held together by a so-called hand-shake motif of protein structure. The histone octamer is wrapped by 146bp of DNA to complete nucleosome formation.

Figure adapted from Shuaib. 2012[11].

The defining histone of the H3 nucleosome is H3 histone. Both H3.1 and H3.2, with only one different amino acid (S96C), are canonical histones[21]. Both of them are synthesized and then localized to the chromosome in S phase. In addition, there are many sequence and functionally distinct variants of H3 histones including CENP-A (Fig 5)[11,21], which plays a crucial role in cell division (described below).

A. Sequence alignment of human H3 variants

```
H3.1            -ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 59
H3.2            -ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 59
H3t             -ARTKQTARKSTGGKAPRKQLATKVARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 59
H3.3            -ARTKQTARKSTGGKAPRKQLATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQKSTE 59
H3.5            -ARTKQTARKSTGGKAPRKQLATKAARKSTPSTCGVK-PHRYRPGTVALREIRRYQKSTE 58
H3.X            -ARTKQTARKATAWQAPRKPLATKAARKRASPTGGIKKPHRYKPGTLALREIRKYQKSTQ 59
H3.Y            -ARTKQTARKATAWQAPRKPLATKAAGKRAPPTGGIKKPHRYKPGTLALREIRKYQKSTQ 59
CENP-A          GPRRRSRKPEAPRRRSP-SPTPTPGPSRRGPSLGASSHQHSRRR-QGWLKEIRKLQKSTH 58

H3.1            LLIRKLPFQRLVREIAQDFKT--DLRFQSSAVMALQEACEAYLVGLFEDTNLCAIHAKRV 117
H3.2            LLIRKLPFQRLVREIAQDFKT--DLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRV 117
H3t             LLIRKLPFQRLMREIAQDFKT--DLRFQSSAVMALQEACESYLVGLFEDTNLCVIHAKRV 117
H3.3            LLIRKLPFQRLVREIAQDFKT--DLRFQSAAIGALQEASEAYLVGLFEDTNLCAIHAKRV 117
H3.5            LLIRKLPFQRLVREIAQDFNT--DLRFQSAVVGALQEASEAYLVGLLEDTNLCAIHAKRV 116
H3.X            LLLRKLPFQRLVREIAQAISP--DLRFQSAAIGALQEASEAYLVQLFEDTNLCAIHARRV 117
H3.Y            LLLRKLPFQRLVREIAQAISP--DLRFQSAAIGALQEASEAYLVQLFEDTNLCAIHARRV 117
CENP-A          LLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHLFEDAYLLTLHAGRV 118

H3.1            TIMPKDIQLARRIRGERA---------- 135
H3.2            TIMPKDIQLARRIRGERA---------- 135
H3t             TIMPKDIQLARRIRGERA---------- 135
H3.3            TIMPKDIQLARRIRGERA---------- 135
H3.5            TIMPKDIQLARRIRGERA---------- 134
H3.X            TIMPRDMQLARRLRGEGAGEPTLLGNLAL 146
H3.Y            TIMPRDMQLARRLRREGP---------- 135
CENP-A          TLFPKDVQLARRIRGLEEGLG-------- 139
```

B. Different characteristics of H3 variants

H3 variants		Amino acids	Identity with H3.1 (%)	Related Functions
Canonical	H3.1	136	_ _	DNA replication, repair, chromosome stability
	H3.2	136	99	_ _
Non-canonical (Somatic)	H3.3	136	96	Transcription, Sperm pronucleus decondensation, Incorporation during mammalian meiotic sex inactivation (MSCI),Chromosome inactivation, Pericentric transcription,Telomere silencing
	CENP-A	140	42	Assembly of Kinetochore, Chromosome segregation
	H3.X	147	73	?
	H3.Y	136	77	Response to external stimuli
Non-canonical (Testis Specific)	H3t	136	97	Expressed in testicular cells
	H3.5	135	85	Associated with actively transcribed genes Seminiferous tubules

Figure 5. Sequence alignment and characteristics of H3 variants.

A Amino acid sequence alignment of different human H3 variants (H3.1, H3.2, H3t, H3.3, H3.5, H3.X, H3.Y and CENPA). Identical amino acids are represented in black letters and the amino acids differences among human H3 variants are shown in red letters. The residues of H3.3, H3.5, H3.X and H3.Y corresponding to replication independent deposition are highlighted in gray. The position 31 of H3.3 and H3.5 contain serine residue. Differences of CENP-A from other H3 variants are highlighted in green.

B Different properties of human H3 variants include number of amino acids, % identity with H3.1 and major functions.

Figure adapted from Hamiche et al., 2012[21].

1.3 Centromere & CENP-A nucleosome

The term 'centromere' was first defined in 1936[22]. At M phase, pairs of sister chromatids are connected by centromere at the main constriction region of chromosome, and are divided into two arms - short arm (p) and long arm (q). Unlike the chromatin of chromosomal arm, more commonly, the centromere is mainly composed of highly repetitive heterochromatin. One of the main difference is that the DNA wound outside the heterochromatin is high-order repeated DNA - satellite DNA[23]. Among many eukaryotes, alpha satellite DNA is one of the most prevalent one[24], such as human (*Homo sapiens*). Compared to human, the composition of the repeat-units of chicken (*Gallus gallus*) centromeric DNA is more complex, with each repeat unit being chromosome-specific and three chromosomes containing non-tandem-repetitive sequences[25]. The repetitive sequence of these DNA are not necessary to form centromeres[26]. In addition, the main feature that defines eukaryotic centromere is the presence of a histone H3 variant centromere protein A (CENP-A), also known as centromeric histone H3 (CenH3), which is an epigenetic marker[27–29]. Therefore, the nucleosomal repeat unit in centromere region is a satellite DNA surrounded octamer, containing CENP-A, H4, H2A and H2B histones, with

interspersed H3 nucleosomes[30,31].

Highly specialized centromeres are necessary to maintain genome stability and are essential for chromosome segregation during cell division process. For more than 100 years, scientists have focused on the study of centromere. Although many important research advances have been shown, until now, it is still a big challenge to clarify both functions and specific molecular mechanisms of the centromere in different species and cell cycle phases. The specific molecular mechanism of centromere deposition is one of the basic research questions in this subject, and also one of the most key points that need to be answered.

1.3.1 Centromere

The centromere is an epigenetically specified region in eukaryotes, except for in budding yeast[32,33]. Based on extensive studies of eukaryotes, chromosomes have been classified into two types, monocentric chromosome and holocentric chromosome, based on the distribution status of centromere throughout the chromosome (Fig 6)[33].

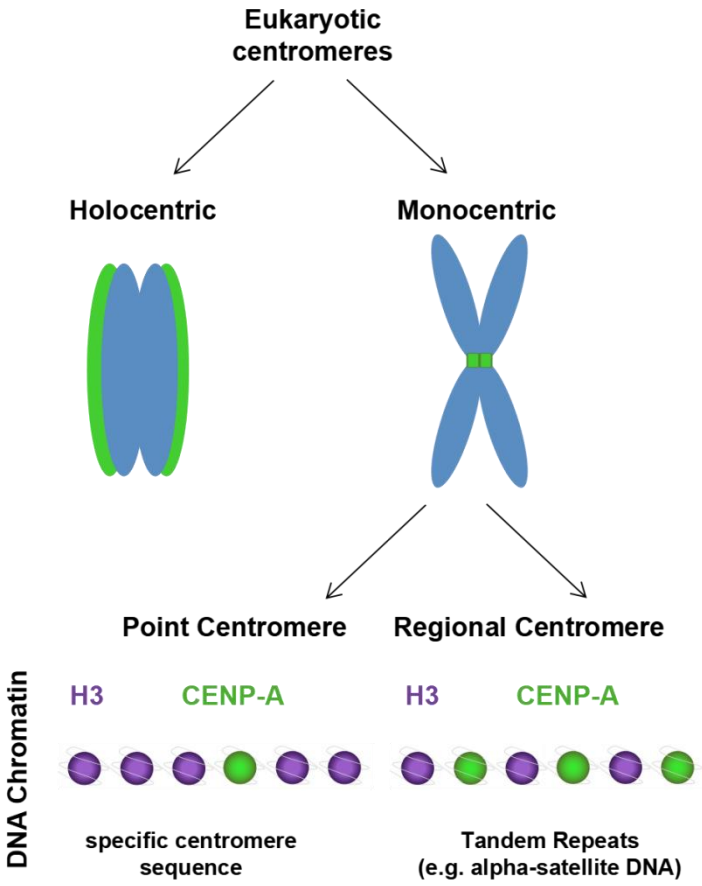


Figure 6. Centromere specification.

Diagram of the diverse types of centromeres found across eukaryotes. Holocentric chromosomes assemble a diffuse centromere across the whole chromosome. Monocentric chromosomes assemble a centromere at a single localized site on the chromosome, which is visible as a constriction between the chromosomes in mitosis (known as the primary constriction). Monocentric chromosomes can be further divided into those with point centromeres and those with regional centromeres. Point centromeres contain a specific DNA sequence that is sufficient for centromere function (here illustrated with the S.cerevisiae DNA architecture), which assembles a single CENP-A nucleosome. Regional centromeres contain large regions of DNA that is often repetitive (such as alpha-satellite DNA in primates), and assemble numerous CENP-A nucleosomes.

Figure adapted from McKinley et al., 2016[33].

For most eukaryotic chromosomes, it is common for the centromere to be concentrated in one region with an monocentric distribution. A monocentric chromosome may posses a point centromere with only one CENP-A nucleosome, such as *S.cerevisiae*, or a regional centromere with more than one CENP-A nucleosomes assembled as in human (*Homo sapiens*)[34,35]. Moreover, based on the location, monocentric distribution can be divided into four forms: metacentric centromere (in the middle), submetacentric centromere (towards one end), acrocentric centromere (near one end) and telocentric chromosomes (at one end)[36]. However, for a few eukaryotes like certain nematodes, there is a phenomenon called holocentricity, where the centromere of such cells is diffusely assembled throughout the chromosome. Chromosome like this is called holocentric chromosomes[37].

Interestingly, centromere does not always assemble on the centromeric region. In 1993, centromere was first detected locating at an abnormal location on the human chromosome, without alpha satellite DNA but unique DNA sequences, but still, shows stable mitotic function[38]. This kind of centromere is named as ‘neocentromere’. The finding of neocentromere provides the best evidence for the sequence-independent epigenetic mechanisms of centromere. Also, neocentromere can be used as a reliable tool for centromeric epigenetic mechanism learning[39].

1.3.2 CENP-A nucleosome

The existence of histone CENP-A defines the centromere in most eukaryotes. In the 1980s, three antigens located in the centromere region of the Hela chromosome were identified by examining "preimmune" serum samples from a patient with scleroderma CREST, and the 17-kd antigen was named CENP (Centromere Protein)) - A[40]. Further research identified that CENP-A, as a histone protein of chromatin, is homologous to H3[29,41]. Importantly, CENP-A is an epigenetic marker of centromere[31,32,42–44].

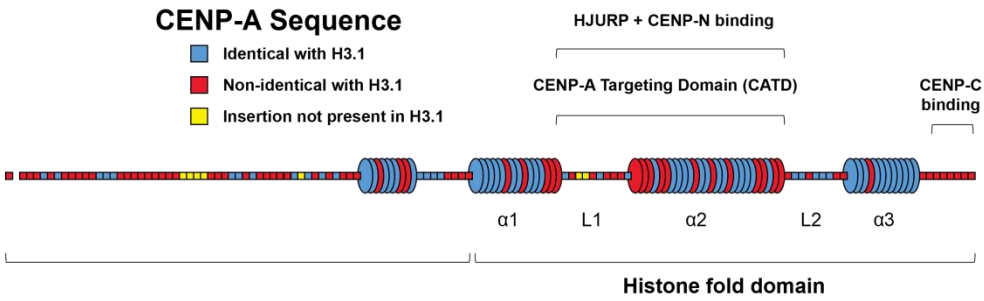


Figure 7. Model of human CENP-A primary and secondary structure showing conservation with histone H3.

Each segment corresponds to a single amino acid, and is colored according to its conservation with human H3.1 as indicated. The first N-terminal amino acid, shown detached, represents the cleaved initiator methionine. Barrels represent alpha helices, and rods represent loops. Within the histone fold domain, the helices are designated alpha1 through alpha3, and the loops are designated L1 and L2. L1 and alpha2 comprise the CENP-A targeting domain, which is sufficient to target CENP-A to centromeres due to its interaction with the CENP-A chaperone, HJURP. This region also binds to CENP-N[45] and is important for CENP-C recruitment[46,47]. CENP-C also binds to the C terminal residues of CENP-A[48–50].

Figure adapted from McKinley et al., 2016[33].

Checking the sequence of histone CENP-A, the histone fold domain and an N-terminal tail are the two main distinct domains from that of human histone H3 (Fig 7)[33,41]. In fact, the CENP-A sequences also differed between species (Fig 8)[51,52]. The multiple sequence alignment among several key species shows the strong conservation of the CENP-A histone fold domain, showing the highest divergency between two species is 60.4 %. While, the N-terminal tail are in higher divergence, as the shortest one is 35 residues in chicken and the longest residues shows 52 in zebrafish[51]. In the histone folding domain, the region consisting of L1 (first loop) and alpha 2 (second alpha helix) is called the CENP-A targeting domain (CATD) which is the key domain to targeting CENP-A to centromere (Fig 7)[33]. The CENP-A nucleosome has different structural properties from the H3 nucleosome with the presence of CATD, which causes the (CENP-A-H4)2 tetramer has a more rigid conformation than the (H3-H4)2 tetramer. The structural difference also results in an altered centromere chromatin structure that is conducive to the realization of

centromere function[53,54]. In addition to the centromere specificity, the sequence differences between CENP-A and H3 also give CENP-A the property of being assembled with kinetochore (described below).

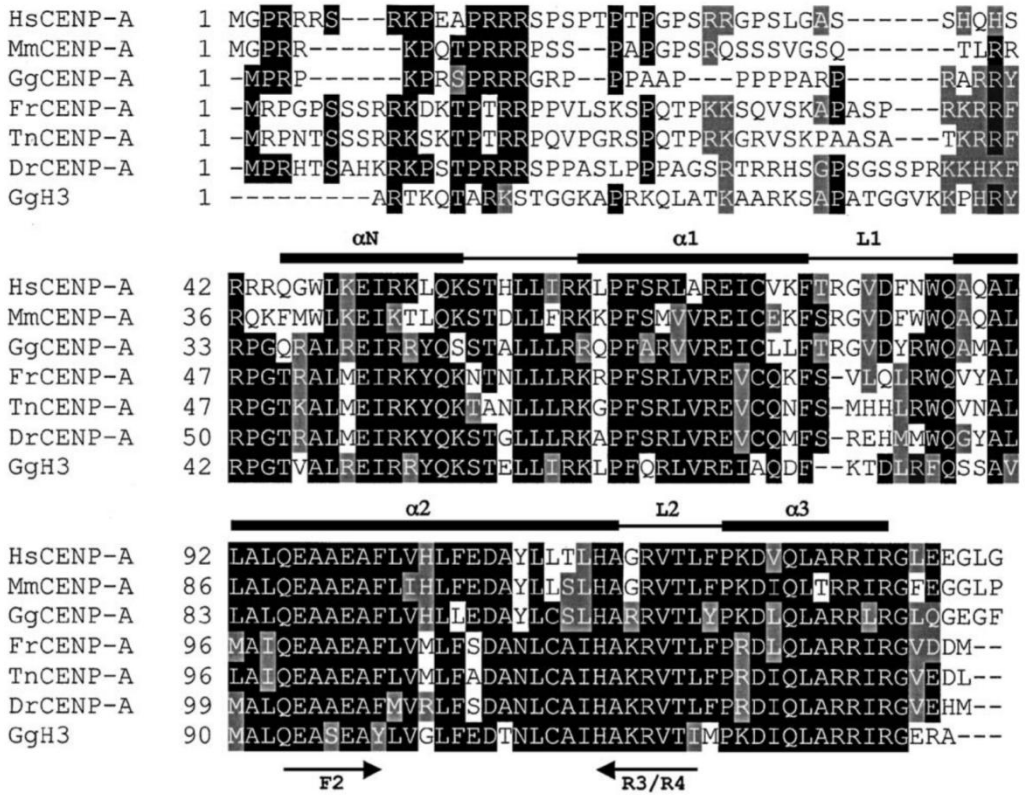


Figure 8. Comparative sequence analysis of vertebrate CENP-A and histone H3 (*Gallus gallus*).

Hs: *Homo sapiens*, Mm: *Mus musculus*, Gg: *Gallus gallus*, Fr: *Fugu rubripes*, Tn: *Tetraodon nigroviridis*, Dr: *Danio rerio*. Pufferfish (Fr and Tn) and zebrafish (Dr) CENP-A were predicted from computational analysis of databases. Identical and similar residues are shown in black and gray, respectively. The location of degenerate primers used to successfully amplify a stretch of chicken CENP-A is depicted on the alignment. Schematic secondary structure (α -helix and loops) in the histone-fold domain is also indicated (structure assignment is from Luger et al., 1997[20]).

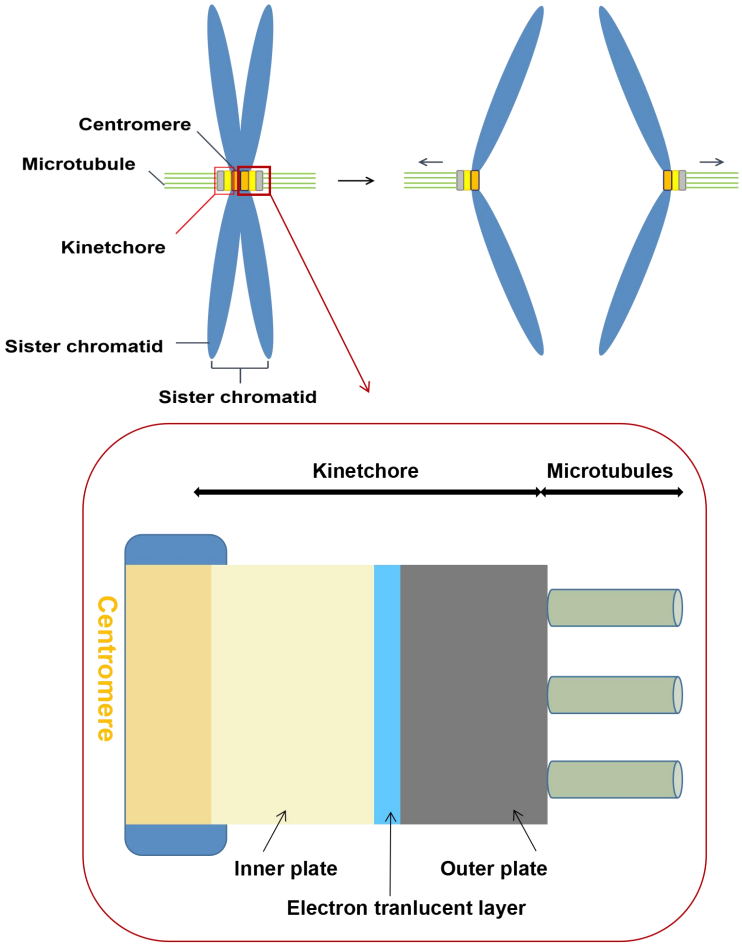
Figure adapted from Regnier et al., 2003[51].

1.4 Kinetochore

For the centromere, in addition to the CENP-A nucleosome and chromatin marks, interactions with kinetochore proteins are necessary to ensure proper mitosis[33,55–58]. In fact, the microtubule-centromere linkage is mediated by kinetochore, which is fully assembled to centromere during mitosis. The term 'kinetochore' (movement place) was first used in a footnote in the book Introduction to Cytology, published in 1934[59].

During mitosis, a complete kinetochore consisting of numerous proteins is formed on each sister chromatid, for example, there are about more than one hundred kinetochore proteins in human cells[60]. The trilaminar structural kinetochore is revealed by Electron microscopy studies, containing electron-dense inner and outer plates, and an electron-translucent middle layer (Fig 9A)[61,62]. Although the composition of each layer of matter is not completely understood, current research provides a basic understanding of the kinetochore. The kinetochore infrastructure model consists of two main parts, constitutive centromere-associated network (CCAN) and Knl1-Mis12-Ndc80 (KMN) network (Fig 9B) [61]. CCAN consists of 16 protein subunits (CENP-C, -H, -I, -K, -L,-M, -N, -O, -P, -Q, -R, -S, -T, -U, -W, and -X) that interact with centromeric chromatin throughout the cell cycle and is a component of the inner plate along with CENP-A nucleosome[63]. CCAN recruits KMN at the M phase, and a KMN component Ndc80, also as a component of the outer plate[64], directly binds to microtubules (MTs)[65,66]. Thus, a linkage from a centromere to microtubules is formed, and the chromosome migration is mediated by the kinetochore. At last, after chromosome segregation, the outer plate disassembles.

A


Centromere
Microtubule
Kinetchore
Sister chromatid
Sister chromatid
Kinetchore
Microtubules
Centromere
Inner plate
Outer plate
Electron tranlucent layer

B

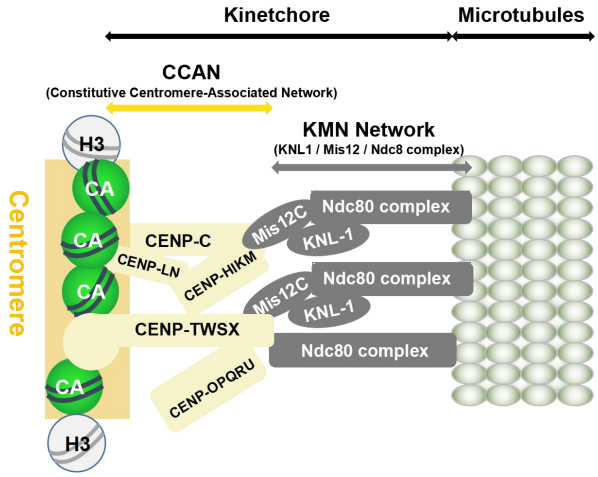

Kinetchore
Microtubules
CCAN
(Constitutive Centromere-Associated Network)
KMN Network
(KNL1 / Mis12 / Ndc8 complex)
H3
CA
CA
CA
CA
H3
Centromere
CENP-C
CENP-LN
CENP-HIKM
CENP-TWSX
CENP-OPQRU
Mis12C
KNL-1
Ndc80 complex
Mis12C
KNL-1
Ndc80 complex
Ndc80 complex

Figure 9. The kinetochore in M-phase.

A Mitotic chromosome segregation and a trilaminar structural model for the kinetochore. The kinetochore is a large protein complex built on the centromere of each sister chromatid (chromosome). The kinetochores attach to microtubules emanating from each opposing spindle pole, segregating chromatids. Electron microscopy studies revealed that the kinetochore has three layers in its structure: electron-dense inner and outer plates, and an electron-translucent middle layer.

B A model of basic kinetochore structure. The main structure of the kinetochore is formed of constitutive centromere-associated network (CCAN) composed of 16 protein subunits and the KMN (Knl1, Mis12, and Ndc80 complexes: Knl1C, Mis12C, Ndc80C) network (KMN). CCAN interacts with the centromeric chromatin epigenetically marked with the CENP-A nucleosome during the entire cell cycle. CCAN recruits KMN, which directly binds to microtubules, in M-phase, forming a linkage between the centromere and microtubules.

Figure adapted from Hara et al., 2020[61].

Furthermore, the spindle assembly checkpoint (SAC) proteins use kinetochore as a binding platform to ensure that all chromosomes are attached to the poles of the spindle before the cell enters anaphase[60,67,68]. Also, these SAC proteins are regulated by the kinetochores, which corrects incorrect microtubule attachment and coordinates cell cycle progression[69,70].

