Biochemical and biophysical investigations of the dependences of chromatin/nucleosome structure on linker histone and histone H4K20-trimethylation

Wang, Sai

2018

Wang, S. (2018). Biochemical and biophysical investigations of the dependences of chromatin/nucleosome structure on linker histone and histone H4K20-trimethylation. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73338>

<https://doi.org/10.32657/10356/73338>

Biochemical and Biophysical Investigations of the Dependences of
Chromatin/Nucleosome Structure on Linker Histone and Histone H4K20-
trimethylation

SAI WANG

SCHOOL OF BIOLOGICAL SCIENCES

2018

**Biochemical and Biophysical Investigations of the Dependences of
Chromatin/Nucleosome Structure on Linker Histone and Histone H4K20-
trimethylation**

SAI WANG

SCHOOL OF BIOLOGICAL SCIENCES

**A thesis submitted to the Nanyang Technological University in partial fulfillment of the
requirement for the degree of Doctor of Philosophy**

2018

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my supervisor Prof. Lars Nordenskiöld for offering me this exciting project, for his continuous support, patience, and immense knowledge, and for the warm and encouraging smile every time we met.

My heartfelt appreciation must go to Dr. Nikolay Korolev for his profound knowledge and invaluable suggestions in solving experimental problems, and tremendous efforts in revising this thesis, as well as for his affable attitude, support and encouragement every time I seek help from him.

My debt of gratitude also goes to lab members (both previous and present) Dr. Abdollah Allahverdi for his patient guidance in experimental techniques; Dr. Terry Sun Tiedong for his help with computation clusters; I am also thankful to the other lab mates Luo Di, Chen Qinming, Fan Yanping, Nikolay Berezhnoy, Shi Xiangyan, Chinmayi, Aghil, Jaclyn, Vishal, Alexander, Surya, working with them is a very pleasant experience.

My sincere appreciation goes to Dr. Sara Sandin for guidance in cryo-Electron Microscopy; and the EM manager Dr. Andrew Wong for training on grid preparation and operation of EM, as well as help on data collection; I thank Dr. Eugene Chua for comments on the EM part of this thesis. Many thanks go to other EM group members and EM users Wee Mei Ling, Vinod, Oviya Inian, Tofayel, Feng Xiaoyu.

I thank Dr. Yang Renliang for the preparation of H4K20-trimethylation histone sample.

I thank Dr. Daniela Rhodes for kindly providing the protocol for H1 purification.

I thank Dr. Curt Davey for the generous gift of linker histone construct, Zenita Adhireksan's suggestions on H1 purification, and Dr. Bao Qiuye's suggestions on cloning of 601 array DNA constructs.

Last but never the least, I am grateful to my dear family members who always understand and support me. From them, I obtain endless power, faith and courage to go through those rough days of my Ph.D. study.

Table of contents

Acknowledgements	1
Table of contents	2
List of figures	6
List of tables.....	11
Abbreviations.....	12
Abstract	14
1 Introduction	17
1.1 Nucleosome	17
1.2 Histone octamer	18
1.3 Core histone modifications.....	20
1.3.1 H4K20 Methylation	21
1.3.2 Roles of H4K20 methylations in DNA replication	22
1.4 Linker histone	23
1.4.1 Linker histone variants	25
1.5 Chromatosome	29
1.5.1 H1 stoichiometry	31
1.5.2 H1 and other core histones	32
1.5.3 Linker histone interaction with nucleosome	32
1.5.4 Various models for linker histone binding with the nucleosome.....	33
1.5.5 Linker histone and chromatin structure	41
1.6 Chromatin higher-order structure	42

1.6.1	30 nm chromatin fiber, one-start or two-start?	43
1.6.2	Do 30 nm fibers really exist <i>in vivo</i> ?	44
1.7	Objectives	44
2	Materials and methods	46
2.1	Materials	46
2.1.1	Chemicals	46
2.1.2	Buffers	47
2.1.3	<i>Escherichia coli</i> competent cells	49
2.1.4	Culture Media	49
2.1.5	Enzymes and reaction buffers	49
2.1.6	Gel and running buffer	50
2.1.7	Gel ladders	50
2.1.8	Purification columns	50
2.2	Methods	52
2.2.1	Preparation of 601-DNA	52
2.2.2	Preparation of core histone H2A, H2B, H3 and H4	56
2.2.3	Histone octamer refolding and purification	58
2.2.4	Preparation of linker histone H1 ⁰	59
2.2.5	<i>In vitro</i> reconstitution of nucleosome arrays	61
2.2.6	<i>In vitro</i> reconstitution of nucleosomes	62
2.2.7	<i>In vitro</i> reconstitution of chromatosomes	62
2.2.8	Cryo-electron microscopy (Cryo-EM) single particle analysis	63
2.2.9	Precipitation assay (PA)	65

2.2.10	Small-angle X-ray Scattering (SAXS) and X-Ray diffraction	66
2.2.11	Analytical ultracentrifugation (AUC)	68
3	Results and Discussions	71
3.1	Large scale production and purification of 601 DNA	71
3.1.1	12-177-601 DNA preparation	71
3.1.2	601-177 monomer DNA preparation	72
3.2	Preparation of core histones	73
3.3	Histone octamer preparation	75
3.3.1	Purification of <i>Xenopus</i> wild-type histone octamer	75
3.3.2	Purification of <i>Xenopus</i> HO-H4K20Me3	77
3.3.3	Purification of <i>human</i> HO	78
3.4	Linker histone H1 ⁰ large scale expression and purification	79
3.5	Nucleosome/ chromatosome reconstitution	81
3.5.1	177 bp-nucleosome reconstitution	81
3.5.2	177 bp-chromatosome reconstitution	83
3.6	Nucleosome array reconstitution	84
3.6.1	12-177-601 nucleosome array reconstitution	84
3.7	Cryo-Electron Microscopy single particle analysis of 177 bp-chromatosome	88
3.8	Self-association of the 177 bp nucleosome and chromatosome	100
3.8.1	The polyelectrolyte property of 177 bp-nucleosome and chromatosome	100
3.9	X-Ray diffraction investigations of chromatosome phase behavior	105
3.9.1	177 bp chromatosome forms lamella-columnar phase	107
3.9.2	Binding of linker histone changed the nucleosome stacking phase	108

3.10	Solution SAXS studies of 177 bp nucleosome and chromatosome.....	110
3.11	Analytical Ultracentrifugation (AUC) investigation of effects of H4K20Me3 on chromatin folding	113
3.11.1	12-177-wtHO nucleosome array	114
3.11.2	12-177-HO_H4K20Me3 nucleosome array.....	119
3.12	Self-association of the 12-177 nucleosome array with unmodified HO or H4K20Me3 modification.....	127
4	Conclusion.....	132
5	Perspectives	136
	References	138
	Appendix.....	146

