Doctoral Dissertation

博士論文

Structural basis of centromeric chromatin formation involving the CENP-A nucleosome

（CENP-A ヌクレオソームによるセントロメアのクロマチン形成の構造基盤）

A Dissertation Submitted for the Degree of Doctor of Philosophy
July 2023
令和 5 年 7 月博士（理学）申請

Department of Biological Sciences, Graduate School of Science,
The University of Tokyo
東京大学大学院理学系研究科
生物科学専攻

HO Cheng-Han
何　承翰

Contents

Figures and Tables

Abstract

The genomic DNA of the eukaryotic cell is packed into the cell nucleus by forming chromatin, which consists of nucleosomes as the fundamental unit. The nucleosome consists of DNA wrapped around the histone octamer, formed from 2 copies of histone H2A, H2B, H3, and H4. The centromere chromatin is a special region of chromatin that is epigenetically defined by CENP-A, a histone H3 variant, and is responsible for chromosome segregation. In this study, I investigated 2 epigenetic features of the CENP-A nucleosome: the flexible DNA ends and the mono-methylation of H4K20 (H4K20me1).

To investigate the effect of the flexible DNA ends of the CENP-A nucleosome on higher-order chromatin structure, tri-nucleosomes with a center-positioned CENP-A nucleosome or H3 nucleosome were reconstituted *in vitro* to mimic the centromeric chromatin. Subsequently, cryo-EM analyses of the reconstituted tri-nucleosomes were performed. From the cryo-EM structures, it can be seen that the tri-nucleosomes with a center positioned CENP-A nucleosome are less twisted. This may enable the CENP-A nucleosomes to be exposed from surrounding H3 nucleosomes in a poly-nucleosome fiber and contribute to kinetochore assembly.

On the other hand, it has been shown that the H4K20 in the CENP-A nucleosome is mono-methylated in cells, and that the structural characteristics of the CENP-A nucleosome may lead to higher H4K20 mono-methylation rates compared to the canonical H3 nucleosome. The H4K20 mono-methylation reaction is solely catalyzed by the methyltransferase SET8 (also named PR-Set7/KMT5A). However, how SET8 recognizes the nucleosomal H4K20 substrate, and why CENP-A nucleosome increases the mono-methylation rates of the H4 tail remain elusive. To address these issues, the

SET8-CENP-A nucleosome complex and the SET8-H3 nucleosome complex were reconstituted *in vitro* and cryo-EM analyses were performed.

The cryo-EM structures of the SET8-CENP-A nucleosome complex and the SET8-H3 nucleosome complex share a similar appearance. In both structures, SET8 is located above the nucleosome disc and interacts directly with the acidic patch of nucleosome, which is formed by a cluster of negatively charged residues from H2A and H2B. Apart from the acidic patch, SET8 can also be seen interacting with the histone H4 tail. The H4 tail extends from the nucleosome disc surface into the SET domain of SET8.

Importantly, I discovered that the H4 tail of the H3 nucleosome is repositioned upon the binding of SET8. Interestingly, this new conformation of the H4 tail in the SET8-bound nucleosome is the conformation that has been suggested to be preferred by the free CENP-A nucleosome. Therefore, the H4 tail conformation in the free CENP-A nucleosome may be preferred by SET8, leading to a higher methylation rate.

This study provides the basis for understanding how the structural features of the CENP-A nucleosome may affect the higher-order structure and histone modification of centromeric chromatin.

Abbreviations

APM: acidic patch mutant

bp: base pair

CATD: CENP-A targeting domain

CBB: Coomassie brilliant blue

CCAN: constitutive centromere associated network

CDE: centromere DNA element

CENP: centromere protein

CIA: chloroform-isoamyl alcohol

COOT: crystallographic object-oriented toolkit

Cryo-EM: cryogenic electron microscopy

CTF: contrast transfer function

DM: double mutant

DNA: deoxyribonucleic acid

DTT: dithiothreitol

ECL: Enhanced Chemiluminescence

EDTA: ethylenediaminetetraacetic acid

EMDB: Electron Microscopy Data Bank

EMSA: electrophoretic mobility shift assay

EtBr: ethidium bromide

FSC: Fourier shell correlation

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HJURP: Holliday junction recognition proteins

HRP: Horseradish peroxidase

IPTG: isopropyl-β-D-thiogalactopyranoside

LB: lysogeny broth

Ni-NTA: nickel-nitrilotriacetic acid

PAGE: poly-acrylamide gel electrophoresis

PBS-T: phosphate buffered saline-tween

PCR: polymerase chain reaction

PDB: Protein Data Bank

PTM: post-translational modification

PVDF: polyvinylidene difluoride

SD: semi-dry

SDS: sodium dodecyl sulfate

SET: Su(var)3-9, Enhancer-of-zeste and Trithorax

TB: terrific broth

TBE: tris-borate-EDTA buffer

TCEP: tris(2-carboxyethyl)phosphine

Tris: tris(hydroxymethyl)aminomethane

UCSF: University of California, San Fransico

Chapter 1: General Introduction

1.1. The function and structure of chromatin

In eukaryotic cells, the genomic DNA is packed compactly into the cell nucleus by forming chromatin (Fig. 1A) (1–3). Chromatin is a nucleoprotein complex, which consists of nucleosomes as the fundamental unit (4). Nucleosomes are composed of DNA wrapped around a histone octamer, formed by 2 of each core histones H2A, H2B, H3, and H4 (Fig. 1A, lower panel) (5). From electron microscopy images, it can be seen that the nucleosomes form a "beads-on-a-string" structure, in which the nucleosomes are the beads (1,2,6). The nucleosomes are joined by linker DNAs.

Chromatin structure can be divided into loosely compacted euchromatic regions and densely compacted heterochromatic regions (7). The aforementioned structural difference can be regulated by several epigenetic factors, including the incorporation of histone variants, and the post-translational of histones (Fig. 1B&C) (8).

Moreover, some chromatin regions have specific functions that are particularly significant during mitosis. As a cell divides, DNA condenses into the distinctive structures known as mitotic chromosomes. The ends of a chromosome are known as the telomeres, whereas the place where the duplicated chromosomes is joined, is known as the centromere (Fig. 1A). The centromere has an important role in chromosome segregation because it is the place where the kinetochore, a complicated protein complex responsible for attaching the duplicated chromosomes to the mitotic spindle, is formed (9,10).

In this study, I focused on several epigenetic features of the centromeric chromatin, and investigated how these features may affect centromere structure, and how these features are introduced to centromeric chromatin regions.

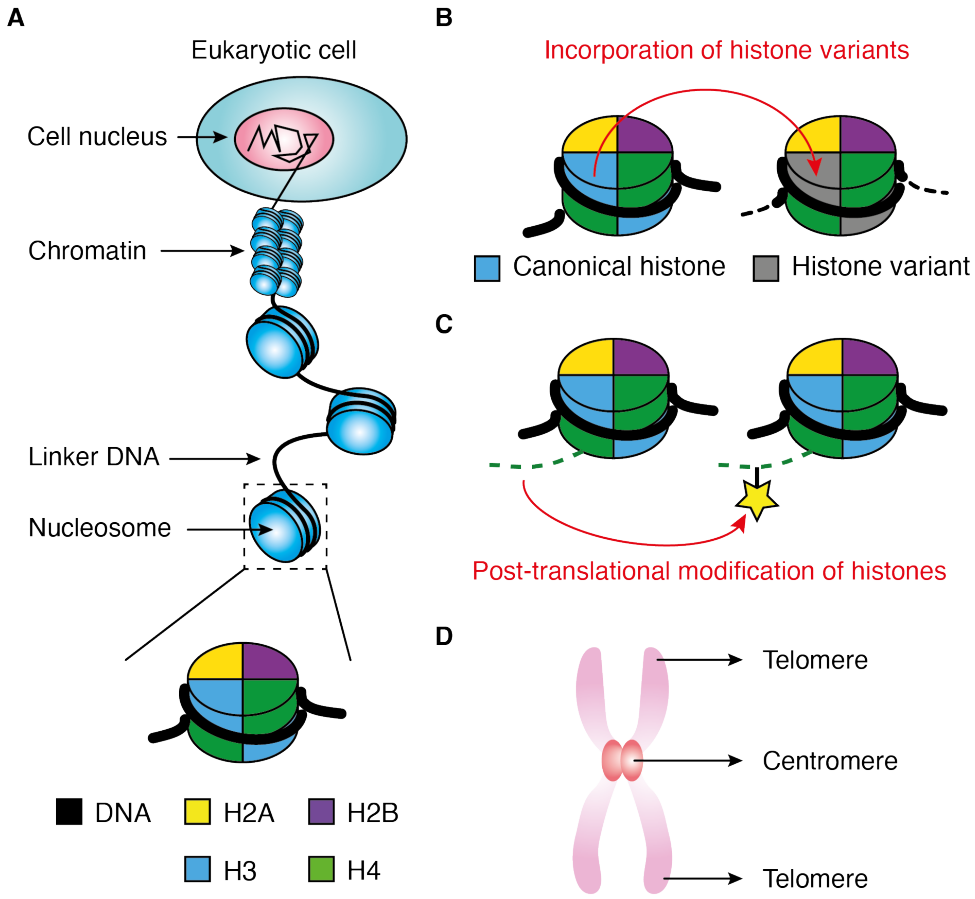


Fig. 1 Regulation of chromatin structure

(A) Schematic model depicting the chromatin hierarchy in the cell nucleus. The fundamental unit of chromatin is the nucleosome, which consists of DNA wrapped around the histone octamer. Nucleosomes are joined by linker DNAs. **(B)** Schematic model depicting the regulation of chromatin structure by the incorporation of histone variants. **(C)** Schematic model depicting the regulation of chromatin structure by the post-translational modification of histones. **(D)** Schematic model depicting the mitotic chromosome. The telomeres and the centromere are shown.

1.2. The nucleosome

The histone core of the nucleosome consists of the 4 core histones, H2A, H2B, H3, and H4. All core histones contain a histone fold domain, which are made up of 3 alpha helices and 2 loops (Fig. 2A) (11). Apart from the histone fold domain, there is also a N-terminal tail in each histone. The histone tail is subject to numerous kinds of post-translational modifications, which can serve as epigenetic marks to recruit their specific interacting partners to chromatin (8,12–15). However, the tails of the histones are often not visible in crystal structures or cyro-EM structures, possibly due to their high movability in regard to the rest of the elements in a nucleosome.

The histone folds enable histone H2A and H2B to form a dimer, through a handshake interaction (Fig. 2B). Similarly, histone H3 and H4 can first form a dimer, and further dimerize in to an $(H3\text{-}H4)_2$ tetramer. Two histone H2A-H2B dimers and one $(H3\text{-}H4)_2$ tetramer can then form the core of the nucleosome, or the histone octamer (11). Finally, approximately 145~147 bp of DNA can wrap around the histone octamer, by the positively charged lysine or arginine residues present within the histones, and the negatively charged DNA backbone, to form the nucleosome (Fig. 2C) (16).

The surface charge of the nucleosome can be displayed based on its crystal structure (Fig. 2D) (17). Apart from the obvious negatively charged DNA backbone, an “acidic patch” can also be observed on the nucleosome face (18,19). This is formed by several negatively charged amino acid residues in histone H2A (E56, E61, E64, D90, E91, E92) and H2B (E105, E113). The acidic patch is a binding spot of many nucleosome-interacting factors (20), including LANA (21), cGAS (22–27), and DOT1L (28–31).

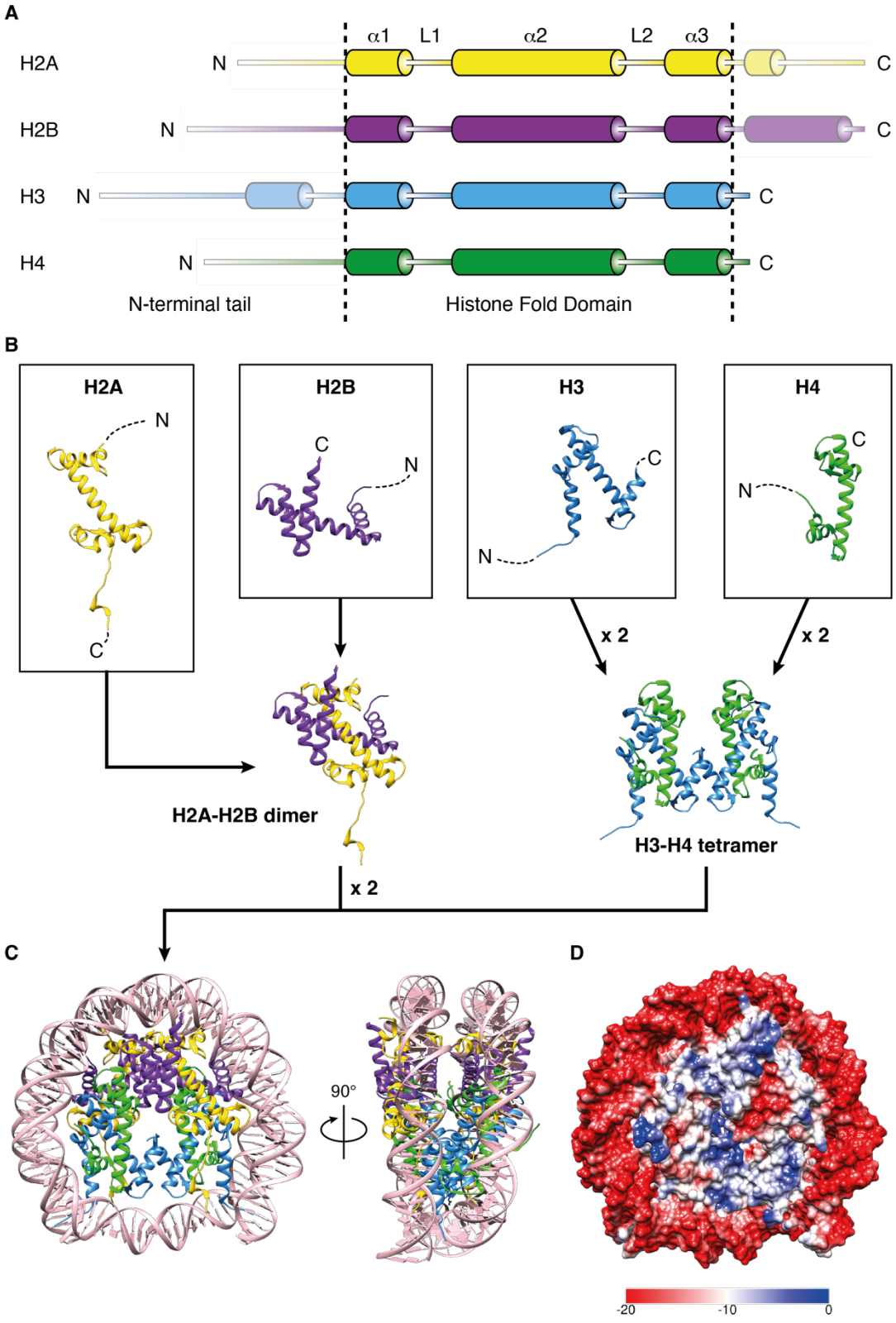

A
α1
L1
α2
L2
α3
H2A
H2B
H3
H4
N
C
N-terminal tail
Histone Fold Domain
B
H2A
H2B
H3
H4
x 2
x 2
H2A-H2B dimer
H3-H4 tetramer
x 2
C
90°
D
-20
-10
0

Fig. 2 Composition and surface charge of the nucleosome

(A) Histone H2A, H2B, H3, and H4 share a common motif known as the histone fold domain. Additionally, they also have disordered N-terminal tails that often cannot be observed in nucleosome structures. **(B)** Different components of the histone octamer found in a nucleosome, including 2 H2A-H2B dimers and 1 $(H3\text{-}H4)_2$ tetramer. The histone components were taken from a nucleosome crystal structure (PDB:3LZ0) and visualized by Chimera. **(C)** Crystal structure of the nucleosome (PDB: 3LZ0), visualized by Chimera. **(D)** Surface charge of the nucleosome (PDB: 3LZ0), visualized by Chimera.

1.3. Histone variants

Apart from the canonical core histones, there also exist histone variants that can substitute for the canonical histones in a nucleosome (Fig. 1B) (32). Each canonical histone (H2A, H2B, H3, and H4) has its corresponding histone variant or variants (32). Canonical histones and histone variants share some similarities, such as both being composed of a histone fold domain and a tail that can be post-translationally modified. On the other hand, they also have several significant differences.

First, the expression and incorporation timing of canonical histones and histone variants are different (32,33). Canonical histones are expressed in the S phase and their incorporation into chromatin is replication coupled. The incorporation and expression of histone variants, however, are not coupled to the cell cycle. Canonical histones and histone variants may also have different chaperones responsible for their deposition on the genome (32).

Second, the gene structure of canonical histones and histone variants are different (32). The gene of canonical histones are intronless and exist as multiple copies in "histone clusters". The transcribed mRNAs of the canonical histones have a special stem-loop structure at the end instead of the typical polyA tail (34). On the other hand, the genes of histone variants are more like other typical genes, containing introns and polyA tails. They could be spliced differently, resulting in isoforms (32). They usually only have one or two genes and exist outside of the histone clusters.

The histone variants can alter nucleosome properties and therefore affect chromatin structure, resulting in a change of gene expression (32,35). The variants may be cell type-specific or occupy a particular genomic region (32). How histone variants contribute to the fine tuning of chromatin structure remains an important issue to be understood.

1.4. Histone post-translational modifications (PTMs)

Another factor that affects chromatin structure and alter gene expression is the post-translational modification (PTM) of the histones (Fig. 1C) (8,13,14). The PTMs are established by enzymes that are termed “writers”, and recognized by other proteins that are termed “readers” (8,15). The PTMs can also be removed by enzymes termed “erasers” (8,15).

Some common modifications include the acetylation and methylation of histones. For example, the acetylation of histones is often linked to the decompaction of the chromatin, resulting in active transcription (8,15). On the other hand, the methylation mark can be both a transcription active or repressive mark, depending on the modified residue or the state of methylation (mono-methylation, di-methylation, or tri-methylation) (8,15). Some active marks include the H3K4me3 mark and the H3K36me3 mark (8,15). Conversely, some repressive marks include the H3K9me3 mark, responsible for the formation of constitutive heterochromatin, and the H3K27me3 mark, responsible for the formation of facultative heterochromatin (8,15,36–39).

Histone PTMs have also be known to be linked to recombination, DNA repair, DNA replication, and genome topology (8). Furthermore, canonical histones and histone variants can be modified differently, adding on to the complexity of the regulatory system of chromatin (8). Consequently, it is crucial to investigate how the PTMs of histones are established by specific enzymes.

1.5. The centromere and its epigenetic features

During cell division, the chromosomes are segregated equally into daughter cells. In this process, the kinetochore can connect the chromosomes to the microtubules of the mitotic or meiotic spindles for distributing the chromosomes. Eukaryotic chromosomes fall into two main types: holocentric and monocentric, depending on where the kinetochore can be assembled (Fig. 3A&B) (9,10).

In the case of a holocentric chromosome, the kinetochore can be assembled along the entire length of the chromosome. On the other hand, in the case of a monocentric chromosome, the kinetochore can only be assembled on a single specific position of the chromosome, the centromere. Most eukaryotes have monocentric chromosomes, which can be further categorized into point centromeres and regional centromeres (Fig. 3C) (40). Budding yeasts, for example, have point centromeres. The centromere is assembled on a single CENP-A^{Cse4} nucleosome that contains a specific DNA sequence of approximately 125 bp, composed of 3 conserved DNA elements CDEI, CDEII, and CDEIII (9,10,40). The CDEs are necessary and sufficient for centromere formation (Fig. 3C, upper panel). CDEI is bound by the CBF1 protein and CDEIII is bound by the CBF3 complex (41,42). The CBF3 complex and CDEIII are essential for centromere function. On the other hand, CDEII is AT rich and serve as the main contacting component of the CENP-A^{Cse4} nucleosome.

Conversely, organisms with regional centromeres assemble their centromeres on repetitive DNA sequences, which can be several kilobases to several megabases long. Humans also have regional centromeres, that forms along α-satellite repeats (43). The α-satellite repeats are composed of monomers that are approximately 170 bp in length (44). The adjacent monomers have low similarity, but they can form higher-order repeats (Fig.

3C, lower panel) (9,10).

These DNA sequences alone are not sufficient for centromere function in organisms with regional centromeres. Instead, regional centromeres are marked and defined by a special histone H3 variant found at the centromere, CENP-A (Fig. 3C) (45–47). CENP-A-containing nucleosomes (CENP-A nucleosomes) are spread among H3-containing nucleosomes (H3 nucleosomes) (9,48–51). The ratio of H3 nucleosomes to CENP-A nucleosomes vary upon species, and upon different chromosomes within the same species. In human, the ratio is estimated to be approximately 1 CENP-A nucleosome per 4~5 nucleosomes, or approximately 1 CENP-A nucleosome per 25 nucleosomes, depending on the method for measuring (9,52,53).

The CENP-A nucleosome that is found in centromeric chromatin has 2 epigenetic features, which will be discussed in more detail in the following chapters. First, compared to canonical H3 nucleosomes, the CENP-A nucleosome has more flexible DNA ends (which will be introduced in more detail in Section 2.1.1) (54). Second, the H4 in the CENP-A nucleosome of centromeric chromatin is mono-methylated at the 20th lysine (H4K20) (55). This H4K20me1 mark is essential for kinetochore assembly (which will be introduced in more detail in Section 3.1). However, how the flexible DNA ends of the CENP-A nucleosome affect centromeric chromatin structure remains elusive. Additionally, how the H4K20me1 mark is installed to the CENP-A nucleosome remains unclear. Therefore, in this study, I applied cryo-EM analysis to address these 2 epigenetic features from a structural point of view.

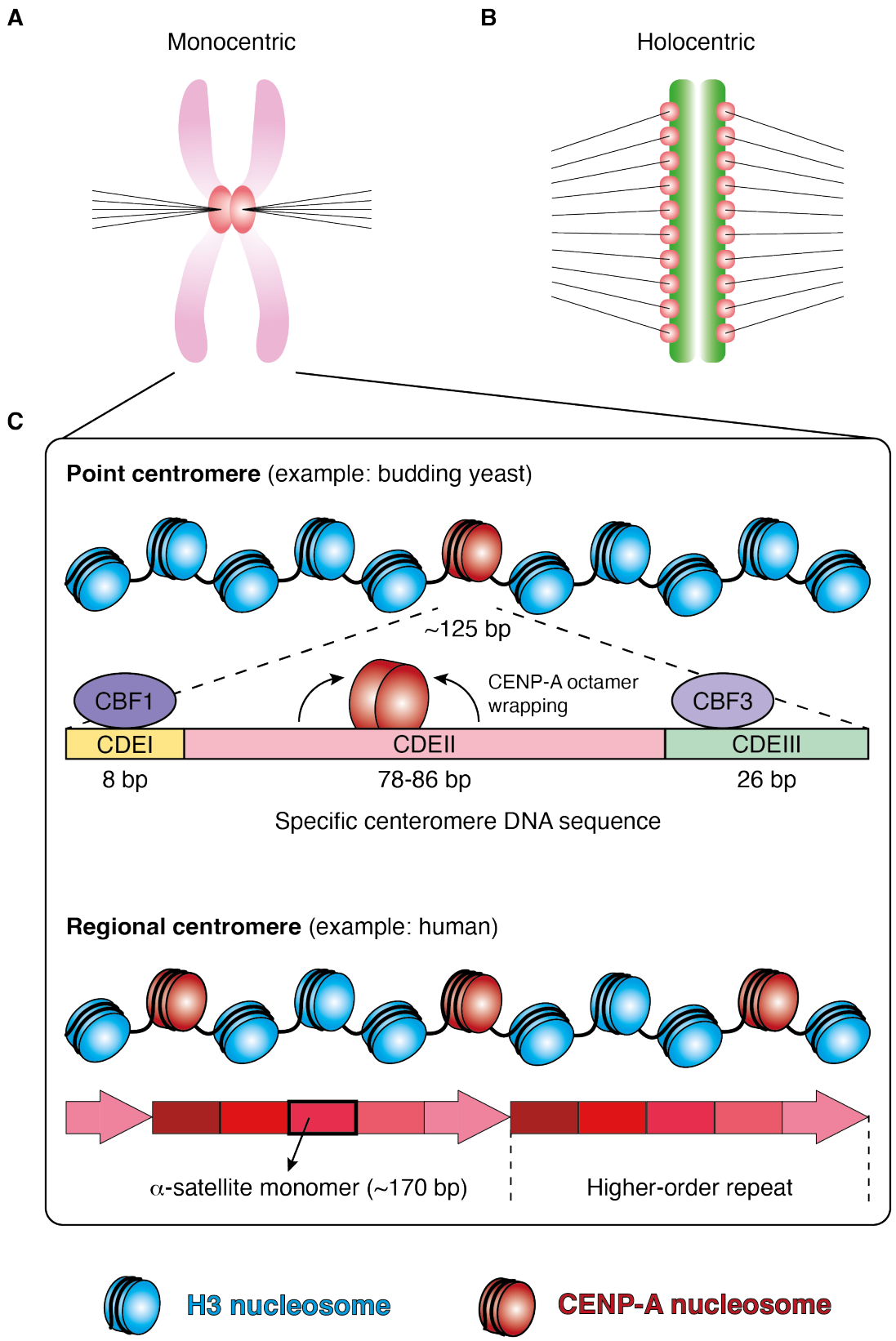

A
Monocentric
B
Holocentric
C
Point centromere (example: budding yeast)
~125 bp
CBF1
CENP-A octamer wrapping
CBF3
CDEI
CDEII
CDEIII
8 bp
78-86 bp
26 bp
Specific centeromere DNA sequence
Regional centromere (example: human)
α-satellite monomer (~170 bp)
Higher-order repeat
H3 nucleosome
CENP-A nucleosome

Fig. 3 Eukaryotic centromeres

(A) A monocentric chromosome. **(B)** A holocentric chromosome. **(C)** In monocentric chromosomes, the centromere can be further divided into the point centromere and the regional centromere. Point centromeres (the upper panel) can be found in budding yeast, and they only contain a single CENP-A nucleosome. They also contain a specific DNA sequence (CDEI~III) that is sufficient for centromere function. CBF1 can bind CDEI, and the CBF3 complex can bind CDEIII. Regional centromeres (the lower panel), on the other hand, contain numerous CENP-A nucleosomes and assemble on a large DNA region, which is often repetitive. In human, the centromeres contain α-satellite monomers, which are approximately 170 bp in length. Adjacent monomers are different to each other, but they form higher-order repeats. CDE: centromere DNA element.

Chapter 2: Structural Analysis of the Fundamental Unit of Centromeric Higher-Order Chromatin

2.1. Introduction

2.1.1. Features of the histone H3 variant CENP-A

There are many H3 variants, among which CENP-A is the variant with the lowest similarity. Only approximately 50% amino acid identity is conserved between CENP-A and the canonical H3.1 (which will be referred to H3 in this study for simplicity) (Fig. 4A) (56). CENP-A contains a CENP-A targeting domain (CATD) that is important for many of its functions, such as the centromere localization of CENP-A (57,58), and the deposition of CENP-A by its chaperone HJURP (59–61). The CATD is located in the histone fold domain of CENP-A, and to be specific, the L1 and the α2 helix (Fig. 4A) (58).

2.1.2. Structural features of the CENP-A nucleosome

Our team previously reported the crystal structure of the CENP-A nucleosome (54). Interestingly, we found that it closely resembles the typical H3 nucleosome structure (Fig. 4B&C). Both involve DNA wrapped around a histone octamer. In the case of the CENP-A nucleosome, the H3 components are replaced by CENP-A molecules.

One intriguing distinction, however, lies in the length of the DNA wrapped around the histone octamer in the CENP-A nucleosome. This segment is approximately 13 base pairs shorter than in the H3 nucleosome. This difference occurs despite the preparation of both nucleosomes with DNA of similar length. This suggests that the DNA ends of the CENP-A nucleosome might be more flexible, which could explain why they are not

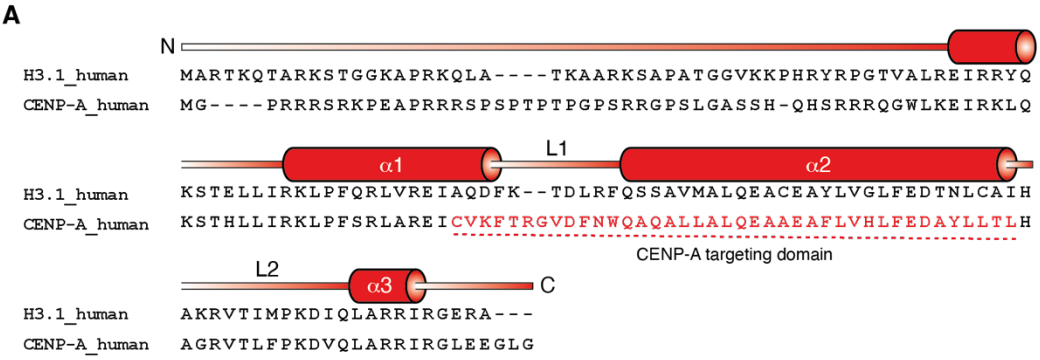
visible in the crystal structure (Fig. 4C).