1.4.1 Kinetochore assembles on the centromere

The centromere is the assembly region of the kinetochore. Since CCAN, the inner kinetochore complex, connects with the centromere through all the cell cycle, an important question is how kinetochore is assembled on centromere? Some CCAN proteins were confirmed directly bind with centromeric DNA[55,71–73], for example, human CENP-LN channel could interact with extra nucleosomal DNA duplex[74]. But as for specific binding, only CENP-C and CENP-N specifically bind with CENP-A nucleosome and they are not in competition with each other[45,49]. Thus, these direct binding contribute to the CCAN specific localization to the centromere region[49,50,75].

On the other hand, another kinetochore sub-complex, KMN network is recruited on CCAN through two parallel pathways: CENP-C pathway and CENP-T pathway[76]. Both of CENP-C and CENP-T are CCAN proteins, they connect the centromere region by their C-terminus[55,73]. And with their N-terminal region, these two proteins bind to the KMN network sub-complex (Fig 10)[73,76–79].

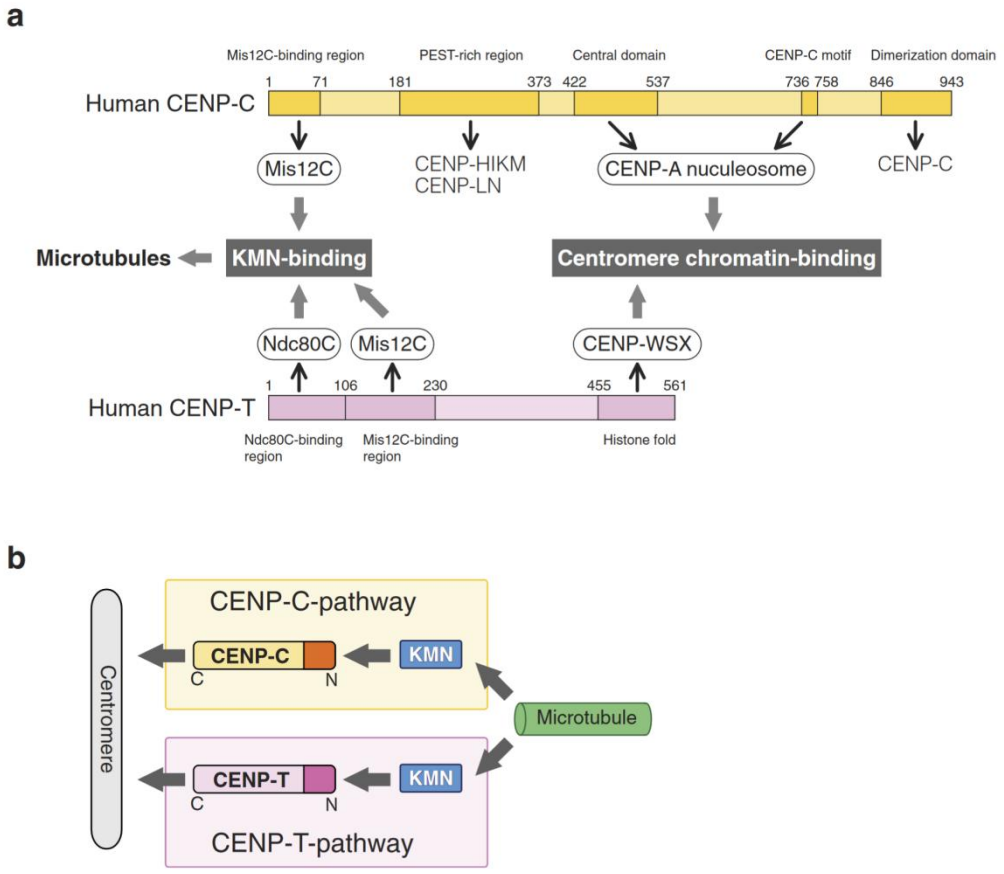


Figure 10. Two-pathways model connecting the centromere to spindle microtubules.

A Schematic representation of human CENP-C and CENP-T. Human CENP-C interacts with various centromere/kinetochore proteins. The Mis12 complex (Mis12C)-binding region in the extreme N-terminus binds to KMN, which associates with the spindle microtubules. The PEST-rich region interacts with the CCAN subcomplexes, CENP-HIKM, and CENP-LN. There are two CENP-A nucleosome-binding regions: central domain and CENP-C motif. The dimerization domain in the extreme C-terminus is thought to promote self-dimerization. Human CENP-T directly binds to the Ndc80 complex (Ndc80C) through the extreme N-terminus (Ndc80C-binding region). Mis12C interacts to the Mis12C-binding region next to the Ndc80C-binding region (Mis12C-binding region). There is a histone fold domain in the C-terminus, which forms a nucleosome like complex with CENP-W, -S and -X, which also contain histone fold domains. The complex interacts with the centromere chromatin.

B Two pathways in the kinetochore that bridge the centromere to microtubules: CENP-C- and CENP-T-pathways. Both CENP-C and CENP-T bind to KMN via their N-terminus and centromere chromatin via their C-terminus, independently, making the two pathways in the kinetochore.

Figure adapted from Hara et al., 2019[76].

For example, using chicken cells, it was found that artificially tethered the CENP-C N terminus to a non-centromeric locus was sufficient to recruit KMN and that the artificially induced functional kinetochore was able to replace the native centromere for chromosome segregation[77,80,81]. With both or either of these two pathways as the scaffold, the complete kinetochore assembles on the centromere and bridge the centromere and microtubules[76].

1.4.2 CENP-C interact with CENP-A

CENP-C protein, belongs to the CCAN complex, is the largest CCAN subunit[82]. Along with CENP-A，CENP-C was first identified by examining "preimmune" serum samples from the scleroderma CREST patient[40,83]. Human CENP-C contains multifunctional domains, and can specifically bind with other proteins: 1) N-terminal Mis12 binding domain (binding with Mis12 complex)[79,84]; 2) mid-conserved domain in the middle-conserved region (binding with CENP-HIKM and CENP-LN complexes)[85,86]; 3) central domain (binding CENP-A nucleosome); 4) CENP-C motif (binding CENP-A nucleosome)[49,50]; and 5) dimerization domain in CENP-C C-terminus (forming CNEP-C dimerization)[87].

As introduced, CENP-C directly binds to CENP-A nucleosome as a key factor of CENP-C pathway[88]. Both CENP-C central domain and CENP-C motif, showing sequence similarity, are recognized by CENP-A nucleosome in human cells[50,89,90]. However, the CENP-C domain is conserved in most mammalian cells, while central domain is not[87]. For example, chicken (Gallus gallus) CENP-C protein only has CENP-C motif, but not the central motif (Fig 11)[91].

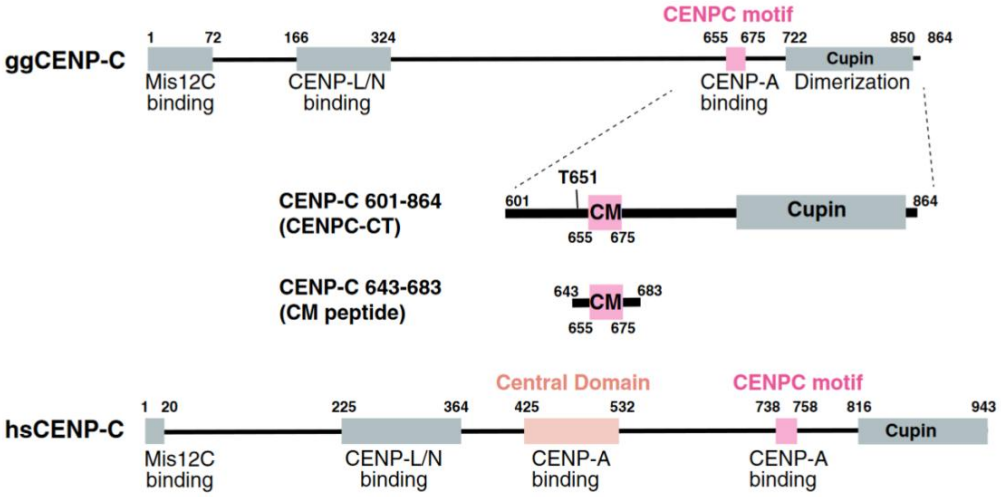


Figure 11. CENP-A nucleosome binding of the C-terminal fragment of CENP-C. A Schematic diagram of the functional regions of chicken and human CENP-C molecules. The canonical CENPC motifs for CENP-A binding (aa 655-675 in chicken CNEP-C, aa 738-758 in human CENP-C) are colored pink. Another CENP-A binding region (central domain) in human CENP-C is colored light pink. The C-terminal fragment (aa 601-864: CENPC-CT) and the CENPC motif-containing peptide (aa 643-683; CM peptide) derived from chicken CNEP-C, which were used for the in vitro CENP-A nucleosome-binding assay, are diagrammed.

Figure adapted from Ariyoshi et al., 2019[91].

A model of kinetochore localization of CENP-C shows the CENP-C regulation mechanism during the cell cycle (Fig 12)[92]. During mitosis, the CENP-C C-terminus binds to CENP-A nucleosome by CDK1 phosphorylation. But in the interphase, CENP-C connects with the centromere region through CENP−HIKM−LN complex, without directly binding with CENP-A[90,92].

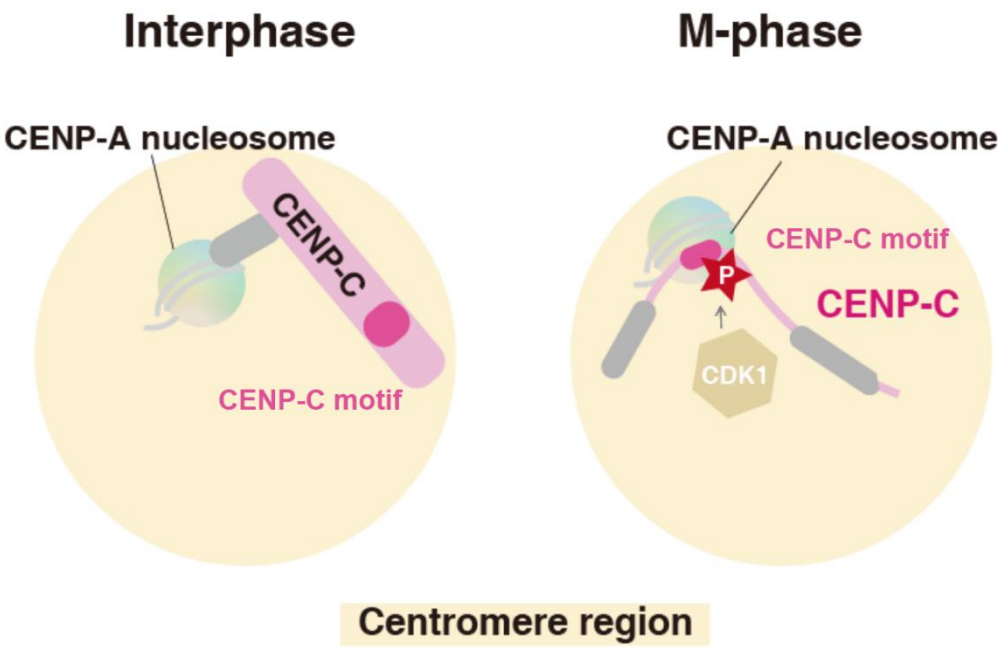


Figure 12. A model of kinetochore localization of CENP-C during interphase and mitosis.

CENP-C localizes to interphase centromeres through interaction of the CENP-C middle region with the CENP–HIKM–LN complex. During mitosis, T651 or T734 in gCENP-C or hCENP-C, respectively, is phosphorylated by CDK1 (depicted as "P"), and this phosphorylation facilitates binding of the CENP-C motif to the CENP-A nucleosome (CENP-A nuc.). Kinetochore localization of CENP-C depends on both CENP–HIKM–NL and the CENP-A nucleosome during mitosis.

Figure adapted from Watanabe et al., 2019[92].

1.5 Mis18 complex & KNL2

In the fission yeast Mis18 mutant which exhibits CENP-A deposition defects, Mis18 gene was identified for the first time, and this Mis18 has human homologs Mis18α and Mis18β[93]. In addition, M18BP1 was identified as human Mis18α and Mis18β binding partner by biochemical screening[94], which shows sequence homology to centromere protein KINETOCHORE NULL2 (KNL2) of C. elegans[95]. In chicken cells, M18BP1/KNL2 (aa 1-1088) contains three domains: SANTA domain, CENP-C like domain and SANTA domain (Fig 13)[96–100]. Of these, SANT and SANTA are conserved domains. SANT domain, locating on C-terminus, is critical for CENP-C binding in mouse cells[101]. Besides, it may be involved in chromatin binding as a Myb-like DNA binding domain[95]. The SANTA domain is known as a SANT-associated domain, however its function is not yet well understood. M18BP1/KNL2 CENP-C-like domain is not quite conserved among species, for example, it was observed in Xenopus and different fish species but not in human[98,99]. To compare the sequences between CENP-C-like motif and CENP-C motif among species, the highly conserved residues are noticed[50,91,92]. As introduced above, the

CENP-C motif in kinetochore protein CENP-C can be recognized by CENP-A nucleosome[91,92], thus, the M18BP1/KNL2 CENP-C-like motif is speculate to interact with CENP-A nucleosome. In addition, the extreme N-terminal region (aa 1-140) is essential for the function of KNL2 and is capable of binding to Mis18α/β[102–104]. Furthermore, in vertebrate cells, all Mis18α, Mis18β and M18BP1/KNL2 together constitute the Mis18 complex, and it is a licensing factor for new CENP-A deposition (described below)[94,99].

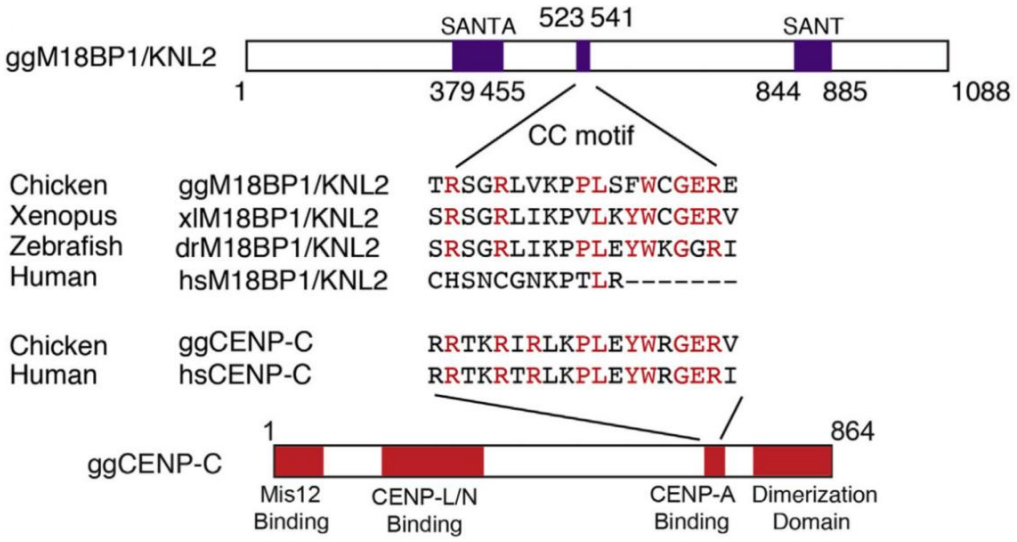


Figure 13. CENP-C-like Motif in Chicken M18BP1/KNL2 Is Essential for Its Centromere Localization and Function

Diagram of chicken M18BP1/KNL2 (top) and CENP-C (bottom). The position of the SANT domain, CENP-C (CC)-like motif, and SANTA domain is shown. Sequences of CENP-C-like motifs in various organisms are shown. Corresponding sequences for chicken and human CENP-C are also shown. Human CENP-C has two CENP-A binding regions and the second region (C terminus) corresponds to chicken CENP-A binding region. Red text indicates highly conserved residues between CENP-C (in CENP-A binding region) and the CENP-C-like motif in M18BP1/KNL2.

Figure adapted from Hori et al., 2017[99].

1.6 CENP-A deposition

To form a functional centromere, canonical H3 histone is replaced by CENP-A at the specialized chromatin region. The molecular mechanism of CENP-A deposition to centromere varies among species.

At present, some progress has been made in the study of the related mechanism in human[105]. New CENP-A deposited at centromeres in the early G1 phase during interphase, not coupled with DNA replication[106]. The exact molecular mechanism is shown in the figure (Fig 14)[105]. Specifically, CENP-C protein can discriminately bind with CENP-A, but not H3 nucleosomes[48,50]. Because the M18BP1/KNL2, one of Mis18 complex component, is recognized by CENP-C directly[101,107,108]. The Mis18 complex is a licensing factor for CENP-A deposition, which is localized to centromere since anaphase before new CENP-A loading[108,109]. For CENP-A timely regulating, Mis18 is positively regulated by PLK1 (polo-like kinase 1)[110] and negatively regulated by CDK (cyclin-dependent kinase)[103,111]. Another key factor HJURP (Holliday junction recognition protein) is a CENP-A-specific chaperone, which specifically deposits two new CENP-A/H4 heterodimers because of its self-dimerization[112–114]. For new CENP-A recruitement[106], HJURP is recruited to the centromere region by cell cycle regulated Mis18 complex[99,104,105,112,113]. After the end of CENP-A deposition, the Mis18 complex dissociated from the centromere because of the cell cycle dependent regulation[94,110,111].

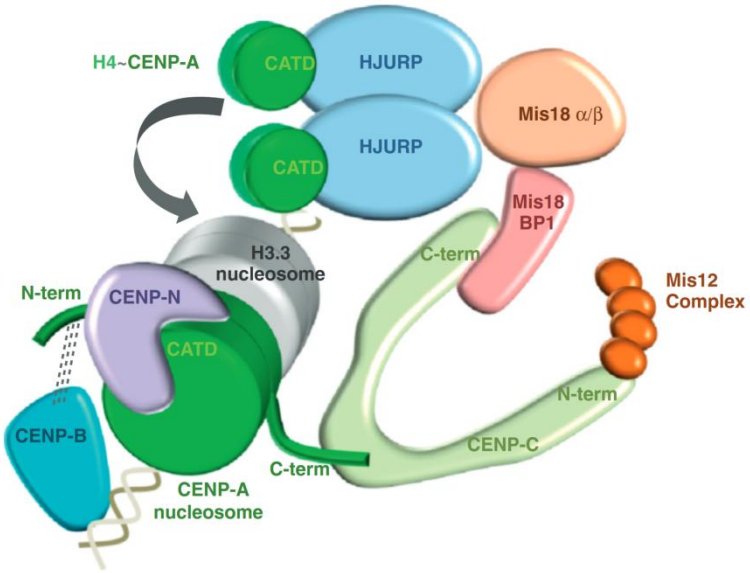


Figure 14. CENP-A recognition and propagation at regional centromeres. CENP-C binds directly to the C-terminus of CENP-A in nucleosomes. The C-terminus of CENP-C recruits the Mis18 complex through Mis18BP1 (also known as Knl2). During replication the CENP-A nucleosomes are distributed equally to each sister-centromere so that CENP-A levels are halved and histone H3.3 is deposited as a placeholder. Free CENP-A/H4 heterodimers associate with a homodimer of HJURP which is recruited to centromeres via the Mis18 complex in telophase, allowing replacement of H3.3 with CENP-A in G1. Once assembled, the CATD within the HFD of CENP-A nucleosomes is recognized by CENP-N allowing recruitment of many other constitutive kinetochore components including the CENP-T/W/S/X complex (not shown). The CENP-B protein is known to bind directly to centromere repeats in mammals but is stabilized via the N-terminus of CENP-A and contributes to kinetochore integrity. The N-terminus of CENP-C associates with the Mis12 complex. For simplicity interactions and nomenclature are only shown for vertebrate proteins.

Figure adapted from Catania et al., 2014[105].

The current CENP-A deposition model in chicken cells is not identical to that in human (Fig 15)[99]. The difference is that the chicken CENP-A localization mode does not involve the CENP-C protein. The Mis18 complex directly interact with existing CENP-A nucleosome through KNL2 without the intermediate of CENP-C protein. And the hexamer of Mis18α/β can be recognized by the CENP-A/H4 containing

HJURP complex to facilitate centromeric CENP-A deposition[102,103,107,115]. Then the new CENP-A carried by HJURP is deposited near the existing CENP-A nucleosome[99].

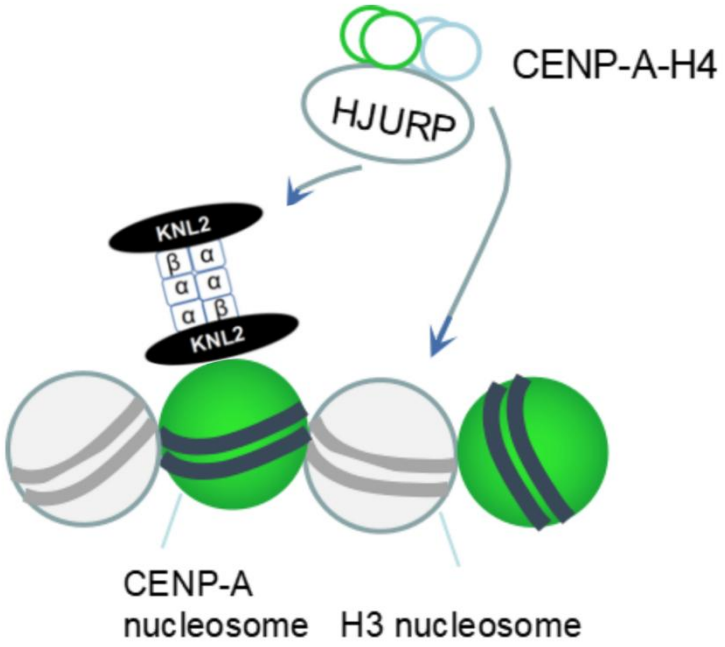


Figure 15. Current model of CENP-A deposition in chicken cells.

The Mis18 complex directly associated with CENP-A nucleosome (old CENP-A) before CENP-A deposition. HJURP, bound with CENP-AH4 (new CENP-A), recognizes the Mis18 complex, and new CENP-A is deposited near existing CENP-A nucleosome. α means Mis18α; β means Mis18β.

Figure adapted from Hori et al., 2017[99] and Pan et al., 2017[103].

Chapter 2: KNL2 directly binds to the CENP-A nucleosome via the conserved CENP-C-like motif region

2.1 Introduction

Eukaryotic cells rely on mitosis for proliferation, and the whole cell cycle is roughly divided into interphase and mitotic (M) phase. The filamentous chromatin condenses into chromosomes during M phase, and the two replicated sister chromatids are linked by centromere and pulled through spindle fiber to the opposite cellular poles, eventually forming two daughter cells[4–6]. In most organisms, centromere is a chromosomal specialized region defined by sequence-independent epigenetic mechanisms[27–29]. CENP-A (centromere protein A), the canonical histone H3 variant, marks the centromere region as a key marker for centromeric specification and maintanance[31,32,43,44,52]. To understand how CENP-A is deposited on the centromere, previous studies have gotten some progress in several species. To compare the current CENP-A deposition model between human and chicken, the significant difference is about the CENP-A recognition protein[99,105]. The human CENP-A deposition model shows that CENP-C binds with existing CENP-A nucleosome, and the Mis18 complex (a licensing factor for new CENP-A deposition) interact with CENP-C to localize to the centromere via a Mis18 complex component KNL2[94]. However, in chicken cells or Xenopus egg extract, KNL2 still localize to the interphase centromere without the existence of CENP-C[46,116]. Actually, the KNL2 directly binds to the CENP-A nucleosome in vitro without the intermediate of CENP-C protein[96,99,100]. An interesting point is, in some species but not human and mice, KNL2 has a CENP-C-like motif which shows quite similar conserved sequence as that of CENP-A binding motif in CENP-C (Fig 16)[50,91,92,96–100]. Therefore, a reasonable speculation is that, in chicken cells, existing centromeric CENP-A nucleosome directly binds to KNL2 but not CENP-C protein for new CENP-A deposition in the G1 phase.