List of figures

Figure 1.1 A schematic overview of DNA packing during mitosis.	17
Figure 1.2 Schematic representations of the nucleosome core particle.....	18
Figure 1.3 The histones “handshake” motif	19
Figure 1.4 Post-translational modifications of core histones.....	20
Figure 1.5 Illustration of three methylation states of lysine	21
Figure 1.6 Structural composition of linker histone.....	23
Figure 1.7 3D structure of linker histone H5 globular domain	24
Figure 1.8 Model of nucleosome-directed folding of linker histone CTD	25
Figure 1.9 Evolutionary tree of linker histones	26
Figure 1.10 Multiple alignment of human linker histone subtypes.	27
Figure 1.11 Illustration of a chromatosome composition	30
Figure 1.12 Electron microscope images of chromatin with or without linker histone	30
Figure 1.13 Linker histone stoichiometry and NRL.....	31
Figure 1.14 Binding of linker histone with nucleosome has geometry preference.....	32
Figure 1.15 Two nucleosomal DNA binding sites of GH5	33
Figure 1.16 Structural model of GH5 binding with the nucleosome	34
Figure 1.17 Three nucleosomal DNA binding sites of GH5.....	35
Figure 1.18 First crystal structure of chromatosome at 3.5 Å resolution.....	35
Figure 1.19 Asymmetric structural model of the Drosophila GH1-nucleosome complex.....	37
Figure 1.20 <i>In vivo</i> mapping of the two nucleosome binding sites on mouse H1.0 globular domain.	37
Figure 1.21 Schematic representation of the 197 bp chromatosome structure.....	38

Figure 1.22 Cryo-EM structure of the 30 nm chromatin fiber.....	39
Figure 1.23 Different linker histone binding modes lead to different chromatin compaction	40
Figure 1.24 Schematic illustration of packing of chromatin into 30 nm fiber assisted by linker histones.	41
Figure 1.25 Schematic illustration of chromatin compaction into condensed fiber.	42
Figure 1.26 Different models of the 30-nm chromatin fiber.....	44
Figure 2.1 Scal restriction endonuclease recognition sequence	55
Figure 2.2 Schematic illustration of the dialysis set-up.....	63
Figure 2.3 Sample curve of the Boltzmann function	66
Figure 2.4 Schematic illustration of the principle of SAXS experiment.....	66
Figure 2.5 Schematic illustration of the AUC double-sector centerpiece	68
Figure 2.6 A collection of 38 scans with 10 mins interval between each scan.....	69
Figure 3.1 Agarose gel analysis of the EcoRV digestion of 12-177-601 plasmid	71
Figure 3.2 Agarose gel analysis of the pellets from PEG precipitation.....	72
Figure 3.3 10 % DNA-PAGE analysis of the purified 601-177 DNA.....	73
Figure 3.4 Size exclusion chromatography of histone H2A	74
Figure 3.5 18 % SDS-PAGE analysis of fractions from the Sephacryl S-200 size exclusion chromatography.....	74
Figure 3.6 Resource S cation exchange chromatography of histone H2A purification	75
Figure 3.7 18 % SDS-PAGE analysis of the H2A peak fraction (Figure 3.6) from the Resource S cation exchange chromatography.....	75
Figure 3.8 Size exclusion chromatography profile of the Xenopus HO	76
Figure 3.9 18 % SDS-PAGE analysis of the fractions of the size exclusion chromatography	76
Figure 3.10 Size exclusion chromatography profile of the Xenopus HO-H4K20Me3	77

Figure 3.11 18 % SDS-PAGE analysis of the fractions of the size exclusion chromatography	77
Figure 3.12 Size exclusion chromatography profile of the <i>human</i> HO	78
Figure 3.13 18 % SDS-PAGE analysis of the fractions of the size exclusion chromatography	78
Figure 3.14 SP Sepharose cation exchange chromatography profile of linker histone H1 ⁰	79
Figure 3.15 18 % SDS-PAGE analysis of the fractions from the SP Sepharose cation exchange chromatography of H1 ⁰	80
Figure 3.16 Resource S chromatography profile of linker histone H1 ⁰	80
Figure 3.17 18 % SDS-PAGE analysis of fractions from the Resource S cation exchange chromatography of H1 ⁰	81
Figure 3.18 EMSA analysis of the 177 bp-nucleosome samples on 6 % native PAGE	82
Figure 3.19 EMSA analysis of the 177 bp chromatosome samples on 5 % native PAGE	83
Figure 3.20 0.8 % agarose gel analysis of nucleosome arrays with different wtHO to 12-177-601 DNA molar ratios	85
Figure 3.21 Analysis of the Scal digestion product of the 12-177-wtHO nucleosome array samples on 6 % native PAGE	85
Figure 3.22 0.8 % agarose gel analysis of nucleosome arrays with different HO_H4K20Me3 to 12-177-601 DNA molar ratios	87
Figure 3.23 Analysis of the Scal digestion product of the 12-177-HO_H4K20Me3 nucleosome array samples on 6 % native PAGE	87
Figure 3.24 Cryo-EM micrograph of 177 bp-chromatosome	89
Figure 3.25 2D class averaging results	90
Figure 3.26 Selection of “H1-bound” 2D class averages	91
Figure 3.27 Preparation of the initial reference map for 3D classification	91
Figure 3.28 The three classes from the 3D classification	92

Figure 3.29 Refined density maps of class 2 and 3 from the 3D classification	93
Figure 3.30 Fourier Shell Correlation (FSC) plots for the EM density map	94
Figure 3.31 The atomic structure of 167 bp-chromatosome fitted into the class 3 map of 177 bp chromatosome after post-processing	96
Figure 3.32 Orientation of linker histone globular domain within the chromatosome structure	96
Figure 3.33 Modification of the 197 bp chromatosome crystal structure to the “model-177 bp chromatosome”	97
Figure 3.34 The atomic structure of the “model-177 bp chromatosome” fitted into the class 3 map of 177 bp chromatosome after post-processing	97
Figure 3.35 Binding of linker histone confers asymmetry to the nucleosome	99
Figure 3.36 177 bp-chromatosome map contains densities that are suggested to be assigned to the linker histone C-terminal domain	99
Figure 3.37 Precipitation assays of 177 bp-nucleosome and chromatosome	100
Figure 3.38 Cartoon illustration of the charge distribution on human linker histone hH1 ⁰	102
Figure 3.39 Resolubilization of Mg ²⁺ induced precipitation assay	103
Figure 3.40 Resolubilization of 177 bp-nucleosome and chromatosome induced by Mg ²⁺	104
Figure 3.41 The SAXS spectrum of 177 bp chromatosome	105
Figure 3.42 Schematic presentation of different NCP phases	106
Figure 3.43 SAXS spectra of 177 bp chromatosome samples	107
Figure 3.44 SAXS spectra of 177 bp nucleosome samples	109
Figure 3.45 Structural comparison of 177 bp chromatosome with nucleosome core particle ..	110
Figure 3.46 Solution SAXS spectra and Guinier plots of 177 bp nucleosomes	111
Figure 3.47 vHW curves of 12-177-wtHO nucleosome array with different Mg ²⁺ concentrations	114