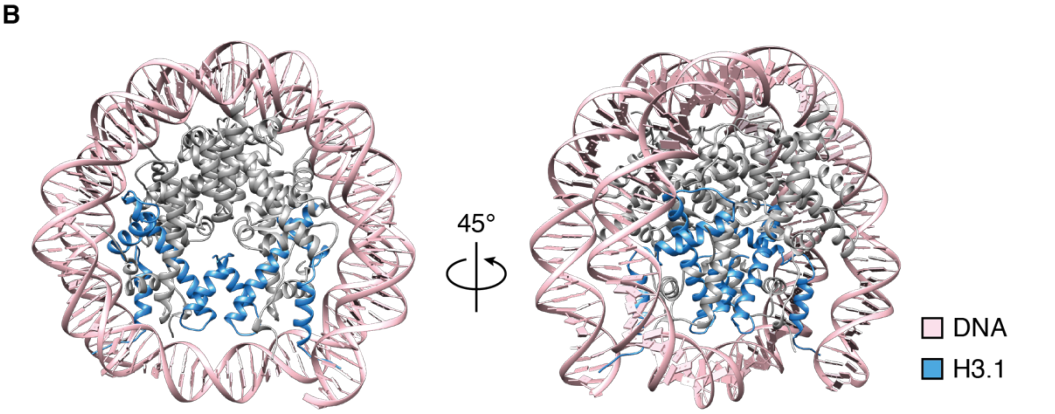
To confirm this hypothesis, further analysis was conducted using the exonuclease assay. The result showed that the DNA ends of CENP-A nucleosomes are more susceptible to exonuclease III digestion then those of the H3 nucleosomes (54). This proved that the DNA ends of the CENP-A nucleosome are indeed less protected than those of the H3 nucleosome (54). Based on these findings, we concluded that the CENP-A nucleosome features more flexible DNA at the entry and exit sites.

2.1.3. Centromeric proteins that interact with the CENP-A nucleosome

The kinetochore can be divided into the outer kinetochore and the inner kinetochore (Fig. 5). The outer kinetochore is responsible for binding to the microtubule ends. It is composed of 10 proteins that form the KMN network, and can be divided into 3 subcomplexes, namely the Knl1 complex, the Mis12 complex, and the Ndc80 complex (10,50).

On the other hand, the inner kinetochore directly interacts with the centromeric chromatin. The main component of the inner kinetochore is the constitutive centromere associated network (CCAN), which contains 16 protein subunits (10,50,62). CCAN has several functions, including CENP-A binding, CENP-A recruitment, KMN recruitment, and chromosome congression/oscillation (10).Two CCAN subunit proteins can interact directly with the CENP-A nucleosome and exhibit specificity for CENP-A nucleosomes over H3 nucleosomes, namely CENP-C and CENP-N (Fig. 5) (10,50). CENP-C interacts with the acidic patch of the nucleosome and the C terminal of CENP-A and connects CENP-A nucleosome with CCAN and KMN network. On the other hand, CENP-N binds directly to the CATD of CENP-A (50).


A
N
H3.1_human MARTKQTARKSTGGKAPRKQLA----TKAARKSAPATGGVKKPHRYRPGTVALREIRRYQ
CENP-A_human MG----PRRRSRKPEAPRRRSPSPTPTPGPSRRGPSLGASSH-QHSRRRQGWLKEIRKLQ
α1
L1
α2
H3.1_human KSTELLIRKLPFQRLVREIAQDFK--TDLRFQSSAVMALQEACEAYLVGLFEDTNLCAIH
CENP-A_human KSTHLLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHLFEDAYLLTLH
CENP-A targeting domain
L2
α3
C
H3.1_human AKRVTIMPKDIQLARRIRGERA---
CENP-A_human AGRVTLFPKDVQLARRIRGLEEGLG


B
45°
DNA
H3.1

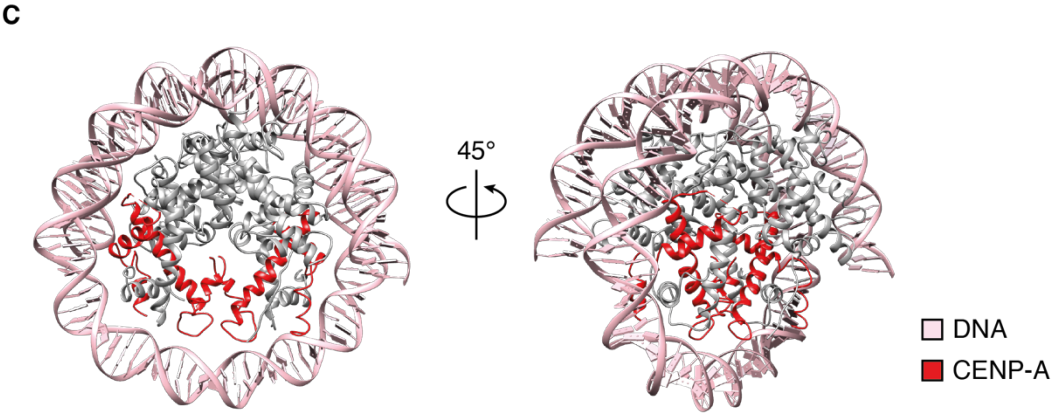

C
45°
DNA
CENP-A

Fig. 4 Structural features of CENP-A and the CENP-A nucleosome

(A) Alignment of the canonical histone H3 and its variant CENP-A. A schematic representation of the helices and loops of CENP-A is displayed above the amino acid sequences. The CATD of CENP-A is colored in red and is located in L1 and α2 of CENP-A. **(B)** Crystal structure of the H3 nucleosome (PDB:3LZ0). **(C)** Crystal structure of the CENP-A nucleosome (PDB: 3AN2).

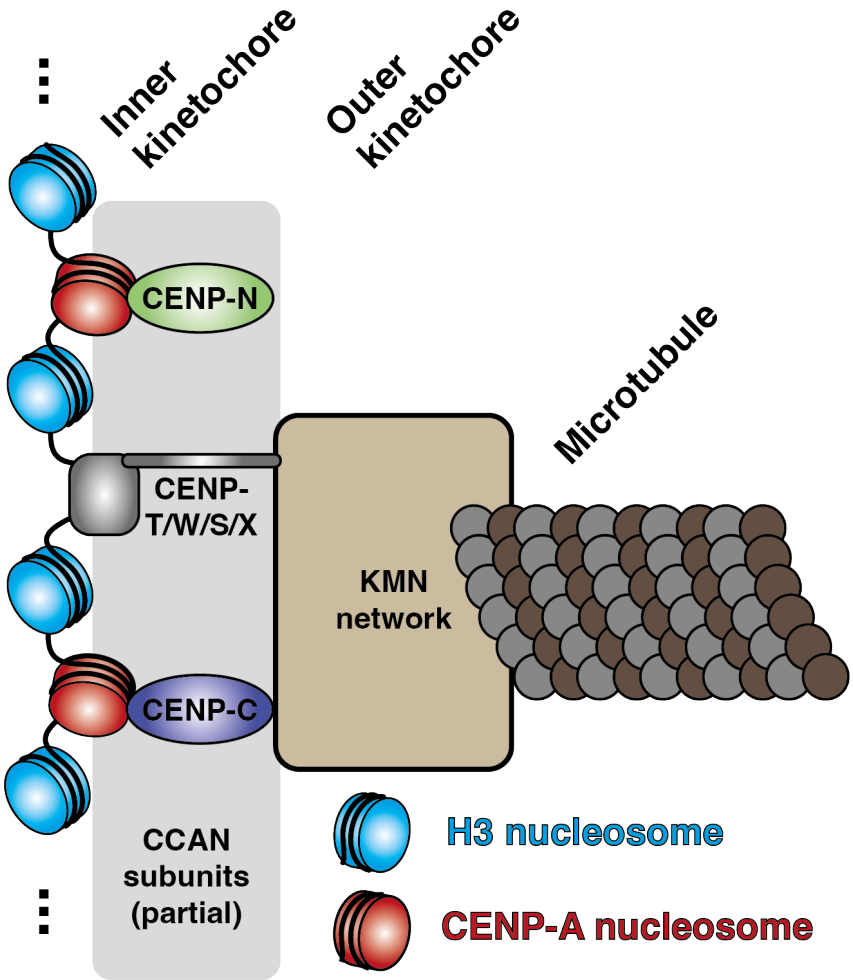


Fig. 5 A simplified schematic model of the kinetochore and centromere

The kinetochore can be divided into the inner and outer kinetochore. The inner kinetochore consists of the CCAN, while the outer kinetochore consists of the KMN network. CCAN subunits CENP-N and CENP-C of the CCAN can bind CENP-A nucleosomes directly. CCAN subunits that contribute to the recruitment of the microtubule-binding interfaces, CENP-N and CENP-T/W/S/X, are highlighted. Other CCAN subunits are excluded for clarity.

2.2. Research Aims

The dynamic nature of CENP-A nucleosome structure, which is characterized by flexible DNA ends, has been reported (54). However, the mechanism by which these flexible DNA ends influence the higher-order chromatin structure within the centromere remains largely unexplored. Additionally, the interspersed arrangement of CENP-A nucleosomes amid an abundance of H3 nucleosomes raises further questions (9,48–51). Specifically, the processes through which CENP-A nucleosomes are identified by centromeric proteins such as CENP-C and CENP-N for subsequent kinetochore assembly remains unclear. Accordingly, this study hypotheses that an examination of the centromeric higher-order chromatin structure, with specific emphasis on its fundamental component, the CENP-A nucleosome, may shed light on these critical biological phenomena.

In pursuit of these objectives, I prepared a variety of tri-nucleosomes, which are the fundamental components of the higher-order chromatin structure. This was followed by a thorough analysis using cryo-EM. This investigation helped to uncover the influence of the CENP-A nucleosome and the length of linker DNA on the structure of chromatin. In summary, the obtained cryo-EM structures of the tri-nucleosomes, which contain CENP-A nucleosomes, contributed to a structural model for the basis of centromeric chromatin.

2.3. Methods

2.3.1. Purification of human histones

The individual preparation of Histones H2A, H2B, H3, H4, and CENP-A was carried out as previously outlined (63). Briefly, histones featuring an N-terminal His_6-tag were expressed in Escherichia coli cells via incubation at 37 °C overnight. The harvested cells underwent lysis through sonication, after which the insoluble fractions were isolated. These fractions were then mixed with the denaturing buffer and subjected to rotation at 4 °C overnight. Following centrifugation, the supernatants were combined with Ni-NTA agarose beads, facilitated by rotation at 4 °C for an hour. These Ni-NTA agarose beads were subsequently relocated to Econo-Columns (Bio-Rad), washed with 5 mM imidazole, and eluted under denaturing conditions using a 5-500 mM imidazole gradient.

The eluted fractions were then dialyzed to clear away urea. Subsequently, the His_6-tags were excised by thrombin protease, and the His_6-tags and histones were separated using Mono S chromatography (GE Healthcare). Lastly, the histones were subjected to dialysis against water prior to freeze-drying. The resultant purified histones were stored at 4 °C.

2.3.2. *In vitro* reconstitution and purification of the histone octamers

The canonical histone octamer was reconstituted by using H2A, H2B, H3, and H4 histone proteins, while the CENP-A histone octamer was reconstituted with H2A, H2B, CENP-A, and H4 histone proteins. The histone proteins were combined in an equimolar ratio within the denaturing buffer (20 mM Tris-HCl (pH 7.5), 7 M guanidine hydrochloride, and 2 mM or 20 mM 2-mercaptoethanol) and subjected to rotation at 4 °C for a duration of 1.5 hours. Following this, the mixtures underwent dialysis against 500

mL of the refolding buffer (10 mM Tris-HCl (pH 7.5), 2 M NaCl, 1 mM EDTA, and 5 mM 2-mercaptoethanol) for four separate times.

Lastly, using HiLoad 16/600 Superdex 200 pg (GE Healthcare) column chromatography, the reconstituted histone octamers were purified. The purified histone octamers were preserved at -80 °C.

2.3.3. *In vitro* reconstitution and purification of the tri-nucleosomes

In order to prepare the tri-nucleosome that contains a center positioned CENP-A nucleosome (H3-CENP-A-H3 tri-nucleosome) or the H3-H3-H3 tri-nucleosome, the ligation method was utilized (Fig. 6). This method allows for the independent reconstitution of the three mono-nucleosomes prior to ligation, enabling the preparation of tri-nucleosomes with different nucleosome compositions.

The DNAs utilized for the ligation of tri-nucleosomes, namely 601a DNA, 601b DNA (linker DNA 22 bp or 30 bp), and 601c DNA, were first prepared (Fig. 6A, Table 1). Subsequently, the 601a, 601b, and 601c mono-nucleosomes were initially reconstituted using the salt dialysis method and then purified via non-denaturing polyacrylamide gel electrophoresis (PAGE) (6%, 0.5 x TBE, gel height=5.5 cm), utilizing a Prep Cell model 491 apparatus (BioRad) (Fig. 6B). The elute buffer used consists of 20 mM HEPES-KOH (pH 7.8).

Following this process, mono-nucleosomes containing sticky ends were ligated utilizing T4 DNA ligase in a specific ratio (601a nucleosome: 601b nucleosome: 601c nucleosome=1:0.9:1) (Fig. 6C). It is important to note that the DNA fragments were precisely designed, ensuring that 601a could only be ligated to 601b, and 601c could only be ligated to 601b (Fig. 6A). This specially designed sequence effectively prevents the

ligation of 601a to 601c.

Upon completion of the ligation process, the tri-nucleosomes were subjected to additional purification via non-denaturing polyacrylamide gel electrophoresis (PAGE) (4%, 0.5 x TBE, gel height=5.5 cm) utilizing a Prep Cell model 491 apparatus (BioRad), with an elution buffer composed of 20 mM HEPES-KOH (pH 7.8). This process ensures the removal of any remaining mono-nucleosomes and di-nucleosomes. Finally, the purified tri-nucleosomes were concentrated, stored at 4 °C, and ready for subsequent analysis.

2.3.4. Cryo-EM sample preparation

For the cryo-EM analysis, C-flat 2/1 400-mesh grids (Protochips, USA) were utilized. Prior to use, the grids were subjected to plasma cleaning via a Solarus Plasma Cleaner (Gatan) for 10 seconds at 20 W, utilizing a gas mix comprising 23% H2 and 77% O2. Grid preparation was performed with the assistance of a Vitrobot Mark IV (Thermo Fisher Scientific).

In the case of tri-nucleosomes that lacked Mg^{2+}, a 2.5 µL volume of the tri-nucleosomes (1 mg/mL) was applied once to the grids.

Conversely, in order to prepare tri-nucleosomes containing Mg^{2+}, equal volumes of tri-nucleosomes (0.06 mg/mL) absent of $MgCl_2$ and a 20 mM HEPES-KOH (pH 7.8) buffer supplemented with 2 mM $MgCl_2$ were mixed together. This resulted in a final sample of tri-nucleosomes (0.03 mg/mL) in a buffer containing 20 mM HEPES-KOH (pH 7.8) and 1 mM $MgCl_2$. The sample solution was applied to the grids twice. Initially, a 2.5 µL volume of the sample was applied to the grid and then withdrawn using the pipette that was utilized for the application. Following this, an additional 2.5 µL volume

of the sample was reapplied to the grids.

For both the Mg^{2+}-containing and Mg^{2+}-free tri-nucleosomes, the blotting time was set at 3 seconds, maintaining both the temperature and humidity at 16 °C and 100%, respectively. Following blotting, the grids were promptly plunged into liquid ethane.

2.3.5. Cryo-EM analysis and data collection

By utilizing in-focus phase plate imaging, vitrified samples of the tri-nucleosomes (H3-H3-H3 22 bp, H3-CENP-A-H3 22 bp, and CENP-A-H3-CENP-A with 30 bp), in the absence or presence of 1 mM $MgCl_2$, were analyzed. To be specific, the analyses were performed at liquid nitrogen temperature using a Titan Krios cryo-EM (Thermo Fisher Scientific), which was operated at 300 kV. The Titan Krios cryo-EM was also equipped with a Quantum GIF imaging filter (Gatan), utilizing the energy-filtered transmission electron microscopy (EFTEM) mode with a slit width of 20 eV. The calibrated magnification was set at 35,971x. Digital micrographs of the samples were captured using a K2 Summit direct electron detector in the counting mode, yielding a pixel size of 1.39 Å/pixel at the specimen level.

For the H3-H3-H3 22 bp and H3-CENP-A-H3 22 bp tri-nucleosomes, the images were manually collected with a Volta phase plate (VPP) in-focus, following the methodology described (64). For the CENP-A-H3-CENP-A 30 bp, the images were automatically collected using EPU software (Thermo Fisher Scientific). The microscope setup included a 50 mm C2 condenser aperture, with the illumination established in the nanoprobe mode within the parallel beam range.

On the other hand, defocused data collection of H3-H3-H3 22 bp, the H3-H3-H3 30 bp, the H3-CENP-A-H3 22 bp, and the H3-CENP-A-H3 30 bp tri-nucleosome (all in the

presence of 1 mM $MgCl_2$), was conducted without the phase plate. by a Talos Arctica cryo-EM (Thermo Fisher Scientific) operating at 200 kV. A Falcon 2 direct electron detector was applied for the H3-CENP-A-H3 22 bp sample, while the Falcon 3 detector in the linear mode was applied for the other samples. Data collection was conducted automatically using the EPU software (Thermo Fisher Scientific). Details can be found in Table 2.

2.3.6. Image processing

After data collection, the frames were first aligned by MOTIONCOR2 with dose weighting (65). Subsequently, CTFFIND4 was utilized to estimate the defocus value and objective lens astigmatism by fitting the contrast transfer function (CTF) from micrographs without dose weighting (66).

The following procedures were conducted by using the software RELION 2.1 (67). Several rounds of 2D classification were employed to discard junk particles, followed by several rounds of 3D classification. The best classes were selected and subjected to 3D refinement. The final maps were normalized by MAPMAN (68) and visualized by UCSF Chimera (69). Details for processing are summarized in Table 2.

2.3.7. Model building

The process of model fitting and refinement for the nucleosome samples, H3-CENP-A-H3 (22 bp) and H3-H3-H3(22 bp), shared common methods. CENP-A nucleosomes (PDB: 3AN2) and/or H3 nucleosomes (3LZ0) were manually fitted into each cryo-EM maps using UCSF Chimera (69), positionally refined with rigid-body docking with colors, by the aid of the SITUS program package. Distinctively, the H3-CENP-A-H3 sample

involved the fitting of two copies of the H3 nucleosome (PDB: 3LZ0) and a single copy of the CENP-A nucleosome (PDB: 3AN2) into the 20.8 Å resolution cryo-EM map. Conversely, for the H3-H3-H3 sample, three copies of the nucleosome (PDB: 3LZ0) were fitted into the 12.4 Å resolution cryo-EM map.

Additionally, linker DNA regions missing in the partial models were modeled as double-stranded DNA, with gaps manually modeled by COOT through real-space refinement and regularization. Human histones were used to replace the amino acid sequence and coordinates of the crystal structure of the *Xenopus laevis* nucleosome (PDB:3LZ0) (17).

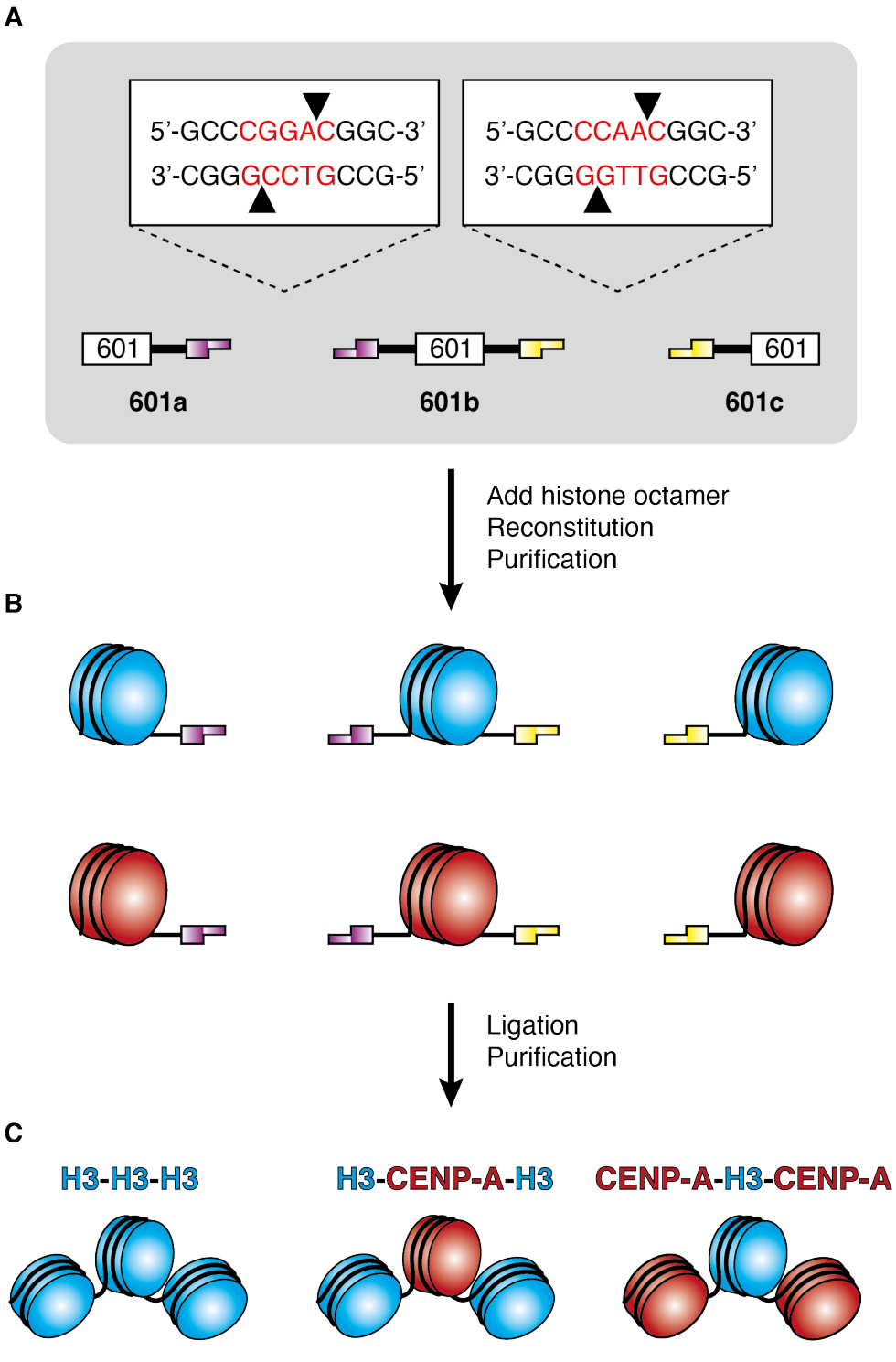

A
5'-GCCCGGACGGC-3'
3'-CGGGCCTGCCG-5'
5'-GCCCCAACGGC-3'
3'-CGGGGTTGCCG-5'
601
601
601
601a
601b
601c
Add histone octamer
Reconstitution
Purification
B
Ligation
Purification
C
H3-H3-H3
H3-CENP-A-H3
CENP-A-H3-CENP-A

Fig. 6 Preparation of tri-nucleosomes by the ligation method

(A) Preparation of the DNA sequences for ligation. The DNA ends were processed by BglI, which can recognize the 5'-GCCNNNNNGGC-3' sequence. The cohesive ends were designed differently so that 601a can only be ligated to 601b, and 601c can only be ligated to 601b. Note that 601a cannot be ligated to 601c. There are 2 kinds of 601b sequences [601b (22 bp) and 601b (30 bp)], for preparing the tri-nucleosomes with linker DNA 22 bp and 30 bp. **(B)** H3 or CENP-A mono-nucleosomes can be reconstituted by using different histone octamers. **(C)** After ligation of different kinds of mono-nucleosomes, different tri-nucleosomes can be obtained.

Table 1 DNA sequences for preparing tri-nucleosomes by the ligation method

DNA	Sequence
601a DNA	ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAG ACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCC CCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCC AGGCACGTGTCAGATATATACATCGATTGGATAGGCCC
601b DNA (22 bp)	GGACGGCCTGGATAATCAGAATCCCGGTGCCGAGGCCGC TCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCA CGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGAT TACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCG ATTGGATAGGCCCCAA
601b DNA (30 bp)	GGACGGCCATCTATATTGGATAATCAGAATCCCGGTGCCG AGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTT AAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCA AGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATA TACATCGATTGGATATATATCTAGGCCCCAA
601c DNA	CGGCCTGGATAATCAGAATCCCGGTGCCGAGGCCGCTCA ATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGT ACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTAC TCCCTAGTCTCCAGGCACGTGTCAGATATATACATCGAT

601 sequence BglI recognition: GCCNNNNNGGC

Table 2 Details of data collection and image processing of tri-nucleosomes prepared by the ligation method

Tri-nucleosome sample	**H3-H3-H3 22 bp**	**H3-CA-H3 22 bp**	**H3-H3-H3 30 bp**	**H3-CA-H3 30 bp**	**CA-H3-CA 30 bp**
Cryo-EM	Talos Arctica	Talos Arctica	Talos Arctica	Talos Arctica	Titan Krios
Camera	Falcon3	Falcon2	Falcon3	Falcon3	Falcon3
Pixel size (A/pix)	1.40	1.40	1.40	1.40	1.39
Exposure time (sec)	2	1	2	2	2
Total dose (e/A^2)	80	25	80	80	68
Micrographs (no.)	4,512	4,515	8,035	10,351	808
Particles for 2D class (no.)	118,395	92,674	109,525	63,365	39,060
Particles for 3D class (no.)	100,725	43,008	20,953	22,515	
Particles in the final map (no.)	7,312	4,999	4,534	4,574	
Symmetry	C2	C2	C2	C1	
Final resolution	12.3 Å	18.7 Å	15.7 Å	19.6 Å	

H3-H3-H3: H3-H3-H3 tri-nucleosomes

H3-CA-H3: H3-CENP-A-H3 tri-nucleosomes

2.4. Results

2.4.1. *In vitro* reconstitution and purification of mono-nucleosomes (22 bp)

In order to prepare tri-nucleosomes by the ligation method, mono-nucleosomes were first prepared. Specifically, different combinations of 601a, 601b, or 601c DNAs with either H3 or CENP-A histone octamers were used to create the 601a H3 mono-nucleosome, the 601b H3 mono-nucleosome, the 601b CENP-A mono-nucleosome, and the 601c H3 mono-nucleosome (Fig. 7A). The qualities of the purified mono-nucleosomes were confirmed by SDS-PAGE and non-denaturing PAGE (Fig. 7B&C).

The SDS-PAGE results displayed 4 distinct histone bands for each nucleosome type: H2A-H2B-H3-H4 for H3 nucleosomes and H2A-H2B-CENP-A-H4 for CENP-A nucleosomes (Fig. 7B). Crucially, these bands exhibited equal intensities, indicating a balanced representation of each histone component in their respective nucleosome types.

The non-denaturing PAGE showed only 1 band for each nucleosome sample, indicating they have been highly purified (Fig. 7C). Collectively, these results show that the mono-nucleosomes have been prepared successfully.

2.4.2. Preparation of the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes (22 bp)

After preparing the mono-nucleosomes successfully, they were ligated and purified to form the H3-H3-H3 or H3-CENP-A-H3 tri-nucleosomes with 22 bp of linker DNAs (Fig. 7D). The qualities of the purified tri-nucleosomes were confirmed by SDS-PAGE and non-denaturing PAGE (Fig. 7E&F).

In the SDS-PAGE analysis, the H3-H3-H3 22 bp tri-nucleosome displayed bands of H2A, H2B, H3, and H4, each with equivalent intensities, which was the anticipated outcome (Fig. 7E). Meanwhile, for the H3-CENP-A-H3 22 bp tri-nucleosome, five

distinct bands were observed, corresponding to H2A, H2B, H3, CENP-A, and H4 (Fig. 7E). Notably, the intensity of the CENP-A band was slightly less than that of the H3 band, indicating a lower quantity of CENP-A compared to H3 in the sample, consistent with the expected ratio in the H3-CENP-A-H3 trinucleosome.

The non-denaturing PAGE analysis revealed a single band for each nucleosome sample, indicating high purity and the absence of mono- or di-nucleosomes (Fig. 7F). Collectively, these results confirm the successful preparation of the H3-CENP-A-H3 bp tri-nucleosome and H3-H3-H3 22 bp tri-nucleosome.

2.4.3. Cryo-EM analysis of the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes (22 bp)

The purified H3-CENP-A-H3 tri-nucleosomes and H3-H3-H3 with 22 bp linker DNAs were then observed with a Titan Krios cryo-EM equipped with a Volta phase plate. Through the obtained micrographs, a noticeable stacking of side-positioned nucleosomes in the tri-nucleosome was observed upon the addition of 1 mM Mg^{2+} (Fig. 8, compare B to C and E to F). This stacking stabilized the overall structure of the tri-nucleosomes, leading to the decision to conduct structural analysis in the presence of 1 mM Mg^{2+}. Data collections of the two tri-nucleosome types, H3-CENP-A-H3 22 bp and H3-H3-H3 22 bp were conducted using a Talos Arctica cryo-EM.

After data collection, single particle analyses of the 2 kinds of tri-nucleosomes were conducted by using RELION. Surprisingly, the 2D classification results led to the discovery of two distinct linker DNA path conformations in the tri-nucleosomes (Fig. 9). The first conformation, or the "outward-path" conformation, is characterized by linker DNAs joining to the outer side of the stacked nucleosomes (Fig. 9A, B, and D). On the

other hand, the second conformation, termed the "inward-path" conformation, displayed linker DNAs joining to the inner side of the stacked nucleosomes (Fig. 9E).

In the 2D classification result of the H3-CENP-A-H3 22 bp tri-nucleosome, only the outward-path can be observed (Fig. 9C). However, in the outward-path conformation, 2 subtypes exist, which can be distinguished by the orientation of the center-positioned CENP-A nucleosome (Fig. 9A&B). On the other hand, in the 2D classification result of the H3-H3-H3 22 bp tri-nucleosome, both the outer-path and inner-path conformations are present (Fig. 9F). A closer inspection in Fig. 9F shows that 2 subtypes of the outward-path conformation may exist, similar to the case of the H3-CENP-A-H3 22 bp tri-nucleosome.