To answer this hypothesis, I conducted a series of studies using biochemistry, molecular biology and structural biology (described in chapter 3), which eventually led to a specific molecular mechanism for the binding of KNL2 to the CENP-A nucleosome.

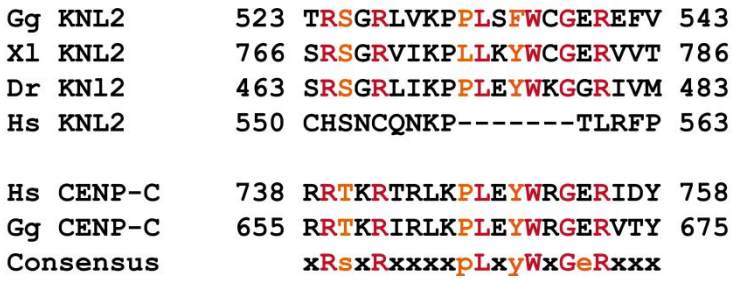


Figure 16. Sequence alignment of the CENP-C-like motif of KNL2 or the CENP-C motif of CENP-C from various species: Gg, chicken; Xl, frog; Dr, zebrafish; and Hs, human.

The identical and similar residues are colored in red and orange, respectively.

2.2 Results and discussion

Chicken KNL2 contains three major domains, SANTA motif, CENP-C-like motif and SANT motif from N-terminal to C-terminal, respectively (Fig 13)[97–100]. To bind to the CENP-A nucleosome, the CENP-C-like motif (CC; amino acid residues 523-543) was shown as the most likely binding region in previous study[99]. Also, as this CENP-C-like motif shows sequence similarity to one of the CENP-A binding motif in CENP-C (Fig 16), known as CENP-C motif, it worth to examine the CENP-A binding ability focusing around this region. Since the N-terminal SANTA motif position is close to the CENP-C-like motif, it was also be examined. Within the ggKNL2 region containing amino acids sequence 347 to 560 (aa 347-560), five various constructs were designed in total, with or without SANTA or CENP-C-like motifs. All the recombinant protein fragments are tagged with GST (GST-KNL2$^{347\text{-}560}$, containing both SANTA and CENP-C-like motifs; GST-KNL2$^{456\text{-}560}$, containing the CENP-C-like motif and its N-terminal upstream (CC-upstream) region but not the SANTA domain; GST-KNL2$^{518\text{-}560}$, containing CENP-C-like motif only; GST-KNL2$^{347\text{-}455}$, containing SANTA motif only; and GST-KNL2$^{456\text{-}516}$, containing CC-upstream region only) for further purification and biochemical assays (Fig 17).

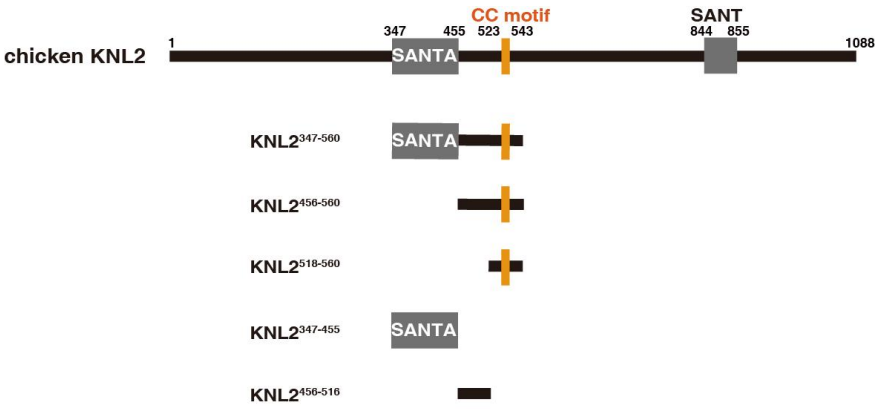


Figure 17. ggKNL2 recognizes the CENP-A nucleosome via the CENP-C-like motif.

A Schematic diagram of the domain organization of chicken KNL2 (ggKNL2). The CENP-C-like motif for CENP-A binding (aa 523 – 543) is colored in orange. The identical and similar residues are colored in red and orange, respectively. The KNL2 fragments, which were used for in vitro CENP-A nucleosome-binding assays, are diagramed.

A chimeric CENP-A, in which the native ggCENP-A N-terminal region was replaced by the corresponding region of histone H3 (Fig 18), was prepared to reconstitute the CENP-A nucleosome, together with histone H4, H2A,H2B and 601 DNA[117,118], in vitro. Because the full-length CENP-A reconstitution was failed, the chimeric CENP-A is a reliable alternative histone for biochemical and structural checking[91,92,99]. As a control, canonical H3 nucleosomes were also prepared.

To determine the exact KNL2 binding region with the CENP-A nucleosome, I performed both EMSAs and BLI assays to examine the binding of CENP-A nucleosome to each of the KNL2 fragment mentioned above.

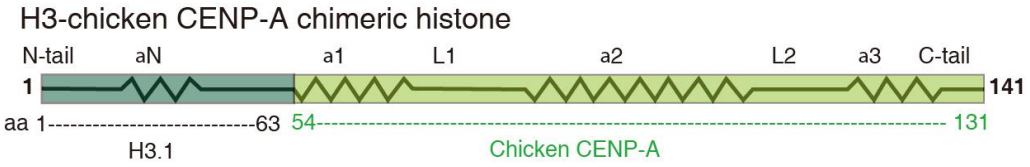


Figure 18. A schematic representation of the H3-chicken CENP-A chimeric histone (141 amino acid residues) used in biochemical and structural studies.

The boundary between histone H3 and chicken CENP-A is indicated. The N-terminal region derived from histone H3 (aa 1–63) is colored in dark green.

2.2.1 KNL2 CC containing fragments stably bind to the CENP-A nucleosome in vitro

EMSA and BLI assays results demonstrate that all the CENP-C-like motif containing KNL2 fragments (GST-KNL2$^{347-560}$, GST-KNL2$^{456-560}$, GST-KNL2$^{518-560}$) specifically bound to CENP-A nucleosome (Fig 19A-C; Fig 20A-C and F). Both of KNL2$^{347-560}$ and KNL2$^{456-560}$ fragments were observed to form a complex at a lower molar ratio to the CENP-A nucleosome than that of KNL2$^{518-560}$, indicating stronger CENP-A binding ability. For comparing, the KNL2$^{456-560}$ and KNL2$^{518-560}$ fragments do not bind to the H3 nucleosome, while KNL2$^{347-560}$ was shown clear bound shift in the EMSA result (Fig 19A-C) (Refer to the next section). To examine the dissociation constant (K_D) between CENP-A and these CENP-C motif containing KNL2 fragments, KNL2$^{518-560}$ BLI assay was performed, and CC only fragment was estimated to be 5.5 ± 1.1 μ M, while both the CC containing CC-upstream region with or without SANTA motif (KNL2$^{347-560}$ and KNL2$^{456-560}$) were shown quite similar K_D value (16.4±5.9 nM and 22.0±1.8 nM, respectively) (Fig 20A-C). Comparing these data, these K_D values remind that both KNL2$^{347-560}$ and KNL2$^{456-560}$ possess similar binding affinity to the CENP-A nucleosome, which of both were higher than that of the KNL2$^{518-560}$ fragment with only KNL2 CC.

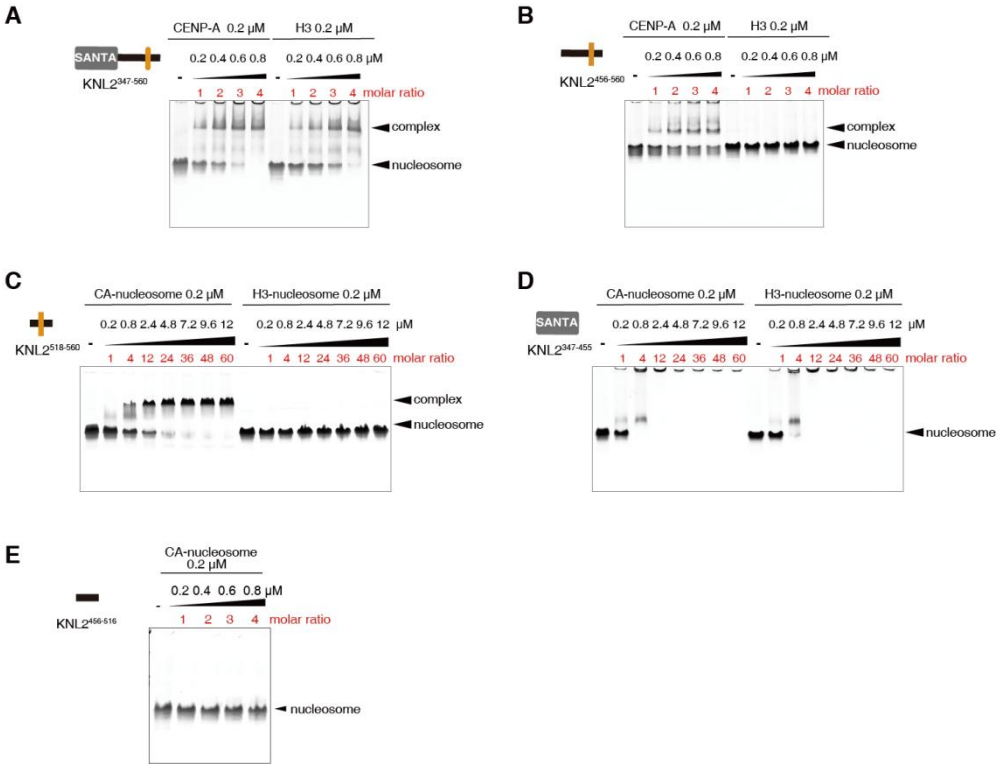


Figure 19. Electrophoretic mobility shift assay (EMSA) to examine binding affinities of the GST-KNL2 fragments for the CENP-A or H3 nucleosome.

A GST-KNL2$^{347–560}$; B GST-KNL2$^{456–560}$; C GST-KNL2$^{518–560}$; D GST-KNL2$^{347–455}$; E GST-KNL2$^{456–516}$ (CC–upstream). The molar ratio of the KNL2 fragment against nucleosomes are indicated.

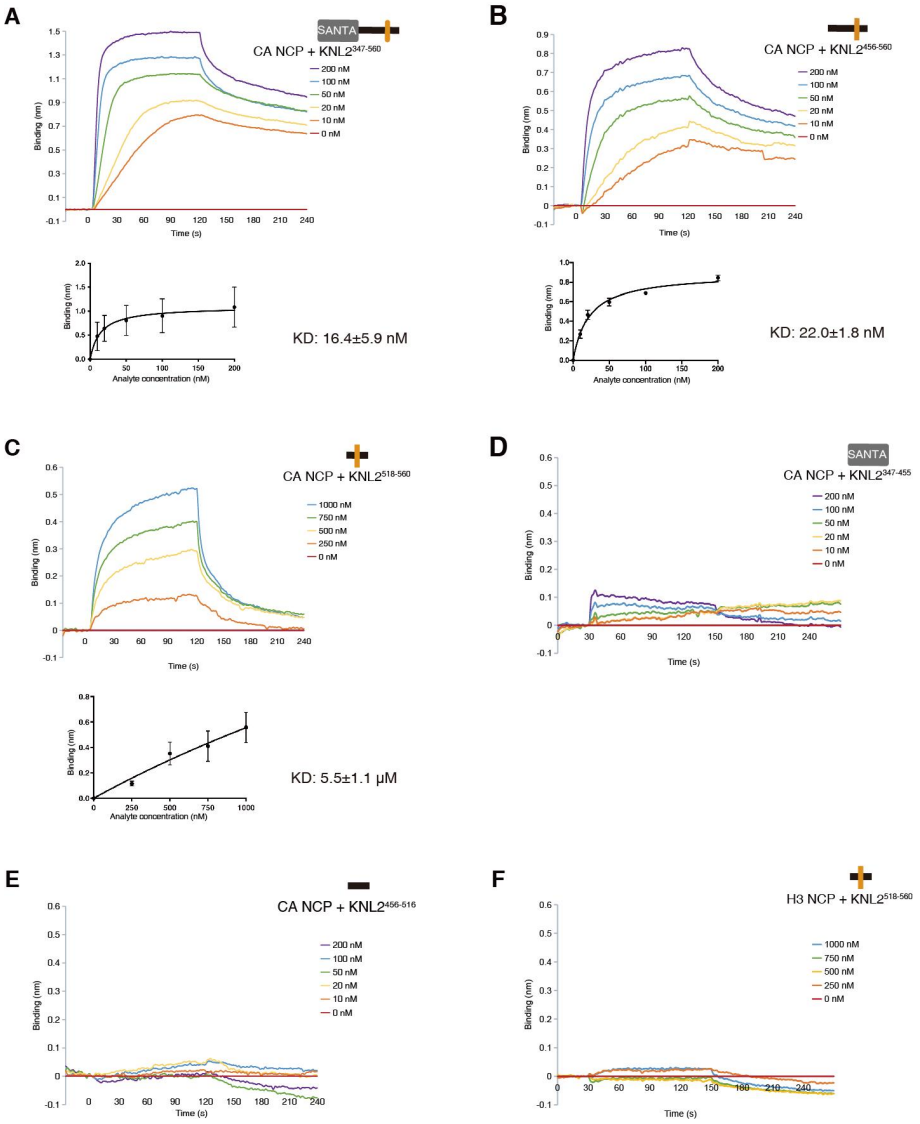


Figure 20. Biolayer interferometry (BLI) assays to examine the interaction between the GST-KNL2 fragments (immobilized ligand) and CENP-A nucleosome (A-E) or H3 nucleosome (F) (analyte):

(A) GST-KNL2$^{347–560}$; (B) GST-KNL2$^{456–560}$; (C) GST-KNL2$^{518–560}$; (D) GST-KNL2$^{347–455}$; (E) GST-KNL2$^{456–516}$ (CC–upstream). Representative BLI sensorgrams measured with analyte concentrations ranging from 0 to 1 µM in (B, F) or from 0 to 200 nM in (A,C-E) are plotted in the upper panel. The BLI sensorgrams were measured at association time intervals of 120 s and dissociation time intervals of 120 s. Binding assays were technically replicated three times. BLI-derived steady-state analysis of the binding response (nm) as a function of the concentration of each KNL2 fragment is presented in the lower panel of (A–C). Data are mean with standard deviations calculated from the technical replicates, n = 3.

2.2.2 KNL2 SANTA only fragment shows non-specific binding to the CA nucleosome

The KNL2 fragment containing both SANTA and CENP-C-like motif is shown to bind to both CENP-A and H3 nucleosomes in vitro, while the other CC containing fragments ($KNL2^{456-560}$ and $KNL2^{518-560}$) can only bind to the CENP-A nucleosome but not H3 nucleosome (Fig 19A-D; Fig 20A-C). Does it mean that the SANTA motif is specifically recognized by the nucleosomes? The relevant K_D value shown similar affinities of $KNL2^{347–560}$ and $KNL2^{456–560}$ for the CENP-A nucleosome, which implied that SANTA would not contribute to improving the binding affinity in the CENP-A nucleosome–KNL2 interaction (Fig 20A, B). Besides, comparing the band shift condition between CENP-A nucleosme-KNL2 and 601 DNA-KNL2 interaction in the EMSA assay, the KNL2 SANTA only fragment could definitely bind to the histone octamer binding DNA — the ‘601 DNA’ was selected from synthetic random sequences and discovered by Lowary and Widom (Fig 19D-Fig 21)[117,118]. Also, the binding response of $KNL2^{347–455}$ shows quite low level with a decreasing tendency, which means this SANTA containing KNL2 fragment does not specifically bind to the CENP-A nucleosome (Fig 20D). Therefore, my conclusion is that the SANTA motif of KNL2 is not involved in the nucleosome-related binding and it might nonspecifically bind to the artificial 601 DNA sequence, which is consistent with the previous data that SANTA domain is dispensable for the KNL2 centromere localization[99].

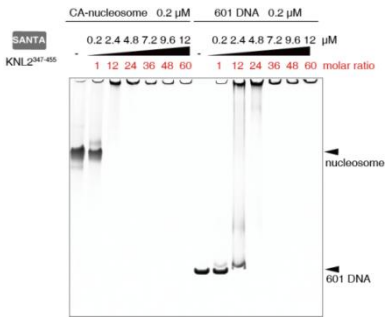


Figure 21. Electrophoretic mobility shift assay (EMSA) to examine binding affinities of the GST-$KNL2^{347–455}$ fragments for the CENP-A or 601 DNA.

The molar ratio of the KNL2 fragment against nucleosomes are indicated.

2.2.3 KNL2$^{456-516}$ fragment does not bind to the CENP-A nucleosome

As shown in the BLI data, comparing with 5.5 ± 1.1 μ M K_D value of KNL2$^{518-560}$, the smaller value of KNL2$^{456-560}$ (22.0±1.8 nM) indicated that KNL2$^{456-560}$ bound to the CENP-A nucleosome more stably. However, the only difference is KNL2$^{456-560}$ contains the middle region between SANTA and CENP-C-like motifs with the amino acid (aa) residues from 456 to 516. Intriguingly, the KNL2 fragment with this CC-upstream region only, shown as KNL2$^{456–516}$, did not bind to the CENP-A nucleosome in the binding assays (Fig 19E, Fig 20E). These data suggest that KNL2$^{456–516\ (CC\text{-}upstream)}$ may facilitate the CNEP-A-KNL2 complex binding by some unknown mechanisms, but without directly binding to the CENP-A nucleosome.

2.2.4 Discussion

To sum up, KNL2$^{518–560}$, the KNL2 fragment with CENP-C-like motif only, is the core region which is sufficient for CENP-A specific binding without the existence of any other motifs, while this stable binding can be facilitated by the CC upstream region of KNL2$^{456-516}$ for the more stable interaction. Studying how the ggKNL2 upstream region facilitates the interaction of CENP-C-like motif with the CENP-A nucleosome is a topic that is needed to address in the future. In addition, although DT40 cells without SANT and SANTA structural domains are still viable, cKO-KNL2 cells expressing KNL2 without SANT or SANTA structural domains grow slower than the expressing of WT KNL2 in previous study[99], suggesting that these structural domains may contribute to proper KNL2 function to some extent. This topic is also needed to study.

2.3 Materials and methods

Recombinant protein expression

Recombinant expression plasmid: The selected cDNA fragments encoding the C-terminal fragments of chicken KNL2 were each inserted into a pGEX-6p bacterial expression vector (Cytiva). Each fragment is tagged by both GST tag and FLAG-tag.

cDNA fragments: C-terminal fragments of chicken KNL2 (aa 347–560, KNL2$^{347–560}$; aa 456–560, KNL2$^{456–560}$; aa 518-560, KNL2$^{518–560}$; aa 347–455, KNL2$^{347–455}$; aa 456–516, KNL2$^{456–516}$) (Fig 17)

Tag: glutathione S transferase (GST) tag (N-terminus); FLAG tag (C-terminus)

Expression: Each KNL2 fragment was expressed in E. coli (Escherichia coli) RosettaTM2 (DE3) as a GST-fused recombinant protein. E. coli cells were grown in LB media at 37°C until OD600 reached 0.6, then the protein expression was induced by addition of isopropyl-b-D-thiogalactoside (IPTG) to a final concentration of 0.2 mM, and culture was continued at 17°C overnight or 37°C for 2 h for each KNL2 fragments.

Host cell: E. coli RosettaTM2 (DE3) (Novagen)

Media: LB-Lennox media

Temperature: 37 ° C for cell growth; 17 ° C (KNL2 fragments expression: GST-KNL2$^{347 – 560}$ and GST-KNL2$^{347 – 455}$) or 37 ° C (KNL2 fragments expression: GST-KNL2$^{456–560}$; GST-KNL2$^{518-560}$; GST-KNL2$^{456–516}$) for protein expression.

Recombinant protein purification

Sample collection: The GST-binding buffer was used to suspend cell pellet, which was lysed by sonication on ice. Then the lysate was centrifuged for clarification.

GST-binding buffer: 20 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol and 2 mM Dithiothreitol [DTT], supplemented with protease inhibitor cocktail (Roche).

GST affinity purification: The clarified supernatant was applied to glutathione Sepharose FF (Cytiva) and was pre-equilibrated with GST-binding buffer. After extensive column washing with GST-binding buffer, the immobilized GST-fusion

protein was eluted from the column with GST elution buffer.

Beads: glutathione Sepharose FF (Cytiva)

GST elution buffer: 50 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol and 10 mM reduced glutathione

ION exchange column chromatography purification: Base on the biochemical properties of the aim KNL2 fragment, applying the eluted fractions to a cation exchange column or a anion exchange column. Each GST-KNL2 fragment was eluted with a linear NaCl gradient from 70 to 700 mM. Peak fractions containing the GST-KNL2 fragment were combined and concentrated to 5 ml.

ION exchange column: HiTrap SP HP cation exchange column (Cytiva) for GST-KNL2$^{347–560}$ and GST-KNL2$^{347–455}$; Resource Q anion exchange column (Cytiva) for GST-KNL2$^{456–560}$; GST-KNL2$^{518-560}$; GST-KNL2$^{456–516}$

A Buffer: 20 mM HEPES pH 7.5, 1 M NaCl, 5% glycerol and 2 mM DTT.

B Buffer: 20 mM HEPES pH 7.5, 5% glycerol and 2 mM DTT.

Size-exclusion column chromatography purification: Further, to purify the KNL2 fragments with size-exclusion column in a buffer containing 300 mM NaCl. Peak fractions containing the GST-KNL2 fragment were combined, concentrated (typically to 35 mg/ml), and stored at 80°C until further use for biochemical analyses.

Size-exclusion column: HiLoad 16/60 Superdex 200 pg column (Cytiva)

Buffer: 20 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol and 2 mM DTT.

Chimeric chicken CENP-A (H31-63-CAa1-end) expression[91]

Recombinant expression plasmid: pET15b-H31-63-CAa1-end[92]

Chimeric chicken CENP-A (H31-63-CAa1-end) composition: Removing N-terminal tail (aa 1–53) from chicken CENP-A and replacing it with canonical H3 (aa 1–63 in human H3).

Tag: His-tag (N-terminus)

Expression: His-tag chimeric chicken CENP-A (H31-63-CAa1-end) was expressed in E. coli which was transformed with the pET15b-H31-63-CAa1-end plasmid[92] and was produced as inclusion body in bacteria cells. The cells are cultured in LB media overnight at 37°C.

Host cell: E-coli strain BL21 (DE3)

Media: LB-Lennox media

Temperature: 37°C

Chimeric chicken CENP-A (H31-63-CAa1-end) purification

Sample collection: Cells were resuspended in lysis buffer and were lysed using sonication. After centrifugation (33,000 g for 20 min), pellet was solubilized in unfolding buffer with overnight incubation at 37°C.

Lysis buffer: 50 mM Tris pH 7.5, 0.5 M NaCl, 5% glycerol, supplemented with protease inhibitor cocktail (Roche)

Unfolding buffer: 50 mM Tris pH 7.5, 7 M guanidine-HCl, and 5% glycerol

Ni-NTA affinity purification: The unfolding sample was then applied to a Ni-NTA agarose column equilibrated with His-binding buffer. After extensive wash of the column with His-binding buffer containing 50 mM imidazole, the immobilized proteins were eluted with His-elution buffer.