Figure 3.48 Sedfit curves of 12-177-wtHO nucleosome array with different Mg^{2+} concentrations	115
Figure 3.49 vHW curves of 12-177-wtHO nucleosome array with different Na^+ concentrations	117
Figure 3.50 Sedfit curves of 12-177-wtHO nucleosome array with different Na^+ concentrations	118
Figure 3.51 vHW curves of 12-177-HO_H4K20Me3 nucleosome array with different amount of histone octamers in TEK buffer	119
Figure 3.52 Sedfit curves of 12-177-HO_H4K20Me3 nucleosome array with different amount of histone octamers at 0 salt condition	120
Figure 3.53 vHW curves of 12-177-HO_H4K20Me3 nucleosome array in the presence of 1 mM Mg^{2+}	122
Figure 3.54 Sedfit curves of 12-177-HO_H4K20Me3 nucleosome array in the presence of 1 mM Mg^{2+}	123
Figure 3.55 vHW curves of 12-177-HO_H4K20Me3 nucleosome array in the presence of 100 mM Na^+	123
Figure 3.56 Sedfit curves of 12-177-HO_H4K20Me3 nucleosome array in the presence of 100 mM Na^+	124
Figure 3.57 Comparison of the compaction of wild-type and modified nucleosome array	125
Figure 3.58 Precipitation assays of unmodified and modified (H4K20Me3) 12-177 nucleosome array	128
Figure 3.59 Resolubilization of Mg^{2+} induced precipitation assay	129
Figure 3.60 Resolubilization of 12-177 nucleosome array induced by Mg^{2+}	130

List of tables

Table 1.1 Human H1 subtypes.....	28
Table 3.1 Summary of EC50 values	101
Table 3.2 EC50 of the Mg^{2+} induced 177 bp-nucleosome and chromatosome resolubilization	104
Table 3.3 Summary of nucleosome/ chromatosome geometry.....	112
Table 3.4 Summary of S-values of 12-177-wtHO nucleosome array folded in different Mg^{2+} ...	114
Table 3.5 Summary of Sedfit calculation of S-values of 12-177-wtHO nucleosome array folded in different Mg^{2+}	115
Table 3.6 Summary of S-values of 12-177-wtHO nucleosome array folded in different Na^+	117
Table 3.7 Summary of Sedfit calculation of S-values of 12-177-wtHO nucleosome array folded in different Na^+	118
Table 3.8 Summary of S-values of 12-177-HO_H4K20Me3 nucleosome array with different amount of histone octamers.....	119
Table 3.9 Summary of Sedfit calculation of S-values of 12-177-HO_H4K20Me3 nucleosome array with different amount of histone octamers at 0 salt condition	120
Table 3.10 S-values of 12-177-HO_H4K20Me3 nucleosome array with 1 mM Mg^{2+}	122
Table 3.11 Sedfit calculation of S-values of 12-177-HO_H4K20Me3 nucleosome array in 1 mM Mg^{2+}	123
Table 3.12 S-values of 12-177-HO_H4K20Me3 nucleosome array in 100 mM Na^+	123
Table 3.13 Sedfit calculation of S-values of 12-177-HO_H4K20Me3 nucleosome array in 100 mM Na^+	124
Table 3.14 Summary of EC50 values	128
Table 3.15 EC50 of the Mg^{2+} induced 12-177 modified and unmodified nucleosome array resolubilization.....	130

Abbreviations

APS	Ammonium persulfate
AUC	Analytical ultracentrifugation
CAP	Catabolite gene activator protein
cryo-EM	cryo-electron microscopy
CoHex ³⁺	cobalt (III) hexammine
CTD	C-terminal domain
CTF	Contrast transfer function
Dmax	Maximum particle dimension
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
ES cells	Embryonic stem cells
FRET	Fluorescent resonance energy transfer
FSC	Fourier shell correlation
GD	Globular domain
GH5	Globular domain of Linker histone H5
G phase	Gap phase of cell cycle
HI-HII	Helix I and Helix II
H4K20	Histone H4 lysine 20
HO	Histone octamer
IPTG	Isopropyl-thiogalactopyranoside
KDM	Lysine demethylase
KMT	Lysine methyltransferase
Me1	Mono-methylation
Me2	Di-methylation

Me ₃	Tri-methylation
MES	4-Morpholinoethanesulfonic acid
NaAc	Sodium acetate
NCP	Nucleosome core particle
NMR	Nuclear magnetic resonance
NRL	Nucleosome repeat length
NTD	N-terminal domain
OD ₆₀₀	Optical density measured at wavelength of 600 nm
PA	Precipitation assay
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
PEG	Polyethylene glycol
PMSF	Phenylmethanesulfonyl fluoride
R _g	Radius of gyration
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulfate
S phase	Synthesis phase of cell cycle
SUV4-20H1	Histone-lysine N-methyltransferase
SUV4-20H2	Histone-lysine N-methyltransferase
S-value	Sedimentation coefficient
TEM	Transmission electron microscopy
TEMED	Tetramethylethylenediamine
v _{HW}	van Holde-Weischet
VPP	Volta phase plate

Abstract

The dynamic structure of chromatin plays critical roles in genomic activities within living cells. Chromatin has different levels of compaction to cope with various cellular activities. The structure of chromatin could be loose euchromatin to expose DNA for the access of molecular machines during DNA repair, replication or gene transcription, or condensed heterochromatin in non-coding regions and chromosomes during cell mitosis. Various factors contribute to the formation and maintenance of the diversity of chromatin structures.

This thesis investigates structural and biophysical properties of the primary building blocks of eukaryotic chromatin: nucleosome, chromatosome (complex of nucleosome and linker histone), and nucleosome array with techniques of cryo-electron microscopy single particle analysis (cryo-EM), X-Ray diffraction, small-angle X-ray scattering (SAXS), precipitation assay (PA), and analytical ultracentrifugation (AUC).