2.4.4. Cryo-EM structures of the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes (22 bp)

To further characterize the two different conformations, the 3D cryo-EM maps of the tri-nucleosomes were reconstructed (Fig. 10). More details of the cryo-EM maps can be found in Fig. 14. In the case of the H3-CENP-A-H3 22 bp tri-nucleosome, although 2 subtypes of the outward-path were observed in the result of 2D classification (Fig. 9A&B), only the cryo-EM map corresponding to the class seen in Fig. 9A was obtained (Fig. 10A). In this structure, the orientation of the center positioned CENP-A nucleosome is perpendicular to the orientation of the stacked H3 nucleosomes at the bottom.

On the other hand, in the case of the H3-H3-H3 22 bp tri-nucleosome, although both the outward-path and inward-path conformation were observed in the result of 2D classification (Fig. 9D&E), only the cryo-EM map corresponding to the outward conformation (Fig. 9D) was obtained (Fig. 10B) for comparison with the H3-CENP-A-

H3 22 bp tri-nucleosome structure. In this structure, the orientation of the center positioned H3 nucleosome is parallel to the orientation of the stacked H3 nucleosomes at the bottom.

Upon comparing the cryo-EM structures of the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes with 22 base-pair linker DNAs, differences in the orientation of the center-positioned nucleosomes were noted (Fig. 10). The H3-CENP-A-H3 22 bp tri-nucleosome structure was found to be less twisted than the H3-H3-H3 22 bp tri-nucleosome. This led to the suggestion that the flexible DNA ends of the CENP-A nucleosome could potentially restrict its capability to twist adjacent nucleosomes within the tri-nucleosome structure.

To further investigate this hypothesis, I next decided to obtain the cryo-EM structure of tri-nucleosomes with extended linker DNA lengths (30 bp on each side).

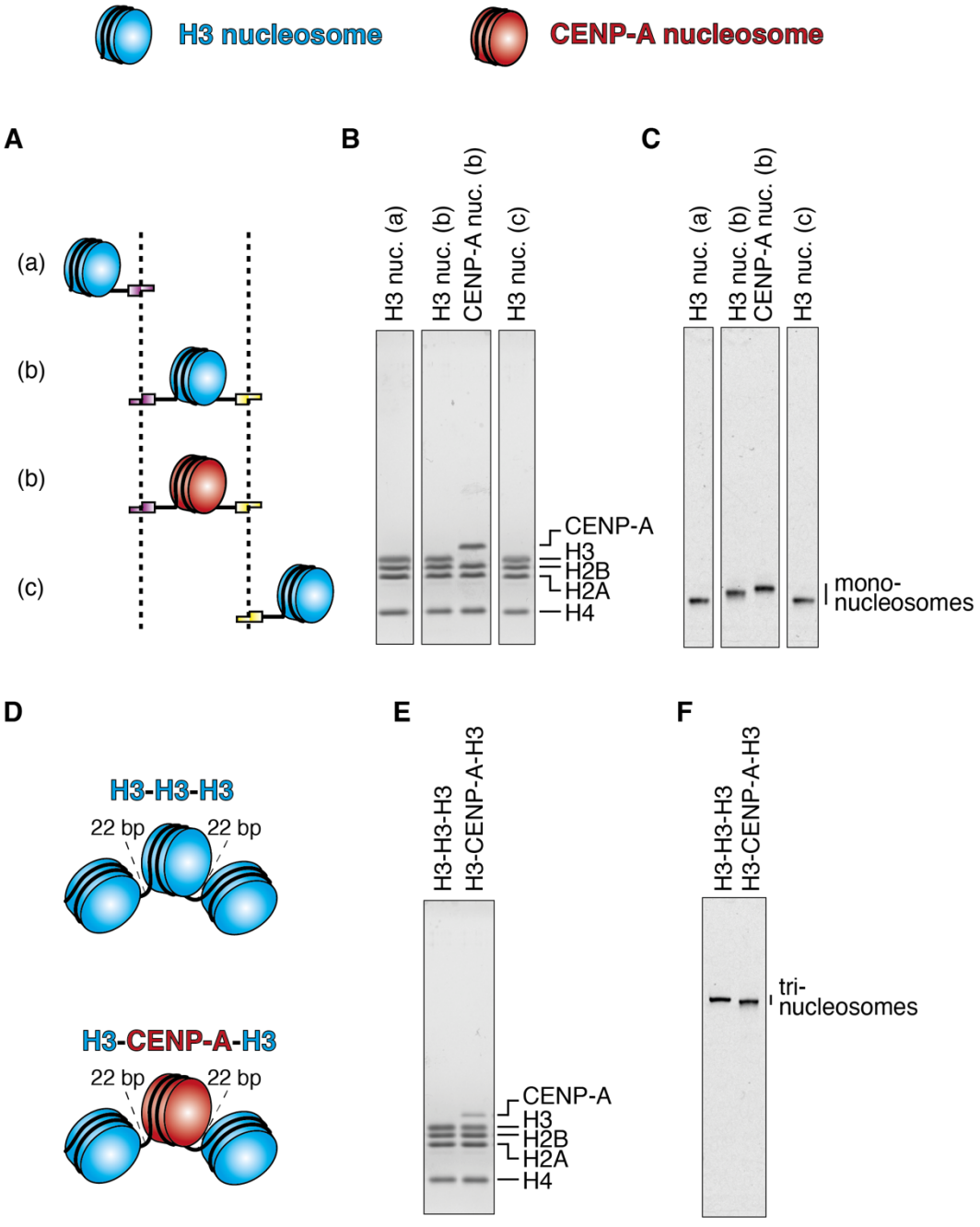

H3 nucleosome
CENP-A nucleosome
A
(a)
(b)
(b)
(c)
B
H3 nuc. (a)
H3 nuc. (b)
CENP-A nuc. (b)
H3 nuc. (c)
CENP-A
H3
H2B
H2A
H4
C
H3 nuc. (a)
H3 nuc. (b)
CENP-A nuc. (b)
H3 nuc. (c)
mono-
nucleosomes
D
H3-H3-H3
22 bp
22 bp
H3-CENP-A-H3
22 bp
22 bp
E
H3-H3-H3
H3-CENP-A-H3
CENP-A
H3
H2B
H2A
H4
F
H3-H3-H3
H3-CENP-A-H3
tri-
nucleosomes

Fig. 7 Preparation of the tri-nucleosomes with 22 bp linker DNAs

(A) Schematic representation of the mono-nucleosomes. **(B)** The purified mono-nucleosomes were analyzed by SDS-PAGE and stained by CBB, visualizing protein bands. **(C)** The purified mono-nucleosomes were analyzed by non-denaturing PAGE and stained by EtBr, visualizing DNA bands. **(D)** Schematic representation of the tri-nucleosomes. **(E)** The purified tri-nucleosomes with 22 bp linker DNAs were analyzed by SDS-PAGE and stained by CBB, visualizing protein bands. **(F)** The purified tri-nucleosomes with 22 bp linker DNAs were analyzed by non-denaturing PAGE and stained by EtBr, visualizing DNA bands.

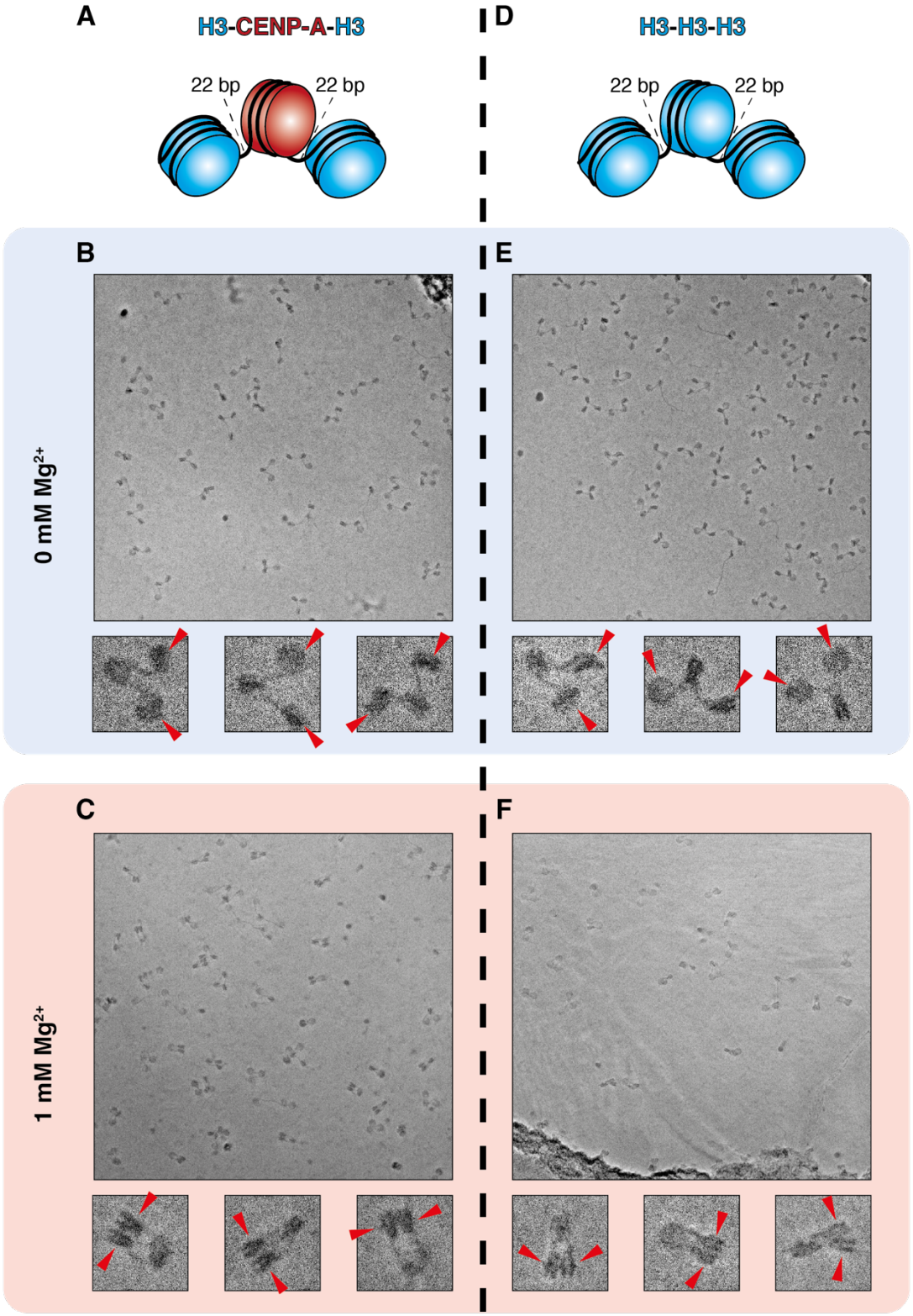

A
H3-CENP-A-H3
22 bp
22 bp
D
H3-H3-H3
22 bp
22 bp
B
E
0 mM Mg2+
C
F
1 mM Mg2+

Fig. 8 Observation of the tri-nucleosomes with 22 bp linker DNAs in the absence and presence of Mg^{2+}

Smaller boxes of on the bottom of each micrograph show magnified particles in the micrograph. The side-positioned nucleosomes are indicated by red arrowheads. **(A&D)** Schematic representation of the H3-CENP-A-H3 22 bp tri-nucleosome (A) and H3-H3-H3 22 bp tri-nucleosome (D). **(B&E)** Micrographs of samples observed in the absence of Mg^{2+}. **(C&F)** Micrographs of samples observed in the presence of 1mM Mg^{2+}.

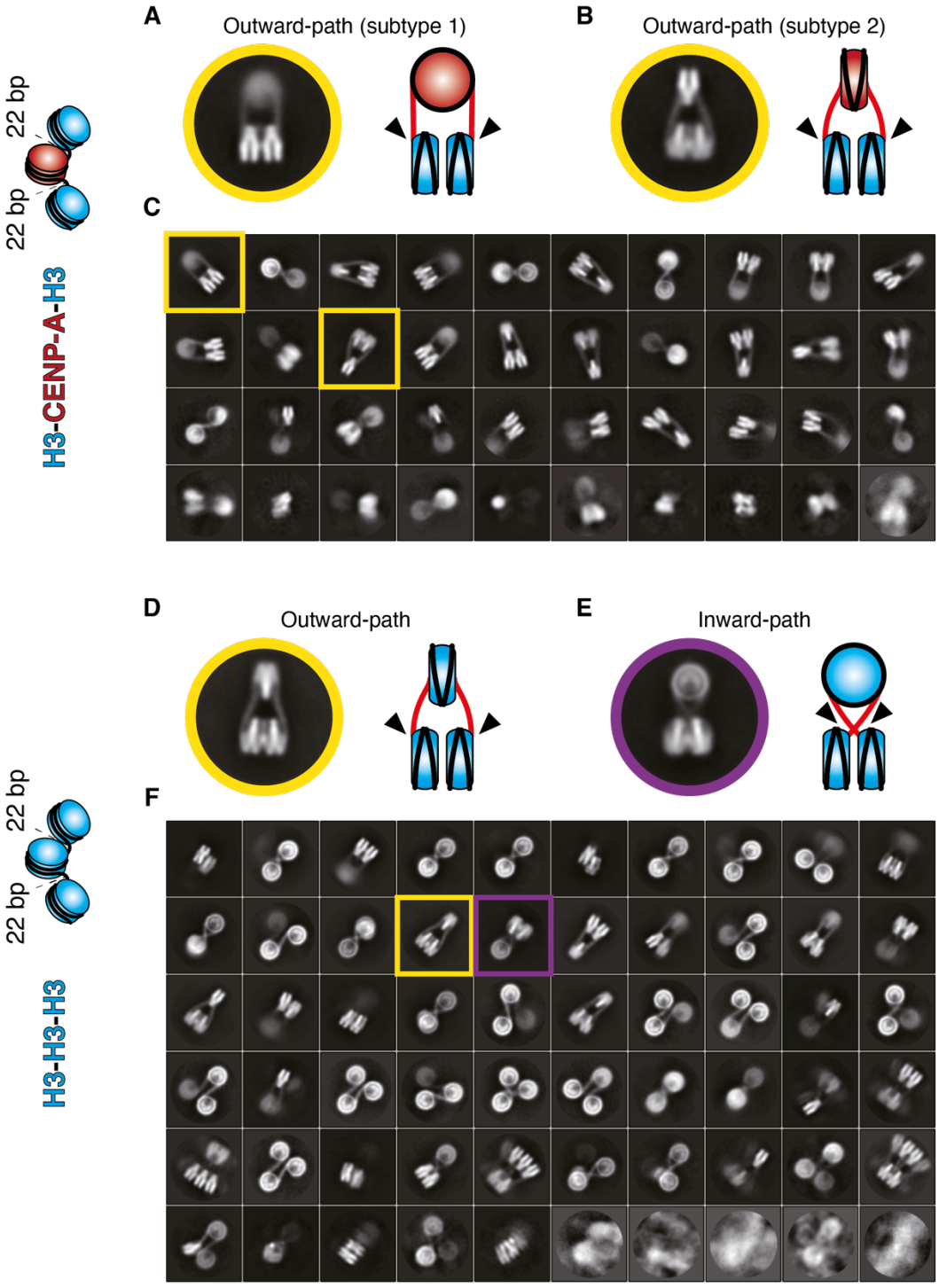

A
Outward-path (subtype 1)
B
Outward-path (subtype 2)
22 bp
22 bp
H3-CENP-A-H3
C
D
Outward-path
E
Inward-path
22 bp
22 bp
H3-H3-H3
F

Fig. 9 2D classification results of tri-nucleosomes with 22 bp linker DNAs

Two different linker DNA conformations can be observed. The “outward-path” conformation, featuring linker DNAs that are joined to the outer side of the stacked nucleosomes at the bottom, is shown in yellow. On the other hand, the “inward-path” conformation, featuring linker DNAs that are joined to the inner side of the stacked nucleosomes at the bottom, is shown in purple. **(A-C)** 2D classification result of the H3-CENP-A-H3 22 bp tri-nucleosome. Squared boxes in (C) indicate the classes that are cropped out and shown in (A) and (B). **(D-F)** 2D classification result of the H3-H3-H3 22 bp tri-nucleosomes. Squared boxes in (F) indicate the classes that are cropped out and shown in (D) and (E).

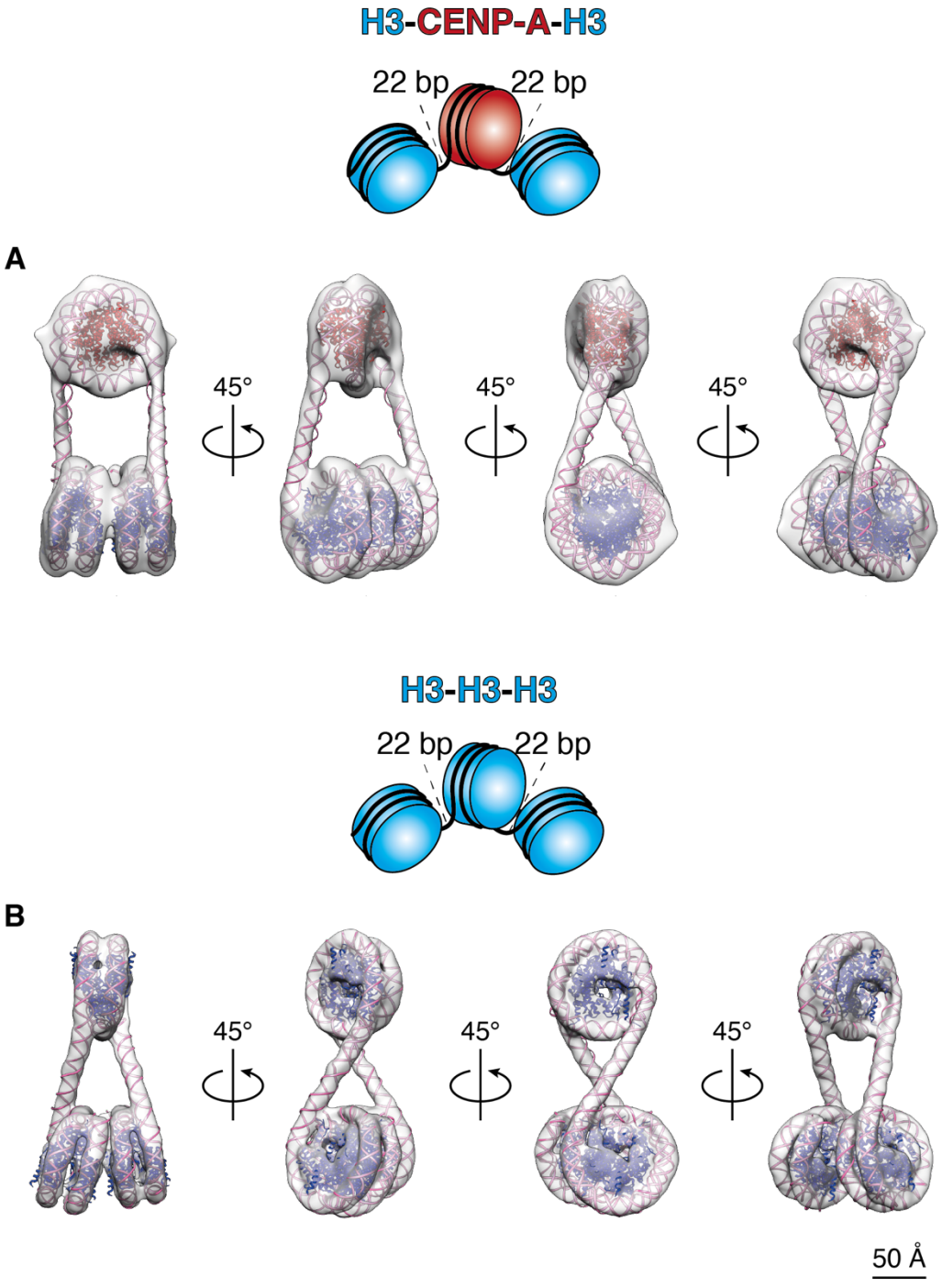

H3-CENP-A-H3
22 bp
22 bp
A
45°
45°
45°
H3-H3-H3
22 bp
22 bp
B
45°
45°
45°
50 Å

Fig. 10 Cryo-EM structures of tri-nucleosomes with 22 bp linker DNAs

The corresponding models are fitted into the cryo-EM map. The scale bar is 50 Å. **(A)** The cryo-EM map of the H3-CENP-A-H3 22 bp tri-nucleosome, with a final resolution of 18.7 Å. **(B)** The cryo-EM map of the H3-H3-H3 22 bp tri-nucleosome, with a final resolution of 12.3Å.

2.4.5. Preparation of the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes (30 bp)

The ligation method has been proved to work from the previous analysis on tri-nucleosomes with 22 bp linker DNAs. We therefore extended the DNA fragment of 601b, so that after ligation the tri-nucleosomes will have linker DNAs of 30 bp (Table 1). The mono-nucleosomes and tri-nucleosomes were subsequently prepared similarly (Fig. 11A). The qualities of the purified tri-nucleosomes were confirmed by SDS-PAGE and non-denaturing PAGE (Fig. 11E&F).

The SDS-PAGE analysis affirmed the successful preparation of both the H3-H3-H3 and H3-CENP-A-H3 tri-nucleosomes with 30 bp linker DNAs. While the former exhibited even intensities across bands for H2A, H2B, H3, and H4 (Fig. 11E), the latter showed an additional CENP-A band with a slightly weaker intensity than the H3 band, suggesting a lesser amount of CENP-A in line with theoretical expectations (Fig. 11E). Furthermore, the absence of mono- or di-nucleosomes was confirmed through non-denaturing PAGE analysis, which demonstrated a single band for each sample, indicating high sample purity (Fig. 11F). Collectively, these results indicate that the H3-H3-H3 and H3-CENP-A-H3 tri-nucleosomes with 30 bp linker DNAs have been prepared successfully.

2.4.6. Cryo-EM analysis of the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes (30 bp)

Data collections and single particle analyses of the H3-CENP-A-H3 30 bp tri-nucleosome and H3-H3-H3 30 bp tri-nucleosome were conducted similarly to the tri-nucleosomes with 22 bp linker DNAs. From the result of 2D classification, both the outward-path conformation and the inward-path conformation can be observed in the H3-

CENP-A-H3 30 bp tri-nucleosome (Fig. 12A-C), while only the inward-path conformation is observed in the H3-H3-H3 30 bp tri-nucleosome (Fig. 12D&E).

This is different from the 2D classification results of the tri-nucleosomes with 22 bp linker DNAs. In the case of H3-CENP-A-H3 tri-nucleosome, only the outward-path conformation is present when linker DNA is short (Fig. 9A&B), but the inward-path conformation emerged upon the extension of the linker DNA (Fig. 12A&B). On the other hand, in the case of H3-H3-H3 tri-nucleosome, both conformations are present when the linker DNA is short (Fig. 9D&E), but only the inward-path conformation remained when the linker DNA is extended (Fig. 12D).

2.4.7. Cryo-EM structures of the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes (30 bp)

Next, the cryo-EM structures for the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes with 30 bp were obtained (Fig. 13). More details of the cryo-EM maps can be found in Fig. 14. Only the outward-path conformation structure was obtained for the H3-CENP-A-H3 30 bp tri-nucleosome (Fig. 13A), although both conformations are present in the 2D classification result (Fig. 12A&B). This structure is essentially similar to the structure of the H3-CENP-A-H3 22 bp tri-nucleosome in Fig. 10A, only with the linker DNA extended.

For comparison, the inward-path conformation structure of the H3-H3-H3 30 bp tri-nucleosome was obtained (Fig. 13B), making it the only 3D cryo-EM map of the inward-path conformation obtained in this study. The cryo-EM map showed that the orientation of the center-positioned H3 nucleosome is perpendicular to the stacked H3 nucleosomes at the bottom.

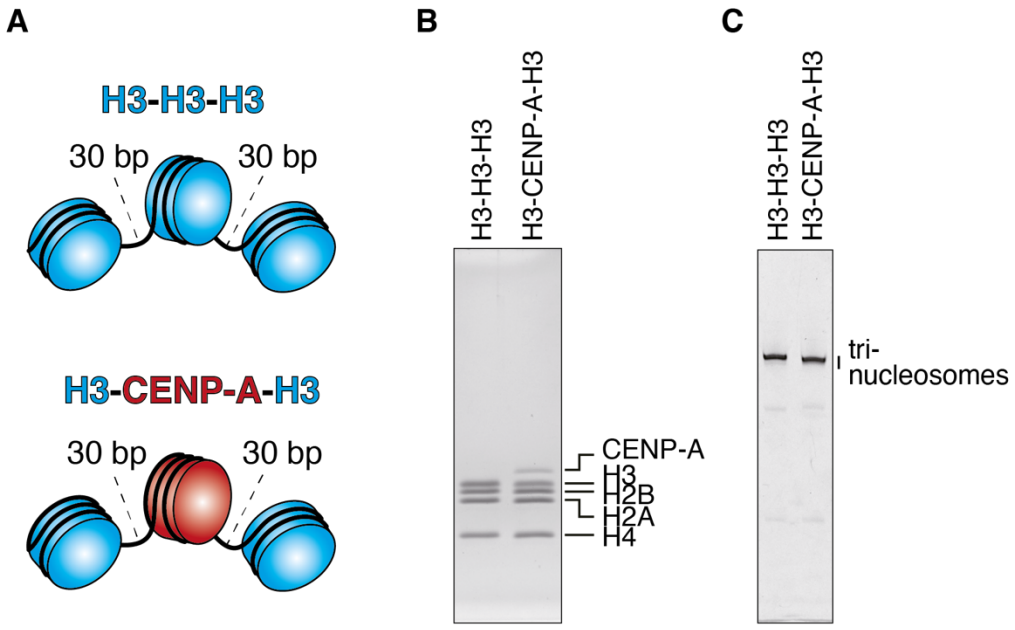


Fig. 11 Preparation of the tri-nucleosomes with 30 bp linker DNAs

(A) Schematic representation of the tri-nucleosomes with 30 bp linker DNAs. **(B)** The purified tri-nucleosomes with 30 bp linker DNAs were analyzed by SDS-PAGE and stained by CBB, visualizing protein bands. **(C)** The purified tri-nucleosomes with 30 bp linker DNAs were analyzed by non-denaturing PAGE and stained by EtBr, visualizing DNA bands.

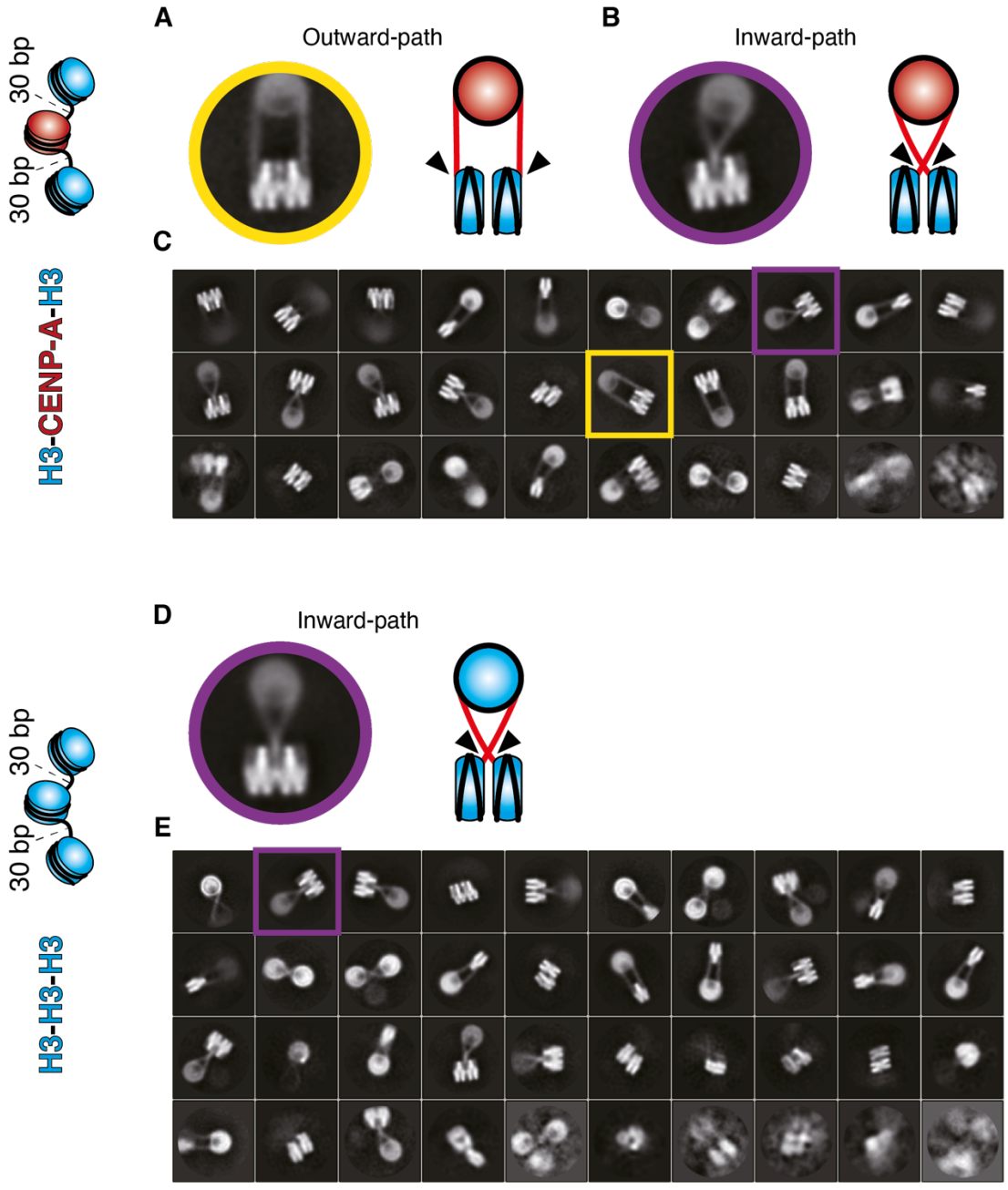

A
Outward-path
B
Inward-path
30 bp
30 bp
C
H3-CENP-A-H3
D
Inward-path
30 bp
30 bp
E
H3-H3-H3

Fig. 12 2D classification results of H3-H3-H3 and H3-CENP-A-H3 tri-nucleosome (30 bp)

Two different linker DNA conformations can be observed. The "outward-path" conformation is shown in yellow, while the "inward-path" conformation is shown in purple. (**A-C)** 2D classification result of the H3-CENP-A-H3 30 bp tri-nucleosome. Squared boxes in (C) indicate the classes that are cropped out and shown in (A) and (B). **(D&E)** 2D classification result of the H3-H3-H3 30 bp tri-nucleosomes. Squared boxes in (E) indicate the classes that are cropped out and shown in (D).