Beads: Ni-NTA agarose (Qiagen)

His-binding buffer: 50 mM Tris pH 7.5, 0.5 M NaCl, 6 M urea, 10 mM imidazole

His-elution buffer: 50 mM Tris pH 7.5, 0.5 M NaCl, 6 M urea, 300 mM imidazole

Dialysis and His-tag cleavage: The eluted sample was dialyzed against dialysis buffer and treated with thrombin protease overnight at 4°C to remove the His-tag.

Dialysis buffer: 10 mM Tris buffer (pH 7.5), 5 mM DTT

Protease: thrombin protease (Wako; 2 U thrombin/1 mg Histone)

ION exchange column chromatography purification: Untagged H31–63-CAa1-end was further purified by ion exchange column chromatography using a HiTrap SP column with a linear gradient buffer of 200–900 mM NaCl.

ION exchange column: HiTrap SP column (GE Healthcare)

A Buffer: 1 M NaCl, 20 mM sodium acetate buffer (pH 5.2) , 1 mM EDTA, 6 M urea, and 5 mM DTT

B Buffer: 20 mM sodium acetate buffer (pH 5.2) , 1 mM EDTA, 6 M urea, and 5 mM DTT

Dialysis: The peak fractions were combined and dialyzed against water.

Lyophilization: H31 – 63-CAa1-end was lyophilized using a vacuum centrifugal concentrator (TOMY).

Histone H3.2 expression[92]

Recombinant expression plasmid: pHCE-AMPFREE-gH3.2 (from pHCE-AMPFREE (Takara))

Tag: His-tag

Expression: His-tagged chicken histone H3.2 was expressed using pHCE-AMPFREE-gH3 in E. coli (BL21 (DE3)). (Refer to Chimeric CENP-A above)

Host cell: E-coli strain BL21 (DE3)

Media: LB-Lennox media

Temperature: 37°C

Sample collection: Cells expressing H3.2-His were resuspended in column buffer A and lysed by sonication on ice. The cell lysates were clarified by centrifugation (35,000 ×g) at 4°C. The pellet was resuspended with the buffer A and lysed by sonication on ice. Then the lysates were clarified by centrifugation (35,000 ×g) at 4°C, and the supernatants were removed. The pellet was resuspended with buffer B and lysed by sonication on ice. The lysate was incubated with agitation at 4°C overnight. The lysate was clarified by centrifugation (35,000 ×g) at 4°C and the pellet was

removed.

Lysis buffer A: 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% Glycerol, 5 mM DTT, and 1× complete EDTA-free proteinase inhibitor [Roche]

Lysis buffer B: 50 mM Tris-HCl, pH 8.0, 7 M Guanidine-HCl, 5% Glycerol, and 5 mM DTT

Histone H3.2 purification (Refer to Chimeric CENP-A above)

Histone H4 expression

Plasmid: pET15-b-gH4

Tag: His-tag

Expression: His-tagged chicken histone H4 was expressed using pET15-b-gH4 in E. coli (BL21 (DE3)). (Refer to Chimeric CENP-A above)

Host cell: E-coli strain BL21 (DE3)

Media: LB-Lennox media

Temperature: 37°C

Histone H4 purification

Sample collection: Cells expressing H4-His were resuspended in column buffer A and lysed by sonication on ice. The cell lysates were clarified by centrifugation (35,000 ×g) at 4°C. The pellet was resuspended with the buffer A and lysed by sonication on ice. Then the lysates were clarified by centrifugation (35,000 ×g) at 4°C, and the supernatants were removed. The pellet was resuspended with buffer B and lysed by sonication on ice. The lysate was incubated with agitation at 4°C overnight. The lysate was clarified by centrifugation (35,000 ×g) at 4°C and the pellet was removed.

Lysis buffer A: 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% Glycerol, 5 mM DTT, and 1× complete EDTA-free proteinase inhibitor [Roche]

Lysis buffer B: 50 mM Tris-HCl, pH 8.0, 7 M Guanidine-HCl, 5% Glycerol, and 5

mM DTT

Ni-NTA affinity purification: The supernatant was incubated with Ni-NTA agarose for 30 min at 4 °C. Ni-NTA agarose was washed with washing buffer four times. Proteins were eluted with elution buffer.

Beads: Ni-NTA agarose (Qiagen)

Washing buffer: 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 6 M Urea, 5% Glycerol, and 5 mM Imidazole, pH 8.0

Elution buffer: 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 6 M Urea, 5% Glycerol, 300 mM Imidazole, pH 8.0, and 1 mM DTT

Dialysis and His-tag cleavage: The eluted sample was dialyzed with dialysis buffer. His-tag of proteins were cleaved with Thrombin for 75 min at RT.

Dialysis buffer: 10 mM Tris-HCl, pH 7.5, and 5 mM DTT

Protease: thrombin protease (Wako; 2 U thrombin/1 mg Histone)

ION exchange column chromatography purification: The cleaved proteins were applied to 5 ml HisTrap HP column in salt gradient with a linear gradient of imidazole from 20 to 500 mM. The peak fractions of His-gH4 were dialyzed against dialysis buffer and applied to a HiTrap SP column (GE Healthcare).

ION exchange column: HisTrap HP column (GE Healthcare) and HiTrap SP column (GE Healthcare)

Dialysis buffer: 10 mM Tris buffer (pH 7.5) containing 5 mM DTT

Dialysis: The peak fractions were combined and dialyzed against water four times.

Lyophilization: H4 was lyophilized using a vacuum centrifugal concentrator (TOMY) and stocked as powder at 4°C.

<u>CA/H4 tetramer preparation</u>

Sample refolding: The CENP-A (H31-63-CAa1-end) and His-H4 powders were suspended in unfolding buffer, then mixed at 1:1 molar ratio in refolding buffer and assembled as tetramers by three dialysis steps in which NaCl concentration was decreased stepwise to 200 mM.

Unfolding buffer: 20 mM Tris-HCl, pH 7.5, 7 M Guanidine-HCl, and 20 mM DTT

Folding buffer A: 10 mM Tris pH 7.5, 1 M NaCl, 1 mM EDTA and 5 mM DTT

Folding buffer B: 10 mM Tris pH 7.5, 500 mM NaCl, 1 mM EDTA and 5 mM DTT

Folding buffer C: 10 mM Tris pH 7.5, 200 mM NaCl, 1 mM EDTA and 5 mM DTT

CA/H4 tetramer purification: The tetramer was further purified by the size exclusion column in folding buffer C, and concentrated using Amicon Ultra centrifugal filter (Merck). Fractions were pooled and frozen in liquid nitrogen and kept at −80°C.

Size exclusion column: HiLoad 16/60 Superdex 200 pg column (Cytiva)

H3/H4 tetramer preparation (Refer to CA/H4 tetramer above)

Histones H2A and H2B expression

Recombinant expression plasmid: pETDuet-His-SUMO-H2A/H2B (made from a in-house–modified version of pETDuet-1)

Tag: 6xHis-SUMO-tagged H2A (N-terminus)

Expression: To generate H2A/H2B dimer, H2A and H2B were co-expressed in E. coli as a homodimer from a pETDuet-His-SUMO-H2A/H2B plasmid which generates N-terminally 6xHis-SUMO-tagged H2A and untagged H2B.

Host cell: E. coli RosettaTM2 (DE3)

Histones H2A and H2B purification

Sample collection: Cells expressed H2A/H2B dimer were resuspended in high-salt

buffer and disrupted by sonication.

Lysis buffer: 20 mM HEPES pH 7.5, 2 M NaCl, 5% glycerol, and 5 mM tris (2-carboxyethyl) phosphine (TCEP)

Ni-NTA affinity purification: The clarified lysate was applied to a Ni-NTA agarose column. After extensive column washing, bound proteins were eluted from the column with suspension buffer containing 300 mM imidazole.

Beads: Ni-NTA agarose (Qiagen)

His-binding buffer: 50 mM Tris pH 7.5, 0.5 M NaCl, 6 M urea, 10 mM imidazole

His-elution buffer: 50 mM Tris pH 7.5, 0.5 M NaCl, 6 M urea, 300 mM imidazole

His-tag cleavage: The elution was incubated with SENP7 protease to remove the HisSUMO-tag.

Protease: in-house-made SENP7 protease (Fukagawa Lab)

Size exclusion column chromatography purification: The complex was further purified using size exclusion column chromatography in the buffer. The purified H2A/H2B dimer was concentrated using Amicon Ultra centrifugal filter (Merck). Relevant fractions were pooled, frozen in liquid nitrogen, and kept at −80°C.

Size exclusion column: Hi-Load Superdex 16/60 75 pg (GE Healthcare)

Buffer: 10 mM Tris pH 7.5, 200 mM NaCl, 1 mM EDTA and 5 mM DTT

601 DNA purification[119]

Expression: 601 DNA was expressed in E. coli cells which was transformed with the pGEM- 601- 145 bp x 8 plasmid. The cells are cultured in LB media overnight at 37° C.

Plasmid: pGEM- 601- 145 bp x 8

Host cell: E-coli DH5α

Media: LB-Lennox media

Temperature: 37°C

601 DNA preparation

Sample collection: The 145-bp 601containing plasmid was purified with isopropanol and treated with RNase. The plasmid was extracted with phenol twice and further purified with PEG and ethanol precipitation. The plasmid was digested with EcoRV, and the digested 601 DNA was purified with PEG and ethanol precipitation. The 601 DNA was suspended with TE buffer.

Enzyme: EcoRV

Reconstitution of chicken CENP-A nucleosome

H2A/H2B dimers, chicken (CENPA/H4)2 hetero-tetramers and 601 DNA were mixed with a molar ratio of 2:1:0.8 at 0.75 mg/ml of final DNA concentration in the presence of 2 M KCl. A gradient dialysis to low salt buffer (2–0.2 M KCl) was performed at 4° C more than 21 h[120]. Finally, the nucleosome solution was dialysis against final buffer (50 mM KCl). Assembled nucleosomes were then uniquely positioned on the DNA by a thermal shift for 1 h at 37 °C. Nucleosome formation was examined using native polyacrylamide gel electrophoresis (Native PAGE) and DNA was detected via EtBr. The concentration of each nucleosome sample was determined using absorbance at 260 nm.

Molar ratio: H2A/H2B dimers: chicken (CENPA/H4)2 hetero-tetramers: 601 DNA=2:1:0.8

High salt buffer: 2 M KCl in 10 mM Tris pH 7.5, 1 mM EDTA, 5 mM DTT

Low salt buffer: 0.2 M KCl in 10 mM Tris pH 7.5, 1 mM EDTA, 5 mM DTT

Final salt buffer: 50 mM KCl in 10 mM Tris pH 7.5, 1 mM EDTA, 5 mM DTT

Reconstitution of H3 nucleosome (Refer to CENP-A nucleosome above)

Electrophoretic mobility shift assay (EMSA)

Each GST-KNL2 fragments was mixed with CENP-A or H3 nucleosome in binding buffer for 60 min on ice. The mixtures were analyzed by Native PAGE using

SuperSep 5–20% gels. Native PAGE was typically performed at constant 200 V for 100 min at 4°C in 0.5xTAE buffer. The gels were stained with GelRedTM (Biotium, Inc) and visualized by UV illumination at 260 nm.

Binding buffer: 20 mM HEPES pH 7.5, 1 mM EDTA, 100 mM NaCl, 5 mM TCEP, 5% glycerol, and supplemented with protease inhibitor cocktail [Roche].

Gel: SuperSep 5–20% gels (FUJIFILM WAKO Chemicals)

Biolayer interferometry (BLI) assays

The association and dissociation between each GST-KNL2 fragment and CENP-A nucleosome was measured using the BLItz biolayer interferometer. Each GST-fused KNL2 fragment was loaded onto the anti-GST biosensors (Forte Bio) as follows: Biosensor probes were hydrated in BLI-binding buffer for 10 min followed by loading 0.6 μM of each GST-fused KNL2 fragment (ligand) and were incubated for 2 min. The ligand-bound sensor tips were washed with the BLI-binding buffer for 2 min. The CENP-A nucleosome (analyte) dissolved in BLI binding buffer was loaded onto the GST-fused KNL2 fragment sensor tips at each concentration indicated in the BLI figures. The association phase (2 min) was measured and followed by a dissociation phase (2 min) in BLI-binding buffer. All the BLI measurements were performed at room temperature. Sensorgrams were analyzed by BIAevaluation 4.1 (BIACORE) to perform steady-state affinity fitting to estimate apparent equilibrium dissociation constant.

BLI-binding buffer: 20 mM HEPES-NaOH (pH 7.5), 200 mM NaCl, 2 mM TCEP, 1 mM EDTA, 0.5% CHAPS

Equipment: BLItz biolayer interferometer (Forte Bio, Fremont, CA, USA)

Chapter 3: Cryo-EM structural analysis reveals the specific binding mechanism of the CENP-A nucleosome to the CENP-C-like motif of ggKNL2

3.1 Results and discussion

Since the KNL2 CENP-C-like motif region is sufficient for CENP-A specific binding, a synthetic KNL2 peptide with amino acid residues from 517 to 560 was used for further analyses of molecular mechanism of binding of ggKNL2 to the CENP-A nuclesome (Fig 22). To determine the specific binding residues for binding of ggKNL2 to the CENP-A nucleosome, considering properties such as molecular size, the cryo-EM single-particle analysis is a most suitable method (Fig 23; Table 1).

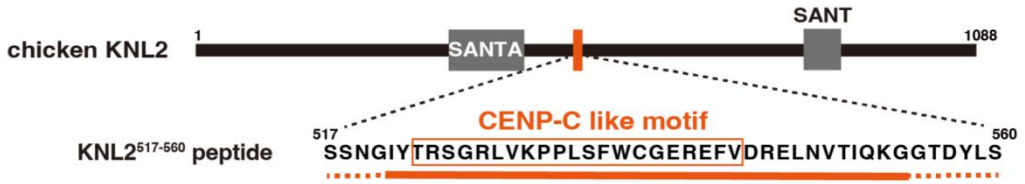


Figure 22. Schematic diagram showing functional elements in chicken KNL2 (ggKNL2).

Amino acid sequence of the chicken KNL2$^{517-560}$ peptide (aa 517-560) used for cryo-EM analysis is presented. The region observed in Cryo-EM structure is underlined (aa 521-554). The conserved CENP-C-like motif is enclosed in the orange box.

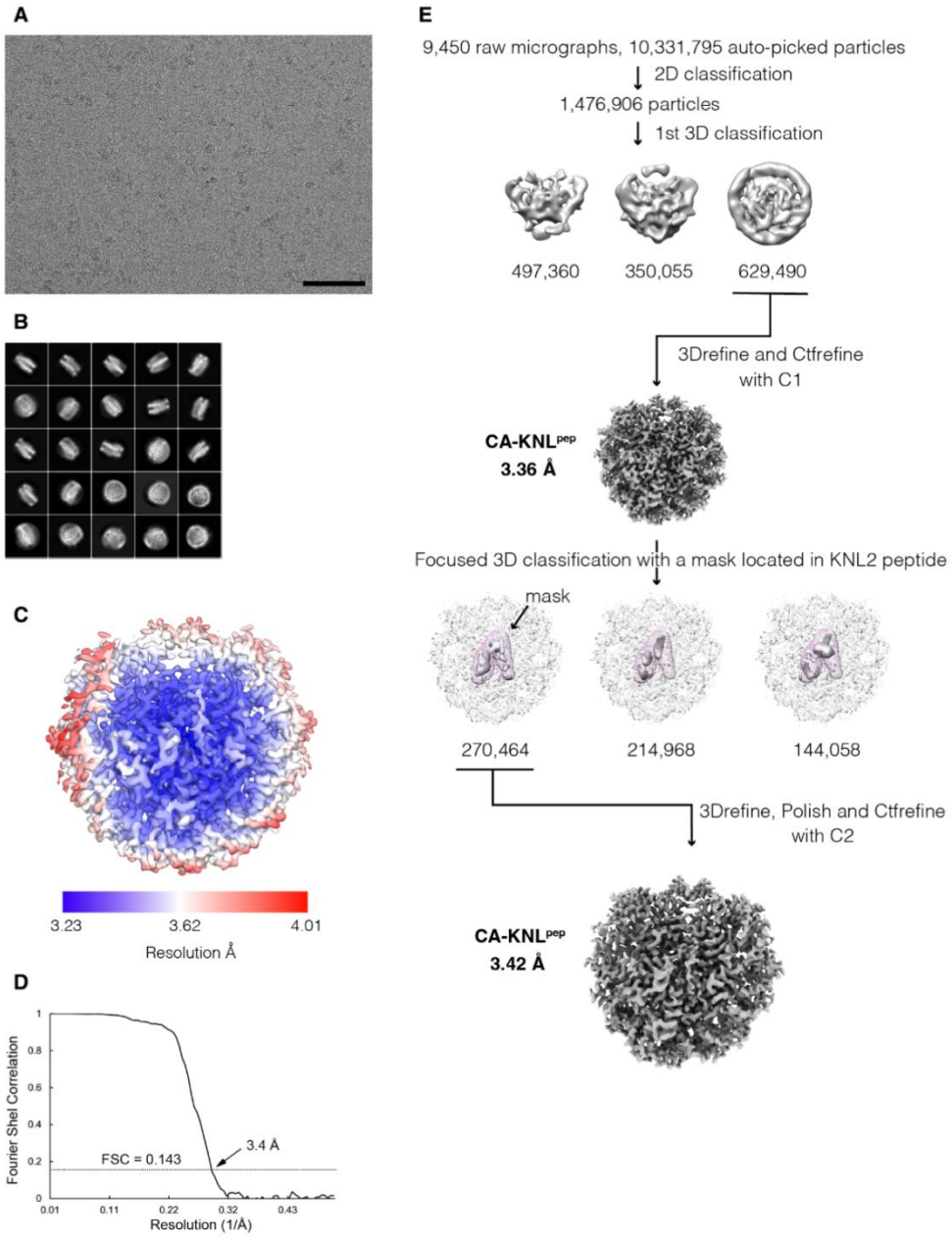


Figure 23. Cryo-EM single-particle image analysis of the CENP-A nucleosome–KNL2$^{517–560}$ complex (related to Fig 24).

A A representative micrograph of the CENP-A nucleosome in complex with the ggKNL2$^{517–560}$ peptide. Scale bar indicates 100 nm.

B Representative 2D class averages of the CENP-A nucleosome – KNL2$^{517-560}$ complex.

C Cryo-EM density map of the chicken CA-KNL2$^{517-560}$ complex at a 3.4 Å resolution.

colored according to the local resolution estimated by RELION.

D Gold-standard Fourier shell correlation (FSC) curve of the Cryo-EM density map is displayed. Reported resolution (3.42 Å) was based on the FSC = 0.143 criterion indicated by the dot line.

E Flow chart showing the image processing pipeline for the cryo-EM single-particle image analysis of the chicken CENP-A nucelsome-KNL2$^{517-560}$ complex.

In vitro, the synthetic KNL2 peptide is mixed with the reconstituted CENP-A nucleosome at certain molar ratio (5:1) to form a complex. After performing cryo-EM single-particle analysis, the CENP-A-KNL2$^{517-560}$ complex is reconstructed at three-dimensional (3D) level at the 3.42 Å resolution (Fig 23; Tab 1). The cryo-EM map reveals that each side of the CENP-A nucleosome, as a flat plate structure wrapped by the 601 DNA, is associated with the KNL2 peptide residues from 521 to 554 with a stable binding interface (Fig 22, 24A and 25).

Table 1. Statistics for cryo-EM single particle image analysis and structure refinement.

Cryo-EM density map	KNL2_CC—CA nucleosome (3.42 Å) (EMD-33666) (PDB 7Y7I)
Data collection and processing	
Sample	KNL2_CENP-C like motif peptide bound to the CENP-A nucleosome
Magnification	50,000
Voltage (kV)	300
Electron exposure (e^{-}/Å^2)	40
Defocus range (μm)	−1.0 to −2.5
Pixel size (Å)	0.98
Symmetry	C2
Initial particle images (no.)	10,331,795
Final particle images (no.)	270,464
Map resolution (Å)	3.42
FSC threshold	0.143
Refinement	
CCmap_model	0.82
Model compositions	
Nonhydrogen atoms	12,447
Protein residues	824
Nucleic acids	286
R.m.s. deviations	
Bond lengths (Å)	0.004
Angles (°)	0.577
Ramachandran plot	
Favored (%)	98.13
Allowed (%)	1.87
Outliers (%)	0.00
Rotamer outliers (%)	0.29
Clash score	9.14
Cβ outliers (%)	0.00
CaBLAM outliers (%)	1.28

FSC, Fourier shell correlation; R.m.s. deviations, root-mean-square deviations.

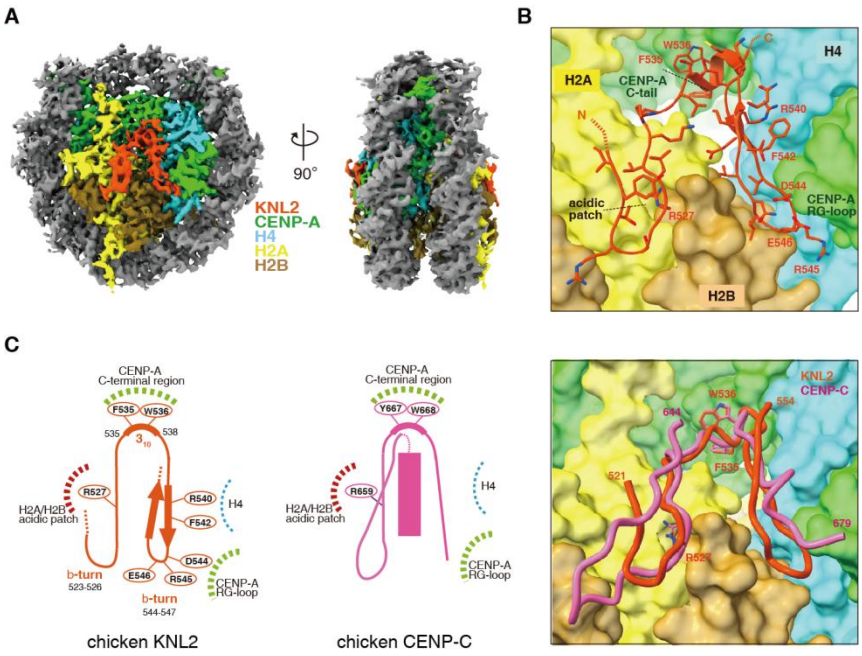


Figure 24. Cryo-EM structure of the CENP-C-like motif of ggKNL2 bound to the CENP-A nucleosome.

A Cryo-EM density map of the CENP-A nucleosome in complex with the KNL2$^{517-560}$ peptide at 3.42 Å resolution. The side views of the complex along the two-fold axis are shown. The density corresponding to each molecule in the complex is color coded as indicated in the fig.

B The structure of the KNL2$^{517-560}$ peptide bound to the CENP-A nucleosome. The surface model of the CENP-A nucleosome was color coded as indicated in (A). The key residues to CENP-A nucleosome binding are labeled.

C Structural comparison of the CENP-A nucleosome binding between KNL2 and CENP-C. Left panel, schematic representation of the KNL2$^{517-560}$ peptide structure. Dot circles indicate the recognition sites for the H2A/H2B acidic patch, the CENP-A C-terminal region, and the CENP-A RG-loop. The key residues in each site are indicated. Middle panel, schematic representation of the structure of chicken CENP-C (ggCENP-C) bound to CENP-A nucleosome. Right panel, structural comparison of the CENP-A nucleosome-binding interface between ggKNL2 and ggCENP-C. The backbone structures of ggKNL2 (residues 521–554, orange) and ggCENP-C (residues 642-679, PDB ID: 7BY0, magenta) are superimposed. The CENP-A nucleosome is presented in a surface model.