Cryo-EM method is used to determine the structure of 177 bp chromatosome at 11 Å resolution. The two 15 bp linker DNA arms are positioned close to each other, together with the nearby nucleosome dyad region, forming a triangle area. The densities in the middle of this area are suggested to be from the bound linker histone. The linker histone C-terminal domain (CTD) is proposed to localize in the vicinity of the two linker DNA ends. The data supports the idea that the highly positively charged CTD acts as an “electrostatic bridge” to screen the negative charges of the two linker DNA leading to a more rigid and compacted overall structure.

Self-association and resolubilization of the 177 bp chromatosome and 177 bp nucleosome in the presence of Mg^{2+} and cobalt (III) hexamine ($CoHex^{3+}$) was studied by precipitation assay. Aggregation of the chromatosome proceeds at lower concentrations of Mg^{2+} and $CoHex^{3+}$ in agreement with electrostatic rules (presence of linker histone decreases chromatosome negative charge by 25 %). Resolubilization of 177 bp nucleosome and chromatosome was observed at high concentration of Mg^{2+} . Binding of linker histone increased stability of the aggregated state, thus higher amount of Mg^{2+} is required for resolubilization. The linker histone and Mg^{2+} coordinated regulation delivers possible biological significance in control of chromatin accessibility.

X-Ray diffraction of the chromosome precipitates allowed determination of dimensions of the 177 bp chromosome (thickness 58.7 Å, diameter 111.2 Å). The phase of precipitated chromosomes showed presence of bilayer organization with bilayer-bilayer distance dramatically different from the one reported in literature for the bilayers of 145-148 bp nucleosomes. The difference in bilayer-bilayer distance reflects distinct features of chromosome shape and binding of the linker histone.

Solution SAXS is applied to calculate the radius of gyration (R_g) of 177 bp nucleosome and chromosome and to evaluate the degree of unwrapping of nucleosomal DNA. The larger R_g value of 177 bp chromosome compare to the nucleosome corresponds to the increase of the chromosome size caused by binding of linker histone. The deeper characteristic unwrapping dip in the chromosome spectrum compared to that of the nucleosome suggests full wrapping of DNA in chromosome and some degree of unwrapping in nucleosome. The results indicate that repulsion of the linker DNAs in the nucleosome might cause partial DNA unwrapping. Presence of the linker histone reduces DNA-DNA repulsion and increases DNA wrapping.

AUC was applied to study the influence of heterochromatin-specific histone modification, trimethylation of Lys20 of the histone H4, on folding of the 12-177 nucleosome arrays. Both unmodified and modified (H4K20Me3) arrays achieved maximal compaction to ~50 S-value, and showed a tendency of self-association. Similar folding properties of the unmodified and H4K20Me3 arrays support the idea that while H4K20Me3 promotes the compaction of nucleosome arrays with longer NRLs, it has no additive effect on the salt-dependent folding of nucleosome arrays with shorter NRLs.

Self-association of the modified (H4K20Me3) and unmodified 12-177 nucleosome array in the presence of Mg^{2+} and $CoHex^{3+}$ was studied by precipitation assay. No difference was observed in the salt-induced precipitation between these two arrays. In coincidence with 177 bp nucleosome and chromosome, Mg^{2+} induced resolubilization was observed for both modified and unmodified nucleosome arrays. The resolubilization processes were almost the same between the two arrays. Together, the results indicate that H4K20Me3 has no influence in the 12-177 nucleosome array self-association. In the electrostatic view, the similar polyelectrolyte

properties of the modified and unmodified 12-177 nucleosome arrays provide explanations to the observations above.

1 Introduction

In human cells, the genetic material DNA, when completely stretched out is as long as 2 meters, are organized by histones into compacted chromatin and fitted within the confined space of nucleus which is as small as 5-7 μm in diameter (Figure 1.1). The dynamic structure of chromatin plays a fundamental role in the regulation of all kinds of cell activities related with DNA, for example, DNA replication, gene transcription, and DNA repair.

Figure 1.1 A schematic overview of DNA packing during mitosis. DNA is packaged and compressed into chromosomes that fit within the nucleus of a cell (Collaborative, 2014).

1.1 Nucleosome

Nucleosome core particle is the basic subunit of chromatin, it contains a minimum of 145 bp DNA wrapping 1.75 left-handed superhelical turns around a histone octamer which is composed of two molecules each of the core histones H2A, H2B, H3 and H4 (Finch et al., 1977;

Kornberg, 1974) (Figure 1.2). The dimension of the NCP is about 110 x 110 x 57Å (Finch et al., 1977).

Figure 1.2 Schematic representations of the nucleosome core particle. From the crystal structure of nucleosome core particle at 1.9 Å resolution (PDB ID: 1KX5) (Davey et al., 2002). A) View down the DNA superhelix axis showing the organization of the four kinds of core histones and DNA. The DNA strands are shown in grey, histone H2A in yellow, H2B in red, H3 in blue, H4 in green. The dyad is indicated by red arrow. B) View perpendicular to the DNA superhelix axis showing the disc-like shape of the nucleosome core particle. Histone octamer is shown in pale green, DNA is shown in dark grey.

1.2 Histone octamer

The histone octamer has a 2-fold symmetry and adopt shapes as a wedge or a flat disk. The regular ridges running on the outer surface enable it to serve as a left-handed helical spool, with the pitch of $\sim 28\text{\AA}$, onto which a maximum of two superhelical turns of DNA wrap around (Finch et al., 1977; Klug et al., 1980). The eight histones are in a tripartite assembly. The two copies of H3 and H4 histones together form a disk-shaped tetramer organizing the central turn of the core DNA. H2A and H2B form two H2A-H2B dimers, with one dimer sitting on each side of the H3-H4 tetramer disk and defining half of a DNA superhelical turn (Arents et al., 1991; Burlingame et al., 1985; Klug et al., 1980; Richmond et al., 1984). The DNA superhelix has

several sharp bends, making large numbers of interactions with the histone octamer onto which it wraps around (Richmond et al., 1984).

Within the H2A-H2B and H3-H4 histone pairs, the two histones clap with each other, appearing as a “handshake” motif (Figure 1.3 A). There is a “histone fold” motif shared by all the four core histones. The histone fold is composed of a long helix in the center and two short helices on each side connected by a short loop (Arents et al., 1991; Arents and Moudrianakis, 1995) (Figure 1.3 B). In addition to the histone fold, each core histone contains a positively charged N-terminal tail (Arents et al., 1991). Besides, H2B has a histone-fold extension at the C-terminus, name α C; H3 has one α -N histone-fold extension at N-terminus (Davey et al., 2002; Luger et al., 1997).