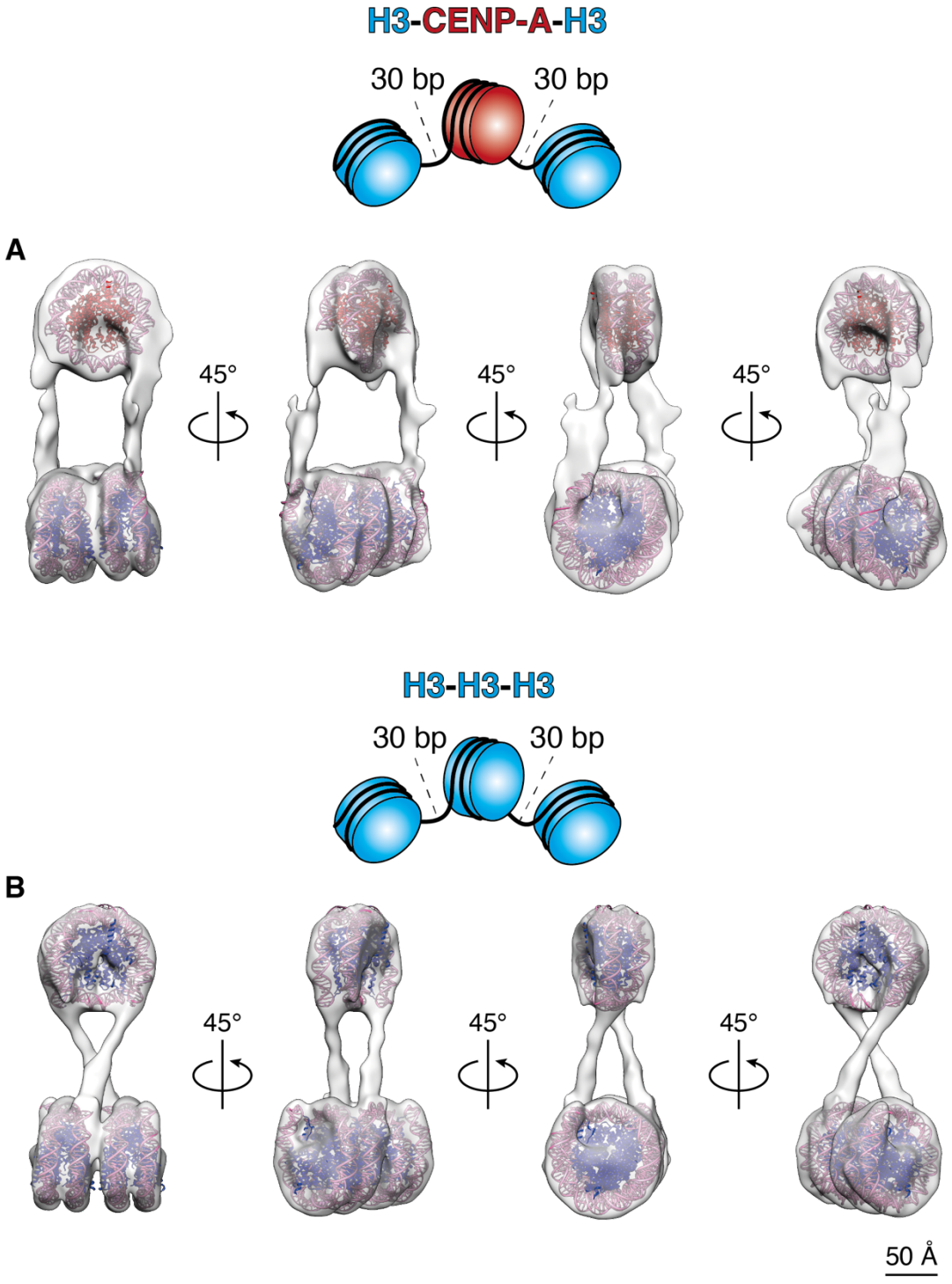

H3-CENP-A-H3
30 bp
30 bp
A
45°
45°
45°
H3-H3-H3
30 bp
30 bp
B
45°
45°
45°
50 Å

Fig. 13 Cryo-EM structures of the tri-nucleosomes with 30 bp linker DNAs

Crystal structures of the H3 nucleosome (PDB:3LZ0) and the CENP-A nucleosome (3AN2) were placed into the corresponding positions. **(A)** The cryo-EM map of the H3-CENP-A-H3 30 bp tri-nucleosome. **(B)** The cryo-EM map of the H3-H3-H3 30 bp tri-nucleosome.

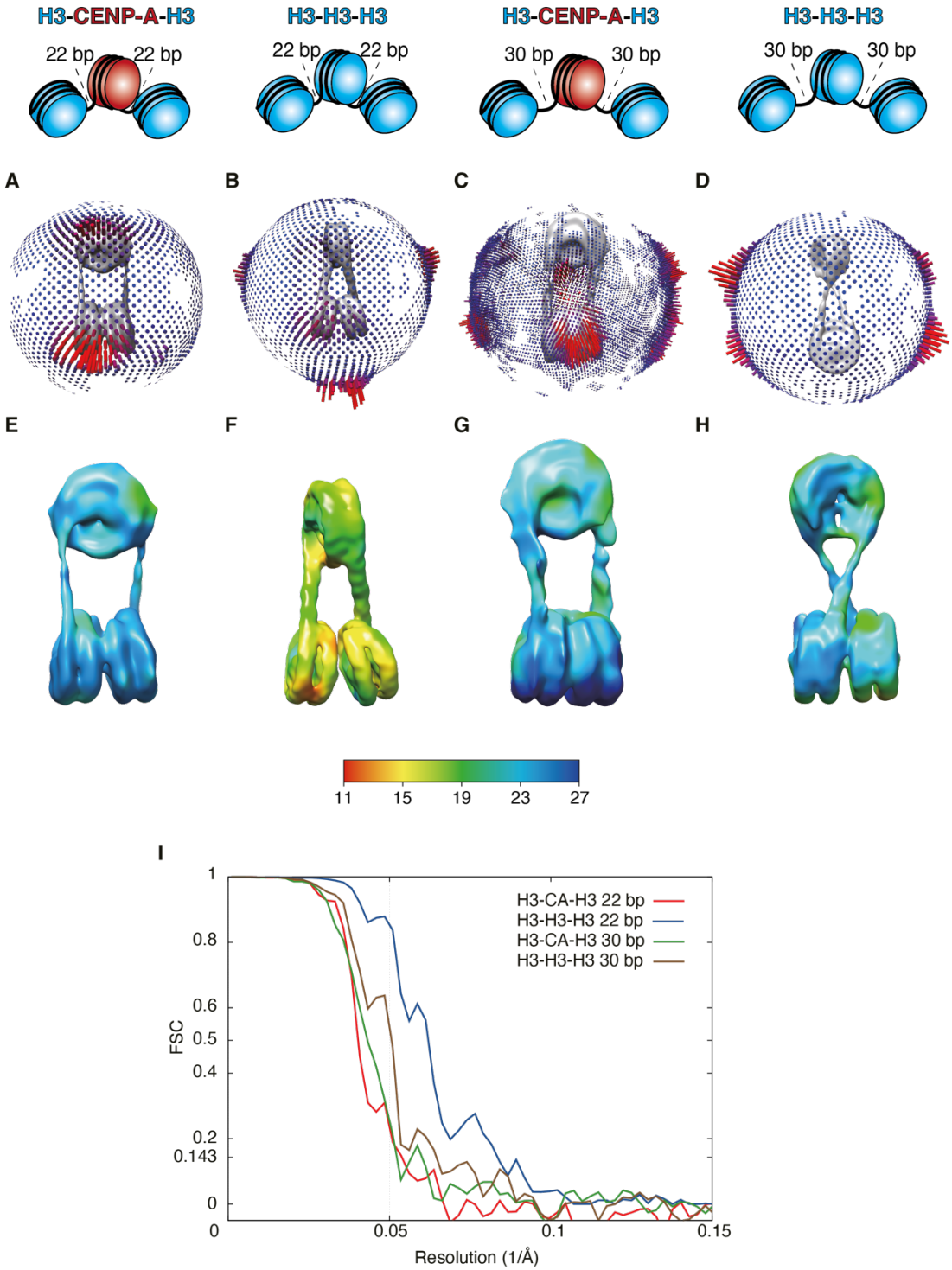

H3-CENP-A-H3
22 bp
22 bp
H3-H3-H3
22 bp
22 bp
H3-CENP-A-H3
30 bp
30 bp
H3-H3-H3
30 bp
30 bp
A
B
C
D
E
F
G
H
11
15
19
23
27
I
H3-CA-H3 22 bp
H3-H3-H3 22 bp
H3-CA-H3 30 bp
H3-H3-H3 30 bp
FSC
1
0.8
0.6
0.5
0.4
0.2
0.143
0
0
0.05
0.1
0.15
Resolution (1/Å)

Fig. 14 Euler angle distribution maps, local resolutions, and Fourier Shell Correlation (FSC) curves of the tri-nucleosomes

(A-D) Euler angle distribution maps of the H3-CENP-A-H3 22 bp tri-nucleosomes (A), H3-H3-H3 22 bp tri-nucleosomes (B), H3-CENP-A-H3 30 bp tri-nucleosomes (C), and H3-H3-H3 30 bp tri-nucleosomes (D). **(E-H)** Local resolutions of the H3-CENP-A-H3 22 bp tri-nucleosomes (E), H3-H3-H3 22 bp tri-nucleosomes (F), H3-CENP-A-H3 30 bp tri-nucleosomes (G), and H3-H3-H3 30 bp tri-nucleosomes (H). **(I)** FSC curves of the tri-nucleosomes. H3-CA-H3: H3-CENP-A-H3.

2.4.8. Preparation of the CENP-A-H3-CENP-A tri-nucleosome (30 bp)

The previous results of the H3-CENP-A-H3 and H3-H3-H3 tri-nucleosomes with 22 bp and 30 bp linker DNAs suggest that the center-positioned CENP-A nucleosome might influence the overall structure of the tri-nucleosome. I next decided to investigate if the position of the CENP-A nucleosome is important for the determination of the tri-nucleosome structure. To do so, I prepared the CENP-A-H3-CENP-A tri-nucleosome with 30 bp linker DNAs and performed cryo-EM analysis. If the center-positioned nucleosome type, and not the side-positioned nucleosome, is critical for determining the linker DNA conformation of the tri-nucleosome, then the CENP-A-H3-CENP-A 30 bp tri-nucleosome should have a similar linker DNA conformation as the H3-H3-H3 30 bp tri-nucleosome, or the inward-path conformation.

The CENP-A-H3-CENP-A 30 bp tri-nucleosome was prepared by reconstituting side-positioned mono-nucleosomes with the CENP-A histone octamer, and center-positioned H3 nucleosome with the H3 histone octamer (Fig. 15A). These were then ligated to from the CENP-A-H3-CENP-A 30 bp tri-nucleosome.

2.4.9. Cryo-EM analysis of the CENP-A-H3-CENP-A tri-nucleosome (30 bp)

Different from the previous cryo-EM analyses on H3-CENP-A-H3 or H3-H3-H3 tri-nucleosomes, the cryo-EM analysis of the CENP-A-H3-CENP-A was conducted by Titan Krios. The stacking of the side-positioned CENP-A nucleosomes can be observed in the presence of Mg^{2+} (Fig. 16B). This suggests that both H3 and CENP-A nucleosomes have the capacity to stack, which is a plausible observation given that nucleosome stacking is prompted by the interaction between the H4 tail of one nucleosome and the acidic patch of another. This interaction should remain unaffected by a change from H3 to CENP-A.

Subsequent data collection was conducted in the presence of Mg^{2+}.

From the result of 2D classification, only classes with the inward-path conformation were observed (Fig. 16C&D). This result is similar to that of the H3-H3-H3 tri-nucleosome with 30 bp of linker DNAs (Fig. 12D&E), indicating that the type of nucleosome positioned at the center plays a crucial role in determining the DNA path of the tri-nucleosome. If the center-positioned nucleosome is an H3 nucleosome, then the tri-nucleosome may favor the inward-path conformation. Alternatively, if the center-positioned nucleosome is an H3 nucleosome, then the tri-nucleosome may favor the outward-path conformation.

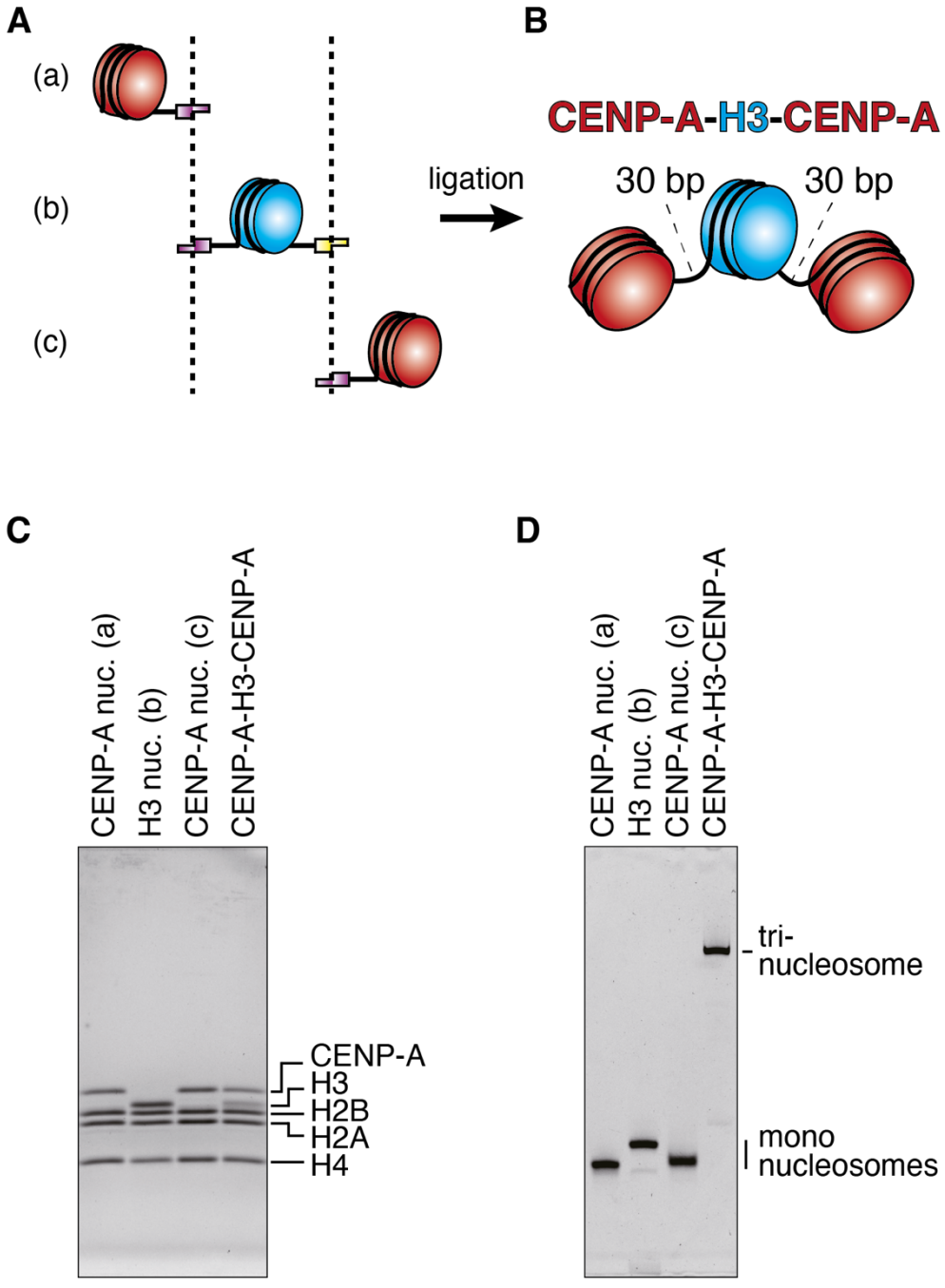

A
(a)
(b)
(c)
B
CENP-A-H3-CENP-A
ligation
30 bp
30 bp
C
CENP-A nuc. (a)
H3 nuc. (b)
CENP-A nuc. (c)
CENP-A-H3-CENP-A
CENP-A
H3
H2B
H2A
H4
D
CENP-A nuc. (a)
H3 nuc. (b)
CENP-A nuc. (c)
CENP-A-H3-CENP-A
tri-
nucleosome
mono
nucleosomes

Fig. 15 Preparation of the CENP-A-H3-CENP-A tri-nucleosome (30 bp)

(A) Schematic representation of the mono-nucleosomes. **(B)** Schematic representation of the tri-nucleosome after ligation. **(C)** The purified mono-nucleosomes and tri-nucleosome were analyzed by SDS-PAGE and stained by CBB, visualizing protein bands. **(D)** The purified mono-nucleosomes and tri-nucleosome were analyzed by non-denaturing PAGE and stained by EtBr, visualizing DNA bands.

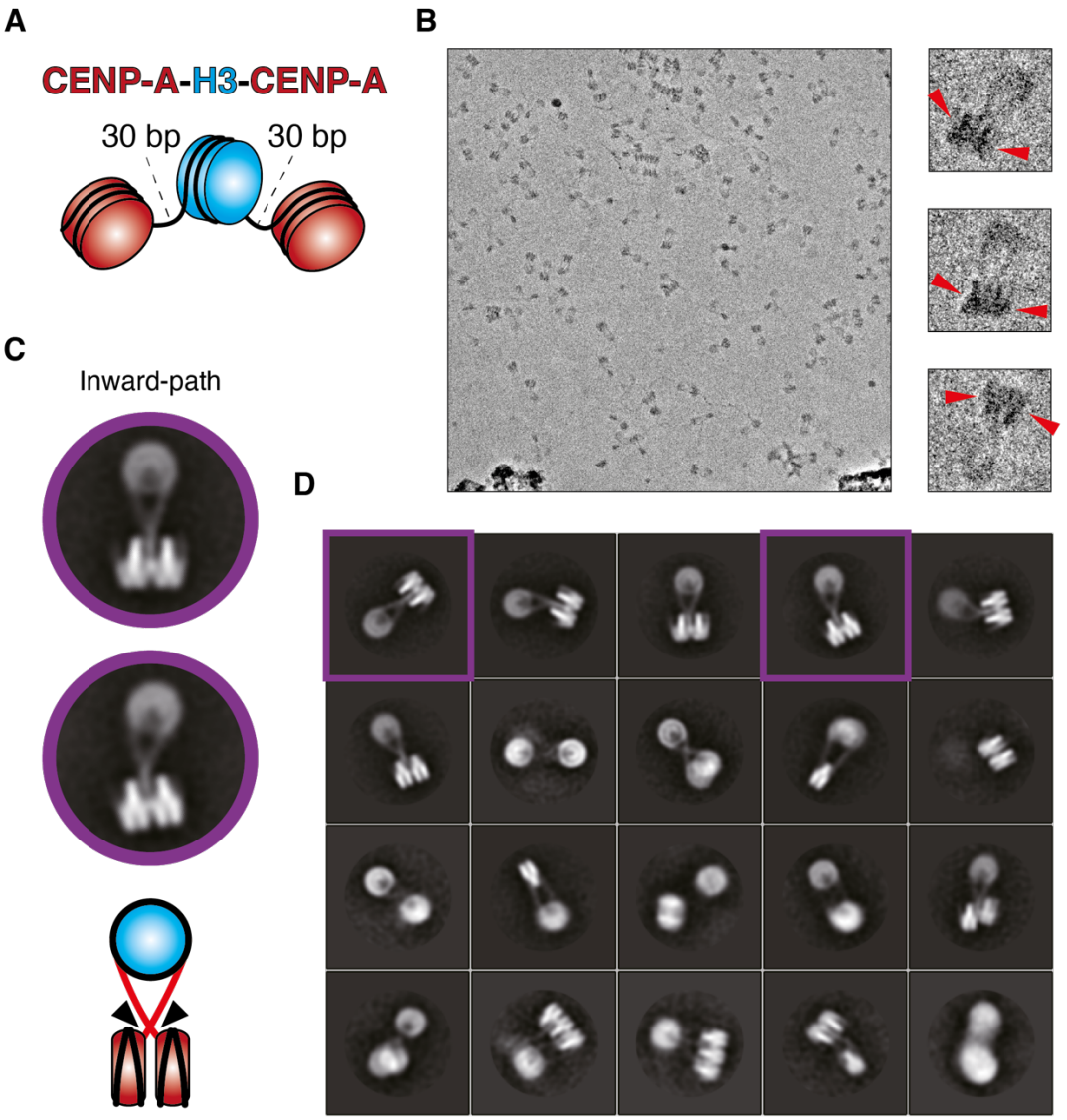

A
CENP-A-H3-CENP-A
30 bp
30 bp
B
C
Inward-path
D

Fig. 16 Cryo-EM Analysis of CENP-A-H3-CENP-A 30 bp

(A) Schematic representation of the CENP-A-H3-CENP-A tri-nucleosome with 30 bp of linker DNAs. **(B)** Micrographs of samples observed in the presence of 1mM Mg^{2+}. Smaller boxes of on the right of the micrograph show magnified particles in the micrograph. The side-positioned nucleosomes are indicated by red arrowheads. **(C)** Representative classes of the CENP-A-H3-CENP-A tri-nucleosome with 30 bp of linker DNAs. **(D)** 2D classification results of the CENP-A-H3-CENP-A tri-nucleosome with 30 bp of linker DNAs. Squared boxes indicate the classes that are cropped out and shown in (C).

2.4.10. Poly-nucleosome models for centromeric and non-centromeric chromatin

Based on the obtained tri-nucleosome structures, poly-nucleosome models of the centromeric chromatin and non-centromeric chromatin were created (Fig. 17). This was accomplished by aligning the side-positioned nucleosomes, or the side and center-positioned nucleosomes, in the tri-nucleosome cryo-EM maps to form nucleosome arrays with 16 nucleosomes. For the non-centromeric chromatin model, only H3-H3-H3 tri-nucleosomes were used for generating the model, which consists of 16 H3 nucleosomes (Fig. 17A). On the other hand, in the centromeric chromatin model, one H3-H3-H3 tri-nucleosome was substituted for one H3-CENP-A-H3 tri-nucleosome, resulting in 1 CENP-A nucleosome and 15 H3 nucleosomes (Fig. 17B).

In the non-centromeric chromatin model, it could be seen that the chromatin fiber is properly aligned and condensed. In this model, the H3 nucleosomes are close to each other, making the nucleosome disc surface seemingly inaccessible to other factors (Fig. 17A).

On the other hand, in the centromeric chromatin model, it could be seen that the incorporation of the H3-CENP-H3 tri-nucleosome disrupts the alignment of the poly-nucleosome array (Fig. 17B). Notably, the CENP-A nucleosome is exposed from the surrounding H3 nucleosomes. In this form, the nucleosomal disc surface of the CENP-A nucleosome may be accessible to and be recognized by other proteins, such as CENP-C and CENP-N. In conclusion, this model provides an explanation on how the flexible DNA ends of the CENP-A nucleosome may influence centromeric chromatin, leading to the increased accessibility of the CENP-A nucleosome and ultimately contributing to kinetochore formation.

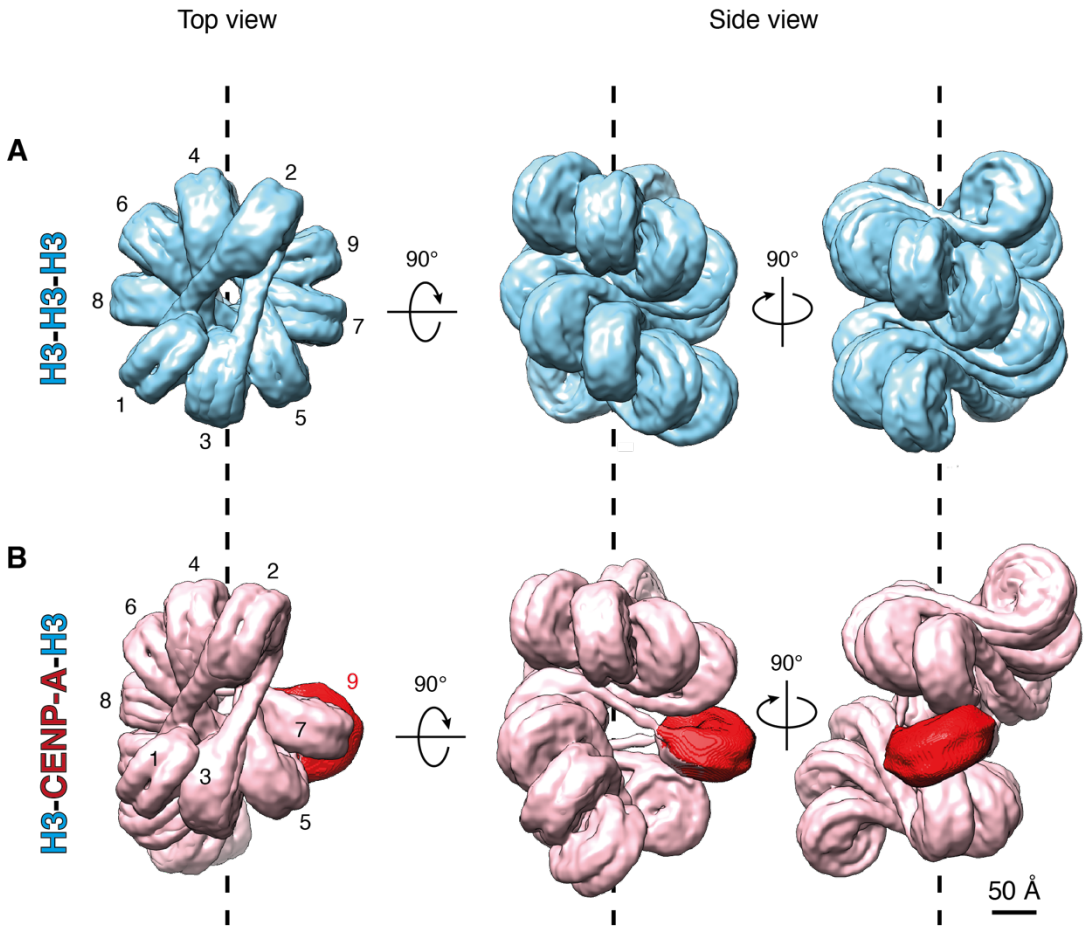


Fig. 17 Poly-nucleosome models of non-centromeric and centromeric chromatin

The assembly orders of the nucleosomes are indicated by numbers in the top view. **(A)** Poly-nucleosome model composed of 16 H3 nucleosomes with 22 bp linker DNAs. The model was generated by the H3-H3-H3 22 bp structure in Fig. 10B. **(B)** Poly-nucleosome model composed of 15 H3 nucleosomes and 1 CENP-A nucleosome with 22 bp linker DNAs. The CENP-A nucleosome is colored in red. The model was generated by the H3-H3-H3 22 bp and H3-CENP-A-H3 22 bp structures in Fig. 10.

2.5. Discussion

Here, I summarize the previous findings on the structures of the tri-nucleosome with 22 bp or 30 bp of linker DNAs. Two kinds of linker DNA conformations were observed, namely the outward-path conformation and the inward-path conformation. The two conformations can be distinguished by which side of the stacked nucleosome the linker DNAs are connected to.

When the linker DNA is 22 bp, which correspond to the linker DNA length found in human, the tri-nucleosome with a center-positioned CENP-A nucleosome (H3-CENP-A-H3 22 bp) only formed the outward-path conformation. This result was different form the H3-H3-H3 22 bp tri-nucleosome, which formed both inward-path and outward-path conformation. Furthermore, the obtained cryo-EM map of the H3-CENP-A-H3 22 bp tri-nucleosome showed that the orientation of the center-positioned CENP-A nucleosome is perpendicular to the stacked H3 nucleosome located at the side, which was different from the H3-H3-H3 22 bp tri-nucleosome. This may have an essential meaning for centromeric higher-order chromatin structure, which will be discussed in detail in the next section (section 2.5.2). When the linker DNA is extended from 22 bp to 30 bp, the tri-nucleosome with a center-positioned CENP-A (H3-CENP-A-H3 30 bp) changed from only exhibiting the inward-path conformation to both the inward-path and outward-path conformation.

Taken together, these results suggest that the inward-path conformation may be favorable by the stacked tri-nucleosome when the linker DNA is long (>=30 bp). Alternatively, the outward-path conformation may be favorable when the linker DNA length is short (<=22 bp). Additionally, tri-nucleosomes with a center positioned CENP-A nucleosome seem to favor the outward-path conformation, because the outward-path

conformation remained even after extending the linker DNA to 30 bp. This suggests that the CENP-A nucleosome may alter the chromatin higher-order structures in centromeres. This also suggests that the linker DNA length is also an important factor for determining the higher-order chromatin structure. Resultingly, the length of the repetitive DNA found in the centromere may also play a key role in shaping the functional structure of centromeric chromatin, which may be meaningful to kinetochore formation.