In total, the KNL2$^{517-560}$ peptide binds to the CENP-A nucleosome at three regions. Firstly, the conserved arginine residue R527 of KNL2$^{517-560}$ binds to the H2A/H2B acidic patch (Fig 24B, C; Fig 26A). Next, the conserved ggKNL2 phenylalanine residue F535 and tryptophan residue W536, locating in a 3_{10} helix, recognize the C-terminal region of CENP-A (Fig 24B, C; Fig 27A). The third one is, aspartic acid residue D544 of KNL2 directly binds to the CENP-A specific RG-loop (Fig 24B, C; Fig 28A). I experimentally confirmed these residues are involved in binding to the CENP-A nucleosome using specific mutant KNL2. I also used single or double binding site mutants of the chicken KNL2 for cell biology experiments.

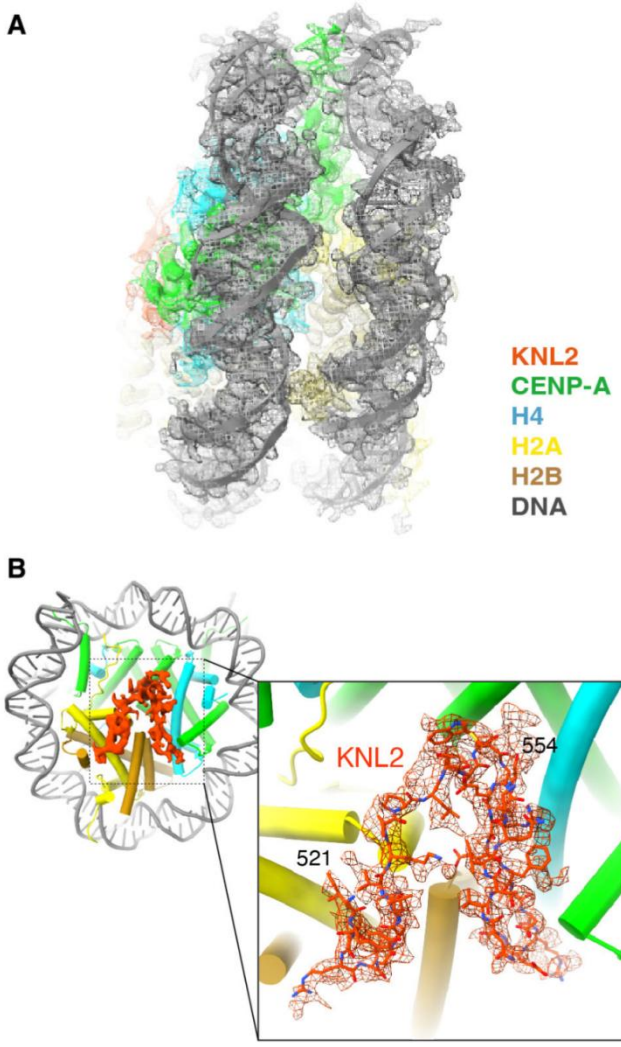


Figure 25. Cryo-EM structure of the CENP-A nucleosome in complex with KNL2$^{517-560}$ (related to Fig 24).

A Cryo-EM density map of the CENP-A nucleosome in complex with chicken KNL2$^{517-560}$ peptide at a 3.42 Å resolution. The EM map for the entire complex is shown as mesh representation together with the ribbon representation of the complex molecules color-coded as indicated in the fig.

B Cryo-EM density map for the KNL2$^{517-560}$ peptide is shown as an orange surface representation together with the cartoon model of the CENP-A nucleosome. An enlarged view of the density map (mesh representation) is presented with the final structure of the KNL2$^{517-560}$ peptide (stick model) in a panel. The final model contains KNL2 peptide residues from 521 to 554.

3.1.1 The conserved arginine residue R527 of KNL2$^{517–560}$ N-terminal region binds to the H2A/H2B acidic patch

The well-defined CENP-A-KNL2$^{517-560}$ complex structure is analyzed in details. The acidic patch of histone H2A/H2B is noticed to be recognized by ggKNL2 within the CENP-C-like motif (aa 523-543) region. The exact binding sites on the acidic patch are aspartic residue D90 and glutamic residue E91 in histone H2A and the ggKNL2 arginine 527 (R527) residue (Fig 26A). To examine the binding contribution of ggKNL2 R527, R527 was mutated to alanine and a recombinant KNL2$^{518–560}$ mutant (KNL2$^{518–560\ R527A}$) fragment was prepared. Then, I examined for its binding affinity to the CENP-A nucleosome, using EMSA and BLI. The WT KNL2$^{518–560}$ was used for binding experiment to the CENP-A nucleosome as a control. A clear band shift can be observed as a binding signal of CENP-A-KNL2$^{518–560}$ complex (Fig 26B). Comparing with the shift band of complex containing wild-type KNL2$^{518–560}$, the KNL2$^{518–560\ R527A}$ mutant did not shown any band shift, but only the CENP-A nucleosome band was observed. This result indicates that the R527A mutation interferes the complex formation between the KNL2 fragment and the CENP-A nucleosome. Further evidence is provided by the BLI assay. Comparing with the KNL2$^{518-560\ WT}$, BLI assay using the KNL2$^{518-560\ R527A}$ mutant fragment and CENP-A nucleosome showed significantly lower wavelength shifts, which indicated the same conclusion as that of

EMSA results (Fig 20C, Fig 26C).

As a conclusion, both the EMSA and BLI assay results indicated that the wild-type ggKNL2 directly recognizes the CENP-A nucleosome.

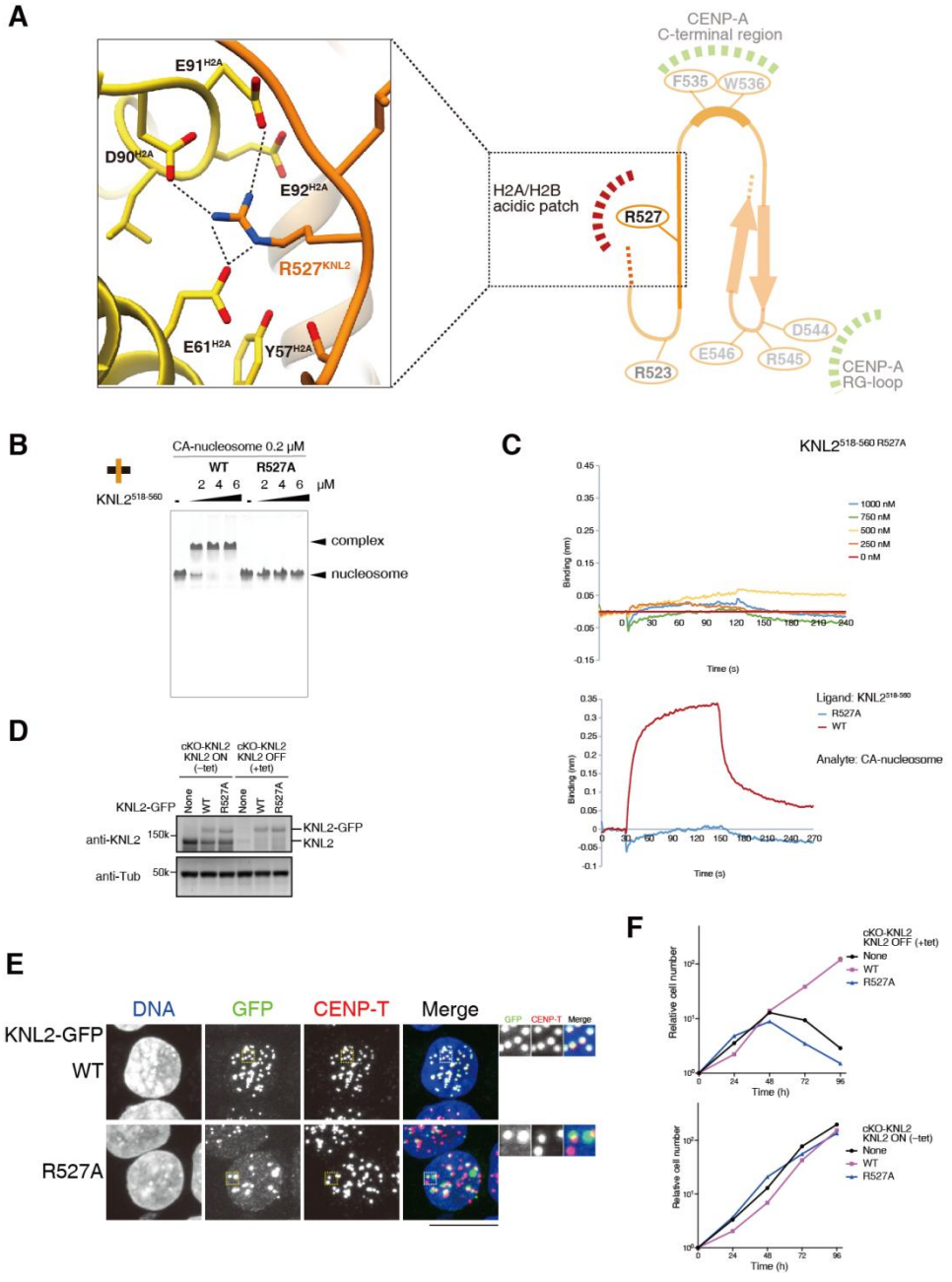


Figure 26. ggKNL2 recognizes the acidic patch of histone H2A/H2B.

A The magnified view of the H2A/H2B binding site of KNL2. R527 of ggKNL2 forms hydrogen bonds with the H2A acidic residues. Side chains of the key residues

are indicated as a stick model. Dotted lines represent the possible hydrogen bonds (< 3.0 Å).

B EMSA to examine contribution of the ggKNL2 R527 residue to the CENP-A nucleosome binding.

C Biolayer interferometry analyses of the CENP-A nucleosome binding of GST-KNL2$^{518\text{-}560\ WT}$ or GST-KNL2$^{518\text{-}560\ R527}$ mutant. Each GST-fused KNL2 fragment was immobilized onto anti-GST prob as ligand, and CENP-A nucleosome is analyte. The BLI sensorgrams were measured at association time intervals of 120 s and dissociation time intervals of 120 s.

D Expression of the WT or R527A mutant GFP-KNL2 in KNL2 conditional knock out chicken DT40 cells (cKO-KNL2). Expression of full-length KNL2 WT was conditionally turned off by tetracycline (tet) addition in cKO-KNL2 cells (None). GFP-fused KNL2 WT or R527A mutant was stably expressed in the cKO-KNL2 cells. These cells were cultured in the presence or absence of tet (+tet: KNL2 OFF or －tet: KNL2 ON) for 48 h. α-Tubulin (Tub) was probed as a loading control.

E Localization analyses of GFP-fused KNL2 WT or R527A mutant (green). CENP-T was stained as a kinetochore marker (red), and DNA was stained using DAPI (blue). Magnified views are shown in insets. Scale bar indicates 10 μm.

F Growth of the cKO-KNL2 cells expressing GFP-fused KNL2 WT or R527A mutant. The upper panel shows examined cell numbers at the indicated time after tet addition (+tet: KNL2 OFF). The cell numbers were normalized to those at 0 h for each line and plotted as relative cell number. The lower panel shows examined untreated cells (－tet: KNL2 ON). Parental cKO-KNL2 chicken DT40 cells (None) were also examined.

3.1.2 The conserved ggKNL2 phenylalanine residue F535 and tryptophan residue W536, locating in a 3_{10} helix, recognize the C-terminal region of CENP-A

In the downstream region of KNL2 R527, two other residues F535 and W536 also bind to the CENP-A nucleosome in the region named CENP-A C-terminal, which is one of the unique structural features of CENP-A (Fig 24B and Fig 27A). The continuous cryoEM map at a higher resolution revealed a two-stranded anti-parallel β -sheet structure in the C-terminal region of $KNL2^{517-560}$ (Fig 25B). Also, a stacking interaction was observed between $KNL2^{W356}$ side chain and the CENP-A R132 side chain (Fig 27A). Same as $KNL2^{R527}$, the function and contribution of the F535 and W536 residues for the CENP-A-KNL2 complex interaction was also required to be evaluated. F535A/W536A double mutations are prepared for binding assay to the CENP-A nucleosome. Again, no binding shift was observed from the EMSA data (Fig 27B), which is consistent with the low binding response from the BLI data (Fig 27C).

Therefore, it is significant that F535 and W536 are essential residues for the KNL2 binding to the CENP-A nucleosome.

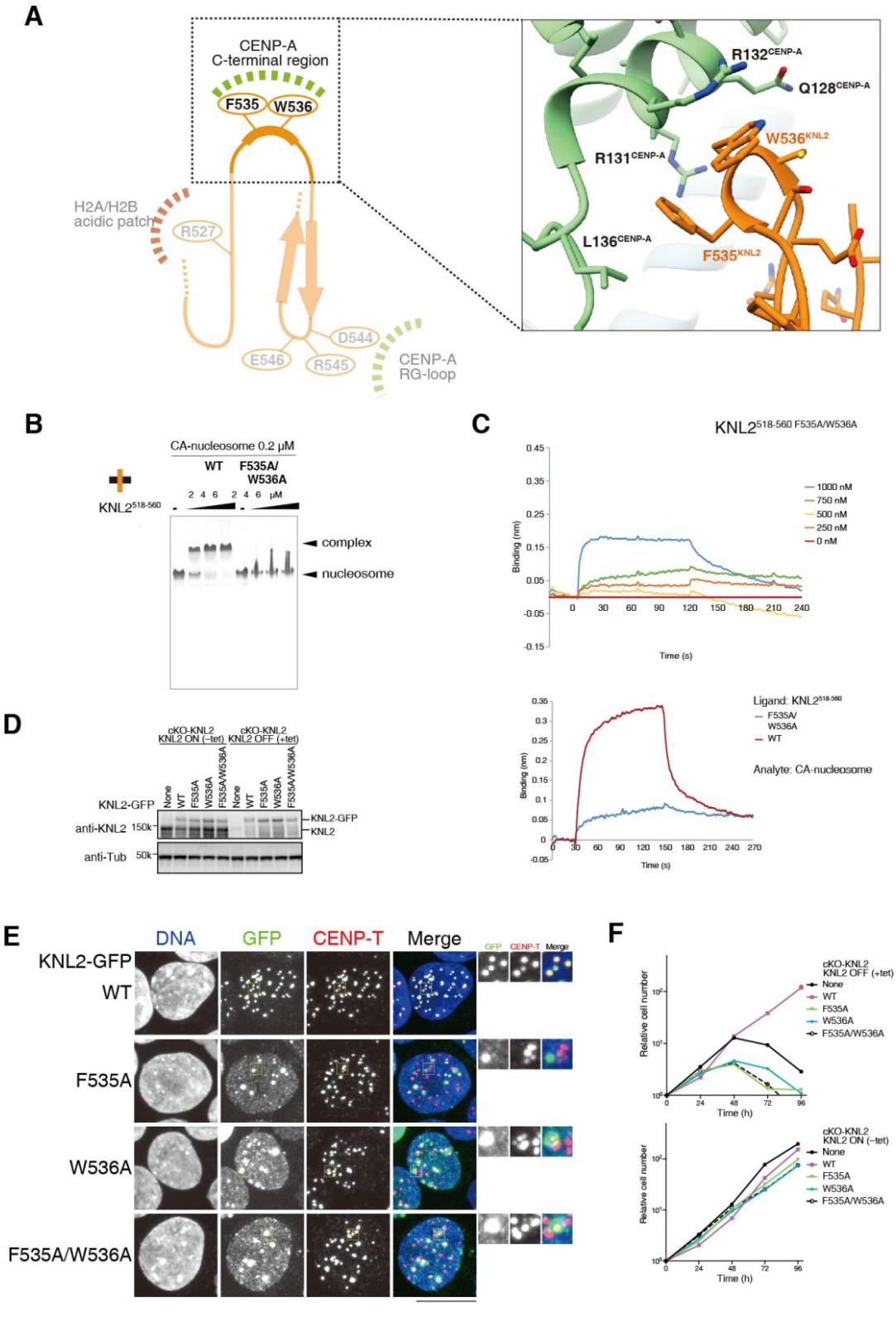


Figure 27. ggKNL2 recognizes the C-terminal region of CENP-A.

A Magnified view of the interface between the CENP-A C-terminal tail and the KNL2$^{517-560}$ peptide. Side chains of the key residues are indicated as a stick model.

B EMSA to examine contribution of the F535 and W536 residues in ggKNL2 to the CENP-A nucleosome binding.

C Biolayer interferometry analyses of the CENP-A nucleosome binding of GST-KNL2$^{518\text{-}560\ WT}$ or GST-KNL2$^{518\text{-}560\ F535/W536}$ mutant. Each GST-fused KNL2 fragment was immobilized onto anti-GST prob as ligand, and CENP-A nucleosome is analyte. The BLI sensorgrams were measured at association time intervals of 120 s and dissociation time intervals of 120 s.

D Expression of GFP-KNL2 WT or CENP-A C-terminal tail-binding mutants in cKO-KNL2 cells. Expression of full-length KNL2 WT was conditionally turned off by tetracycline (tet) addition to the cKO-KNL2 cells (None). GFP-fused KNL2 WT, F535A, W536A, or F535A/W536A was stably expressed in the cKO-KNL2 cells. These cells were cultured in the presence or absence of tet (+tet: KNL2 OFF or tet: KNL2 ON) for 48 h. α-Tubulin (Tub) was probed as a loading control.

E Localization analyses of GFP-fused KNL2 WT and F535A, W536AorF535A/W536A mutant (green). CENP-T was stained as a kinetochore marker (red), and DNA was stained using DAPI (blue). Magnified views are shown in insets. Scale bar indicates 10 lm.

F Growth of the cKO-KNL2 cells expressing GFP-fused KNL2 WT or each CENP-A C-terminal tail binding mutant. The upper panel shows examined cell numbers at the indicated time after tet addition (+tet: KNL2 OFF). The cell numbers were normalized to those at 0 h for each line and plotted as relative cell number. The lower panel shows examined untreated cells (tet: KNL2 ON). Parental cKO-KNL2 chicken DT40 cells (None) were also examined.

3.1.3 A ggKNL2 C-terminal region Aspartic acid residue D544 directly binds to the CENP-A specific RG-loop

Actually, the histone H3 variant CENP-A contains two unique structural features, one is the CENP-A C-terminal region as mentioned above, another one is the RG-loop in

the L1 loop between helices α1 and α2[30,121] (Fig 18). Examining cryo-EM structure of the complex, the KNL2 C-terminal part forms a β-sheet and the KNL2 amino acid residues from 544 to 547 within this β-sheet is shown as a β-turn region which was shown to be located in the vicinity of the CENP-A RG-loop (Fig 24B and C, left panel). β-sheet formation involving the CENP-C-like motif downstream region is a structural feature critical for CENP-A RG-loop recognition. Within the β-turn region, the side chain of KNL2 D544 forms a H-bond interaction (2.5 Å) with the side chain of R80 in the CENP-A RG-loop (Fig 28A). Same as that of other key KNL2 binding residues with CENP-A nucleosome, EMSA and BLI assay were also performed for the KNL2$^{517-560}$ D544A mutant fragments to examine the CENP-A binding ability. As expected, neither the EMSA nor the BLI assay result showed binding of the KNL2 mutant to the CENP-A nucleosome (Fig 28B and C).

Therefore, the ggKNL2 D544 residue is essential for the CENP-A RG-loop recognition to form the CENP-A nucleosome-KNL2.

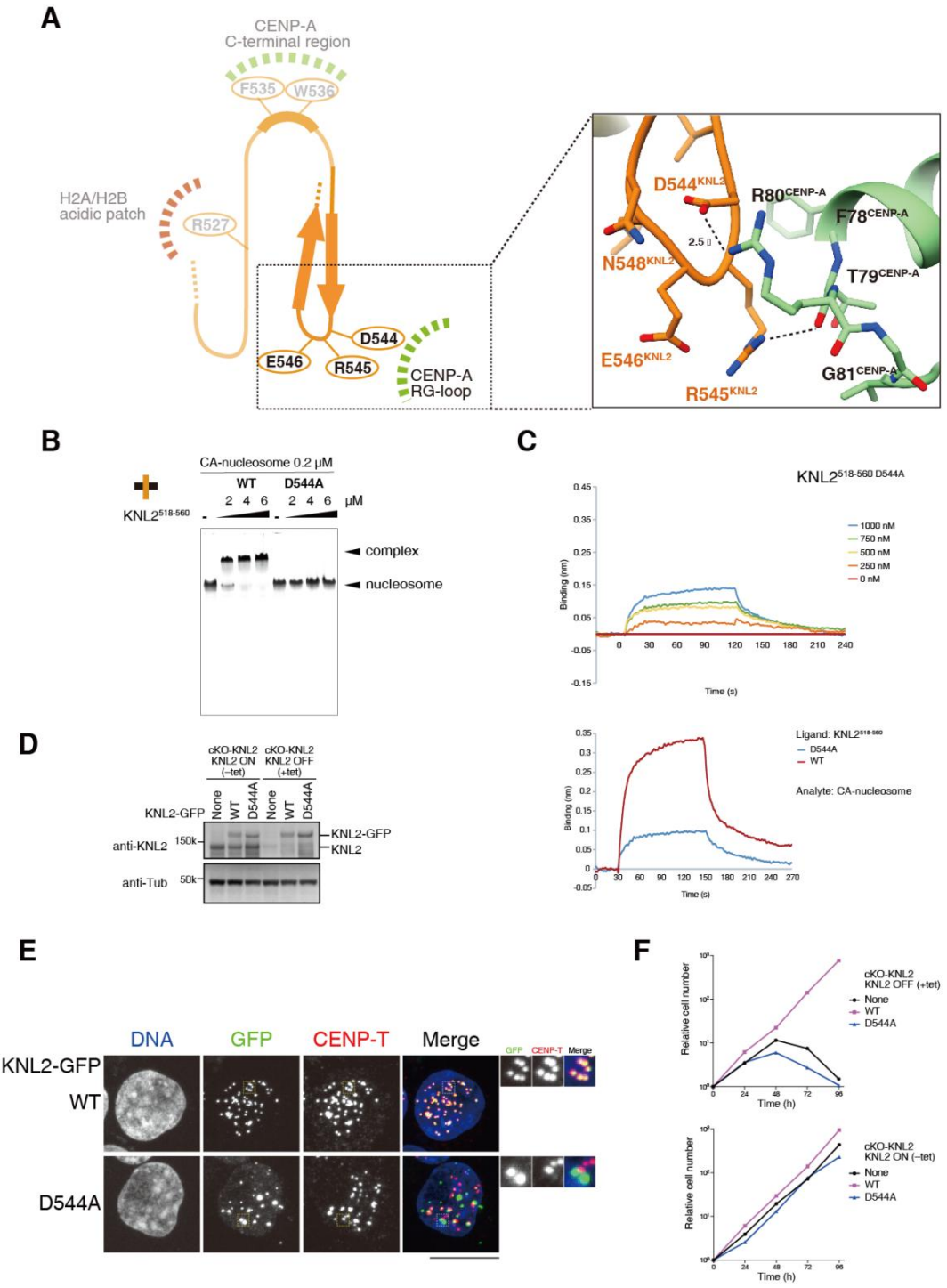


Figure 28. The D544 residue of ggKNL2 associates with the RG-loop of the CENP-A nucleosome.