Figure 1.3 The histones “handshake” motif. A) H2A is shown in yellow, H2B in cyan. The extended chains of H2A and α C extension of H2B are manually excised for a better view. B) Schematic representation of histone fold motif shared by all the four kinds of histones. A long helix lies in the center, with two short helices on each side connected with a short loop. Both figures are derived from the NCP crystal structure (Davey et al., 2002), PDB ID: 1KX5.

The histone folded domains, together with the extended tails and histone-fold extensions define the histone-histone and histone-DNA interactions within the individual nucleosome core particles. The N-terminal tails are able to pass by the DNA superhelix to contact with neighboring NCPs (Luger et al., 1997). When in solutions, the histone-DNA interaction is further stabilized by water molecules through the formation of hydrogen-bond bridges. Besides, the involvement of divalent cations contributes to histone-histone and histone-DNA interactions

between NCPs, whereas monovalent anions cause a hindrance to the electrostatic interaction between histones and DNA (Davey et al., 2002).

1.3 Core histone modifications

The flexible histone tails extending from the nucleosome cores are susceptible to various post-translational modifications, including acetylation of lysine, methylation of lysine and arginine, phosphorylation of serine and threonine, ubiquitination of lysine and ADP-ribosylation. These covalent modifications are mostly found in the tail regions and are believed to influence chromatin structures and functions directly or serve as signals for recruitment of other factors, like chromatin-remodeling complexes and transcriptional machinery. Histone modifications are thought to be associated, directly or indirectly, with various biological processes, including DNA replication, gene expression, transcription, DNA repair, and genome stability (Felsenfeld and Groudine, 2003; Goll and Bestor, 2002; Peterson and Laniel, 2004; Wolffe and Hayes, 1999).

Figure 1.4 Post-translational modifications of core histones (Peterson and Laniel, 2004). Acetylation is indicated by red triangles, methylation by green pentagons, phosphorylation by yellow circles, ubiquitination by blue stars.

1.3.1 H4K20 Methylation

Unlike histone acetylation which is primarily associated with gene activation, histone methylation leads to more diverse situations, it could be either gene activation or repression, depending on methylation positions and states (Berger, 2002). Histone methylation targets either lysine or arginine residues. The most frequent methylation sites in vertebrates are lysine 4, 9 and 27 of histone H3, and lysine 20 of H4 (Wolffe and Hayes, 1999). The author here will focus on histone H4 lysine 20 methylation (H4K20) which is in the N-terminal tail region, and is the major lysine methylation of histone H4 in mammalian cells (Jorgensen et al., 2013). Like any other lysine residues, there are three states of methylation on H4K20, namely mono-, di-, and tri-methylation (Figure 1.5), of which the mono-methylation is mediated by H4K20-specific methyltransferase SET8 (alias PR-Set7) (Fang et al., 2002), and the di- and tri- methylation catalyzed by SUV4-20H1 and SUV4-20H2 (Schotta et al., 2004). H4K20 methylation plays critical roles in the maintenance of genome integrity.

Figure 1.5 Illustration of three methylation states of lysine. Hydrogens of the side chain ϵ -amino group could be replaced by up to three methyl moieties (Gozani, 2017). KMT stands for lysine methyltransferase which catalyzes the methylation of lysine; and KDM for lysine demethylase, which reverses the methylation process.

Mono-methylation (Me1)

The abundance of Me1 is cell cycle-dependent, which is reflected by fluctuation of SET8 level (Pesavento et al., 2008). Me1 plays an important role in regulation of chromatin structure throughout cell cycle. Chromatin is highly compacted during mitosis. The dynamic H4K20

methylation state plays a critical role in the maintenance of cellular homeostasis (Jorgensen et al., 2013). Besides, together with H3K27 trimethylation and H3K9 dimethylation, H4K20 monomethylation is suggested to be a mark of inactive X chromosome.

Di-methylation (Me2)

Me2 is the most abundant methylation state in proliferating cells, which are found on ~80 % of histone H4 molecules (Pesavento et al., 2008). Me2 is widely distributed across the entire genome (Schotta et al., 2004). Both the non-specific distribution and the high abundance of Me2 make it less possible to function in transcriptional regulation. Me2 is suggested to be critical for general chromatin-mediated processes (Jorgensen et al., 2013).

Tri-methylation (Me3)

H4K20 trimethylation is evolutionarily conserved (Schotta et al., 2004). H4K20 trimethylation is an epigenetic mark of transcriptional silence, it is highly enriched in heterochromatic regions like pericentric region, telomeres, and imprinted region (Regha et al., 2007; Schotta et al., 2004). The loss of H4K20me3 has been identified as a hallmark of cancer (Fraga et al., 2005). H3K9 trimethylation is required for the induction of H4K20 trimethylation by SUV4-20H1 and SUV4-20H2 at the pericentric heterochromatin region. A sequential induction mechanism was proposed for the coexistence of H3K9 and H4K20 trimethylation at the pericentric heterochromatin region (Schotta et al., 2004).

1.3.2 Roles of H4K20 methylations in DNA replication

H4K20 methylation plays essential roles in DNA replication. Different methylation states present distinct regulation functions in DNA replication. It is proposed that Me1 is indirectly involved in the replication through the regulation of chromatin structure in the vicinity of ORIs (replication origins) that assist the progression of replication forks, whereas Me2 and Me3 play important roles in the recruitment of the Orc complex to ORIs (Beck et al., 2012). It is found that SET8 depletion leads to massive DNA double-strand breaks (DSBs) and activation of DNA damage checkpoints during DNA replication, which implicates the significant roles of H4K20

mono-methylation in genomic integrity (Jorgensen et al., 2007). Absence of H4k20me2 leads to impaired or delayed development in zebrafish (Kuo et al., 2012) or mice (Schotta et al., 2008).

1.4 Linker histone

Linker histone is one of the five histone families which are components of chromatin in eukaryotic cells, the other four families are core histones H2A, H2B, H3, and H4 which have been introduced above. Linker histone is also called H1 or H5. The discovery of H1 variants in calf thymus was before the discovery of core histone variants (Kinkade and Cole, 1966). Linker histone structure is composed of three domains, in which a globular domain of about 80 amino acids is flanked by two less structured N- and C-terminal tails (Figure 1.6).

Figure 1.6 Structural composition of linker histone. Figure cited from (Lyubitelev et al., 2016). Linker histone structure is composed of 3 domains- the N-terminal domain, globular domain and C-terminal domain shown in red, green and blue, respectively.

The globular domain is highly conserved in sequence among linker histone family. Structure of chicken erythrocyte H5 globular domain (GH5) was solved by X-ray crystallization to 2.5 Å (Ramakrishnan et al., 1993) (Figure 1.7). Unlike core histones, GH5 lacks the histone fold motif, instead it consists of a three α -helix bundle (I-III) and a β -hairpin at the C terminus. In addition, there is a three-residue stretch forming a β -strand between helix I and II, together with the C-terminal beta-hairpin, this short β -strand forms the antiparallel β -sheet.