Intriguingly, the tri-nucleosome with 30 bp linker DNA and side-positioned CENP-A nucleosomes (CENP-A-H3-CENP-A 30 bp) displayed only the inward-path conformation, which was different from the result form H3-CENP-A-H3 tri-nucleosome with the same linker DNA length. This suggests that the proper positioning and spacing of the CENP-A nucleosome in the centromeric chromatin may also be critical for determining the centromeric chromatin structure. Therefore, in further studies, it is important to elucidate how the CENP-A nucleosomes are distributed in the centromeric chromatin to ensure the proper formation of chromatin structure.

Chapter 3: Structural Analysis of the Mono-Methyltransferase SET8 in Complex with the Nucleosome

3.1. Introduction

3.1.1. The H4K20me1 mark and the mono-methyltransferase SET8 (Pr-SET7/KMT5A)

The H4K20 residue has 3 methylation states (mono-, di-, tri-), of which the mono-methylation mark (H4K20me1) is introduced by the mono-methyltransferase SET8 (also named Pr-SET7/KMT5A) (Fig. 18A) (70–74). The H4K20me2/3 marks are introduced by the methyltransferases SUV420H1/2, using H4K20me1 as a substrate (75,76). Therefore, if the H4K20me1 mark is abolished, it may also lead to the disruption of the H4K20me2/3 marks. Indeed, the SET8 knockout in mice results in embryonic lethality (77). Further investigation showed that in embryonic cells, lack of catalytic active SET8 caused a growth arrest due to increased amounts of DNA damage and problems in cell cycle progression (77). Other reported effects of SET8 loss include the decondensation of large-scale chromatin and defects in DNA replication (Fig. 18B) (70,77). Additionally, it has also been reported that SET8 can methylate p53 (70,72,78).

The amount of SET8 and H4K20me1 fluctuates during the cell cycle (Fig. 18C) (70–72). The protein level of SET8 is highest during G2/M and early G1, while being absent in S phase. As a result, H4K20me1 levels show a similar oscillation. This proper levels of SET8 and H4K20me1 are important for the proper progression of the cell cycle.

SET8 is protein that contains a SET domain near the C terminal (Fig. 18D). The crystal structural of the SET domain of SET8 bound to the histone H4 peptide has been reported (Fig. 18E) (79,80). It has been reported that SET8 can mono-methylate the

nucleosomal H4 more efficiently than H4 alone (74). Therefore, it is possible that SET8 may also interact with sites other than the H4K20 residue in the nucleosome, such as the nucleosomal DNA and/or the acidic patch. However, the structural basis of nucleosomal H4K20 mono-methylation by SET8 has not been reported. As a result, the nucleosome sites interacting with SET8 remains unknown.

3.1.2. The H4K20me1 mark in centromeric chromatin

H4K20me1 of the CENP-A nucleosome in centromeric chromatin is essential for kinetochore assembly (55). From the results of ChIP-seq analysis and immunofluorescence analysis, it was shown that the H4K20me1 mark can be found in centromeric chromatin of chicken DT40 cells and human HeLa cells. Depletion of the H4K20me1 mark in centromeric chromatin led to a defect in kinetochore assembly. Additionally, further investigation showed that the mark was primarily installed in the H4 of CENP-A nucleosomes rather than the H4 of H3 nucleosomes.

Interestingly, it has been reported that the SET8 can mono-methylate the H4K20 residue in CENP-A nucleosomes with higher efficiency than the H4K20 residue in H3 nucleosomes *in vitro* (81). This has been accounted for the different orientations of the H4 tails found in the CENP-A nucleosome and the H3 nucleosome. Because the structural basis of the H4K20 mono-methylation by SET8 has not been reported, how the difference in H4 tail orientations affect the efficiency of H4H20me1 installation remains unclear.

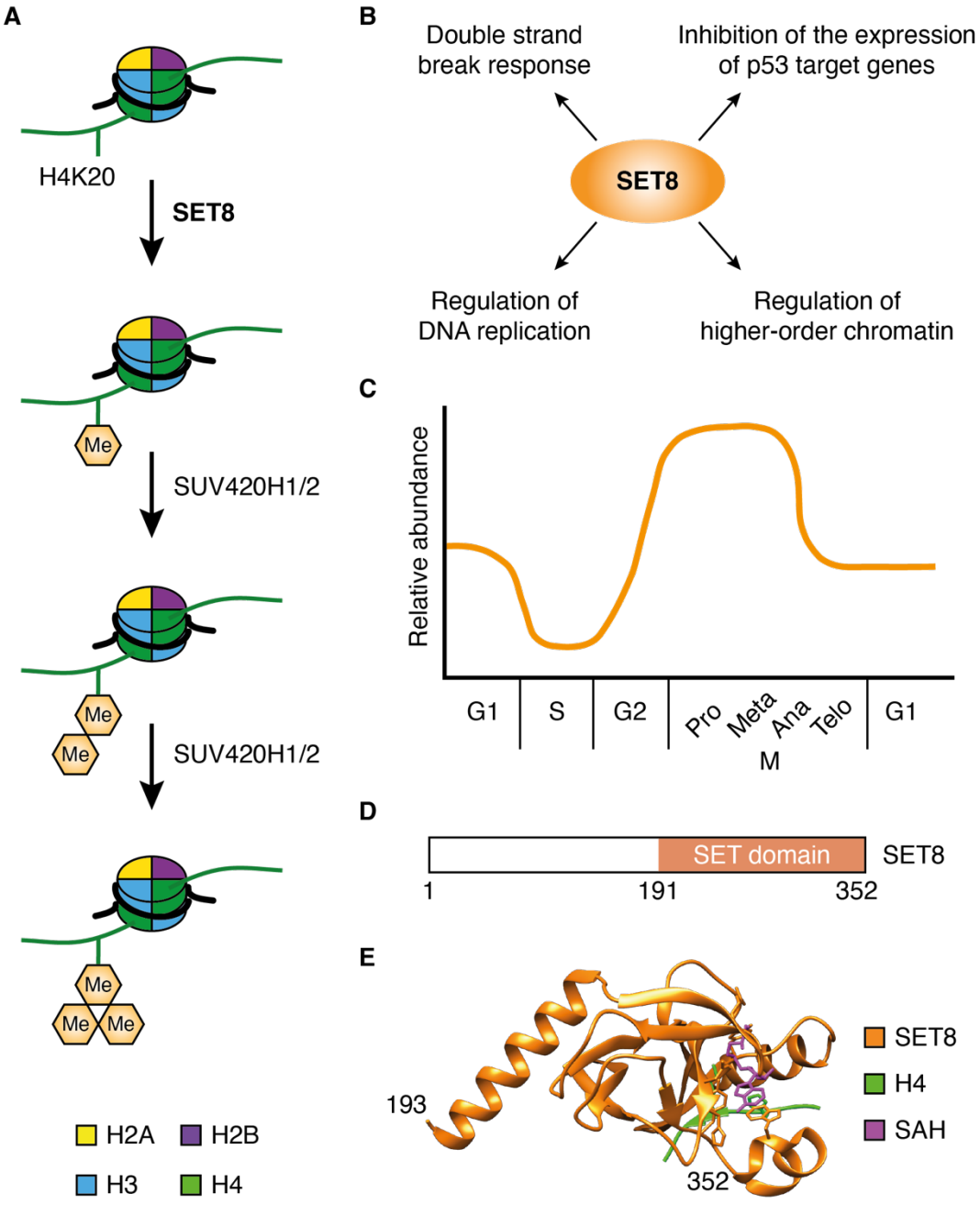

A
H4K20
SET8
Me
SUV420H1/2
Me
Me
SUV420H1/2
Me
Me
Me
H2A
H2B
H3
H4
B
Double strand break response
Inhibition of the expression of p53 target genes
SET8
Regulation of DNA replication
Regulation of higher-order chromatin
C
Relative abundance
G1
S
G2
Pro
Meta
Ana
Telo
M
G1
D
SET domain
SET8
1
191
352
E
193
352
SET8
H4
SAH

Fig. 18 The methyltransferase SET8

(A) H4K20 has three methylation states. **(B)** SET8 is involved in several biological events. **(C)** The amount of SET8 fluctuates throughout the cell cycle. Pro: prophase. Meta: metaphase. Ana: anaphase. Telo: telophase. **(D)** The domain structure of SET8. **(E)** The crystal structure of the SET domain of SET8 and the H4 peptide (PDB: 1ZKK).

3.2. Research Aims

The mono-methylation of H4K20 (H4K20me1), a modification primarily observed in CENP-A nucleosomes within centromeric chromatin, is involved in the functionality of the centromere. This investigation aims to elucidate why CENP-A nucleosomes are subjected to mono-methylation more efficiently than H3 nucleosomes, evaluated from a structural point of view. A comprehensive understanding of this phenomenon could provide a direct link between the structural properties of the CENP-A nucleosome and its biological function within centromeric chromatin.

Moreover, this research aims to identify potential interaction sites between SET8 and the nucleosome, extending beyond the established interaction with H4K20. To this end, the present study determined the cryo-EM structures of the SET8-CENP-A nucleosome complex and the SET8-H3 nucleosome complex. The derived cryo-EM structures yield critical insights into the previously outlined questions.

3.3. Methods

3.3.1. Purification of full length human SET8 and SET8 R188A R192A

In this study, human SET8 (KMT5A Isoform 2; Uniprot ID: Q9NQR1-2) was used. For the expression of the protein, the coding region of SET8 was inserted into a modified pET15b vector. This modified pET15 vector contained a His_6-tag and a PreScission Protease recognition sequence instead of the original thrombin recognition sequence.

The purification process was conducted similarly to a previously described method (81). *E. coli* BL21(DE3) was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to overexpress human SET8. After sonication of the cells, they were centrifuged, and the supernatant containing His_6-tagged SET8 was collected. Subsequently, Ni-NTA affinity chromatography was employed to purify the His_6-tagged SET8. The concentration of the purified His_6-tagged SET8 was determined using the Bradford method. PreScission protease (4 U/mg) was added to the sample, which was subsequently dialyzed against Mono S wash buffer comprising 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10% glycerol, and 2 mM 2-mercaptoethanol. Centrifugation was performed to remove precipitates. His_6-tag removal was confirmed by SDS-PAGE analysis, before further purification using Mono S cation exchange chromatography with Mono S elution buffer consisting of 50 mM Tris-HCl (pH 7.5), 600 mM NaCl, 10% glycerol, and 2 mM 2-mercaptoethanol. Finally, HiLoad 16/600 Superdex 200 pg (GE Healthcare) gel filtration chromatography was employed for the final purification step of SET8. The gel filtration chromatography was performed in a buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 0.2 mM EDTA, 10% glycerol, and 1 mM DTT. The purified SET8 was stored at -80°C before use.

Additionally, SET8 R188A R192A plasmid was generated using PCR site-directed

mutagenesis, and the expression and purification process for SET8 R188A R192A was conducted following a similar procedure.

3.3.2. Purification of human histones and histone mutants

In accordance with previous methods (63), human core histones H2A, H2B, H3, H4, H2A E56T/E61T/E64T/D90S/E91T/E92T (H2A acidic patch mutant, $H2A^{APM}$), and H2B E105T/E113T (H2B acid path mutant, $H2B^{APM}$) were prepared. First, N-terminally His_6-tagged H2A, H2B, H3, H4, $H2A^{APM}$, and $H2B^{APM}$ were expressed in *Escherichia coli* cells and lysed. Then, the insoluble fraction was collected and denatured using a buffer which contained 50 mM Tris-HCl (pH 8.0), 7 M guanidine-HCl, 500 mM NaCl, and 5% glycerol. Next, under denaturing conditions, the His_6-tagged histones were purified by Ni-NTA agarose chromatography (QIAGEN). Thrombin protease was applied to remove the His_6-tags. After removal of the His_6-tags, the histones were further purified using Mono S column chromatography (GE Healthcare). At last, the histones were dialyzed against water, before being freeze-dried. Histone powders were stored at 4°C before use.

3.3.3. *In vitro* reconstitution and purification of the histone octamers

In this project, 3 kinds of histone octamers were used, namely the H2A-H2B-H3-H4 octamer (H3 octamer), the $H2A^{APM}$-$H2B^{APM}$-H3-H4 octamer (APM octamer), and the H2A-H2B-CENP-A-H4 (CENP-A octamer).

The H3 octamer and the APM octamer were prepared similarly, by the following procedure. First, the 4 kinds of required freeze-dried histones (H2A, H2B, H3, and H4 for the H3 octamer; $H2A^{APM}$, $H2B^{APM}$, H3, and H4 for the APM octamer) were mixed at a 1:1:1:1 stoichiometry. The mixed powders were then denatured in a buffer containing

20 mM Tris-HCl buffer (pH 7.5), 7 M guanidine-HCl, and 2 mM 2-mercaptoethanol. Following denaturation, the mixture was dialyzed against a buffer consisting of 10 mM Tris-HCl (pH 7.5), 2 M NaCl, 1 mM EDTA, and 5 mM 2-mercaptoethanol, for refolding. Lastly, HiLoad 16/600 Superdex 200 pg (GE Healthcare) column chromatography was applied for the purification of the histone octamer. The purified histone octamers were stored at -80°C before use.

The CENP-A octamer was also prepared by a similar procedure, by using the freeze-dried H2A, H2B, CENP-A, and H4. However, the composition of the denaturation buffer was different, and consisted of 20 mM Tris-HCl buffer (pH 7.5), containing 7 M guanidine-HCl and 20 mM 2-mercaptoethanol.

3.3.4. *In vitro* reconstitution and purification of the nucleosomes

The salt dialysis method was utilized to reconstitute the H3 nucleosome, the CENP-A nucleosome, and the acidic patch mutant nucleosome (nuc^{APM}). This process involved the use of the 145-base pair Widom 601 DNA and the H3 octamer, the CENP-A octamer, or the APM octamer (82).

Following the dialysis, a Prep Cell Model 491 apparatus (Bio-Rad) was employed to purify the nucleosomes, using a native polyacrylamide gel. An elution buffer that contains 20 mM Tris-HCl with a pH of 7.5 and 1 mM dithiothreitol was used in the process.

3.3.5. EMSA of wild-type SET8 and nucleosomes

The purified H3 nucleosome (0.52 μM) or the purified CENP-A nucleosome (0.52 μM) was mixed with SET8 (0.26, 0.52, 0.78, 1.0, 1.3 μM) in 10 μl of reaction solution. The reaction solution was composed of 10 mM HEPES-KOH (pH 7.8), 18 mM Tris-HCl

(pH 7.5), 50 mM NaCl, 50 mM KCl, 0.10 mM EDTA, 5% glycerol, 0.5 mM DTT, and 0.10 mM S-adenosyl-Lhomocysteine. The mixed sample was incubated at 25°C for 30 minutes and analyzed by 6% native polyacrylamide gel (0.2 TBE).

3.3.6. Purification of the SET8-nucleosome complexes by the GraFix method

To prepare the SET8-H3 nucleosome sample, SET8 (1.0 μM) and the H3 nucleosome (0.52 μM) was mixed in 1 ml of reaction solution, which contained 10 mM HEPES-KOH (pH 7.8), 18 mM Tris-HCl (pH 7.5), 50 mM NaCl, 50 mM KCl, 0.10 mM EDTA, 5% glycerol, 0.5 mM DTT, and 0.10 mM S-adenosyl-Lhomocysteine. To be precise, reaction solution that contained the nucleosome received four separate additions of SET8. After each SET8 addition, the sample mixture was incubated in a water bath at 25°C for 5 minutes. Following the completion of the SET8 addition, the sample mixture was incubated at 25 °C in a water bath for 15 min.

The SET8-CENP-A nucleosome sample was also prepared similarly, except that the concentrations of SET8 was different. SET8 (0.94 μM) was mixed with CENP-A nucleosome (0.52 μM) in this case.

Gradient Master (SKB) was used to prepare the gradient for the GraFix method. The top buffer (light buffer) consists of 10 mM HEPES-NaOH (pH 7.5), 100 mM NaCl, 1 mM DTT, and 5% sucrose, while the bottom buffer (heavy buffer) consists of 10 mM HEPES-NaOH (pH 7.5), 100 mM NaCl, 1 mM DTT, 20% sucrose, and 4% paraformaldehyde. The two buffers were added to Ultra-Clear Centrifuge Tubes (344058; Beckman Coulter) and the short cap was applied to close the tube. The SHORT Sucrose 5-20% (SW28) program was used to prepare the gradient, which was subsequently chilled at 4°C for one hour.

1 ml of the gradient solution was taken from the top of the centrifuge tube, followed by the application of the samples. Subsequently, the centrifuge tubes were put into a SW32Ti swinging bucket rotor (Beckman Coulter) and centrifuged at 27,000 rpm at 4°C for 16 hours. The sample fractions (1 ml each) were then taken from the surface of the gradient solution after the centrifugation.

The fractions were run on a 6% native polyacrylamide gel (0.2 TBE) to decide the fractions to be collected for following procedure. After collection of the required fractions, the buffer was exchanged using chromatography on PD-10 columns (GE Healthcare). The elution buffer was composed of 10 mM HEPES-NaOH (pH 7.5) and 2 mM TCEP (pH 7.5). The samples were concentrated and kept at 4°C.

3.3.7. Cryo-EM sample preparation

Quantifoil R1.2/1.3 200-mesh Cu grids were first glow-discharged. Then, 2.5 μl of the SET8-H3 nucleosome complex or the SET8-CENP-A nucleosome complex was applied to the Quantifoil grid. Using a Vitrobot Mark IV (Thermo Fisher Scientific), the grids were blotted for 5 or 6 seconds without waiting, with the bolt force set to 0, at 100% humidity and 4°C. The grids were immediately plunged into liquid ethane to prevent the formation of crystalline ice.

3.3.8. Cryo-EM data collection

Krios G3i cryo-TEM (Transmission Electron Microscope) (Thermo Fisher Scientific) , running at 300 kV, was used to capture images of the SET8-H3 nucleosome and SET8-CENP-A nucleosome complexes. 2,355 movies for the SET8-nucleosome complex were captured using the EPU (Thermo Fisher Scientific) auto acquisition

program with a 1.05 Å pixel size. On a Falcon 3EC direct electron detector (Thermo Fisher Scientific), digital micrographs were captured with 63 s exposure periods in the electron counting mode, keeping a total of 51 frames with a total dosage of ~52 electron/Å^2.

On the other hand, for the SET8-CENP-A nucleosome, the SerialEM (83) auto acquisition program was used to record 6,075 movies at a pixel size of 1.05 Å. Digital micrographs were captured with exposure times of 7 s on a K3 BioQuantum (Gatan) direct electron detector in the electron counting mode, with a slit width of 25 eV, and the retention of 40 frames with a total dose of ~60 electron/Å^2.

3.3.9. Image processing

For both the SET8-H3 nucleosome and the SET8-CENP-A nucleosome complexes, all movies frames were adjusted using MOTIONCOR2 (65) with dose weighting. The contrast transfer function (CTF) estimation was conducted through CTFFIND4 (66), which utilized micrographs with dose weighting. For subsequent image processing of both complexes, RELION3.0 and RELION3.1 (84) were employed.

Particles with a box size of 180 × 180 pixels were picked semi-automatically. Junk particles were then removed using 2D classification, followed by 3D classification. For the initial model of the 3D classification of the SET8-H3 nucleosome complex, the crystal structure of the NCP (3LZ0) with a lowpass-filter of 60 Å, was used. As for the 3D classification of the SET8-CENP-A nucleosome complex, the ab initio model generated in RELION3.1 served as the initial model.

Following 3D classification for both complexes, particle polishing and several rounds of CTF refinement were carried out. The refined 3D maps of the SET8-H3

nucleosome and SET8-CENP-A nucleosome complexes had resolutions of 3.15 and 3.00 Å, respectively, as estimated by the gold standard Fourier Shell Correlation (FSC) at an FSC = 0.143 (67).

To calculate the local resolutions of both complexes, RELION3.1 was used. The final maps of both complexes were normalized with MAPMAN (68) and visualized using UCSF Chimera (69) and UCSF ChimeraX (85). Further details regarding the processing statistics for the two complexes can be found in Table 3.

3.3.10. Model building and refinement

Using rigid-body fitting in UCSF Chimera (69), the crystal structures of the H3 nucleosome (PDB: 3LZ0), SET8 (PDB: 1ZKK) without ligands, and the atomic model of CENP-A from the CENP-A nucleosome (PDB: 6C0W) were incorporated into the cryo-EM maps of the SET8-H3 nucleosome and the SET8-CENP-A nucleosome complexes. Using COOT (86), the models for both the SET8-H3 nucleosome and the SET8-CENP-A nucleosome complexes were manually constructed. Using Phenix (87), the models were subsequently subjected to real-space refinement. Further details regarding the processing statistics for the two complexes can be found in Table 4.

3.3.11. Binding analysis of the mutant SET8 and the mutant H3 nucleosome

The purified H3 nucleosome or acidic patch mutant nucleosome (nuc^{APM}) (0.52 μM) was mixed with either SET8 or its mutant SET8 DM (1.0 and 2.1 μM) in a 5.0 μl reaction mixture. The reaction mixture contained 10 mM HEPES-KOH (pH 7.8), 16 mM Tris-HCl (pH 7.5), 50 mM NaCl, 50 mM KCl, 0.12 mM EDTA, 5% glycerol, 0.5 mM DTT, 0.10 mM S-adenosyl-L-homocysteine, and 1.5 μM double-stranded DNA 50-mer, which

served as competitor DNA. As a control, mixtures without nucleosomes were also prepared. In these mixtures, the double-stranded 50-mer DNA (1.5 μM) were mixed with SET8 (2.1 and 4.2 μM) in a similar condition.

After incubating the samples at 25°C in a water bath for 30 minutes, they were separated using 6% nondenaturing polyacrylamide gel electrophoresis with a running buffer composed of 0.2× TBE. Finally, gels were stained with EtBr to visualize DNA. Photos of the gels were imaged with an Amersham Imager 680 (GE Healthcare).

3.3.12. Methylation assay of the mutant SET8 and the mutant H3 nucleosome

The purified H3 nucleosome or acidic patch mutant nucleosome (nuc^{APM}) (0.52 μM) was mixed with SET8 or its mutant SET8 DM (0.15 μM) in a 5.0 μl reaction mixture. This mixture contained 10 mM HEPES-KOH (pH 7.8), 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 20 mM KCl, 60 μM EDTA, 2% glycerol, 0.50 mM DTT, 80 μM S-adenosylmethionine, and 1.5 μM of double-stranded DNA 50mer (acting as a competitor). Following this process, the reaction solutions were incubated for either 1 or 3 minutes at 25°C. For the termination of the reaction, 5 μl of 4% SDS solution(0.10 mM Tris-HCl (pH 6.8), 20% glycerol, and 0.2% bromophenol blue) was added. Subsequently, the samples were heated at 95°C for 15 minutes before being fractionated by a 18% SDS-PAGE.

Using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad), the proteins were then transferred to an Amersham Hybond 0.2 μm polyvinylidene difluoride (PVDF) membrane (GE Healthcare). For blocking the membrane, 5% skim milk powder dissolved in phosphate-buffered saline with 0.05% Tween 20 (PBS-T) was applied (1 hour at room temperature). After washing the membrane with PBS-T, it was incubated with primary

antibodies (4°C overnight). The primary antibodies that were used were as following: the mouse monoclonal antibody against monomethylated H4K20 (CMA421; HayashiTakanaka et al, 2015) and the anti-H2B monoclonal antibody (53H3: Cell Signaling). The anti-monomethylated H4K20 antibody was diluted using Can Get Signal solution 1 (TOYOBO) to a final concentration of 1 μg/ml. The anti-H2B antibody was diluted using Can Get Signal solution 1 (TOYOBO) 10,000-fold.

Following another PBS-T wash, the membrane was incubated with the secondary antibody: mouse IgG HRP-linked $F(ab')_2$ fragment from sheep (NA9310: GE Healthcare), at 4°C for 2 hours. The secondary antibody was diluted 10,000-fold with Can Get Signal solution 2 (TOYOBO). After a final wash with PBS-T, Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) was applied to the membrane. Using an Amersham Imager 680 (GE Healthcare), the blot's image was captured through chemiluminescent detection.

Table 3 Data collection and reconstruction details of the SET8-CENP-A nucleosome and the SET8-H3 nucleosome complex

	SET8-CENP-A nucleosome (EMD-30552)	SET8-H3 nucleosome (EMD-30551)
Data Collection		
Electron microscope	Krios G3i	Krios G3i
Camera	K3	Falcon3
Pixel size (A/pix)	1.05	1.05
Defocus range (μm)	1.0-2.5	1.5-3.0
Exposure time (second)	7	63
Total dose (e/A2)	60	52
Movie frames (no.)	40	51
Total micrographs (no.)	6,075	2,355
Reconstruction		
Software	Relion 3.0 & 3.1	Relion 3.0 & 3.1
Particles for 2D classification	3,505,509	981,019
Particles for 3D classification	3,181,015	841,758
Particles in the final map (no.)	380,269	225,437
Symmetry	C1	C1
Final resolution (Å)	3	3.15
FSC threshold	0.143	0.143
Map sharpening B factor (Å2)	-78.86	-25.87

Table 4 Model building, refinement, model composition, validation, and Ramachandran plot details of the SET8-CENP-A nucleosome and the SET8-H3 nucleosome complex

	SET8-CENP-A nucleosome (EMD-30552)	SET8-H3 nucleosome (EMD-30551)
Model building		
Software	Coot	Coot
Refinement		
Software	Phenix	Phenix
Model composition		
Protein	917	934
Nucleotide	272	290
Validation		
MolProbity score	1.78	1.79
Clash score	9.82	8.78
R.m.s deviations		
Bond lengths (Å)	0.005	0.007
Bond angles (°)	0.68	0.717
Ramachandran plot		
Favored (%)	96.11	95.41
Allowed (%)	3.89	4.59
Outluers (%)	0	0

3.4. Results

3.4.1. Purification of the recombinant human SET8

Full length human SET8 was expressed in *E. coli* and cells and purified as a recombinant protein, as described in Section 3-3-1. Briefly, the cells were sonicated, and SET8 was purified by combining Ni-NTA affinity chromatography, Mono S cation exchange chromatography, and HiLoad 16/600 Superdex 200 pg gel filtration chromatography. The quality of the purified SET8 is confirmed by SDS-PAGE (Fig. 19A).

3.4.2. Purification of the CENP-A nucleosome and the H3 nucleosome

The CENP-A nucleosome and the H3 nucleosome were reconstituted *in vitro* by mixing the 145 bp 601 DNA with the CENP-A histone octamer or the H3 histone octamer, respectively, as described in Sections 3-3-2~3-3-4. The qualities of the purified products were confirmed by SDS-PAGE and native PAGE (Fig. 19B-E). From the results of SDS-PAGE, 4 bands with equal intensities can be observed, showing that the nucleosomes contain equimolar ratios of H2A, H2B, CENP-A/H3, and H4 (Fig. 19B&D). From the native PAGE, only a single band can be observed, showing that the purities of the nucleosomes are high (Fig. 19C&E). Together, these results show that the nucleosomes were prepared successfully.