A Magnified view of the interface between the CENP-A RG-loop and ggKNL2. The b-turn structure of KNL2 provided the interface with the CENP-A RG-loop. Side chains of the key residues are indicated as a stick model. Dotted lines represent the possible hydrogen bonds (< 3.0 A).

B EMSA to examine the binding affinities of WT KNL2$^{518-560}$ or KNL2$^{518-560\ D544A}$ mutant to the CENP-A nucleosome.

C Biolayer interferometry analyses of the CENP-A nucleosome binding of GST-KNL2$^{518\text{-}560\ WT}$ or GST-KNL2$^{518\text{-}560\ D544}$ mutant. Each GST-fused KNL2 fragment was immobilized onto anti-GST prob as ligand, and CENP-A nucleosome is analyte. The BLI sensorgrams were measured at association time intervals of 120 s and dissociation time intervals of 120 s.

D The expression of WT ggKNL2 was conditionally turned off by tet addition in cKO-KNL2 cells (None). GFP-fused WT ggKNL2 and its D544A mutant were stably expressed in cKO-ggKNL2 cells. These cells were cultured in the presence or absence of tet (+tet: KNL2 OFF or -tet: KNL2 ON) for 48 h. a-Tubulin (Tub) was probed as a loading control.

E Localization of GFP-fused KNL2 WT and D544A mutant (green). CENP-T was stained as a kinetochore marker (red), and DNA was stained using DAPI (blue). Magnified views are shown in insets. Scale bar indicates 10 μm. GFP and CENP-T signals on kinetochores in mitotic cells were quantified in each cell line.

F Growth of the cKO-KNL2 cells expressing GFP-fused KNL2 WT and D544A mutant. The upper panel shows examined cell numbers at the indicated time after tet addition (+tet: KNL2 OFF) and were normalized to those at 0 h for each line. The lower panel shows examined untreated cells (tet: KNL2 ON). Parental cKO-KNL2 cells (none) were also examined.

3.1.4 Cryo-EM Structural comparison of the CENP-A nucleosome binding between KNL2 and CENP-C

As mentioned above, chicken KNL2 CENP-C-like motif shows sequence similarity to the CENP-C motif of chicken and human CENP-C, and these region were verified to bind to the CENP-A nucleosome by cryo-EM single particle analysis[91,122]. A consequent question is whether CENP-C motif and CENP-C-like motif share the

same CENP-A binding sites. In this study, the structure of CENP-A-ggKNL2 complex (PDB: 7Y7I) was compared with the structures of two CENP-A-CENP-C complex (chicken, PDB: 7BY0; human, PDB: 6MUO). The backbone structures of the CENP-A nucleosome binding interface with ggKNL2 (residues 521–554) and ggCENP-C (residues 642-679) superimposed on the surface model of the CENP-A nucleosome show that the two backbones almost overlap (Fig 24B, C; Fig 29A). In fact, all the ggKNL2, ggCENP-C and hsCENP-C can recognize two of the major binding regions of the CENP-A nucleosome, H2A/B acidic patch binding region and CENP-A C-terminal binding region (Fig 29)[91,122–124], which is consistent with the sequence similarity among their conserved CENP-C motifs (Fig 16). The specific binding mechanism in chicken CENP-A-KNL2 complex structure shows that ggKNL2 R527 interacts with H2A D90 and E91 residues(Fig 26A)，sharing the same H2A/B acidic patch binding sites with chicken CENP-A R659 residue and human CENP-C R522 residue(Fig 29B)[91]. The another sharing binding region, CENP-A C-terminal, is recognized by F535 and W536 which are ggKNL2 counterparts of Y667 and W668 in the CENP-C motif or W530 and W531 in the hsCENP-C central domain (Fig 29C); this binding mode is well conserved. Further, on the CENP-A nucleosome, these aromatic residue pairs of the side chain arrangements are superposed (Fig 29C). Intriguingly, both ggKNL2 and ggCENP-C can specifically recognize the CENP-A RG-loop, but the observed $KNL2^{517\text{-}560}$ two-stranded antiparallel β -sheet structure in the C-terminal region was not identified in the CENP-A-ggCENP-C complex structure (Fig 24C; Fig 25B)[91]. Actually, the site of CENP-C interaction with the RG loop remains unclear due to the low resolution of the CENP-C-CENP-A nucleosome complex cryo-EM structural data[91]. However, it does not influence the key role of the conserved canonical CENP-C motif (in CENP-C) or CENP-C-like motif (in KNL2) downstream regions in CENP-A recognition. However, a similar RG-loop binding region has not been found in another CENP-A binding domain, central domain, identified in human or mouse CENP-C[101,125].

Simply, both the acidic patch of H2A/B and the C-terminal regions of CENP-A are critical regions for both KNL2 and CENP-C recognition.

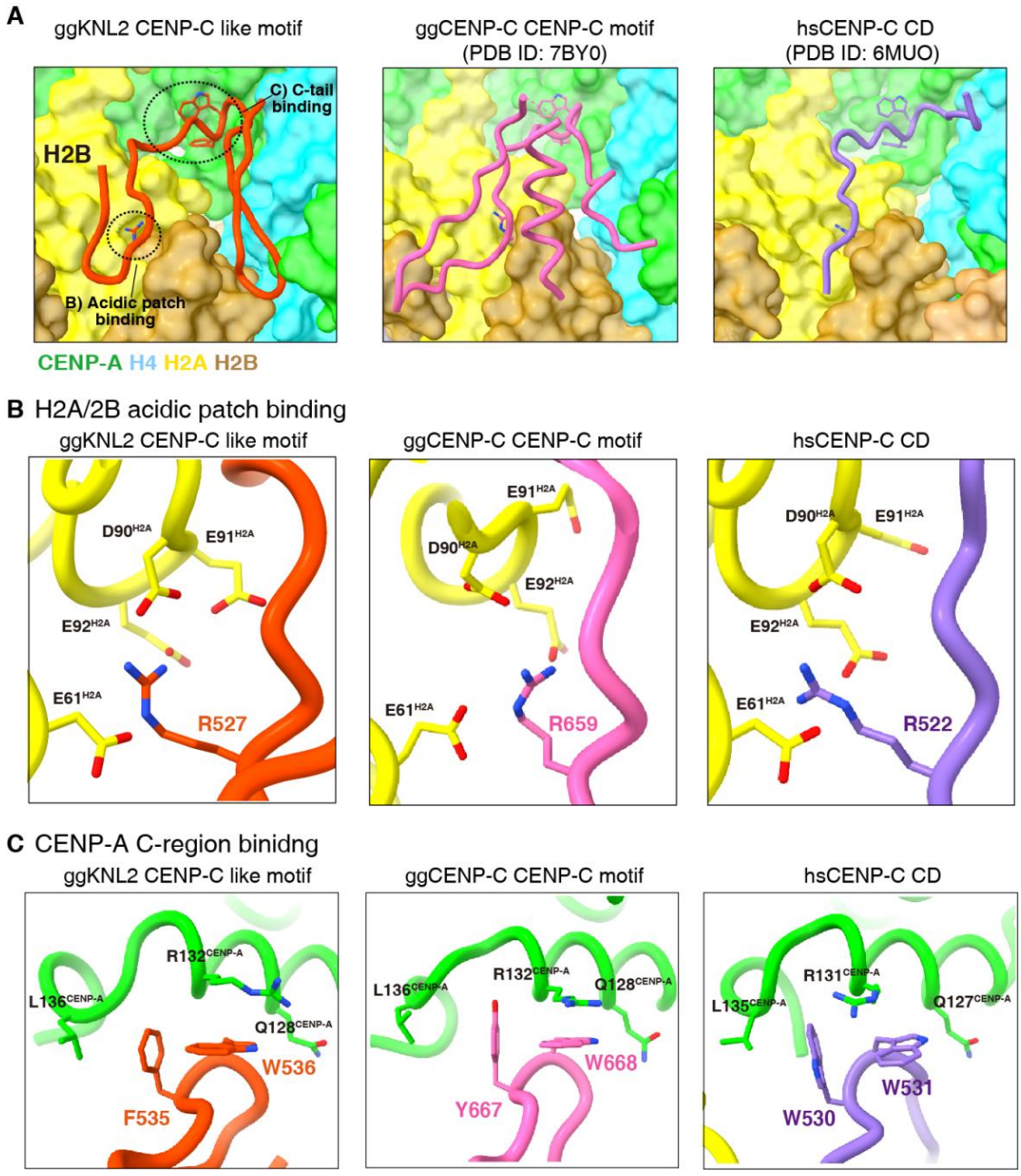


Figure 29. Structural comparison of the CENP-A nucleosome binding between KNL2 and CENP-C (related to Figs 24, 26 and 27).

A Structural comparison of the CENP-A nucleosome-binding interface between KNL2 and CENP-C. The backbone structures of chicken KNL2 (residues 521–554, orange), chicken CENP-C (residues 628–679, PDB ID: 7BY0, magenta), or human CENP-C (residues 520–537 in the central domain (CD), PDB ID: 6MUO) is shown in each panel. The side chains of the conserved residues involved in CENP-A

nucleosome binding are indicated as stick model. The CENP-A nucleosome is presented in a surface model color-coded as indicated in the fig.

B Structural comparison of the H2A/H2B acidic patch recognition mode among ggKNL2 CENP-C-like motif, ggCENP-C CENP-C motif, and the human CENP-C CD. The side chains of the conserved arginine residues are indicated. The yellow stick models indicate histone H2A residues in the acidic patch.

C Structural comparison of the CENP-A C-terminal tail-binding region among the chicken KNL2 CENP-C-like motif, chicken CENP-C motif, and the human CENP-C CD. The side chains of the conserved aromatic residues are indicated together with chimeric chicken CENP-A or human CENP-A residues (green).

3.1.5 Discussion

Until here, the binding mechanism for the CENP-A nucleosome-KNL2 complex is revealed by the cryo-EM structure, biochemical and cell biology in details. Among which, Cryo-EM structural analysis data collection was performed with the guidance and opperation of Dr. Fumiaki Makino; structural analysis was finished in cooperation with Dr. Mariko Ariyoshi, who is co-author of my published paper; and all the cell biology experiments were performed by Dr. Reito Watanabe and Dr. Tetsuya Hori.

Based on the BLI data (Figure 26-28 C), a possible view is that perhaps residue R527 contributes most to the CENP-A binding because of the weak binding signal, suggesting that R527 might facilitate the CENP-A nucleosome binding via electrostatic effect; another point is all the other three CENP-A-KNL2 binding residues, F535, W536, and D544, are likely to equally contribute to the CENP-A binding. As an emphasis, all interaction regions contribute to the KNL2-CENPA nucleosome interaction, as specific mutations introduced in each KNL2 key residue reduced the binding of ggKNL2 to the CENP-A nucleosome. Furthermore, these mutants did not suppress the knockout phenotype of ggKNL2, indicating that each binding site of KNL2 to the CENP-A nucleosome is essential for stable KNL2-CENP-A interaction and KNL2 function.

To further analyze the importance of key residues in the KNL2 function, in vivo experiments ware performed. It should be noted that all in vivo experiments in this paper were done by Dr. Tetsuya Hori and Dr. Reito Watanabe. Each of the GFP-fused $KNL2^{R527A}$, $KNL2^{F535A}$, $KNL2^{W536A}$, $KNL2^{F535A/W536A}$ and $KNL2^{D544A}$ mutants were designed and were introduced into the KNL2 conditional knockout chicken DT40 cells (cKO-KNL2 cells), separately. Two alleles of the KNL2 coding region were disrupted in cKO-KNL2 cells, and KNL2 cDNA was expressed under the control of the tetracycline (tet)-responsive promoter after tet addition[99]. Therefore, in this cell line, endogenous KNL2 expression could be turned off upon tet addition, and the expression of endogenous KNL2 was replaced by that of KNL2-GFP since GFP-fused ggKNL2 cDNA was expressed after tet addition. Using immunoblot, the endogenous KNL2 was confirmed to be replaced by the KNL2-GFP (wild-type: WT or mutant) (Fig 26-28D). To observe and compare the localization condition between $KNL2^{WT}$-GFP and $KNL2^{mutant}$-GFP, CENP-T was used as the centromere localization marker. In interphase, $KNL2^{WT}$-GFP localization is detected overlapped with that of CENP-T in the cKO-KNL2 cells. In contrast, most $KNL2^{mutant}$-GFP did not colocalize with CENP-T in interphase indicating that $KNL2^{mutant}$-GFP did not localize to the centromere region (Fig 26-28E). The mislocalized $KNL2^{mutant}$ protein formed non-centromeric nuclear bodies, which was also observed under the condition of CENP-C-like motif deletion mutant of KNL2[99]. In fact, cKO-KNL2 cells expressing only $KNL2^{mutant}$-GFP caused cell death, and this growth defect could be suppressed by expression of WT KNL2-GFP (Fig 26-28F).

As a conclusion, both the in vivo and in vitro results indicated that the wild-type ggKNL2 directly recognizes the CENP-A nucleosome and is critical for cell viability. Moreover, each of the checked KNL2 residues are key residues for the KNL2 binding. KNL2 R527 binds to the acidic patch region of H2A/B, KNL2 F535A/W536A directly recognize the C-terminal region and KNL2 D544A residue is essential for the CENP-A RG-loop recognition. These key residues facilitate the CENP-A-KNL2 complex formation, which is essential for KNL2 centromere localization and function in DT40 cells.

3.2 Materials and methods

Recombinant KNL2 mutant protein preparation

(Refer to recombinant protein above in Chapter 2 for more details)

Recombinant expression plasmid: The expression vectors of KNL2 mutants were generated based on the pGEX-KNL2$^{518-560}$ vector. The selected cDNA fragments of mutants encoding the C-terminal fragments of chicken KNL2 were each inserted into a pGEX-6p bacterial expression vector (Cytiva). Each fragment is tagged by both GST tag and FLAG-tag.

cDNA fragments: C-terminal fragments of chicken KNL2 (aa 518-560, KNL2$^{518-560R527A}$, KNL2$^{518–560F535A/W536A}$, KNL2$^{518–560D544A}$)

Tag: glutathione S transferase (GST) tag (N-terminus); FLAG tag (C-terminus)

Reconstitution of chicken CENP-A or H3 nucleosome

(Refer to above relevant content in Chapter 2 for more details)

Electrophoretic mobility shift assay (EMSA)

(Refer to above relevant content in Chapter 2 for more details)

Biolayer interferometry (BLI) assays

(Refer to above relevant content in Chapter 2 for more details)

Competitive GST pull-down assay

15 µl of glutathione Sepharose FF beads were equilibrated with pull-down buffer. GST-KNL2$^{456–560}$ (bait) at 1 µM concentration was incubated with CENP-A nucleosome (prey) at 0.25 µM concentration for 30 min at 4°C, and then non-phosphorylated or phosphorylated CENP-C$^{619–690}$ (competitor) was added. The concentration of each CENP-C$^{619–690}$ fragment was increased from 0 to 1 µM. The reaction volume was topped up to 50 µl and incubated for 1 h at 4°C with agitation. Then spun down the beads at 1,000 g for 1 min, and the supernatant was removed.

The beads were washed with 200 µl of pull-down assay buffer 3 times. After completely removing supernatant, the samples were boiled in 15 µl SDS-loading buffer and analyzed on 10–20% SDS–PAGE gel. Bands were visualized with Coomassie brilliant blue staining. Pull down assays were repeated three times. The intensities of the visualized bands were measured with Fiji[126]. The intensities of all histones in the CENP-A nucleosome were measured as the signals for the CENP-A nucleosome and normalized by the intensities of GST-KNL2$^{456–560}$. The intensity ratio of the histones to GST-KNL2$^{456–560}$ of each reaction was plotted in the bar graph. Data are shown as mean ± SD of three independent assays.

Beads: glutathione Sepharose FF beads (Cytiva)

Pull-down buffer: 20 mM HEPES pH 7.5, 200 mM NaCl, 5% Glycerol, 2 mM TCEP, 1 mM EDTA, 0.1% CHAPS

Cryo-EM data collection

The CENP-C-like motif containing peptide(aa 517–560, KNL2$^{517–560}$ peptide) derived from KNL2 was commercially synthesized (Biologica Co. Ltd.) and dissolved in 10 mM Tris–HCl buffer (pH 7.5). In 250 µl binding buffer, the KNL2$^{517–560}$ peptide and CENP-A nucleosome were mixed at a 5:1 molar ratio. The mixture was concentrated using an Amicon Ultra 0.5 ml filter unit (50 kDa molecular weight cut off, Merck). The final concentration of the complex was approximately 11.2 µM. A 2.5 µl protein solution of each complex was applied to grids and frozen in liquid ethane using a Vitrobot IV (FEI, 4 °C and 100% humidity). Data collection of each sample was carried out on a CRYO ARM 300 equipped with a cold field-emission electron gun operated at 300 kV, an Ω-type energy filter with a 20 eV slit width and a K3 Summit direct electron detector camera. An automated data acquisition program, SerialEM[127], was used to collect cryoEM image data. Movie frames were recorded using the K3 Summit camera at a calibrated magnification of × 50,000 corresponding to a pixel size of 0.98 Å with a setting defocus range from −1.0 to −2.5 µm. The data were collected with a total exposure of 3 s fractionated into 40 frames, with a total dose of ~40 electrons Å-2 in counting mode. 9,450 movies were collected.

Binding buffer: 20 mM HEPES pH 7.5, 100 mM NaCl, 0.1% CHAPS, and 5 mM TCEP

Grid: Quantifoil Cu R0.6/1.0 holey carbon grid

Equipment: Vitrobot IV; CRYO ARM 300 (JEOL, Japan); K3 Summit direct electron detector camera (Gatan, USA)

Image processing and 3D reconstruction

Motion correction was carried out by MotionCor2[128] to align all CENP-A nucleosome complex micrographs and the CTF parameters were estimated by Gctf[129]. All CENP-A complexes were automatically selected by Auto-picking using the Laplacian and Gaussian in RELION 3.1[130], and they were extracted into a box of 192 × 192 pixels. 10,331,795 particles were selected after performing 2D classification by RELION 3.1. An initial 3D structure was generated by de novo 3D initial model in Relion 3.1 with C1 symmetry. An initial 3D classification into three classes with C1 symmetry resulted in one class (629,490 particles) that was used as further 3D refinement, CTF refinement procedure and particles polishing. The map for a well-defined nucleosome structure and the KNL2$^{517-560}$ peptide at 3.36 Å resolution was obtained after solvent mask post-processing. Further focused 3D classification with a mask generated for the KNL2$^{517-560}$ peptide resulted in one class (270, 464 particles). The map of the CENP-A nucleosome - KNL2$^{517-560}$ peptide complex at 3.42 Å resolution was obtained after 3D refinement, particle polishing, CTF refinement, and solvent mask postprocessing with C2 symmetry.

Software program: RELION 3.1

Model building

The crystal structure model of the nucleosome containing the chimeric H3/CENP-A (PDB entry, 5Z23) and the cryo-EM structure of the CENP-C fragment in complex with CENP-A nucleosome (PDB entry, 7BY0) were modified to generate an initial model. Except for the canonical CENP-C motif region (aa 654–675), the model of CENP-C was removed and the non-conserved residues were manually replaced by alanine in program Coot[131]. The initial model was fit into the cryo-EM density of the CENP-A nucleosome-KNL2$^{517-560}$ peptide complex using UCSF Chimera[132]. The model of the KNL2$^{517-560}$ peptide manually replaced its sequence by that of chicken KNL2 and modified using Coot. The model for the further region of KNL2 was manually built. The model was subjected to real space refinement in PHENIX[133]. Based on the EM map density, the complex model was iteratively modified in Coot and refined in PHENIX. The final model coordinates and maps were deposited in the protein data bank. The structural figures were produced using UCSF Chimera.

Software program: Coot; UCSF Chimera; PHENIX

Chapter 4: A centromeric ggKNL2 localization mechanism in interphase is different from that in mitotic cells

4.1 Results and discussion

Previous study reveals that ggKNL2 localizes to the centromeres throughout all the cell cycle in chicken cells[99]. In the interphase, ggKNL2 localize to the centromere region through direct binding to the CENP-A nucleosome together with other Mis18 complex components as a licensing factor for new CENP-A deposition. ggKNL2 is an upstream factor for the new CENP-A recruitment in the interphase. Comparing the cryo-EM structures between the CENP-A-KNL2 and CENP-A-CENP-C complexes in chicken cells (Fig 24C), it is known that the CENP-C-like motif of ggKNL2 and the CENP-C motif of ggCENP-C compete the same binding region on CENP-A nucleosome[91]. The function and importance of KNL2 binding to the CENP-A nucleosome in interphase were examined as demonstrated in Chapter 3 (Fig 26-28). Therefore, the specific molecular mechanism of ggKNL2 binding in interphase is clarified. However, in M phase, as proved in previous study, the CENP-A-KNL2 binding elements on the CENP-A nucleosome is occupied by CDK1 phosphorylated protein CENP-C[91,92]. The unknown question is whether ggKNL2 localizes to centromeres via CENP-A nucleosome interaction in the presence of ggCENP-C. To answer this question, a cell biological experiment was designed by Dr. Tetsuya Hori to express WT KNL2-GFP or each KNL2 mutant in the cKO-KNL2 DT40 cells to observe the KNL2 localization. In interphase cells, WT KNL2 centromeric localization was not influenced without the existence of CENP-C. However, for each of the KNL2 mutants, centromeric mislocalization signals were observed (Fig 30), which is consistent with the relevant findings of this study. Another phenomenon is the reduction of CENP-A in each KNL2 mutant cell line (Fig 30) probably due to defects of new CENP-A deposition and CENP-A stability. In addition, recent cryo-EM data indicate that neither CENP-C nor CENP-N binds to CENP-A nucleosome during the interphase in chicken DT40 cells[74,134], indicating the possibility that ggKNL2 binds to CENP-A nucleosomes instead of CENP-C and CENP-N in interphase cells. However, in M phase, the centromere localization of each KNL2 mutant was not affected by these mutations (Fig 30), suggesting that the

localization of ggKNL2 at mitotic centromeres does not involve directly binding to the CENP-A nucleosome.

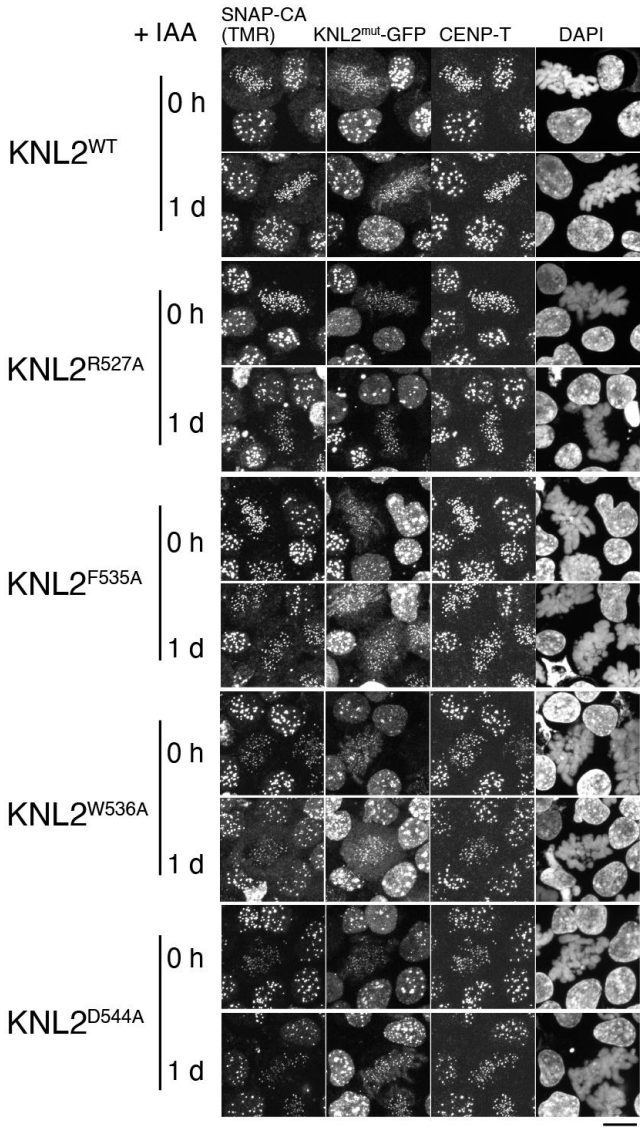


Figure 30. ggKNL2 mutants for binding to the CENP-A nucleosome cause CENP-A reduction at centromeres.