The linker histone N-terminal domain is relatively short, composed of around 45 residues; while the C-terminal domain is almost half the size of the whole molecule, with around 100 residues

and highly enriched in positively charged residues, lysine and Arginine (Happel and Doenecke, 2009). The highly basic C-terminus is suggested to interact with the linker DNA and critical for the high-affinity binding of H1 with nucleosome. In contrast, loss of N-terminal domain only modestly affects the binding affinity (Hendzel et al., 2004). Furthermore, the binding affinity is not evenly distributed along the entire C-terminal domain (CTD), rather mainly comes from residues in either end of CTD (Hendzel et al., 2004).

Figure 1.7 3D structure of linker histone H5 globular domain. The figure derived from crystal structure (PDB ID: 1HST) (Ramakrishnan et al., 1993) shows the three-dimensional composition of the GH5 domain (with color codes from cyan for N terminus to red for C terminus): a three α -helix bundle composing of helix I, II and III; an antiparallel β -sheet consisting of the β -strand in between helix I and II, and the C-terminal β -hairpin.

The CTD is quite flexible and unstructured when linker histone is free (Allan et al., 1980).

Although the amino acid sequences of CTDs among various linker histone isoforms are not conserved, there exists a similar amino acid composition which endows the CTD with intrinsic disorder. It is suggested that the CTD would only adopt secondary structures upon binding with DNA or nucleosome (Roque et al., 2005). Besides, it is proposed that CTD undergoes a coupled binding-and-folding process (Lu and Hansen, 2004; Roque et al., 2005). The random coiled CTD adopts a fully structured conformation upon binding with DNA, with dramatic increase in helical content and formation of TPKK (four amino residues namely threonine, proline, lysine and lysine) turn which has long been proposed as potential DNA binding motif (Vila et al., 2001; Vila

et al., 2000). The FRET (fluorescent resonance energy transfer) experiments (Figure 1.8) show a significant decrease in the distance of the two ends of CTD, which could be explained by the hypothesis that upon binding with nucleosome, the CTD folds into secondary structures (Caterino et al., 2011).

Figure 1.8 Model of nucleosome-directed folding of linker histone CTD. The N-terminal domain and globular domain are simplified as straight line and an oval shape, respectively. The flexible C-terminal domain is shown as dashed curve (upper panel), and helices (lower panel) upon nucleosome binding. Figure obtained from (Caterino et al., 2011).

The site mutation of T152 to glutamic acid (T→E), as a mimic of phosphorylation, causes a greater decrease in the binding affinity than a truncation which results in loss of almost one third of the total positive charges of CTD. Thus, the binding affinity cannot be simply explained by the neutralization effect of the positively charged lysine clusters, whereas the hypothesis that CTD adopts a secondary structure upon the initial non-specific binding with DNA is more acceptable (Hendzel et al., 2004).

1.4.1 Linker histone variants

One prominent feature of linker histone is its heterogeneity. Linker histone is the most divergent histone group compared with the four core histones. The family consists of linker histones with distinct specificity for organisms, tissues and developmental stages (Izzo et al., 2008). The contribution of particular variant to chromatin-mediated processes is currently an open and very intriguing question. Individual linker histones are often referred to as subtypes (within a species) or variants (between different species). Figure 1.9 gives an overview of H1

variants of different species. Human and mouse linker histones have the most diverse subtypes, and they share a common ancestor gene. More details will be given on human linker histones below.

Figure 1.9 Evolutionary tree of linker histones. The number in parentheses indicates the known number of subtypes in a given species (Izzo et al., 2008). Figure from (Wikipedia, 2017)

Comparative analysis shows that the evolutionary origin of linker histone H1 can be traced to eubacteria. The helix-winged central motif arises later in protists and is highly conserved in evolution (Kasinsky et al., 2001), both the N- and C-terminal domains are rather diverse throughout the family (Allan et al., 1980), as also shown in the Figure 1.10 the divergent CTD and NTD domains of human H1 subtypes.

Figure 1.10 Multiple alignment of human linker histone subtypes. Identical residues are highlighted in black. Different grey shades correspond to the degree of amino acid conservation, i.e. the darker the more highly conserved. Above the alignment, the solid lines indicate the winged helix motif of the globular domain. (Happel and Doenecke, 2009).

In human cells, 11 H1 subtypes (Table 1.1) have been described and each is encoded by single gene. According to their temporal and spatial expression, these 11 subtypes can be further categorized into sub-groups. 6 subtypes (H1.1-H1.6) are mainly expressed during S phase of the

cell cycle and hence replication-dependent. They are encoded by genes within histone cluster 1 located on chromosome 6. The other 5 subtypes (H1.0 and H1.7-H1.10) are replication-independent, because they are expressed over the entire cell cycle. Besides, their encoding genes are scattered in the genome. Somatic H1 subtypes (H1.1–H1.5) are present in almost every cell, with the exception that H1.1 is expressed extremely low in several cases. It is also worth to mention that H1.2 and H1.4 are highly expressed in most cell types. Whereas other H1 subtypes, such as H1.0, exist mainly in terminally differentiated cells (Zlatanova and Doenecke, 1994) or are oocyte (H1_{oo}) or testis (H1_t, H1T2 and H1LS1) specific. H1x is currently the least characterized H1 subtype (Happel and Doenecke, 2009; Izzo et al., 2008; Scaffidi, 2016).

Table 1.1 Human H1 subtypes

Human H1 subtypes within cluster of core histone genes (replication-dependent)				
H1 gene (and alias)	H1 protein (and alias)	Gene locus (chromosome)	Length (No. of amino acids)	Expression (cell, tissue)
<i>HIST1H1A</i> (<i>H1F1</i>)	H1.1	6p21.3-22	214	Ubiquitous
<i>HIST1H1B</i> (<i>H1F5</i>)	H1.5	6p21.3-22	225	Ubiquitous
<i>HIST1H1C</i> (<i>H1F2</i>)	H1.2	6p21.3-22	212	Ubiquitous
<i>HIST1H1D</i> (<i>H1F3</i>)	H1.3	6p21.3-22	220	Ubiquitous
<i>HIST1H1E</i> (<i>H1F4</i>)	H1.4	6p21.3-22	218	Ubiquitous
<i>HIST1H1T</i> (<i>H1FT</i>)	H1.6 (H1 _t)	6p21.3-22	206	Spermatocytes
Human H1 subtypes encoded by orphan genes (replication-independent)				
Gene symbol (and alias)	H1 protein symbol (and alias)	Gene locus (chromosome)	Length (amino acids)	Expression (cell, tissue)
<i>H1FO</i> (<i>F1FV</i>)	H1.0 (H1 ⁰)	22q13.1	193	Differentiated cells
<i>H1FNT</i> (<i>HANP1</i>)	H1.7 (H1T2)	12q13.11	233	Spermatids