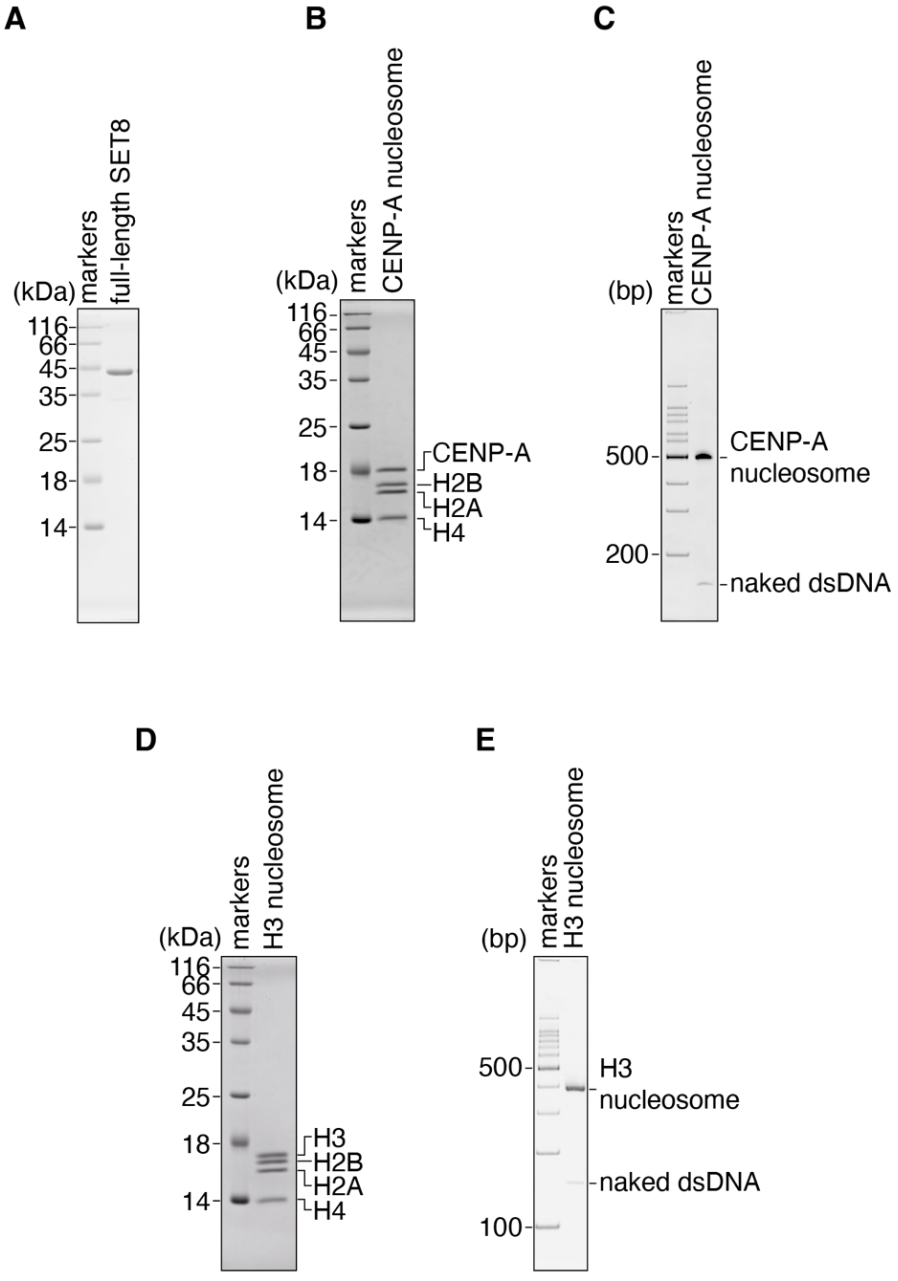

A
markers
full-length SET8
(kDa)
116
66
45
35
25
18
14
B
markers
CENP-A nucleosome
(kDa)
116
66
45
35
25
18
14
CENP-A
H2B
H2A
H4
C
markers
CENP-A nucleosome
(bp)
500
200
CENP-A
nucleosome
naked dsDNA
D
markers
H3 nucleosome
(kDa)
116
66
45
35
25
18
14
H3
H2B
H2A
H4
E
markers
H3 nucleosome
(bp)
500
100
H3
nucleosome
naked dsDNA

Fig. 19 Purification of SET8 and nucleosomes

(A) Purified full-length human SET8 was analyzed by 16% SDS-PAGE and stained by CBB, visualizing protein bands. **(B)** The purified H3 nucleosome was analyzed by 18% SDS-PAGE and stained by CBB, visualizing protein bands. **(C)** The purified H3 nucleosome was analyzed by 6% non-denaturing PAGE and stained by EtBr, visualizing DNA bands. **(D)** The purified CENP-A nucleosome was analyzed by 18% SDS-PAGE and stained by CBB, visualizing protein bands. **(E)** The purified CENP-A nucleosome was analyzed by 6% non-denaturing PAGE and stained by EtBr, visualizing DNA bands.

3.4.3. EMSA of SET8 and the nucleosomes

To confirm if SET8 could bind nucleosomes, EMSA of SET8 and the CENP-A nucleosome/H3 nucleosome was performed. The amount of SET8 was titrated against each kind of nucleosome, and the SET8-nucleosome mixtures were analyzed by non-denaturing PAGE. If SET8 binds the nucleosomal DNA non-specifically, then it is expected that a smear out of the nucleosome band would be observed instead of specific bands.

From the results, two specific bands above the free nucleosome bands can be observed (Fig. 20). This result has two significant meanings. First, it showed that SET8 can bind both the CENP-A nucleosome and the H3 nucleosome. Second, the two discrete bands suggested that SET8 can bind two specific positions of the nucleosome, which are shared between the CENP-A nucleosome and the H3 nucleosome. A possible candidate could be the acidic patch of the nucleosome, which is formed by histones H2A and H2B, and is present in both the CENP-A nucleosome and the H3 nucleosome.

After confirming that SET8 can bind the nucleosome, I proceeded to the preparation of the SET8-CENP-A nucleosome complex and the SET8-H3 nucleosome complex for cryo-EM analysis.

3.4.4. Sample preparation for cryo-EM analysis

The SET8-CENP-A nucleosome complex and the SET8-H3 nucleosome complex was prepared by the GraFix method, as described in Section 3-3-6 (Fig. 21A). Each sample was applied to 2 tubes for ultracentrifugation. The indicated fractions were collected after ultracentrifugation (Fig. 21B-E), buffer exchanged by the PD-10 column, concentrated, and analyzed by 6% non-denaturing PAGE (Fig. 22).

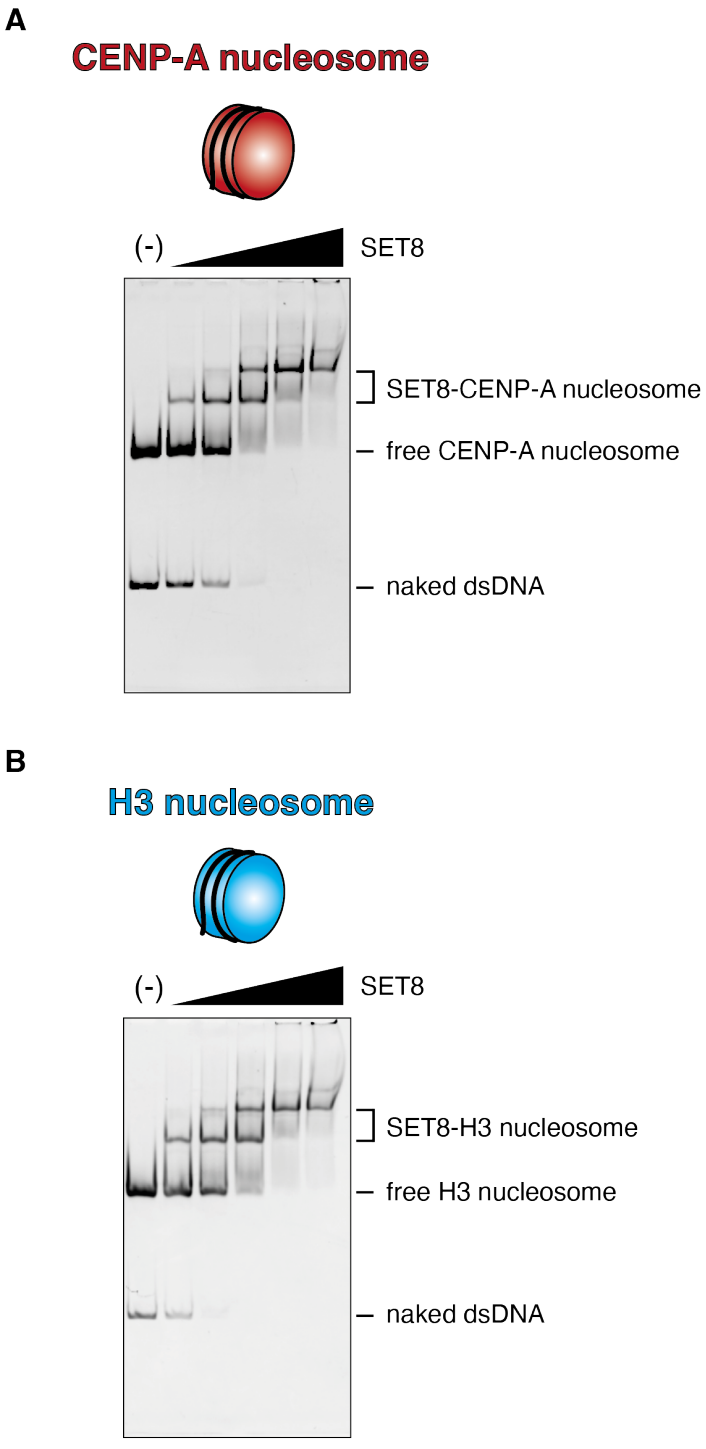


Fig. 20 EMSA of SET8 and the nucleosomes

Samples were analyzed by 6% non-denaturing PAGE and stained by EtBr, staining DNA.

(A) Result of the CENP-A nucleosome. **(B)** Result of the H3 nucleosome.

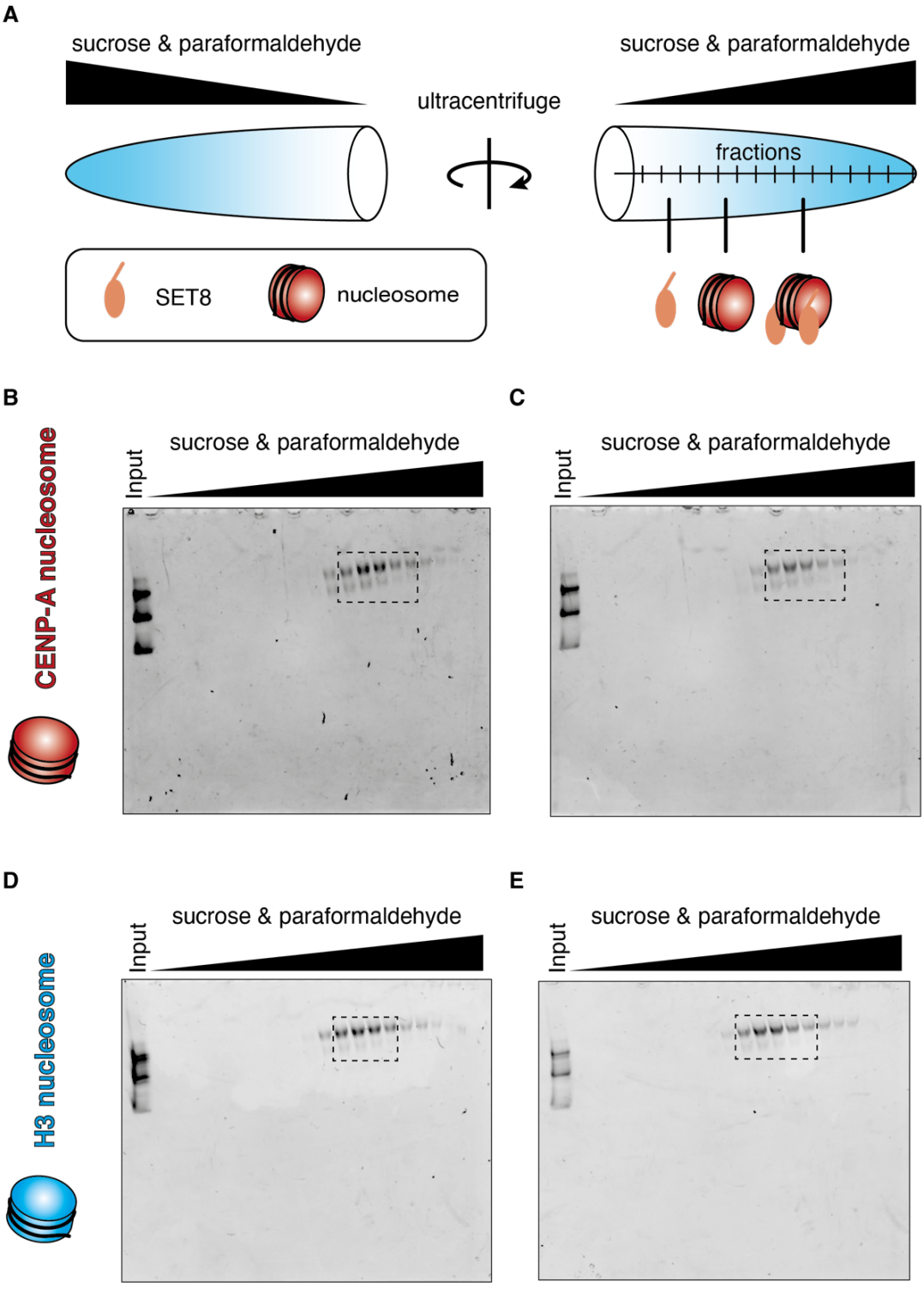

A
sucrose & paraformaldehyde
ultracentrifuge
sucrose & paraformaldehyde
fractions
SET8
nucleosome
B
CENP-A nucleosome
Input
sucrose & paraformaldehyde
C
Input
sucrose & paraformaldehyde
D
H3 nucleosome
Input
sucrose & paraformaldehyde
E
Input
sucrose & paraformaldehyde

Fig. 21 Collected fractions after GraFix

Fractions were analyzed by non-denaturing PAGE following GraFix. Collected fractions are indicated by the dashed box. **(A)** Schematic representation of the GraFix method. **(B&C)** GraFix result of the SET8-CENP-A nucleosome. **(D&E)** GraFix result of the SET8-H3 nucleosome.

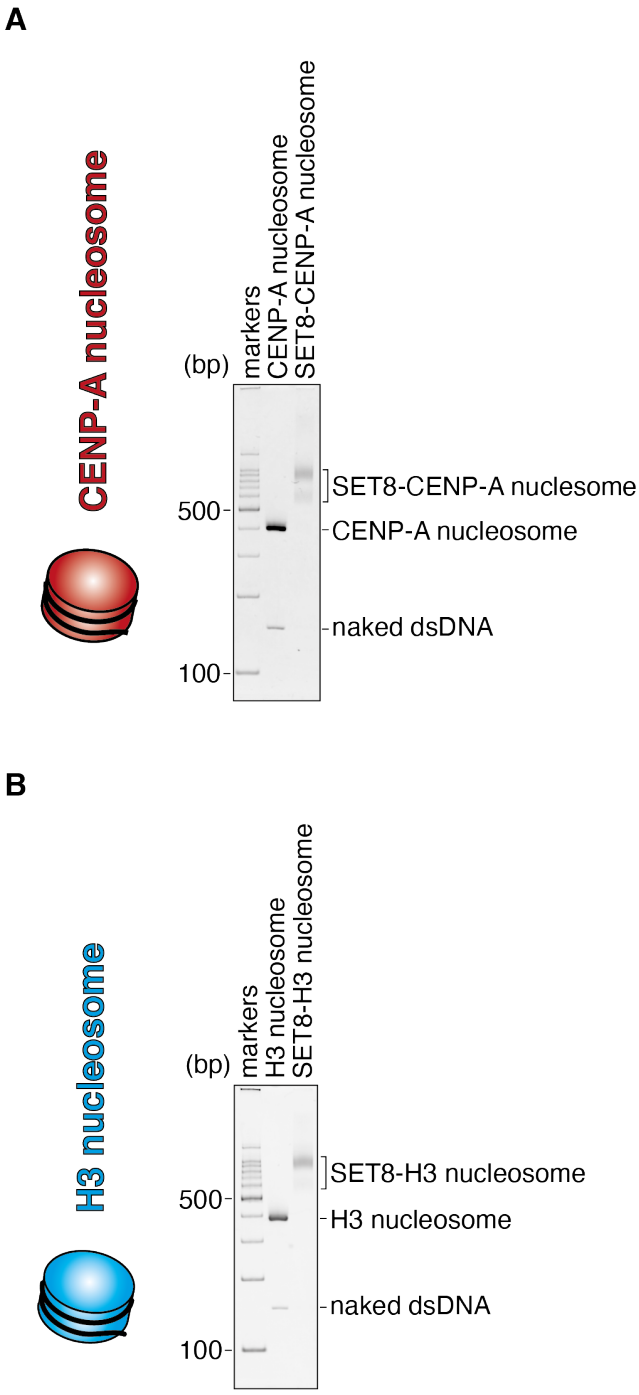


Fig. 22 Purified SET8-nucleosome complexes

Free nucleosomes and complexes were analyzed together. **(A)** The SET8-CENP-A nucleosome complex. **(B)** The SET8-H3 nucleosome complex.

3.4.5. Cryo-EM analysis of the SET8-nucleosome complexes

The purified SET8-CENP-A nucleosome complex and the purified SET8-H3 nucleosome complex were observed by the Krios G4 cryo-Transmission Electron Microscope (cryo-TEM) (Thermo Fisher Scientific) (Fig. 23A&C). The particles had good distribution, so I proceeded to data collection and single particle analysis (SPA).

From the results of 2D classification, density other than the nucleosome core particle can be observed in several classes (Fig. 23B&D). This density is thought to correspond to SET8, and seems to be located above and below the nucleosome disc.

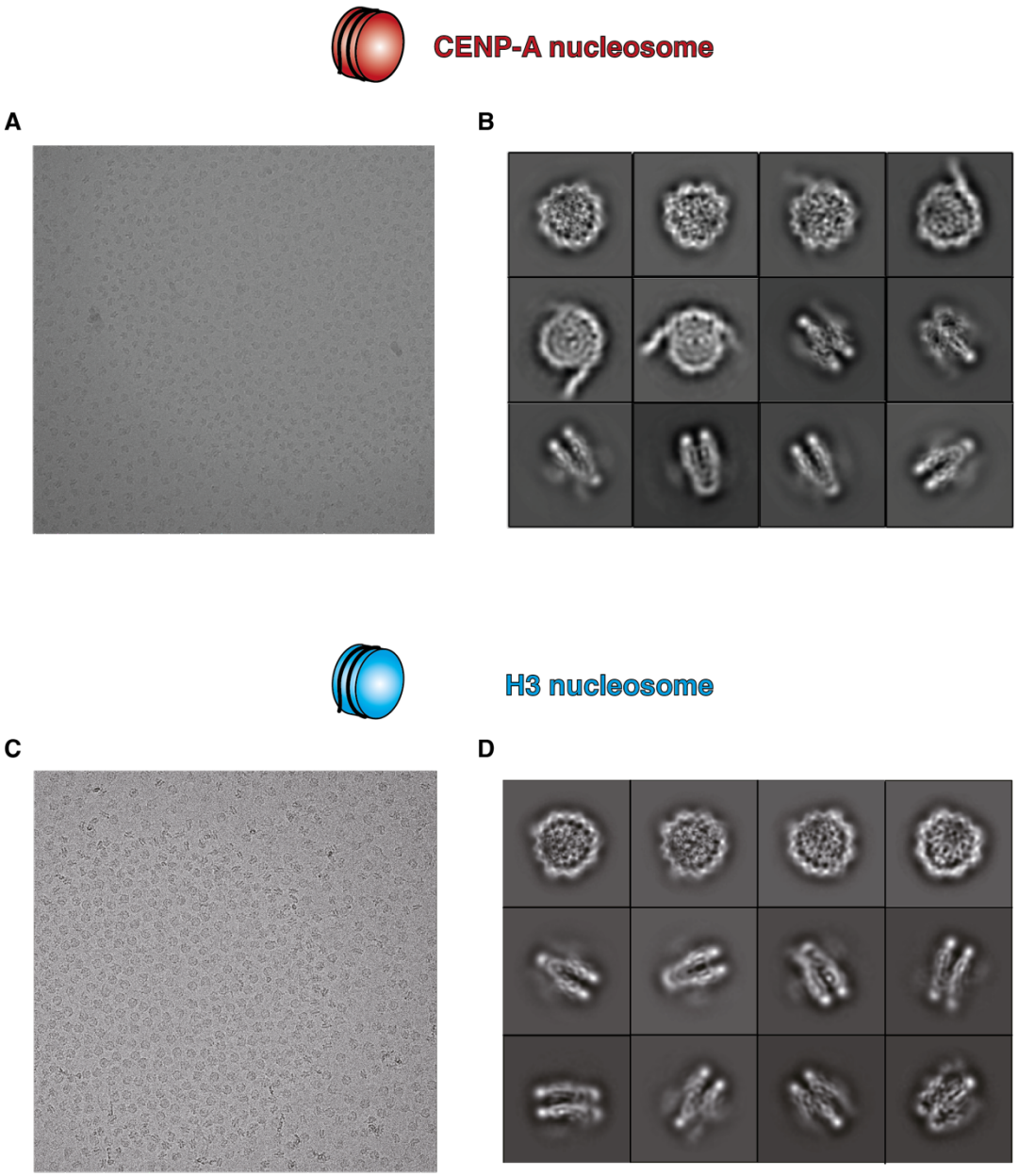


Fig. 23 Micrographs and 2D Classification of SET8-nucleosome complexes

(A) Micrograph of the SET8-CENP-A nucleosome complex. **(B)** 2D classification result of the SET8-CENP-A nucleosome complex. Note that this is from a different data set (grid) from (A). **(C)** Micrograph of the SET8-H3 nucleosome complex. **(D)** 2D classification result of the SET8-H3 nucleosome complex.

3.4.6. Cryo-EM structures of the SET8-nucleosome complexes

Considering that a maximum of two SET8 molecules could bind one nucleosome, and that the purified sample may contain a mixture of one SET8-bound and two SET8-bound particles, a mask was applied to the nucleosome itself and one SET8 on one side of the nucleosome to improve resolution of the final structure. Therefore, in the final structure, only one SET8 molecule is present on one side of the nucleosome, but in solution, two SET8 molecules can bind to both sides of the nucleosome.

Finally, we obtained the cryo-EM structures of the SET8-CENP-A nucleosome complex with 3.00 Å resolution and the SET8-H3 nucleosome complex with 3.15 Å resolution (Fig. 24). Additional information about the cryo-EM structures can be found in Fig. 25 and Fig. 26. Both structures have a similar overall appearance. In the structures, SET8 is located above the nucleosome disc, and the H4 tail can be seen extending from the nucleosome disc into the catalytic pocket of the SET domain in SET8.

Importantly, we noticed that SET8 interacts with the nucleosome acidic patch by several amino acid residues extending from the N-terminus of the SET domain. We were able to build models using the acquired cryo-EM map (Fig. 27A&C). The residues in SET8 interacting with the nucleosome acidic patch were identified as R188 and R192 (Fig. 27B&D). SET8 R188 interacts with the acidic patch residues: H2A E56 and H2B E113, while SET8 R192 interacts with the acidic patch residues: H2AE61 and E92.

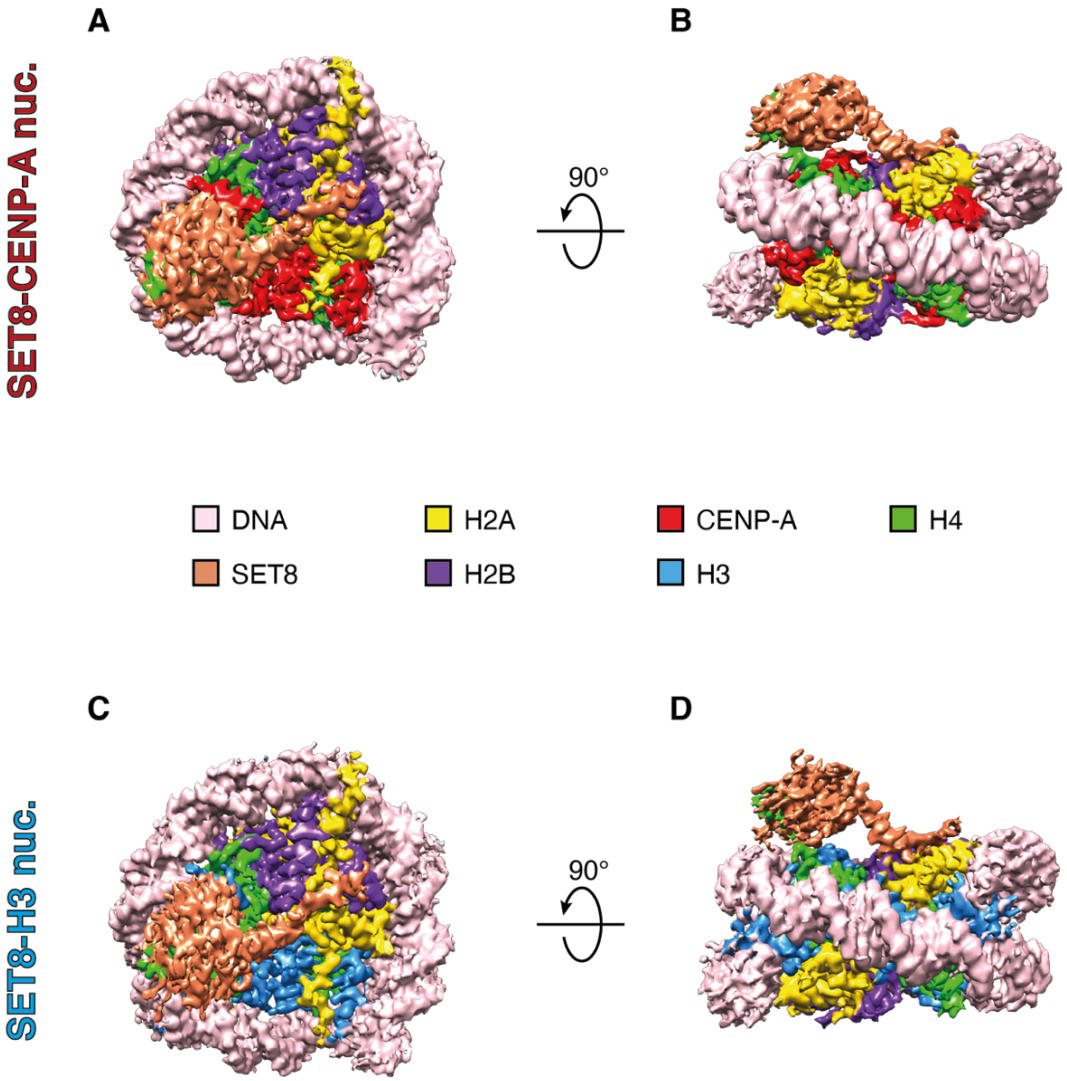


Fig. 24 Cryo-EM structures of the SET8- nucleosome complexes

(A&B) The cryo-EM structures of the SET8-CENP-A nucleosome complex (EMD-30552). The overall resolution of the complex is 3.00 Å. (A) Top view of the complex. (B) Side view of the complex. **(C&D)** The cryo-EM structures of the SET8-H3 nucleosome complex (EMD-30551). The overall resolution of the complex is 3.15 Å. (A) Top view of the complex. (B) Side view of the complex. nuc.: nucleosome.

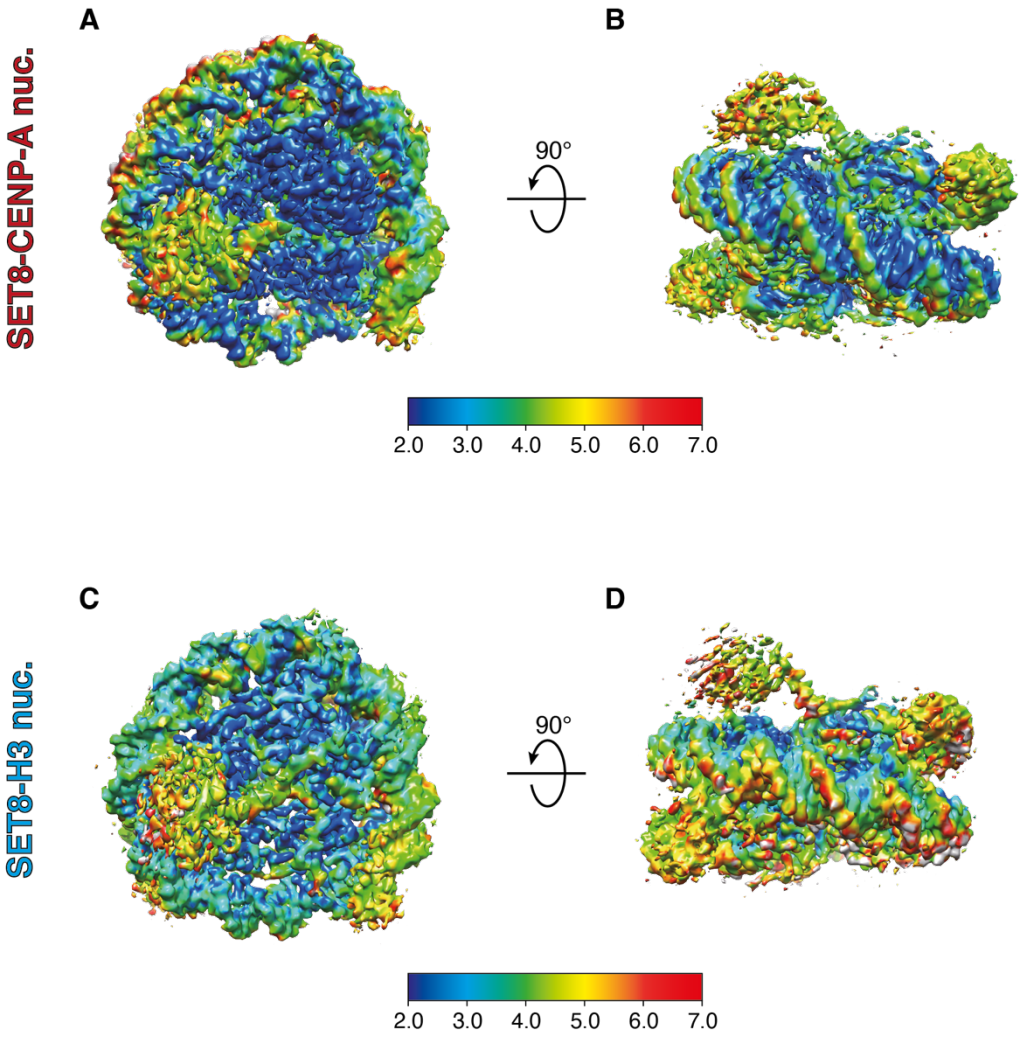


Fig. 25 Local resolutions of the SET8-nucleosome complexes

(A&B) The SET8-CENP-A nucleosome complex. (A) Top view. (B) Side view. **(C&D)** The SET8-H3 nucleosome complex. (C) Top view. (D) Side view. nuc.: nucleosome.