AID-based KNL2 knockout cells expressing GFP fused each KNL2 mutant and SNAP-tagged CENP-A were used. After addition of IAA, SNAP-CENP-A signals were visualized by TMR-Star staining. CENP-T was stained as a kinetochore marker. Representative images were presented. Mitotic centromere localization of each KNL2 mutant was shown as a merged image (Insets): SNAP-CENP-A (red) and each KNL2 mutant (green). DNA was stained with DAPI. Scale bar indicates 10 μm.

Without directly binding to the CENP-A nucleosome, how does ggKNL2 localize to mitotic centromeres? Although in interphase, ggKNL2 directly localizes to centromeres without the existence of CENP-C, in some other species, this binding is intermediated by the CENP-C protein[101,108,135]. An instructive example is that the centromeric localization of KNL2 in Xenopus depends on CENP-C[96]. Reminding the similar binding mechanism that ggKNL2 mitotic centromere localization might be mediated by CENP-C. The ggKNL2 localization was analyzed by Dr. Reito Watanabe in cells without CENP-C expression, with a CRISPR-based genome editing method which introduced the mScarlet sequence into the C-terminal end of the endogenous ggKNL2 gene alleles in the cKO-CENP-C cells[92]. Significantly, mitotic signals of mScarlet-ggKNL2 drastically reduced due to the absence of CENP-C (Fig 31)， for contrast, the interphase signal remained almost no difference, in line with previous study[116].

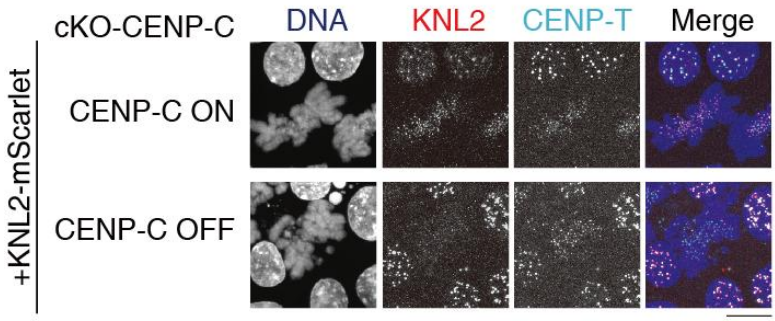


Figure 31. KNL2 (red) localization in CENP-C conditional knockout chicken DT40 cells (cKO-CENP-C) with or without CENP-C expression.

DNA was stained with DAPI (blue). CENP-T was stained as a kinetochore marker (light blue). Scale bar indicates 10 μm.

Therefore, a hypothesis model is the mitotic centromere is recognized by the CDK1 phosphorylated CENP-C, which was occupied on the interphase KNL2 binding sites in CENP-A nucleosomes[91,92]. GST pull-down assay were performed to test this hypothesis (Fig 32). Using GST-fused KNL2$^{456\text{-}560}$ as a bait and the CENP-A nucleosome as a prey in the absence or presence of phosphorylated or

unphosphorylated ggCENP-C fragments containing the CENP-C motif (ggCENP-C$^{619-690}$). The band intensity of CENP-A nucleosome histones after pull down indicates the competition ability between KNL2 and CENP-C fragments. Comparing with the unphosphorylated CENPC$^{619-690}$, the phosphorylated CENP-C$^{619-690}$ was competed more CENP-A nucleosome from the immobilized GST-KNL2$^{456-560}$(Fig 32), implying that the mitotic phosphorylated CENP-C$^{619-690}$ prefers to bind to the CENP-A nucleosome.

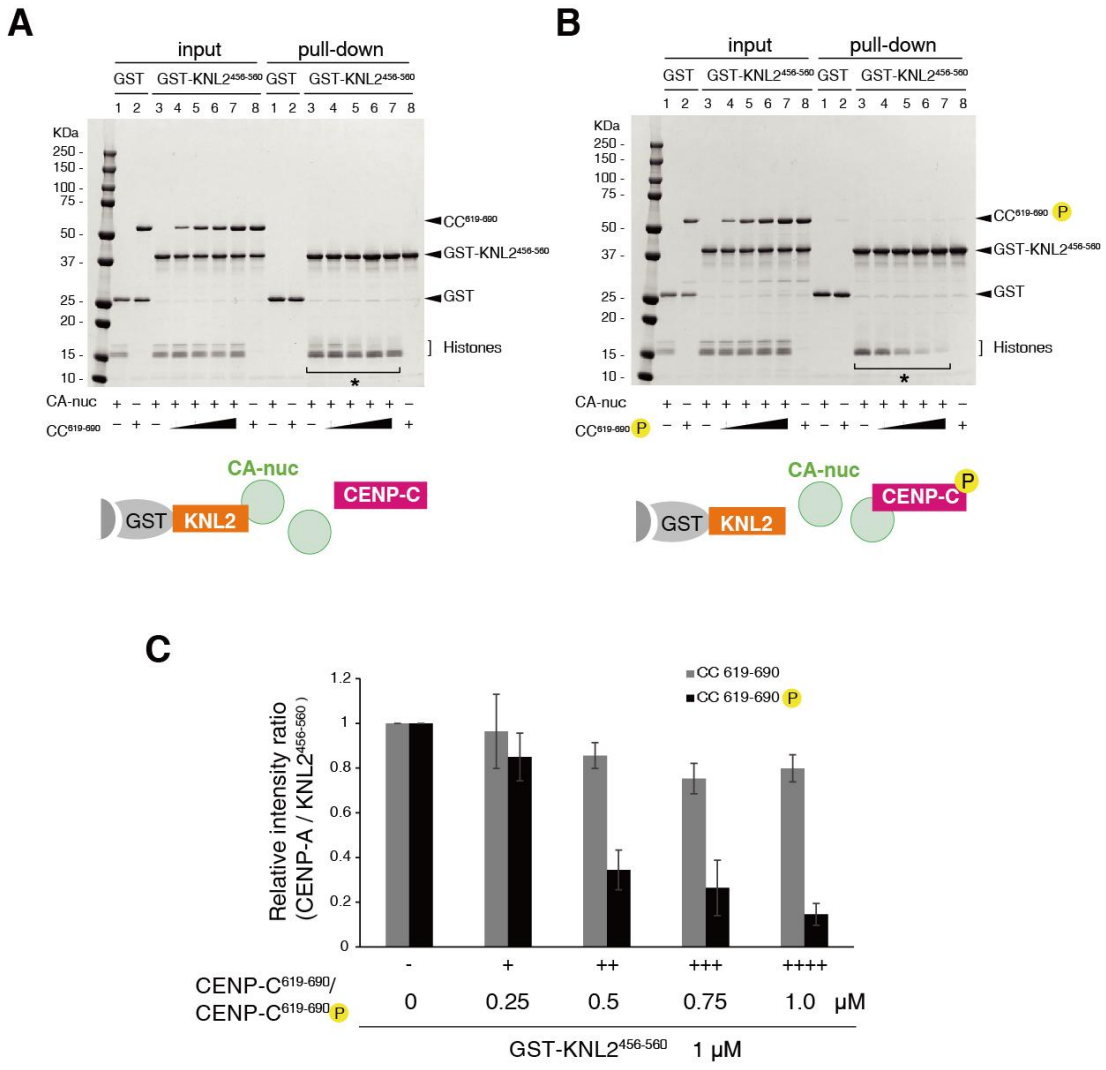


Figure 32. Competitive pull-down assays for the CENP-A nucleosome binding of ggKNL2 and ggCENP-C.

A and B A complex of GST-KNL2$^{456\text{-}560}$ with the CENP-A nucleosome was incubated in the absence or presence of MBP-CENP-C$^{619\text{-}690}$ or phosphorylated MBP-CENP-C$^{619\text{-}690}$ and pulled down by GST affinity. The concentration of CENP-C$^{619–690}$ was increased from 0 μM to1 μM (lanes 3–7).

C The ratio of band intensities of pull-downed histones (*) to GST-KNL2$^{456-560}$ in each lane (lanes 3–7) was estimated and presented in the bar graph. Data are mean with standard deviations calculated from the technical replicates, n = 3.

Considering all the obtained results of molecular biology, cell biology and structural biology, briefly, the cell cycle dependent centromeric KNL2 localization model is mainly described as following (Fig 33). In interphase, ggKNL2 specifically binds to the CENP-A nucleosome via the CENP-C-like motif which is critical for KNL2 function; however, the mitotic centromeric localization of ggKNL2 rely on phosphorylated CENP-C but not on the CENP-A nucleosome (Fig 32). But it is still unclear what is the regulation mechanism of KNL2 with binding partners throughout the cell cycle, and whether the protein ggCNEP-C binds directly to ggKNL2 in mitosis. Both the mechanism of KNL2 localization in mitotic centromeres and its biological significance are intriguing.

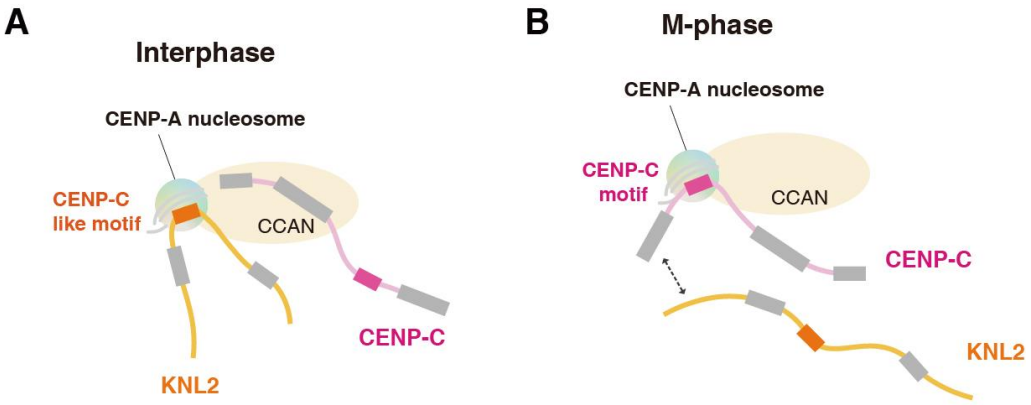


Figure 33. A model of ggKNL2 localization dependency into centromeres in interphase and mitotic cells.

A ggKNL2 localizes at centromeres depending on the CENP-A nucleosome binding in interphase cells.

B In mitotic cells, centromere localization of ggKNL2 is mediated via CENP-C, not via CENP-A. CENP-C localizes at mitotic centromeres via the CENP-A nucleosome interaction.

4.2 Materials and methods

Recombinant KNL2$^{456-560}$ protein preparation

(Refer to recombinant protein above in Chapter 2 for more details)

Recombinant CENP-C$^{619–690}$ protein preparation[91]

Recombinant expression plasmid: A cDNA fragment encoding shorter C-terminal fragments of chicken CENP-C$^{619-690}$ were inserted into a pMAL-c5X-TEV-His plasmid (pMAL-c5X-TEV-CC$^{619–690}$-His).

Plasmid: pMAL-c5X-TEV-His[73]

Tag: maltose-binding protein (MBP) tag (N-terminus); hexa-histidine tag (C-terminus)

Expression: The ggCENP-C fragment (residues 619 – 690; CENP-C$^{619 – 690}$) was expressed in E. coli RosettaTM2 (DE3) as a MBP-fused recombinant protein. E. coli cells were grown in LB at 37 ° C until OD600 reached 0.6; protein expression was induced by addition of IPTG to a final concentration of 0.2 mM, and culture was continued at 17°C overnight.

Host cell: E. coli RosettaTM2 (DE3) (Novagen)

Media: LB-Lennox media

Temperature: 37°C for cell growth; 17°C for protein expression.

Recombinant CENP-C619–690 protein purification

Sample collection: Cells were harvested by centrifugation. The cell pellet was resuspended in lysis buffer and lysed by sonication on ice.

Lysis buffer: 20 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, and 2 mM TCEP supplemented with protease inhibitor cocktail (Roche).

His affinity purification: The column was equilibrated with buffer, then apply the supernatant of lysate into the column. The immobilized MBP-CENP-C$^{619 – 690}$ was eluted from the column with an imidazole gradient from 10 mM to 300 mM.

Column: HisTrap column (Cytiva)

Elution buffer: 50 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol and 10 mM reduced glutathione

ION exchange column chromatography purification: The eluted fractions are applied to a HiTrap SP column. MBP-CENP-$C^{619-690}$ was eluted with a linear NaCl gradient from 200 to 750 mM. Peak fractions were combined, concentrated (typically to 2–5 mg/ ml).

ION exchange column: HiTrap SP HP cation exchange column (Cytiva)

Size-exclusion column chromatography purification: Further, to purify the MBP-CENP-$C^{619–690}$ by size exclusion column with the buffer of 300 mM NaCl. Peak fractions containing MBP-CENP-$C^{619–690}$ were combined, concentrated (typically to 3 – 5 mg/ml), and stored at 80 ° C until further use for structural and biochemical analyses.

Size-exclusion column: HiLoad 16/60 Superdex 200 pg column (Cytiva)

Buffer: 20 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol and 2 mM DTT.

Preparation of the phosphorylated CENP-C619–690 fragments

MBP-CENP-$C^{619-690}$ phosphorylation[92,136,137]**:** MBP-gCENP-$C^{619-690}$ was incubated in the presence or absence of the active cyclin B–CDK1 (relative H1 kinase activity: 1.4 pmol P/min µl) in reaction buffer for 1 h at 25°C in vitro. To examine the phosphorylated proteins, they were separated on SuperSep Ace, 5–20%, or Phos-tag-5.0% SDS-PAGE (25 µM Phos-tag acrylamide and 50 µM MnCl2) and stained with CBB stain one (ready to use; Nacalai Tesque). The reaction was terminated by adding EDTA-NaOH, pH 8.0 (final concentration, 5 mM).

Reaction buffer: 10 mM Tris–HCl, pH 7.5, 2 mM MgCl2, 150 mM NaCl, 100 mM ATP, and 1× complete EDTA-free proteinase inhibitor.

ION exchange column chromatography purification: After the reaction was terminated by adding EDTA-NaOH, pH 8.0 (final concentration, 5 mM), CDK-1treated samples were further purified using a cation exchange column. For

cation exchange column chromatography, the CDK-1-treated sample was applied to HiTrap SP column and eluted with a step-wise gradient of NaCl from 0 to 1,000 mM. The purified samples were concentrated and kept at 80°C.

ION exchange column: HiTrap SP HP cation exchange column (Cytiva)

Buffer: 20 mM HEPES pH 7.5, 5% glycerol and 2 mM DTT, salt concentration increased in steps of 50 mM from 0 to 1000 mM

Cell culture

Chicken DT40 cells were cultured in DMEM medium (Nacalai Tesque) at 38.5 °C with 5% CO2. KNL2 conditional knockout cells[99] were used for examination of centromere localization GFP fused KNL2. Various KNL2 mutant DNAs were cloned into a pEGFP-C3 vector, which were transfected into KNL2 knockout cells with Gene Pulser II electroporator (BioRad). The transfected cells were selected in the medium containing 2 mg/ml G418 (Santa Cruz Biotechnology). AID-based KNL2 conditional knockout cell lines expressing the GFP-tagged wild-type KNL2 ($KNL2^{WT}$-GFP) or each KNL2 mutant ($KNL2^{R527A}$-, $KNL2^{F535A}$-, $KNL2^{W536A}$-, $KNL2^{D544A}$-GFP) were established (Fig 30). First, the knockin construct with the HASNAP-CENP-A expressing cassette was transfected for the beta-actin gene locus (beta-actin-KI) into the AID-based KNL2 conditional knockout cells[99] and selected in the medium containing 25 μg/ml Mycophenolic acid (Wako) with 125 μg/ml Xanthine (SIGMA) to isolate the AID-based KNL2 conditional knockout cells expressing HA-SNAP-CENP-A (KNL2-AID [SNAPCA]). Then, the knockin constructs containing each KNL2 mutant were transfected for the phosphoglycerate kinase 1 (PGK1) gene locus (PGK1-KI) into the KNL2-AID (SNAPCA) cells and selected in the medium containing 2.5 mg/ml Hygromycin B (Wako) to isolate KNL2-AID (SNAPCA) cells expressing each KNL2 mutant (KNL2mut-GFP_KNL2-AID [SNAPCA]).

Vector： pEGFP-C3 vector (Clontech)

Medium: DMEM medium (Nacalaitesque): 10% fetal bovine serum (FBS; Sigma), 1% chicken serum (Gibco), 10 μM 2-mercaptoethanol, Penicillin/ Streptomycin (final: 100 unit/ml and 100 lg/ml, respectively) (Thermo Fisher)

Immunoblotting

Expression of each protein in culture cells was analyzed by immunoblotting. For whole-cell samples, DT40 cells were harvested, washed with PBS, and suspended in 1 × Laemmli sample buffer (LSB; final 104 cells/μl), followed by sonication and heating for 5 min at 96 °C. Proteins were separated on SuperSep Ace, 5 – 20%

(FUJIFILM WAKO Chemicals), and transferred to Immobilon-P (Merck) using HorizeBLOT (ATTO). Both primary and secondary antibodies were used, to increase sensitivity and specificity, Signal Enhancer Hikari (Nacalai Tesque) was used for all antibodies. The antibodies were incubated with the blotted membranes for 1 h at room temperature or for overnight at 4°C. Proteins reacting with antibodies were detected with ECL Prime (Citiva) and visualized with ChemiDoc Touch (Bio-Rad). Acquired images were processed using Image Lab 5.2.1 (Bio-Rad) and Photoshop CC (Adobe).

Primary antibodies: rabbit anti-KNL2; mouse anti-α-tubulin (Sigma)

Secondary antibodies: HRP-conjugated antirabbit IgG (Jackson ImmunoResearch); HRP-conjugated anti-mouse IgG (Jackson ImmunoResearch)

Immunofluorescence and image acquisition

DT40 cells expressing various GFP fused KNL2 were centrifuged onto slide glasses by the Cytospin3 centrifuge (Shandon), and fixed with 4% paraformaldehyde (PFA; Thermo Fisher) in PBS for 10 min at room temperature. The fixed cells were permeabilized in 0.5% NP-40 in PBS for 10 min at room temperature and incubated with rabbit anti-CENP-T (1: 1,000)[55] in 0.5% BSA in PBS for 1 h at 37°C as a primary antibody. The reacted cells were washed with 0.5% BSA (Equitech-Bio Inc) in PBS 3 times and were incubated with Cy3-conjugated mouse anti-rabbit IgG (1:2,000; Jackson ImmunoResearch) in 0.5% BSA in PBS for 30 min at 37°C as a secondary antibody. DNA was stained with 1 μg/ml DAPI in PBS for 1 min and mounted with VECTASHIELD mounting medium (Vector Laboratories). Immunofluorescence images were acquired every 0.2 μm step of Z-slice (total 4–8 μm thick) as a Z-stack image using a Zyla 4.2 sCMOS camera (ANDOR) mounted on a Nikon Eclipse Ti inverted microscope with an objective lens (Plan Apo lambda 100x/1.45 NA; Nikon) with a spinning disk confocal scanner unit (CSU-W1; YOKOGAWA) controlled by NIS-elements (Nikon). The images in figures are the maximum intensity projection of the Z-stack generated with NIS-elements and were processed by Photoshop CC (Adobe).

Measurement for the CENP-A in the KNL2 mutant cells

To evaluate the CENP-A levels at the centromeres in wild-type or each KNL2 mutant cell lines ($KNL2^{WT}$-, $KNL2^{R527A}$-, $KNL2^{F535A}$-, $KNL2^{W536A}$-, $KNL2^{D544A}$-GFP_KNL2-AID [SNAPCA]), IAA (final 500 μM) was added to culture medium at time 0 and the SNAP-tagged CENP-A was labeled with TMR-Star (New England Biolabs; final 3 μM) for 15 min at 0, 5, 24, and 48 h. The TMR-Star labeled cells were fixed with 4% Paraformaldehyde (PFA; Thermo Fisher) in PBS for 5 min and were permeabilized in 0.5% NP-40 in PBS for 5 min. Immunostaining was

performed with rabbit anti-CENP-T (1:1000)[55] as a primary antibody and with Alexa647conjugated goat anti-rabbit IgG (1:1000; Jackson ImmunoResearch) as a secondary antibody. Acquired images were processed and the signal intensities of TMR-Star were measured as a CENP-A amount at the mitotic centromeres using the Imaris software (Bitplane).

References

1. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. How Many Species Are There on Earth and in the Ocean? *PLoS. Biol.* **9**, e1001127 (2011).

2. Baluska, F., Volkmann, D., Menzel, D. & Barlow, P. Strasburger's legacy to mitosis and cytokinesis and its relevance for the Cell Theory. *Protoplasma* **249**, 1151–1162 (2012).

3. Mitchison, T. J. & Salmon, E. D. Mitosis: a history of division. *Nat Cell Biol* **3**, E17–E21 (2001).

4. Peters, J.-M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev.* **22**, 3089–3114 (2008).

5. Uhlmann, F. Chromosome cohesion and segregation in mitosis and meiosis. *Curr Opin Cell Biol* **13**, 754–761 (2001).

6. McIntosh, J. R. Mitosis. *Cold Spring Harb Perspect Biol* **8**, a023218 (2016).

7. Güttinger, S., Laurell, E. & Kutay, U. Orchestrating nuclear envelope disassembly and reassembly during mitosis. *Nat Rev Mol Cell Biol* **10**, 178–191 (2009).

8. Flemming, W. *Zellsubstanz, Kern und Zelltheilung*. (Leipzig : F.C.W. Vogel, 1882).

9. Cremer, T. & Cremer, C. Centennial of Wilhelm Waldeyer's introduction of the term 'chromosome' in 1888. *Cytogenet Cell Genet* **48**, 65–67 (1988).

10. Morgan, T.H., Sturtevant, A.H., & Bridges, C.B. The Mechanism of Mendelian

Heredity. *Nature* **97**, 117–118 (1916).

11. Shuaib, M. Epigenetic mechanism of CENP-A loading to centromeres. (Université de Strasbourg, 2012).

12. Bak, A. L., Zeuthen, J. & Crick, F. H. Higher-order structure of human mitotic chromosomes. *Proc Natl Acad Sci U S A* **74**, 1595–1599 (1977).

13. Olins, A. L. & Olins, D. E. Spheroid Chromatin Units (ν Bodies). *Science* **183**, 330–332 (1974).

14. Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868–871 (1974).

15. Kornberg, R. & Thomas, J. Chromatin Structure - Oligomers of Histones. *Science* **184**, 865–868 (1974).

16. Kornberg, R. Structure of Chromatin. *Annu. Rev. Biochem.* **46**, 931–954 (1977).

17. Oudet, P., Grossbellard, M. & Chambon, P. Electron-Microscopic and Biochemical Evidence That Chromatin Structure Is a Repeating Unit. *Cell* **4**, 281–300 (1975).