<i>H1FOO</i>	H1.8 (H1foo, H1oo)	3q21.3	345	Oocytes
<i>HILS1</i>	H1.9 (HILS1, Hils1)	17q21.33	230	Spermatids
<i>H1FX</i>	H1.10 (H1x, H1X, H1.X)	3q21.3	212	Ubiquitous

Table cited from (Happel and Doenecke, 2009; Izzo et al., 2008)

1.5 Chromatosome

The complex of linker histone and nucleosome is called chromatosome (Simpson, 1978) (Figure 1.11). Chromatosome was first observed as the products of micrococcal nuclease digestion of chromatin. The digestion had several major products, from which one was identified as chromatosome. It contained 166 bp of DNA, a histone octamer, and one molecule of linker histone. The DNA wrapped around the histone octamer by two complete turns, in which an additional 20 bp of DNA was stabilized as compare with nucleosome (i.e. in the absence of H1). Chromatosome is described as the basic repeating structural unit of the chromatin.

Binding of linker histone suppresses the sliding of histone octamer on DNA, which suggests linker histone as a potential inhibitor of nucleosome mobility. It limits the access of chromatin remodeling factors and other regulatory proteins to nucleosome components, and possibly DNA damage response as well, which requires decreased chromatin compaction. The reversible restraint of linker histone on chromatin dynamics is proposed to play a role in local regulation of processes that require access to DNA (Pennings et al., 1994).

There are considerable studies suggesting that linker histones are critical to the higher-order chromatin compaction (discussed in review (Happel and Doenecke, 2009)). The stabilization effect of linker histones on chromatin folding was first demonstrated by EM study (Thoma et al., 1979), which showed that under increasing ionic strength, linker histone-containing nucleosome array went through progressive condensation into a uniform and highly compacted superstructure, whereas H1-depleted nucleosome array adopted a partially folded and extended structure. A more recent study by electron microscopy (EM) also reveals the essential

role of linker histone in the formation of chromatin 30 nm fiber (Figure 1.12) (Routh et al., 2008).

Figure 1.11 Illustration of a chromatosome composition. A chromatosome is made from a linker histone binding to a nucleosome at or near the nucleosome dyad, interacting with one or two linker DNA. (Collaborative, 2014)

Figure 1.12 Electron microscope images of chromatin with or without linker histone. A) The 197 bp nucleosome array as an unfolded “puddles” in the absence of linker histone; B) shows the same array fully folded into the 30 nm fiber in the presence of linker histone with 1:1 stoichiometric amount with nucleosomes. Scale bar is 50nm. Figure from (Routh et al., 2008).

1.5.1 H1 stoichiometry

Linker histone presents on average 1:1 stoichiometry with nucleosome *in vivo* (Widom, 1998). However, single-molecule studies of the H1 and nucleosome binding interactions showed heterogeneous binding stoichiometry (Yue et al., 2016). Besides, H1 availability is suggested as a primary determinant of the nucleosome repeat length (Fan et al., 2005). In addition, the *in vivo* studies of H1 depletion with mouse embryonic stem cells suggested a strong correlation between H1 per nucleosome amount and nucleosome spacing which is commonly indicated by the NRL.

In mouse ES cells with 50 % reduction in the normal H1 amount, the NRL is shortened by ~15 base pairs compared with the wide type (Fan et al., 2005). The strong correlation between linker histone to nucleosome ratio and NRL can be explained in the aspect of electrostatic interactions- reduction in H1 amount leads to the over-production of core histones to compensate for the loss of DNA compaction capacity, so that the balance of positive and negative charges returns to linker histone/nucleosome ratio that is sufficient for DNA compaction and change in this ratio is finally reflected by decrease of the NRL. The mouse data in Figure 1.13 show a liner relation between NRL and H1 per nucleosomes (Woodcock et al., 2006).

Figure 1.13 Linker histone stoichiometry and NRL. The nucleosome repeat lengths of wild-type, H1-depleted mouse tissues (Fan et al., 2003) and embryonic stem (ES) cells (Fan et

al., 2005) were plotted against H1/nucleosome stoichiometry. Number 1-6 represent as following- 1. Wild-type thymus; 2. Wild-type liver; 3. H1-depleted liver; 4. H1-depleted thymus; 5. Wild-type ES cell; 6. H1-depleted ES cell. Dashed line is linear regression for the above data, with extrapolation. Figure from (Woodcock et al., 2006).

1.5.2 H1 and other core histones

It is proposed that the H3 and H2A tails interfere with the binding of H1 C-terminal domain with nucleosome (Allan et al., 1986). On the other hand, histone H3 modifications and N-terminal tail mobility is inhibited by the binding of H1 with nucleosome. What's more, a segregation is observed in the genome-wide level between linker histone occupancy and H3 N-terminal tail modifications. (Stützer et al., 2016). In addition, the level of, especially the charge-related post-translational modification on core histones is also affected by H1 depletion. Acetylation of H4K12 and tri-methylation of H3K27 are decreased by 4 and 2 fold, respectively, presumably to compensate for the loss of H1 which resulted in decreased capability of neutralizing DNA negative charges (Fan et al., 2005).

1.5.3 Linker histone interaction with nucleosome

Linker histones assist chromatin compaction through direct interactions with nucleosomes. A quantitative study suggests that linker histone binds with nucleosome upon recognition of a specific DNA geometry in which two DNA gyres juxtaposing closely with angled arrangement (Figure 1.14), the minimal requirement for this high affinity binding is a single linker DNA arm with the length of at least 18 bp (White et al., 2016).

Figure 1.14 Binding of linker histone with nucleosome has geometry preference. H1 (shown in yellow) prefers cruciform (B) and nucleosomal DNA (C, D) over linear DNA (A);

a single linker DNA arm of ~18 bp is sufficient for high affinity binding. Figure from (White et al., 2016).

The highly conserved globular domain is essential for linker histone binding with nucleosome. The structure of GH5 is strikingly similar to the DNA-binding domain of CAP (catabolite gene activator protein), suggesting helix III as the DNA recognition motif. Two highly conserved clusters of positively charged residues are proposed to be the binding sites with nucleosome, one of which is located on the aforementioned helix III; the second cluster of basic residues are from the loop connecting helix I and II, helix II and C-terminal end of the β -strand (Figure 1.15) (Ramakrishnan et al., 1993).