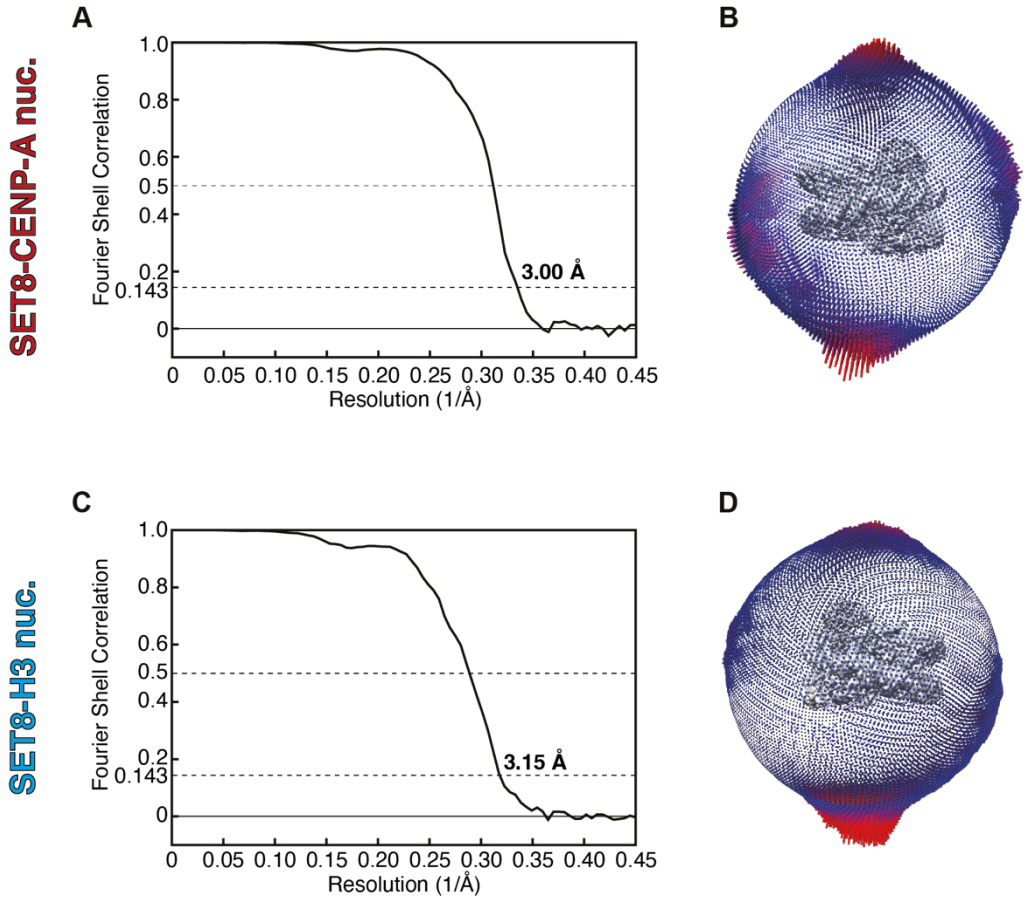


Fig. 26 Fourier Shell Correlation (FSC) curves and Euler angle distribution maps of the SET8-nucleosome complexes

(A&C) FSC curves of the SET8-CENP-A nucleosome (A) and the SET8-H3 nucleosome (C). **(B&D)** Euler angle distribution maps of the SET8-CENP-A nucleosome (B) and the SET8-H3 nucleosome (D).

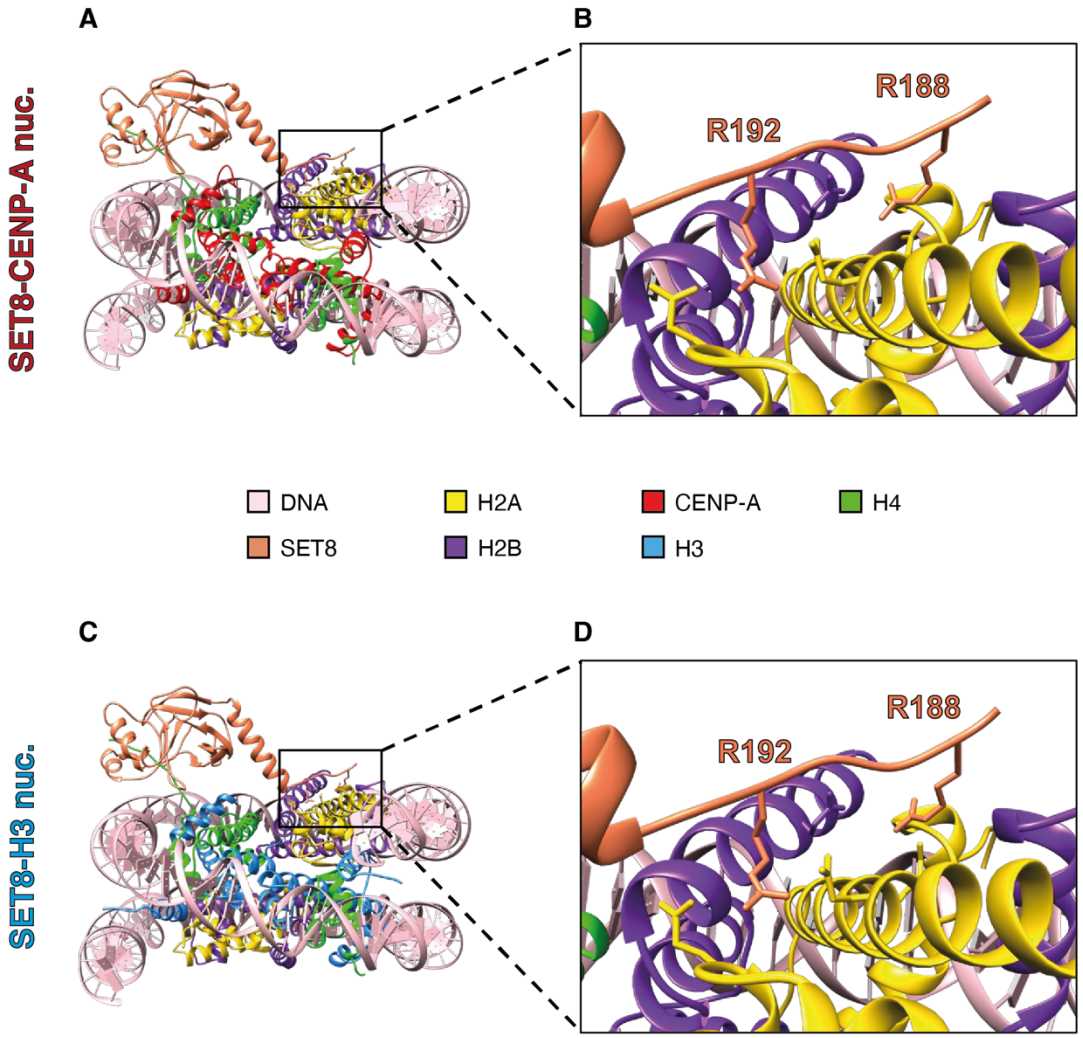


Fig. 27 Interaction between SET8 and the acidic patch of the nucleosome

(A&C) Model of the SET8-CENP-A nucleosome complex (A) (PDB:7D20) and the SET8-H3 nucleosome (C) (PDB: 7D1Z). **(B&D)** Close up of the residues in SET8 that are interacting with the acidic patch of the CENP-A nucleosome (B) and the H3 nucleosome (C). nuc.:nucleosome.

3.4.7. Preparation of the SET8 and nucleosome mutants

To investigate the importance of the SET8-acidic patch interaction on SET8 nucleosome binding ability and SET8 nucleosomal H4 mono-methylation, I proceeded to mutant analysis. In order to do so, I prepared the SET8 R188A R192A double mutant ($SET8^{DM}$) and the nucleosome acidic patch mutant with histone H3 (H3 $nuc.^{APM}$), which contains the H2A E56T/E61T/E64T/D90S/E91T/E92T and H2B E105T/E113T mutations (Fig. 28). Because the structure of the SET8-CENP-A nucleosome complex has a similar overall appearance as the structure of the SET8-H3 nucleosome complex, I chose to prepare the nucleosome mutant with the H3 nucleosome.

3.4.8. Binding analysis of the SET8 and nucleosome mutants

First, I tested the nucleosome binding ability of SET8 by comparing the wild-type SET8 or nucleosome to the mutants (Fig. 29). As expected, in the case of wild-type SET8 and H3 $nuc.^{APM}$, a large amount of free nucleosome band remained upon the titration of SET8 concentration (lanes 1-3 in Fig. 29A-C). This suggests that wild-type SET8 binding decreased significantly in H3 $nuc.^{APM}$ (lanes 4-6 in Fig. 29A-C). In the case of $SET8^{DM}$ and wild-type nucleosome, the discrete bands above the free nucleosome band were replaced by a smear, suggesting that the specific binding of SET8 to nucleosome is disrupted (lanes 7-9 in Fig. 29A-C). In the case of $SET8^{DM}$ and H3 $nuc.^{APM}$, a combined effect of the previous 2 cases was observed (lanes 9-12 in Fig. 29A-C).

To sum up, these results indicate that the SET8-acidic patch interaction is indeed necessary for the specific nucleosome binding of SET8. The results also suggest that the SET8-H4 tail interaction alone may not be sufficient for stable SET8-nucleosome binding, or SET8 cannot bind the H4 tail efficiently without the SET8-acidic patch interaction.

3.4.9. Methylation assay of the SET8 and nucleosome mutants

Next, I tested the mono-methylation activity in the mutants. A time course was taken for 0, 1, or 3 minutes of reaction time, followed by western blotting (Fig. 30). The H4K20me1 band and the H2B band (as a control) were detected. From the result, it can be observed that the H4K20me1 gradually increased over the time in the case of wild-types (lanes 1-3 in Fig. 30). However, in the case of SET8 DM or nuc^{APM}, the H4K20me1 band was not observed (lanes 4- in Fig. 30). This shows that the SET8-acidic patch interaction is important for the efficient H4K20 mono-methylation reaction.

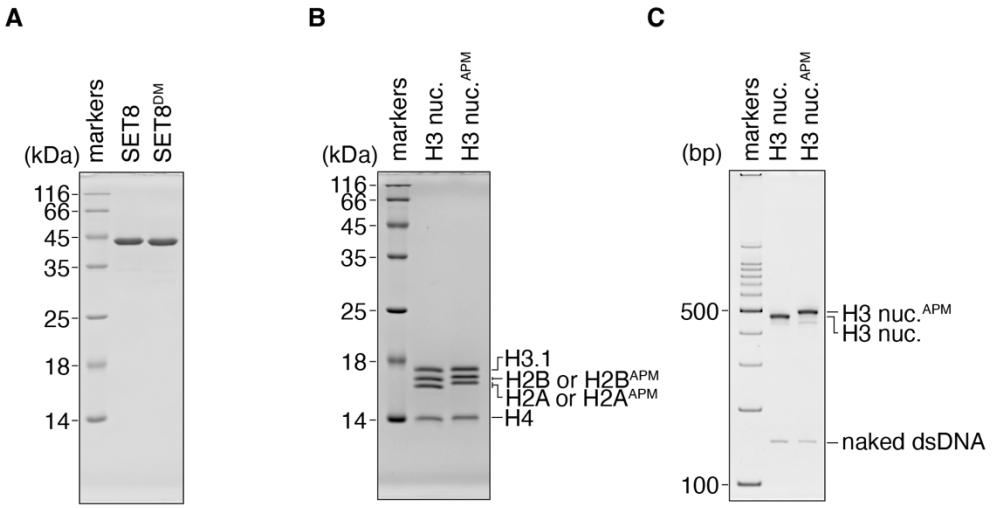


Fig. 28 Preparation of SET8 and H3 nucleosome acidic patch mutants

(A) The SET8 double mutant (R188A R192A) was prepared as a recombinant protein and analyzed by SDS-PAGE. **(B&C)** The H3 nucleosome acidic patch mutant was reconstituted *in vitro* and purified. The qualities of the nucleosomes are confirmed by SDS-PAGE (B) and non-denaturing PAGE (C). The SDS-PAGE was stained by CBB, visualizing protein bands; the non-denaturing PAGE was stained by EtBr, visualizing DNA bands. DM: double mutant, APM: acidic patch mutant.

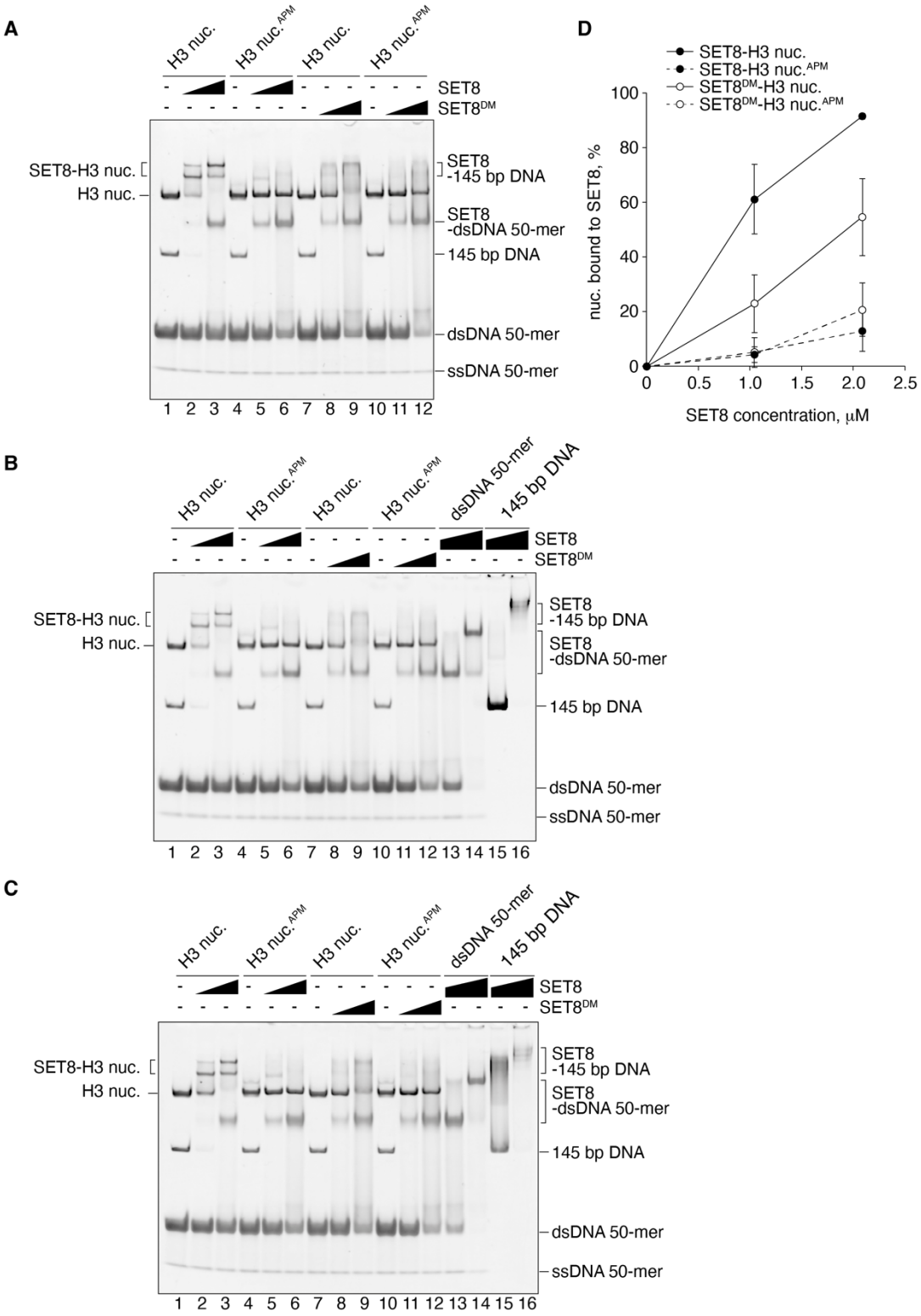

A
H3 nuc.
H3 nuc.APM
H3 nuc.
H3 nuc.APM
SET8
SET8DM
SET8-H3 nuc.
H3 nuc.
SET8
-145 bp DNA
SET8
-dsDNA 50-mer
145 bp DNA
dsDNA 50-mer
ssDNA 50-mer
1 2 3 4 5 6 7 8 9 10 11 12
D
SET8-H3 nuc.
SET8-H3 nuc.APM
SET8DM-H3 nuc.
SET8DM-H3 nuc.APM
nuc. bound to SET8, %
100
80
60
40
20
0
0 0.5 1.0 1.5 2.0 2.5
SET8 concentration, μM
B
H3 nuc.
H3 nuc.APM
H3 nuc.
H3 nuc.APM
dsDNA 50-mer
145 bp DNA
SET8
SET8DM
SET8-H3 nuc.
H3 nuc.
SET8
-145 bp DNA
SET8
-dsDNA 50-mer
145 bp DNA
dsDNA 50-mer
ssDNA 50-mer
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
C
H3 nuc.
H3 nuc.APM
H3 nuc.
H3 nuc.APM
dsDNA 50-mer
145 bp DNA
SET8
SET8DM
SET8-H3 nuc.
H3 nuc.
SET8
-145 bp DNA
SET8
-dsDNA 50-mer
145 bp DNA
dsDNA 50-mer
ssDNA 50-mer
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Fig. 29 Binding assay of SET8 and the H3 nucleosome (wild-types and mutants)

(A-C) Binding assays by non-denaturing PAGE. A short DNA fragment (50-mer) was added to prevent non-specific binding of SET8. The DNA-only lanes in (A) were cropped. **(D)** Quantification of the results from (A)-(C). The quantification was done by measuring the intensity of the free-nucleosome band. The percentage of SET8 bound to nucleosome was calculated by the amount of free-nucleosome that decreased upon SET8 addition.

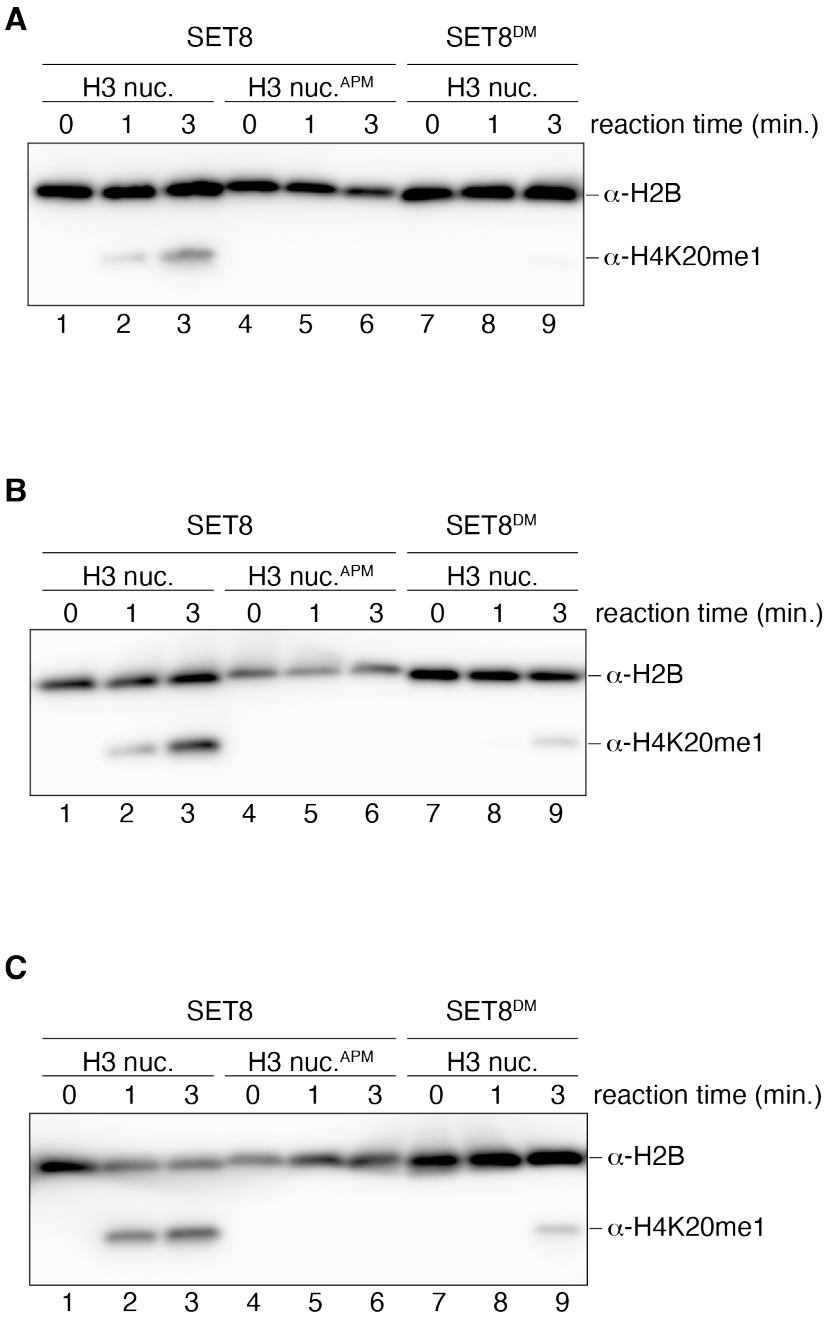


Fig. 30 Methylation assay of SET8 and the H3 nucleosome (wild-types and mutants)

The assay was repeated 3 times in (A), (B), and (C).

3.5. Discussion

3.5.1. Comparison of H4 tail conformations in free and SET8-bound nucleosomes

The objective of this study is to investigate why CENP-A nucleosomes are mono-methylated more efficiently than H3 nucleosome from a structural point of view. This could reveal the connection between the structural characteristics of the CENP-A nucleosome and its biological role in centromeric chromatin. However, the overall structure of the SET8-H3 nucleosome complex and the SET8-CENP-A nucleosome complex is similar. This alone cannot explain the difference in methylation efficiency between H3 nucleosomes and CENP-A nucleosomes. Therefore, I decided to compare the histone H4 tail conformation in free and SET8-bound nucleosomes.

First, I focused on the H3 nucleosome. In the crystal structure of the free H3 nucleosome, where SET8 is not bound, the observed H4 tail is pointed upwards (Fig. 31A). Interestingly, I found that the H4 tail of the SET8-H3 nucleosome, which was elucidated in this study, is pointed in a downward direction (Fig. 31A). From this, I presume that the upward conformation of the H4 tail is likely difficult for SET8 to grasp, and SET8 might grasp it after bending it downwards.

Having determined that the upward conformation of the H4 tail is hard to grasp, I next planned to investigate the H4 tail of the free CENP-A nucleosome. Unfortunately, in the previously reported crystal structure of the free CENP-A nucleosome, the H4 tail was not visible (Fig. 31B). Nevertheless, in our group's previous research, a chimera H3 was created, where a part of histone H3 was replaced with the characteristic domain of CENP-A (CATD) (Fig. 4A). In the crystal structure of the nucleosome containing this chimera H3, which was named $H3^{CATD}$, the H4 tail can be observed (Fig. 31C). In the $H3^{CATD}$ nucleosome, the H4 tail is bended downwards. This is because two critical amino acid

residues in H3 that interacts with the H4 tail, namely H3 Q76 and D77, are not present in CENP-A. This causes the H4 tail in the $H3^{CATD}$ nucleosome to be able to form a different conformation from the H3 nucleosome.

Upon comparing the direction of the H4 tail in the $H3^{CATD}$ nucleosome with that in the SET8-CENP-A nucleosome obtained in this study, I found that they are both oriented in the downward conformation. From this, it can be assumed that the H4 tail in the free CENP-A nucleosome is naturally facing downwards, which makes it more prone to methylation by SET8, compared to the H3 nucleosome.

3.5.2. Comparison of nucleosome binding mechanisms of H4K20 methyltransferases

SET8 is responsible for the mono-methylation of H4K20, while SUV420H1/2 are responsible for the di- and tri-methylation of H4K20me. Recently, the cryo-EM structures of SUV420H1-H2A /H2A.Z nucleosome complexes have been reported (88). In the structures, the catalytic domain, which consists of an N-domain, SET domain, and post-SET domain, can be seen bound to the nucleosome H4 tail. Interestingly, a C terminal sequence extending beyond the catalytic domain of SUV420H1 is seen bound to the nucleosome acidic patch.

In this study, I have shown that SET8 binds the nucleosome acidic patch. Therefore, it may be possible that the interaction with the acidic patch of the nucleosome is a conserved mechanism among H4K20 methyltransferases. This may have several reasons. First, unlike the tail of histone H3, which protrudes between the 2 strands of DNA, the tail of histone H4 extends from the nucleosome disc (Fig. 32A). This makes binding of H4K20 methyltransferases to the acidic patch of the nucleosome a seemingly reasonable

choice for interacting with the H4 tail. Second, the H4 tail in a nucleosome can interact with the acidic patch of another nucleosome, promoting the stacking between nucleosomes. By interacting with the acidic patch, the H4K20 methyltransferases can mask the acidic patch to free the H4 tail of adjacent nucleosomes, further promoting H4K20me of adjacent nucleosomes (Fig. 32B).

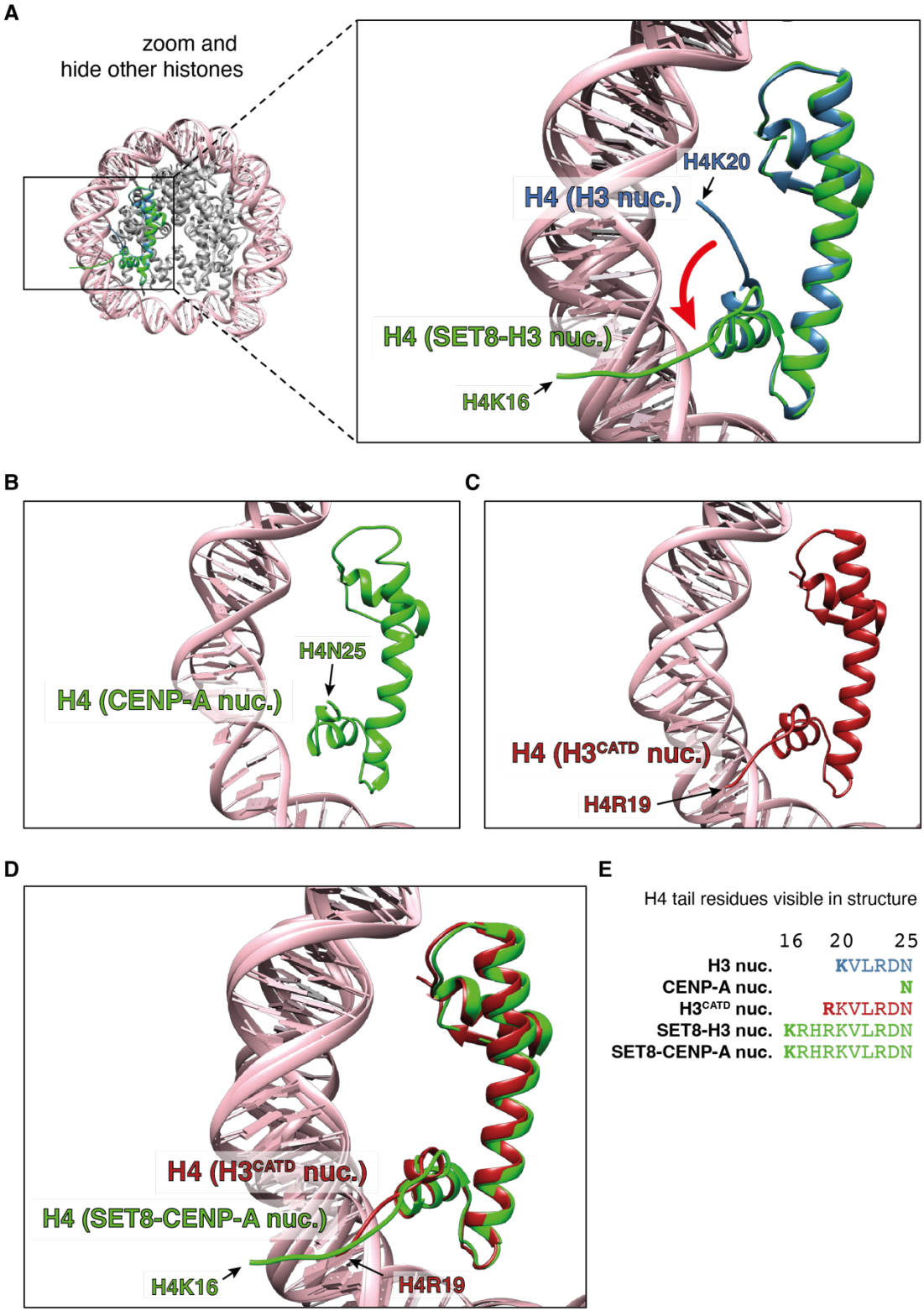

A
zoom and
hide other histones
H4K20
H4 (H3 nuc.)
H4 (SET8-H3 nuc.)
H4K16
B
H4N25
H4 (CENP-A nuc.)
C
H4 (H3CATD nuc.)
H4R19
D
H4 (H3CATD nuc.)
H4 (SET8-CENP-A nuc.)
H4K16
H4R19
E
H4 tail residues visible in structure
16 20 25
H3 nuc. KVLRDN
CENP-A nuc. N
H3CATD nuc. RKVLRDN
SET8-H3 nuc. KRHRKVLRDN
SET8-CENP-A nuc. KRHRKVLRDN

Fig. 31 Comparison of H4 tail conformations in free and SET8-bound nucleosomes

(A) Comparison of the H4 tail conformation in the free H3 nucleosome (PDB: 3LZ0) (blue) and the SET8-H3 nucleosome complex (PDB:7D1Z) (green). In the zoom up picture around the H4 tail, other histones are hidden for clarity. The red arrow indicates the change in the H4 tail conformation in free and SET8-bound nucleosomes. **(B)** Zoom up of H4 in the free CENP-A nucleosome (3AN2). Histones other than 1 H4 are hidden for clarity. The H4 tail is not visible, and the tip of histone H4 in the model is indicated by a red arrow. **(C)** Zoom up of H4 in the free $H3^{CATD}$ nucleosome (5Z23), in which the characteristic domain of CENP-A (CATD) is integrated into H3. Histones other than 1 H4 are hidden for clarity. The H4 tail is visible in this H3 chimera nucleosome. **(D)** Comparison of the H4 tail conformation in the free $H3^{CATD}$ nucleosome (PDB: 5Z23) (crimson) and the SET8-CENP-A nucleosome complex (PDB:7D20) (green). Histones other than 1 H4 in each nucleosome are hidden for clarity. **(E)** H4 tail residues visible in each structure are shown.