18. Finch, J. *et al.* Structure of Nucleosome Core Particles of Chromatin. *Nature* **269**, 29–36 (1977).

19. Olins, D. E. & Olins, A. L. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* **4**, 809–814 (2003).

20. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J.

Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).

21. Hamiche, A. & Shuaib, M. Chaperoning the histone H3 family. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1819**, 230–237 (2012).

22. Darlington, C. D. & Hall, A. D. The external mechanics of the chromosomes I—The scope of enquiry. *Proceedings of the Royal Society of London. Series B - Biological Sciences* **121**, 264–273 (1997).

23. Csink, A. K. & Henikoff, S. Something from nothing: The evolution and utility of satellite repeats. *Trends Genet.* **14**, 200–204 (1998).

24. Fukagawa, T. Centromere DNA, proteins and kinetochore assembly in vertebrate cells. *Chromosome Res* **12**, 557–567 (2004).

25. Shang, W. H. *et al.* Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res* **20**, 1219–1228 (2010).

26. Marshall, O. J., Chueh, A. C., Wong, L. H. & Choo, K. H. A. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet* **82**, 261–282 (2008).

27. Wong, C. Y. Y., Lee, B. C. H. & Yuen, K. W. Y. Epigenetic regulation of centromere function. *Cell Mol Life Sci* **77**, 2899–2917 (2020).

28. Palmer, D., Oday, K., Wener, M., Andrews, B. & Margolis, R. A 17-Kd

Centromere Protein (cenp-a) Copurifies with Nucleosome Core Particles and with Histones. *J. Cell Biol.* **104**, 805–815 (1987).

29. Palmer, D., Oday, K., Trong, H., Charbonneau, H. & Margolis, R. Purification of the Centromere-Specific Protein Cenp-a and Demonstration That It Is a Distinctive Histone. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3734–3738 (1991).

30. Tachiwana, H. *et al.* Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **476**, 232-U135 (2011).

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

32. Fukagawa, T. & Earnshaw, W. C. The centromere: chromatin foundation for the kinetochore machinery. *Dev Cell* **30**, 496–508 (2014).

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

34. Manuelidis, L. Chromosomal localization of complex and simple repeated human DNAs. *Chromosoma* **66**, 23–32 (1978).

35. Schueler, M. G. & Sullivan, B. A. Structural and functional dynamics of human centromeric chromatin. *Annual Review of Genomics and Human Genetics* **7**, 301–313 (2006).

36. Erwinsyah, R., Riandi & Nurjhani, M. Relevance of Human Chromosome Analysis Activities against Mutation Concept in Genetics Course. *IOP Conf. Ser.:*

Mater. Sci. Eng. **180**, 012285 (2017).

37. Guerra, M. *et al.* Neocentrics and Holokinetics (Holocentrics): Chromosomes out of the Centromeric Rules. *Cytogenet Genome Res* **129**, 82–96 (2010).

38. Voullaire, L. E., Slater, H. R., Petrovic, V. & Choo, K. H. A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet* **52**, 1153–1163 (1993).

39. Fukagawa, T. & Earnshaw, W. C. Neocentromeres. *Curr Biol* **24**, R946-947 (2014).

40. Earnshaw, W. C. & Rothfield, N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* **91**, 313–321 (1985).

41. Sullivan, K. F., Hechenberger, M. & Masri, K. Human CENP-A Contains a Histone H3 Related Histone Fold Domain That Is Required for Targeting to the Centromere. *J Cell Biol* **127**, 581–592 (1994).

42. Allshire, R. C. & Karpen, G. H. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat Rev Genet* **9**, 923–937 (2008).

43. Perpelescu, M. & Fukagawa, T. The ABCs of CENPs. *Chromosoma* **120**, 425–446 (2011).

44. Westhorpe, F. G. & Straight, A. F. Functions of the centromere and kinetochore in chromosome segregation. *Curr. Opin. Cell Biol.* **25**, 334–340 (2013).

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

46. Westhorpe, F. G., Fuller, C. J. & Straight, A. F. A cell-free CENP-A assembly system defines the chromatin requirements for centromere maintenance. *Journal of Cell Biology* **209**, 789–801 (2015).

47. Logsdon, G. A. *et al.* Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. *J Cell Biol* **208**, 521–531 (2015).

48. Guse, A., Carroll, C. W., Moree, B., Fuller, C. J. & Straight, A. F. In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature* **477**, 354–358 (2011).

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

50. Kato, H. *et al.* A Conserved Mechanism for Centromeric Nucleosome Recognition by Centromere Protein CENP-C. *Science* **340**, 1110–1113 (2013).

51. Régnier, V., Novelli, J., Fukagawa, T., Vagnarelli, P. & Brown, W. Characterization of chicken CENP-A and comparative sequence analysis of vertebrate centromere-specific histone H3-like proteins. *Gene* **316**, 39–46 (2003).

52. Black, B. E. *et al.* Structural determinants for generating centromeric chromatin.

Nature **430**, 578–582 (2004).

53. Panchenko, T. *et al.* Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *Proceedings of the National Academy of Sciences* **108**, 16588–16593 (2011).

54. Geiss, C. P. *et al.* CENP-A arrays are more condensed than canonical arrays at low ionic strength. *Biophys J* **106**, 875–882 (2014).

55. Hori, T. *et al.* CCAN Makes Multiple Contacts with Centromeric DNA to Provide Distinct Pathways to the Outer Kinetochore. *Cell* **135**, 1039–1052 (2008).

56. Foltz, D. R. *et al.* The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* **8**, 458-U77 (2006).

57. Izuta, H. *et al.* Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. *Genes Cells* **11**, 673–684 (2006).

58. Okada, M. *et al.* The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat. Cell Biol.* **8**, 446-U61 (2006).

59. Sharp, L. W. *Introduction to cytology*. 1–592 (McGraw-Hill Book Company, inc, 1934). doi:10.5962/bhl.title.6429.

60. Cheeseman, I. M. The kinetochore. *Cold Spring Harb Perspect Biol* **6**, a015826 (2014).

61. Hara, M. & Fukagawa, T. Dynamics of kinetochore structure and its regulations during mitotic progression. *Cell Mol Life Sci* **77**, 2981–2995 (2020).

62. McEwen, B. F., Hsieh, C.-E., Mattheyses, A. L. & Rieder, C. L. A new look at kinetochore structure in vertebrate somatic cells using high-pressure freezing and freeze substitution. *Chromosoma* **107**, 366–375 (1998).

63. Maiato, H., DeLuca, J., Salmon, E. D. & Earnshaw, W. C. The dynamic kinetochore-microtubule interface. *Journal of Cell Science* **117**, 5461–5477 (2004).

64. DeLuca, J. G. *et al.* Hec1 and Nuf2 Are Core Components of the Kinetochore Outer Plate Essential for Organizing Microtubule Attachment Sites. *MBoC* **16**, 519–531 (2005).

65. Cheeseman, I. M., Chappie, J. S., Wilson-Kubalek, E. M. & Desai, A. The Conserved KMN Network Constitutes the Core Microtubule-Binding Site of the Kinetochore. *Cell* **127**, 983–997 (2006).

66. Varma, D. & Salmon, E. D. The KMN protein network--chief conductors of the kinetochore orchestra. *J Cell Sci* **125**, 5927–5936 (2012).

67. Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and Kinetochores: From Epigenetics to Mitotic Checkpoint Signaling. *Cell* **112**, 407–421 (2003).

68. Sarangapani, K. K. & Asbury, C. L. Catch and release: how do kinetochores hook the right microtubules during mitosis? *Trends in Genetics* **30**, 150–159 (2014).

69. Musacchio, A. & Desai, A. A Molecular View of Kinetochore Assembly and

Function. *Biology (Basel)* **6**, 5 (2017).

70. Lampson, M. A. & Grishchuk, E. L. Mechanisms to Avoid and Correct Erroneous Kinetochore-Microtubule Attachments. *Biology (Basel)* **6**, 1 (2017).

71. Sugimoto, K., Yata, H., Muro, Y. & Himeno, M. Human centromere protein C (CENP-C) is a DNA-binding protein which possesses a novel DNA-binding motif. *J Biochem* **116**, 877–881 (1994).

72. Sugimoto, K., Kuriyama, K., Shibata, A. & Himeno, M. Characterization of internal DNA-binding and C-terminal dimerization domains of human centromere/kinetochore autoantigen CENP-C in vitro: role of DNA-binding and self-associating activities in kinetochore organization. *Chromosome Res* **5**, 132–141 (1997).

73. Nishino, T. *et al.* CENP-T provides a structural platform for outer kinetochore assembly. *The EMBO Journal* **32**, 424–436 (2013).

74. Yatskevich, S. *et al.* Structure of the human inner kinetochore bound to a centromeric CENP-A nucleosome. *Science* **376**, 844–852 (2022).

75. Pentakota, S. *et al.* Decoding the centromeric nucleosome through CENP-N. *eLife* **6**, e33442 (2017).

76. Hara, M. & Fukagawa, T. Where is the right path heading from the centromere to spindle microtubules? *Cell Cycle* **18**, 1199–1211 (2019).

77. Hori, T., Shang, W. H., Takeuchi, K. & Fukagawa, T. The CCAN recruits

CENP-A to the centromere and forms the structural core for kinetochore assembly. *J Cell Biol* **200**, 45–60 (2013).

78. Huis In 't Veld, P. J. *et al.* Molecular basis of outer kinetochore assembly on CENP-T. *Elife* **5**, e21007 (2016).

79. Screpanti, E. *et al.* Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr Biol* **21**, 391–398 (2011).

80. Gascoigne, K. E. *et al.* Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. *Cell* **145**, 410–422 (2011).

81. Rago, F., Gascoigne, K. E. & Cheeseman, I. M. Distinct organization and regulation of the outer kinetochore KMN network downstream of CENP-C and CENP-T. *Curr Biol* **25**, 671–677 (2015).

82. Saitoh, H. *et al.* CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell* **70**, 115–125 (1992).

83. Moroi, Y., Peebles, C., Fritzler, M. J., Steigerwald, J. & Tan, E. M. Autoantibody to centromere (kinetochore) in scleroderma sera. *Proc Natl Acad Sci U S A* **77**, 1627–1631 (1980).

84. Petrovic, A. *et al.* The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *J Cell Biol* **190**, 835–852 (2010).

85. Klare, K. *et al.* CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. *J Cell Biol* **210**, 11–22 (2015).

86. McKinley, K. L. *et al.* The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. *Mol Cell* **60**, 886–898 (2015).

87. Cohen, R. L. *et al.* Structural and Functional Dissection of Mif2p, a Conserved DNA-binding Kinetochore Protein. *MBoC* **19**, 4480–4491 (2008).

88. Pesenti, M. E., Weir, J. R. & Musacchio, A. Progress in the structural and functional characterization of kinetochores. *Curr Opin Struct Biol* **37**, 152–163 (2016).

89. Guo, L. Y. *et al.* Centromeres are maintained by fastening CENP-A to DNA and directing an arginine anchor-dependent nucleosome transition. *Nat Commun* **8**, 15775 (2017).

90. Nagpal, H. *et al.* Dynamic changes in CCAN organization through CENP-C during cell-cycle progression. *Mol Biol Cell* **26**, 3768–3776 (2015).

91. Ariyoshi, M. *et al.* Cryo-EM structure of the CENP-A nucleosome in complex with phosphorylated CENP-C. *EMBO J* **40**, e105671 (2021).

92. Watanabe, R. *et al.* CDK1-mediated CENP-C phosphorylation modulates CENP-A binding and mitotic kinetochore localization. *J Cell Biol* **218**, 4042–4062 (2019).

93. Hayashi, T. *et al.* Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* **118**, 715–729 (2004).

94. Fujita, Y. *et al.* Priming of Centromere for CENP-A Recruitment by Human hMis18α, hMis18β, and M18BP1. *Developmental Cell* **12**, 17–30 (2007).

95. Maddox, P. S., Hyndman, F., Monen, J., Oegema, K. & Desai, A. Functional genomics identifies a Myb domain–containing protein family required for assembly of CENP-A chromatin. *Journal of Cell Biology* **176**, 757–763 (2007).

96. French, B. T. & Straight, A. F. CDK phosphorylation of Xenopus laevis M18BP1 promotes its metaphase centromere localization. *The EMBO Journal* **38**, e100093 (2019).

97. Sandmann, M. *et al.* Targeting of Arabidopsis KNL2 to Centromeres Depends on the Conserved CENPC-k Motif in Its C Terminus. *Plant Cell* **29**, 144–155 (2017).

98. Kral, L. Possible identification of CENP-C in fish and the presence of the CENP-C motif in M18BP1 of vertebrates. *F1000Res* **4**, 474 (2015).

99. Hori, T. *et al.* Association of M18BP1/KNL2 with CENP-A Nucleosome Is Essential for Centromere Formation in Non-mammalian Vertebrates. *Developmental Cell* **42**, 181-189.e3 (2017).

100. French, B. T., Westhorpe, F. G., Limouse, C. & Straight, A. F. Xenopus laevis M18BP1 Directly Binds Existing CENP-A Nucleosomes to Promote Centromeric Chromatin Assembly. *Developmental Cell* **42**, 190-199.e10 (2017).

101. Dambacher, S. *et al.* CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. *Nucleus* **3**, 101–110 (2012).

102. Spiller, F. *et al.* Molecular basis for Cdk1-regulated timing of Mis18 complex assembly and CENP-A deposition. *EMBO Rep* **18**, 894–905 (2017).

103. Pan, D. *et al.* CDK-regulated dimerization of M18BP1 on a Mis 18 hexamer is necessary for CENP-A loading. *eLife* **6**, e23352 (2017).

104. Pan, D. *et al.* Mechanism of centromere recruitment of the CENP-A chaperone HJURP and its implications for centromere licensing. *Nat. Commun.* **10**, 4046 (2019).

105. Catania, S. & Allshire, R. C. Anarchic centromeres: deciphering order from apparent chaos. *Curr Opin Cell Biol* **26**, 41–50 (2014).

106. Jansen, L. E. T., Black, B. E., Foltz, D. R. & Cleveland, D. W. Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.* **176**, 795–805 (2007).

107. Nardi, I. K., Zasadzińska, E., Stellfox, M. E., Knippler, C. M. & Foltz, D. R. Licensing of Centromeric Chromatin Assembly through the Mis18α-Mis18β Heterotetramer. *Mol Cell* **61**, 774–787 (2016).

108. Moree, B., Meyer, C. B., Fuller, C. J. & Straight, A. F. CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *Journal of Cell Biology* **194**, 855–871 (2011).

109. Barnhart, M. C. *et al.* HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J Cell Biol* **194**, 229–243

(2011).

110. McKinley, K. L. & Cheeseman, I. M. Polo-like kinase 1 licenses CENP-A deposition at centromeres. *Cell* **158**, 397–411 (2014).

111. Silva, M. C. C. *et al.* Cdk Activity Couples Epigenetic Centromere Inheritance to Cell Cycle Progression. *Developmental Cell* **22**, 52–63 (2012).

112. Dunleavy, E. M. *et al.* HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485–497 (2009).

113. Foltz, D. R. *et al.* Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* **137**, 472–484 (2009).

114. Zasadzińska, E., Barnhart-Dailey, M. C., Kuich, P. H. J. L. & Foltz, D. R. Dimerization of the CENP-A assembly factor HJURP is required for centromeric nucleosome deposition. *The EMBO Journal* **32**, 2113–2124 (2013).

115. Subramanian, L. *et al.* Centromere localization and function of Mis18 requires Yippee-like domain-mediated oligomerization. *EMBO Rep.* **17**, 496–507 (2016).

116. Perpelescu, M. *et al.* HJURP is involved in the expansion of centromeric chromatin. *Mol. Biol. Cell* **26**, 2742–2754 (2015).

117. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol* **403**, 1–10 (2010).

118. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to

histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**, 19–42 (1998).

119. Arimura, Y., Tachiwana, H., Oda, T., Sato, M. & Kurumizaka, H. Structural Analysis of the Hexasome, Lacking One Histone H2A/H2B Dimer from the Conventional Nucleosome. *Biochemistry* **51**, 3302–3309 (2012).

120. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. in *Chromatin and Chromatin Remodeling Enzymes, Pt A* (eds. Allis, C. D. & Wu, C.) vol. 375 23–44 (Elsevier Academic Press Inc, 2004).

121. Fang, J. *et al.* Structural transitions of centromeric chromatin regulate the cell cycle-dependent recruitment of CENP-N. *Genes Dev* **29**, 1058–1073 (2015).

122. Allu, P. K. *et al.* Structure of the Human Core Centromeric Nucleosome Complex. *Curr Biol* **29**, 2625-2639.e5 (2019).

123. Arimura, Y. *et al.* The CENP-A centromere targeting domain facilitates H4K20 monomethylation in the nucleosome by structural polymorphism. *Nat Commun* **10**, 576 (2019).

124. Kato, H., Zhou, B.-R., Feng, H. & Bai, Y. An evolving tail of centromere histone variant CENP-A. *Cell Cycle* **12**, 3133–3134 (2013).

125. Ali-Ahmad, A., Bilokapić, S., Schäfer, I. B., Halić, M. & Sekulić, N. CENP-C unwraps the human CENP-A nucleosome through the H2A C-terminal

tail. *EMBO reports* **20**, e48913 (2019).

126. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

127. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36–51 (2005).

128. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

129. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

130. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

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

132. Pettersen, E. F. *et al.* UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

133. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D-Struct. Biol.* **75**, 861–877 (2019).

134. Pesenti, M. E. *et al.* Structure of the human inner kinetochore CCAN

complex and its significance for human centromere organization. *Molecular Cell* **82**, 2113-2131.e8 (2022).

135. Stellfox, M. E., Nardi, I. K., Knippler, C. M. & Foltz, D. R. Differential Binding Partners of the Mis18α/β YIPPEE Domains Regulate Mis18 Complex Recruitment to Centromeres. *Cell Rep* **15**, 2127–2135 (2016).

136. Okumura, E., Sekiai, T., Hisanaga, S., Tachibana, K. & Kishimoto, T. Initial triggering of M-phase in starfish oocytes: a possible novel component of maturation-promoting factor besides cdc2 kinase. *J Cell Biol* **132**, 125–135 (1996).

137. Hara, M., Ariyoshi, M., Okumura, E.-I., Hori, T. & Fukagawa, T. Multiple phosphorylations control recruitment of the KMN network onto kinetochores. *Nat Cell Biol* **20**, 1378–1388 (2018).

Acknowledgement

November 8, 2019 was the day I first came to Japan to meet Prof. Fukagawa. We met hastily and happily at the elevator at 7:00 p.m., and then the lab students kindly invited me to dinner. This was my first impression of Japan and the Fukagawa Lab. I felt the kindness of prof. Fukagawa and the enthusiasm of the lab members, and as a foreigner coming to Japan for the first time, I was grateful for this warm and reassuring start.

I was very happy to meet Ariyoshi-sensei, a teacher who took care of me like a family member in school and life after that, for the first time on the first working day afterwards, and I started to work on my doctoral thesis under the professional and kind guidance of Prof. Fukagawa and Ariyoshi-sensei. Every day, the lab members would chat with me kindly and gently. In such a process of getting along, I am grateful that people are considerate like friends and make me feel like family.

All these warm memories are clear and beautiful, and now it seems like it happened yesterday when I think about it. Time flies, and suddenly I was shocked to realize that almost four years have passed since the first day I remembered.

There is a Japanese proverb that goes like this, '一期一会'. I love this quote and will carry this state of mind, cherish the fate of meeting and getting along with everyone here, and be grateful for all the kindness and goodness I encounter here in Fukagawa-lab.

I would like to thank Prof. Fukagawa for recognizing and accepting me and giving me the opportunity to receive the most professional guidance and advice in a cutting-edge laboratory in the field of chromosome biology. I am also very grateful to Prof. Fukagawa for all the help he has provided me over the years, and I have learned from him and lab members the traits that a good researcher needs to have.

I would like to express my deep gratitude to Ariyoshi-sensei for her professional guidance on the details of my experiments, and her and Prof. Fukagawa's excellent academic ability for supporting me to better grasp the progress of my project.

Thanks to Hara-sensei, Hori-sensei, Takenoshita-sensei, Watanabe-san, Shreyas-san,

Yamaguchi-san, Hirano-sansei, Fukui-san, Cao-san, Kong-san, Miao-san, Sha-san, zhou-san and Li-san et al., for their kind and helpful academic guidance and help.

Thanks to Fukagawa-san, Oshimo-san, Fukuoka-san and Kubota-san for their professional help in basic experiments, I have more time to focus on experimental work.

Thanks to my old and new friends who have accompanied me on this journey of growth over the years.

I would like to express my special thanks to Makino-san and all the seniors of Namba lab who helped me to learn cryo-electron microscopy.

Finally, I would like to thank Ueda-sensei, Hirose-sensei and Okamoto-sensei for their valuable comments and guidance on the defense of my doctoral dissertation. And pecial thanks to the China scholarship council (CSC) scholarship for supporting my life and research in Japan.

As my four years of valuable doctoral studies in Japan are coming to an end, this is the last answer that I have seriously completed as a student. I am honored to have all of you mentioned above or not mentioned above to witness my growth during this journey.

I will face the rest of my life journey with more maturity and return to my loving family who have always supported me and loved me deeply. And I look forward to meeting you all in Japan again in the future!

Here, I look forward to being a contributing researcher in this path of scientific research.

Grateful to meet, grateful to walk together.

Achievement

PUBLICATION

1. Hong Hui Jiang, Mariko Ariyoshi, Tetsuya Hori, Reito Watanabe, Fumiaki Makino, Keiichi Namba, Tatsuo Fukagawa. The cryo-EM structure of the CENP-A nucleosome in complex with ggKNL2. EMBO J (Article). 2023;42(6):e111965. doi:10.15252/embj.2022111965.

2. Hong Hui Jiang, Bo Li, Yue Ma, Su Ying Bai, Thomas D. Dahmer, Adrian Linacre, Yan Chun Xu. Forensic Validation of a Panel of 12 SNPs for Identification of Mongolian Wolf and Dog. Sci Rep 10, 13249 (2020). https://doi.org/10.1038/s41598-020-70225-5.

3. Honghui Jiang, Bo Li, Yue Ma, Yanchun Xu. Feasibility Analysis of Species Identification of Dogs and Wolves by Partial Sequence of mtDNA Control Region. Chinese Journal of Wildlife, 2018, 039(002):277-285.

POSTER PRESENTATION

1. 2022 The 45th Annual Meeting of the Molecular Biology Society of Japan (MBSJ)- poster presentation. Member of the Molecular Biology Society of Japan.

2. 2021 The 44th Annual Meeting of the Molecular Biology Society of Japan (MBSJ)- poster presentation. Member of the Molecular Biology Society of Japan.

ORAL PRESENTATION

2018 Excellent Academic Report, The Ninth Symposium on Animal Research and Conservation in Northeast provinces, China.

AWARDS

1. 2021-2023 China Scholarship Council (CSC) Chinese Government Scholarship

2. 2018 Excellent Academic Report, The Ninth Symposium on Animal Research and

Conservation in Northeast provinces, China.

MAIN CONTRIBUTION AS A CO-AUTHOR

I prepared materials and performed all biochemical experiments under supervision of Dr. Mariko Ariyoshi; Dr. Mariko Ariyoshi, Dr. Fumiaki Makino and I performed all cryo-EM experiments and Dr. Keiichi Namba helped cryo-EM single particle image; Dr. Tetsuya Hori and Dr. Reito Watanabe performed DT 40 experiments; Prof. Tatsuo Fukagawa supervised the entire project; I wrote this manuscript and Prof. Tatsuo Fukagawa and Dr. Mariko Ariyoshi supervised my revision.