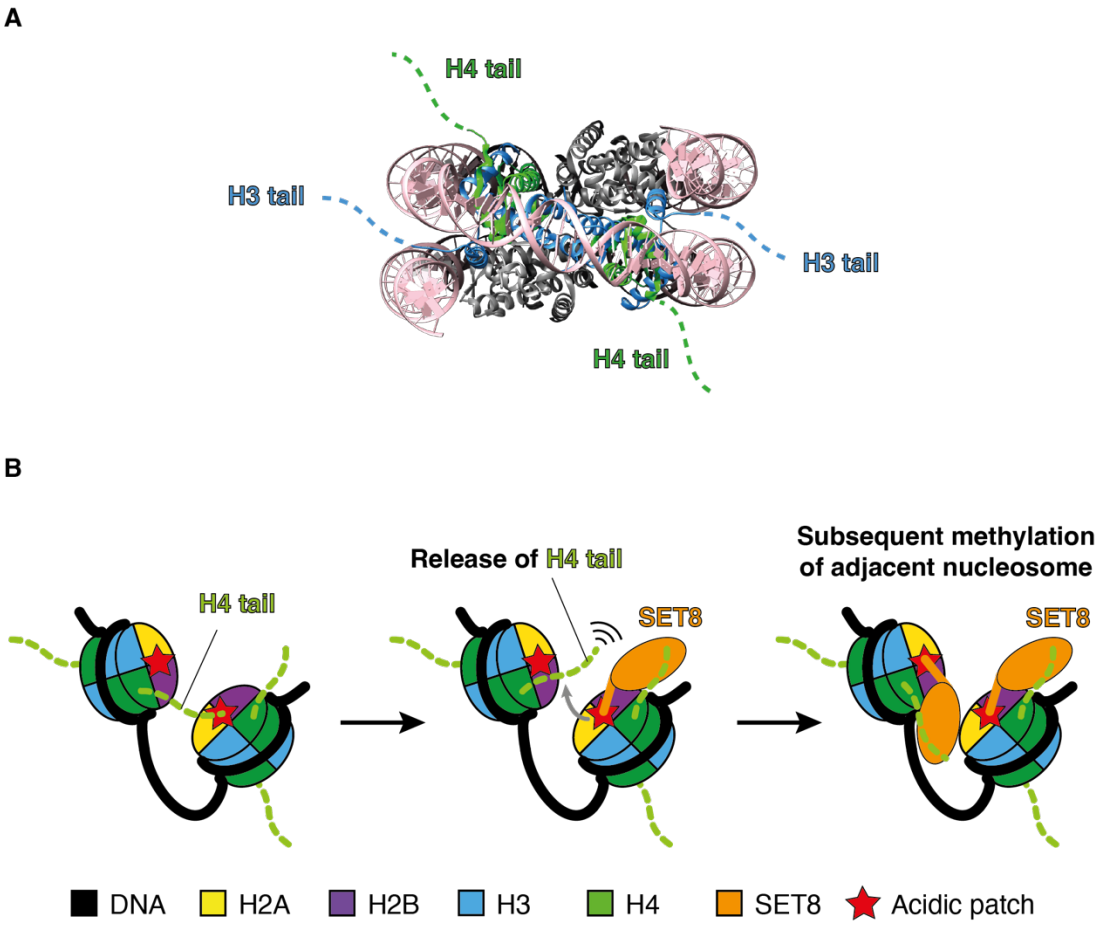


Fig. 32 Nucleosome H4 tail and the acidic patch recognition by SET8

(A) Crystal structure of the nucleosome (PDB: 3LZO), highlighting the positions where the H3 and H4 tails stem from. H2A and H2B are colored in dark gray for simplicity. **(B)** Binding of SET8 to one nucleosome may promote the release of the H4 tail from the other nucleosome, leading to the subsequent methylation of the released H4 tail.

Chapter 4: General Discussion

In this study, I focused on elucidating the role of the CENP-A nucleosome in the formation of a functional centromere. In a functional centromere, centromere proteins are required to identify the CENP-A nucleosomes that are buried among H3 nucleosomes. Moreover, it has been reported that H4K20me1 of the CENP-A nucleosome is necessary for the formation of a functional centromere, and that the H4K20 of CENP-A nucleosomes can be mono-methylated more efficiently than H3 nucleosomes. Consequently, I decided to investigate on the mechanism underlying these facts by cryo-EM single particle analysis.

First, I decided to analyze how the flexible DNA end of the CENP-A nucleosome may influence the structure of the centromeric chromatin. To this end, I performed cryo-EM analysis on tri-nucleosomes that contained CENP-A nucleosome. As a result, I found out that the CENP-A nucleosome may alter chromatin structure in a way such CENP-A nucleosomes may be exposed from surrounding H3 nucleosomes. This can in turn contribute to the recognition of CENP-A nucleosomes by other centromeric chromatin proteins, such as CENP-C and CENP-N, and therefore the formation of kinetochore (Fig. 33A).

Second, I decided to elucidate how the H4K20me1 is introduced to CENP-A nucleosome, to understand why the CENP-A nucleosome is mono-methylated more efficiently than the H3 nucleosome. In order to do so, I performed cryo-EM analysis on the SET8-CENP-A nucleosome and the SET8-H3 nucleosome. Resultingly, I found that SET8 can bind to the acidic patch of the nucleosome. Furthermore, the H4 tail in the SET8-CENP-A nucleosome is oriented in a similar conformation to that of the free

CENP-A nucleosome, but in a conformation different to that of the free H3 nucleosome. This suggests that the different H4 tail orientations in free CENP-A nucleosomes and H3 nucleosomes may affect the mono-methylation efficiency of H4K20.

By combining the results from the first and second part of this study, it can be further inferred that the characteristic centromeric chromatin structure may also contribute to the mono-methylation of H4K20 in CENP-A nucleosomes, which is necessary for the formation of the kinetochore (Fig. 33B). Importantly, the amount of SET8 fluctuates throughout the cell cycle, and peaks during the G2/M phase (Fig. 18C) (70–72). The fact that CENP-A is exposed from surrounding H3 nucleosomes may also promote the binding of SET8 to the acidic patch and H4 tails of CENP-A nucleosomes. The acidic patch of H3 nucleosomes, on the other hand, are seemingly inaccessible to SET8 in the poly-nucleosome model. This may serve as the first level of H4K20me1 regulation. The different H4 tail conformations may then serve as the second level of H4K20me1 regulation. Therefore, both the increased accessibility to the CENP-A nucleosome over the H3 nucleosomes and the H4 tail conformation in the CENP-A nucleosome contribute to the methylation efficiency of SET8 (Fig. 33B).

Although this study is focused on the regional centromeres of human, the less twisted centromeric chromatin higher-order structure may also be conserved in organisms that contain point centromeres, such as budding yeast. Recent studies reported the cryo-EM structure of the yeast centromeric nucleosomes, and showed that they have flexible DNA ends, similar to the human CENP-A nucleosome (89,90). Therefore, the yeast centromeric DNA may also form the outward-path conformation, possibly promoting the binding of CBF1 and CBF3 to CDEI and CDEIII, respectively (90). This could provide a connection between the epigenetic features (less twisted structure) of the point centromere to its

genetic features (DNA sequence specificity).

Some questions remain. For example, the reason why the H4K20me1 mark is required in centromeric chromatin remains unclear (Fig. 33). In the future, more investigation can be conducted to identify possible reader proteins that may read the centromeric H4K20me1 mark and promote CCAN assembly. Subsequently, the structural analysis of the identified protein and the CENP-A nucleosome carrying the H4K20me1 mark can be performed.

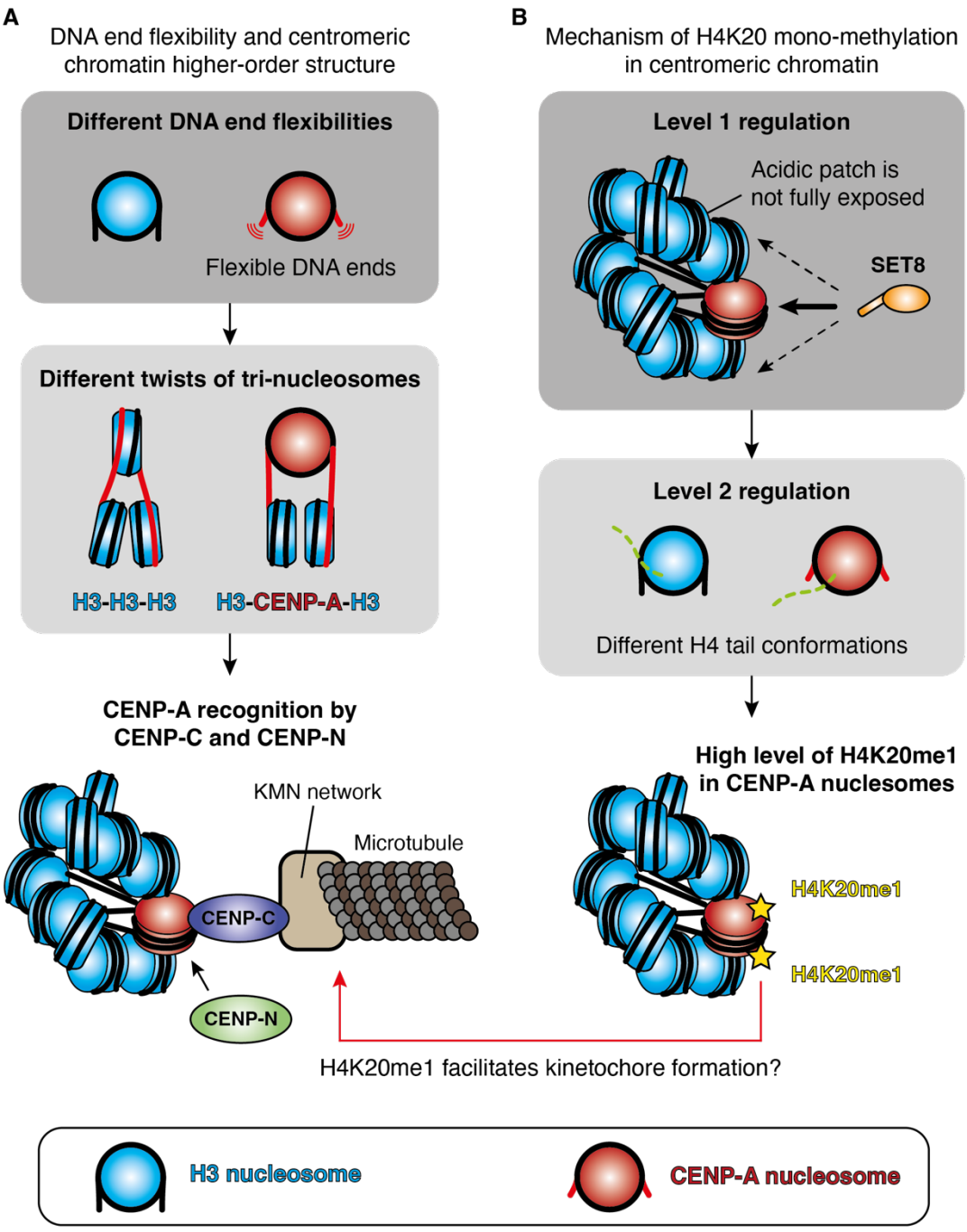

A
DNA end flexibility and centromeric chromatin higher-order structure
Different DNA end flexibilities
Flexible DNA ends
Different twists of tri-nucleosomes
H3-H3-H3
H3-CENP-A-H3
CENP-A recognition by CENP-C and CENP-N
KMN network
Microtubule
CENP-C
CENP-N
B
Mechanism of H4K20 mono-methylation in centromeric chromatin
Level 1 regulation
Acidic patch is not fully exposed
SET8
Level 2 regulation
Different H4 tail conformations
High level of H4K20me1 in CENP-A nuclesomes
H4K20me1
H4K20me1
H4K20me1 facilitates kinetochore formation?
H3 nucleosome
CENP-A nucleosome

Fig. 33 Summary of how structural features of the CENP-A nucleosome promotes the formation of a functional centromere

(A) The different DNA end flexibilities of the H3 nucleosome and CENP-A nucleosome contribute to CENP-A recognition by CENP-C and CENP-N, leading to the formation of the kinetochore. **(B)** The H4K20me1 mark in the centromeric chromatin is regulated at 2 separate levels, ensuring high level of H4K20me1 mark in CENP-A nucleosomes. However, the reason why the H4K20me1 mark is required for kinetochore formation remains unclear.

References

1. Olins AL, Olins DE. Spheroid Chromatin Units (ν Bodies). Science. 1974 Jan 25;183(4122):330–2.

2. Kornberg RD. Chromatin Structure: A Repeating Unit of Histones and DNA: Chromatin structure is based on a repeating unit of eight histone molecules and about 200 DNA base pairs. Science. 1974 May 24;184(4139):868–71.

3. Kornberg RD. Structure of Chromatin. Annu Rev Biochem. 1977 Jun;46(1):931–54.

4. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell. 1975 Apr 1;4(4):281–300.

5. Wolffe A. Chromatin: Structure and Function [Internet]. 3rd ed. San Diego: Academic Press; 1998. 447 p. Available from: https://www.elsevier.com/books/chromatin/wolffe/978-0-08-092660-5

6. Woodcock CLF, Safer JP, Stanchfield JE. Structural repeating units in chromatin. Exp Cell Res. 1976 Jan;97(1):101–10.

7. Heitz E. Heterochromatin der Moose. Jahrb Wiss Bot. 1928;69.

8. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat Rev Genet. 2016 Aug;17(8):487–500.

9. Fukagawa T, Earnshaw WC. The Centromere: Chromatin Foundation for the Kinetochore Machinery. Dev Cell. 2014;30(5):496–508.

10. McKinley KL, Cheeseman IM. The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol. 2016;17(1):16–29.

11. Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EN. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci. 1991 Nov 15;88(22):10148–52.

12. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000;403(6765):41–5.

13. Kouzarides T. Chromatin Modifications and Their Function. Cell. 2007;128(4):693–705.

14. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21(3):381–95.

15. Millán-Zambrano G, Burton A, Bannister AJ, Schneider R. Histone post-translational modifications — cause and consequence of genome function. Nat Rev Genet. 2022 Sep;23(9):563–80.

16. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. 1997;389(6648):251–60.

17. Vasudevan D, Chua EYD, Davey CA. Crystal Structures of Nucleosome Core Particles Containing the '601' Strong Positioning Sequence. J Mol Biol. 2010 Oct;403(1):1–10.

18. McGinty RK, Tan S. Nucleosome Structure and Function. Chem Rev. 2015 Mar 25;115(6):2255–73.

19. Koyama M, Kurumizaka H. Structural diversity of the nucleosome. J Biochem (Tokyo). 2018 Feb 1;163(2):85–95.

20. McGinty RK, Tan S. Principles of nucleosome recognition by chromatin factors and enzymes. Curr Opin Struct Biol. 2021 Dec;71:16–26.

21. The Nucleosomal Surface as a Docking Station for Kaposi's Sarcoma Herpesvirus LANA [Internet]. [cited 2023 Jun 21]. Available from: https://www.science.org/doi/epdf/10.1126/science.1120541

22. Kujirai T, Zierhut C, Takizawa Y, Kim R, Negishi L, Uruma N, et al. Structural basis for the inhibition of cGAS by nucleosomes. Science. 2020 Oct 23;370(6515):455–8.

23. Pathare GR, Decout A, Glück S, Cavadini S, Makasheva K, Hovius R, et al. Structural mechanism of cGAS inhibition by the nucleosome. Nature. 2020 Nov 26;587(7835):668–72.

24. Boyer JA, Spangler CJ, Strauss JD, Cesmat AP, Liu P, McGinty RK, et al.

Structural basis of nucleosome-dependent cGAS inhibition. Science. 2020 Oct 23;370(6515):450–4.

25. Zhao B, Xu P, Rowlett CM, Jing T, Shinde O, Lei Y, et al. The molecular basis of tight nuclear tethering and inactivation of cGAS. Nature. 2020 Nov 26;587(7835):673–7.

26. Michalski S, De Oliveira Mann CC, Stafford CA, Witte G, Bartho J, Lammens K, et al. Structural basis for sequestration and autoinhibition of cGAS by chromatin. Nature. 2020 Nov 26;587(7835):678–82.

27. Cao D, Han X, Fan X, Xu RM, Zhang X. Structural basis for nucleosome-mediated inhibition of cGAS activity. Cell Res. 2020 Dec;30(12):1088–97.

28. Anderson CJ, Baird MR, Hsu A, Barbour EH, Koyama Y, Borgnia MJ, et al. Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L Methyltransferase. Cell Rep. 2019;26(7):1681-1690.e5.

29. Jang S, Kang C, Yang HS, Jung T, Hebert H, Chung KY, et al. Structural basis of recognition and destabilization of the histone H2B ubiquitinated nucleosome by the DOT1L histone H3 Lys79 methyltransferase. Genes Dev. 2019;

30. Valencia-Sánchez MI, De Ioannes P, Wang M, Vasilyev N, Chen R, Nudler E, et al. Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination. Mol Cell. 2019 Jun 6;74(5):1010-1019.e6.

31. Worden EJ, Hoffmann NA, Hicks CW, Wolberger C. Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. Cell. 2019;176(6):1490-1501.e12.

32. Martire S, Banaszynski LA. The roles of histone variants in fine-tuning chromatin organization and function. Nat Rev Mol Cell Biol [Internet]. 2020; Available from: http://dx.doi.org/10.1038/s41580-020-0262-8 http://www.nature.com/articles/s41580-020-0262-8

33. Henikoff S, Smith MM. Histone variants and epigenetics. Cold Spring Harb Perspect Biol. 2015;7(1):1–26.

34. Dominski Z, Marzluff WF. Formation of the 3′ end of histone mRNA. Gene. 1999

Oct 18;239(1):1–14.

35. Talbert PB, Henikoff S. Histone variants on the move: Substrates for chromatin dynamics. Nat Rev Mol Cell Biol. 2017;18(2):115–26.

36. Janssen A, Colmenares SU, Karpen GH. Heterochromatin: Guardian of the Genome. Annu Rev Cell Dev Biol. 2018 Oct 6;34(1):265–88.

37. Maison C, Almouzni G. HP1 and the dynamics of heterochromatin maintenance. Nat Rev Mol Cell Biol. 2004;5(4):296–305.

38. Grewal SIS, Jia S. Heterochromatin revisited. Nat Rev Genet. 2007;8(1):35–46.

39. Simon JA, Kingston RE. Mechanisms of Polycomb gene silencing: Knowns and unknowns. Vol. 10, Nature Reviews Molecular Cell Biology. 2009. 697–708 p.

40. Pluta AF, Mackay AM, Ainsztein AM, Goldberg IG, Earnshaw WC. The Centromere: Hub of Chromosomal Activities. Science. 1995 Dec 8;270(5242):1591–4.

41. Biggins S. The Composition, Functions, and Regulation of the Budding Yeast Kinetochore. Genetics. 2013 Aug;194(4):817–46.

42. Westermann S, Drubin DG, Barnes G. Structures and Functions of Yeast Kinetochore Complexes. Annu Rev Biochem. 2007 Jun 7;76(1):563–91.

43. Manuelidis L. Chromosomal localization of complex and simple repeated human DNAs. Chromosoma. 1978;66(1):23–32.

44. Manuelidis L. Complex and simple sequences in human repeated DNAs. Chromosoma. 1978;66(1):1–21.

45. Earnshaw WC, Rothfield N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma. 1985;91(3–4):313–21.

46. Palmer D, O'Day 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. 1987;104(4):805–15.

47. Earnshaw WC, Allshire RC, Black BE, Bloom K, Brinkley BR, Brown W, et al. Esperanto for histones: CENP-A, not CenH3, is the centromeric histone H3 variant. Chromosome Res. 2013 Apr;21(2):101–6.

48. Ribeiro SA, Vagnarelli P, Dong Y, Hori T, McEwen BF, Fukagawa T, et al. A super-resolution map of the vertebrate kinetochore. Proc Natl Acad Sci. 2010 Jun 8;107(23):10484–9.

49. Müller S, Almouzni G. Chromatin dynamics during the cell cycle at centromeres. Nat Rev Genet. 2017;18(3):192–208.

50. Musacchio A, Desai A. A Molecular View of Kinetochore Assembly and Function. Biology. 2017;6(4):5.

51. Nishimura K, Komiya M, Hori T, Itoh T, Fukagawa T. 3D genomic architecture reveals that neocentromeres associate with heterochromatin regions. Journal of Cell Biology. 2019 Jan 7;218(1):134–49.

52. Altemose N, Maslan A, Smith OK, Sundararajan K, Brown RR, Mishra R, et al. DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide. Nat Methods. 2022 Jun;19(6):711–23.

53. Bodor DL, Mata JF, Sergeev M, David AF, Salimian KJ, Panchenko T, et al. The quantitative architecture of centromeric chromatin. eLife. 2014;3(3):1–26.

54. Tachiwana H, Kagawa W, Shiga T, Osakabe A, Miya Y, Saito K, et al. Crystal structure of the human centromeric nucleosome containing CENP-A. Nature. 2011;476(7359):232–5.

55. Hori T, Shang WH, Toyoda A, Misu S, Monma N, Ikeo K, et al. Histone H4 Lys 20 Monomethylation of the CENP-A Nucleosome Is Essential for Kinetochore Assembly. Dev Cell. 2014;29(6):740–9.

56. Tachiwana H, Kagawa W, Kurumizaka H. Comparison between the CENP-A and histone H3 structures in nucleosomes. Nucleus. 2012 Jan;3(1):6–11.

57. Black BE, Jansen LET, Maddox PS, Foltz DR, Desai AB, Shah JV, et al. Centromere Identity Maintained by Nucleosomes Assembled with Histone H3 Containing the CENP-A Targeting Domain. Mol Cell. 2007;25(2):309–22.

58. Black BE, Foltz DR, Chakravarthy S, Luger K, Woods VL, Cleveland DW. Structural determinants for generating centromeric chromatin. Nature. 2004 Jul;430(6999):578–82.

59. Foltz DR, Jansen LET, Bailey AO, Yates JR, Bassett EA, Wood S, et al. Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. Cell. 2009;137(3):472–84.

60. Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, Nakamura Y, et al. HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. Cell. 2009;137(3):485–97.

61. Sanchez-Pulido L, Pidoux AL, Ponting CP, Allshire RC. Common Ancestry of the CENP-A Chaperones Scm3 and HJURP. Cell. 2009 Jun;137(7):1173–4.

62. McAinsh AD, Meraldi P. The CCAN complex: Linking centromere specification to control of kinetochore–microtubule dynamics. Semin Cell Dev Biol. 2011 Dec 1;22(9):946–52.

63. Kujirai T, Arimura Y, Fujita R, Horikoshi N, Machida S, Kurumizaka H. Methods for Preparing Nucleosomes Containing Histone Variants. In: Methods in Molecular Biology [Internet]. 2018. p. 3–20. Available from: http://link.springer.com/10.1007/978-1-4939-8663-7_1

64. Machida S, Takizawa Y, Ishimaru M, Sugita Y, Sekine S, Nakayama J ichi, et al. Structural Basis of Heterochromatin Formation by Human HP1. Mol Cell. 2018;69(3):385-397.e8.

65. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods. 2017;14(4):331–2.

66. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol. 2015;192(2):216–21.

67. Scheres SHW. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. In: Methods in Enzymology [Internet]. 1st ed. Elsevier Inc.; 2016. p. 125–57. Available from: http://dx.doi.org/10.1016/bs.mie.2016.04.012 https://linkinghub.elsevier.com/retrieve/pii/S0076687916300301

68. Kleywegt GJ, Harris MR, Zou J yu, Taylor TC, Wählby A, Jones TA. The Uppsala Electron-Density Server. Acta Crystallogr D Biol Crystallogr. 2004;60(12):2240–9.

69. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera?A visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605–12.

70. Beck DB, Oda H, Shen SS, Reinberg D. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. Genes Dev. 2012;26(4):325–37.

71. Wu S, Rice JC. A new regulator of the cell cycle. Cell Cycle. 2011;10(1):68–72.

72. Jørgensen S, Schotta G, Sørensen CS. Histone H4 Lysine 20 methylation: Key player in epigenetic regulation of genomic integrity. Nucleic Acids Res. 2013;41(5):2797–806.

73. Fang J, Feng Q, Ketel CS, Wang H, Cao R, Xia L, et al. Purification and Functional Characterization of SET8, a Nucleosomal Histone H4-Lysine 20-Specific Methyltransferase. Curr Biol. 2002;12(13):1086–99.

74. Nishioka K, Rice JC, Sarma K, Erdjument-Bromage H, Werner J, Wang Y, et al. PR-Set7 Is a Nucleosome-Specific Methyltransferase that Modifies Lysine 20 of Histone H4 and Is Associated with Silent Chromatin. Mol Cell. 2002;9(6):1201–13.

75. Schotta G, Sengupta R, Kubicek S, Malin S, Kauer M, Callen E, et al. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev. 2008;22(15):2048–61.

76. Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. 2004 Jun 1;18(11):1251–62.

77. Oda H, Okamoto I, Murphy N, Chu J, Price SM, Shen MM, et al. Monomethylation of Histone H4-Lysine 20 Is Involved in Chromosome Structure and Stability and Is Essential for Mouse Development. Mol Cell Biol. 2009;29(8):2278–95.

78. Shi X, Kachirskaia I, Yamaguchi H, West LE, Wen H, Wang EW, et al. Modulation of p53 Function by SET8-Mediated Methylation at Lysine 382. Mol Cell. 2007

Aug;27(4):636–46.

79. Couture JF, Collazo E, Brunzelle JS, Trievel RC. Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. Genes Dev. 2005;19(12):1455–65.

80. Xiao B. Specificity and mechanism of the histone methyltransferase Pr-Set7. Genes Dev. 2005;19(12):1444–54.

81. Arimura Y, Tachiwana H, Takagi H, Hori T, Kimura H, Fukagawa T, et al. The CENP-A centromere targeting domain facilitates H4K20 monomethylation in the nucleosome by structural polymorphism. Nat Commun. 2019;10(1):3–8.

82. Tachiwana H, Kagawa W, Osakabe A, Kawaguchi K, Shiga T, Hayashi-Takanaka Y, et al. Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. Proc Natl Acad Sci. 2010;107(23):10454–9.

83. Mastronarde DN. Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol. 2005;152(1):36–51.

84. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJH, Lindahl E, et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife [Internet]. 2018;7. Available from: https://elifesciences.org/articles/42166

85. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 2018;27(1):14–25.

86. Emsley P, Cowtan K. Coot : model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60(12):2126–32.

87. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX : a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66(2):213–21.

88. Abini-Agbomson S, Gretarsson K, Shih RM, Hsieh L, Lou T, Ioannes PD, et al. Catalytic and non-catalytic mechanisms of histone H4 lysine 20 methyltransferase SUV420H1 [Internet]. Biochemistry; 2023 Mar [cited 2023 May 15]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.03.17.533220

89. Migl D, Kschonsak M, Arthur CP, Khin Y, Harrison SC, Ciferri C, et al. Cryoelectron Microscopy Structure of a Yeast Centromeric Nucleosome at 2.7 Å Resolution. Structure. 2020 Mar 3;28(3):363-370.e3.

90. Guan R, Lian T, Zhou BR, He E, Wu C, Singleton M, et al. Structural and dynamic mechanisms of CBF3-guided centromeric nucleosome formation. Nat Commun. 2021 Mar 19;12(1):1763.

Original paper

Takizawa Y, Ho CH, Tachiwana H, Matsunami H, Kobayashi W, Suzuki M, et al. Cryo-EM Structures of Centromeric Tri-nucleosomes Containing a Central CENP-A Nucleosome. Structure. 2020;28(1):44-53.e4.

Ho CH, Takizawa Y, Kobayashi W, Arimura Y, Kimura H, Kurumizaka H. Structural basis of nucleosomal histone H4 lysine 20 methylation by SET8 methyltransferase. Life Sci Alliance. 2021 Apr;4(4):e202000919.

Acknowledgements

I would like to express my deepest appreciation to my supervisor, Professor Hitoshi Kurumizaka. Prof. Kurumizaka has always been extraordinarily tolerant and supportive and given me insightful comments and suggestions when I have questions about my research. Without his guidance and continuous support, these works would not have been completed. I would also like to express my deepest gratitude to Associate Professor Yoshimasa Takizawa. Without his support, cryo-EM analysis would have been impossible. Additionally, this endeavor would not have been possible without the financial support provided by Japan Society for the Promotion of Science (JSPS) and WINGS-LST.

I would also like to acknowledge Postdoctoral Researcher Wataru Kobayashi for supporting me on experimental operations. I am also grateful for Professor Matthias Wolf for the help on the cryo-EM analysis of tri-nucleosomes. I would like to extend my sincere thanks to Researcher Hiroaki Tachiwana at the Cancer Institute, who has given me insightful comments on the work of tri-nucleosomes, and Ms. Midori Suzuki, who has supported me on the preparation of tri-nucleosomes. Thanks should also go to Postdoctoral Associate Yasuhiro Arimura and Mr. Hiroki Takagi, who were pioneers in the investigation of SET8 and the CENP-A nucleosome. I am also thankful for the help of all Kurumizaka Laboratory members, especially Ms. Yukari Iikura, Ms. Yasuko Takeda, and Dr. Mariko Dacher for their countless assistance.

Lastly, I would like to mention my family for the support in my life and studying abroad in Japan.

Cheng-Han Ho