Figure 1.15 Two nucleosomal DNA binding sites of GH5. The 3D structure of GH5 was obtained from the crystal structure (PDB ID: 1HST) (Ramakrishnan et al., 1993). The secondary structural components- Helix I, II, III, loop between Helix I and II, C- and N-terminus are indicated. Residues of site 1 are colored in blue, site 2 in lime.

1.5.4 Various models for linker histone binding with the nucleosome

Various structural models for nucleosome interacting with linker histones have been proposed. In this section, the models discussed below involve different linker histone variants and linker DNA lengths. Various linker histone localizations and orientations are demonstrated.

A binding model with three binding sites was proposed in (Cui and Zhurkin, 2009). In addition to the aforementioned two binding sites in Ramakrishnan's work (Figure 1.15), the "wing"- loop connecting the two β strands- containing a highly conserved tetra-peptide (GVGA) nonpolar

region was proposed as the third nucleosome binding site (Figure 1.16 A, site 3 in violet). The two basic residues on helix III (Figure 1.16 A&B, site 1 in blue) interact with the nucleosome DNA and position the linker histone in the DNA major groove which is 5 bp (i.e. half of DNA helical turn) away from the nucleosome dyad. The hydrophobic interactions (Figure 1.16 B, circled in black) between the GH5 “wing” motif and the thymine methyl groups of linker DNA AT-rich region bend the linker DNA by 20° towards the major groove at the exit/entry point. Binding site 2 (Figure 1.16 A&B, in lime) interacts with the same linker DNA together with site 3 (the “wing” in violet).

Figure 1.16 Structural model of GH5 binding with the nucleosome. A) The 3D structure of GH5 showing three DNA-binding sites, coded with different colors- site 1 in blue, site2 in lime, site 3 in violet; critical residues of the 3 sites are designated and shown as “ball-and-stick”. B) Molecular model of GH5-nucleosome complex. The 3 binding sites are represented with the same color code as in (A). Only one linker DNA is bound with GH5, interacting with GH5 binding site 2 and 3; whereas the other linker DNA remains “free”. The bound linker DNA was bent by 20° towards the major groove of nucleosomal DNA through the Interaction with site 3. Figure cited from (Cui and Zhurkin, 2009).

A study by computational docking came out with the same three DNA binding sites on GH5 as proposed in Cui and Zhurkin’s work (Figure 1.16). However, the position of GH5 on nucleosome and the relative orientation are different. The “wing” motif (Figure 1.17), rather than the helix III (Figure 1.16), interacts with the nucleosome dyad; whereas helix III contact ~15 bp of one linker DNA, HI-HII loop together with the basic residues at the C-terminus contacts ~5 bp of the other linker DNA (Fan and Roberts, 2006).

Figure 1.17 Three nucleosomal DNA binding sites of GH5 (Fan and Roberts, 2006). The 3D structure was adopted from the crystal structure of GH5 (PDB ID: 1HST). The secondary structural components- Helix I, II, III, loop between Helix I and II (HI-HII loop), wing motif, C- and N-terminus are indicated. Residues of site 1 are colored in blue, site 2 in lime, site 3 in violet.

Figure 1.18 First crystal structure of chromosome at 3.5 Å resolution. A) Presentation of the chromosome structure adopted from PDB ID 4QLC (Zhou et al., 2015). The 10 bp of linker DNA on each side are highlighted in yellow. The chicken linker histone globular domain is seated at the nucleosome dyad; the “wing” motif interacts with the nucleosomal DNA, helix III and HI-HII loop each interacts with one linker DNA. The red squared region is zoomed in shown in figure B) for a close-up view of the interaction details.

The structure of linker histone globular domain and nucleosome complex was solved by X-ray crystallography at 3.5 Å (Zhou et al., 2015), which for the first time gives structural details of interactions between linker histone and nucleosome. The chromatosome studied is the complex of GH5 and a 167 bp nucleosome with 10 bp of linker DNA on each side. The structure presents a symmetric linker histone binding mode (Figure 1.18) in which the 'wing' motif (loop between the two β strands) is seated at the nucleosome dyad interacting with the nucleosomal DNA; the helix I-II loop (Figure 1.18 B, in lime) and helix III (in blue) each interacts with one of the linker DNA. The chromatosome model proposed here in general agrees with Fan and Roberts' (Fan and Roberts, 2006) computational model in both GH5 relative location and orientation, except the length of linker DNAs involved in the interactions- 10 bp of each linker DNA in Zhou's crystal structure versus 15 bp and 5 bp in Fan and Roberts' computational model. In addition to the chicken linker histone H5, other linker histone variants from fruit fly, mouse, human, and frog were also involved in the study of chromatosome structure. The divergent mode of interactions of various linker histones with nucleosome is very impressive.

It is reported that the *Drosophila melanogaster* H1 globular domain (GH1) binds with nucleosome asymmetrically (Figure 1.19) through two clusters of lysine residues which are located on the surface of GH1 opposite to each other- one from the helix III, the other from the C-terminal end of helix I and β -hairpin as shown in Figure 1.19. Helix III faces the nucleosomal DNA near the dyad, whereas the other binding site contacts 10 bp of one linker DNA. Also indicated in the same study is that two discrete lysine-rich short regions of the C-terminal domain- one immediately following the globular domain, one in the middle of the CTD- are also involved in the chromatosome formation (Zhou, 2013).

The single-base resolution hydroxyl radical mapping experiments (Syed et al., 2010) suggested that the *Mus musculus* linker histone H1.5 globular domain interacts with either mono- or di-nucleosomes with a 3-contact mode. The GH1 binds to the minor groove at the dyad symmetrically with 10 bp of nucleosomal DNA involved, and 10 bp of each linker DNA. In addition, a region of seven residues from the C-terminal domain plays an important role in the formation of the stem structure of linker DNAs.

Figure 1.19 Asymmetric structural model of the *Drosophila* GH1-nucleosome complex. A) Representation of the two binding sites of GH1. Binding residues from the two sites are highlighted and shown in stick. B) A close-up view of the GH1 localization and orientation in the nucleosome. The nucleosome dyad is shown in red; binding site 1 is shown in magenta, site 2 in orange; the 10 bp linker DNA bound with residues from site 2 is shown in lime. Figure obtained from (Zhou, 2013).

Figure 1.20 *In vivo* mapping of the two nucleosome binding sites on mouse H1.0 globular domain. A) Mapping of the binding residues is represented on the atomic structure of GH5 (PDB ID: 1HST) (Ramakrishnan et al., 1993). To highlight, the binding residues are shown in stick; residues of site 1 are colored in blue, site 2 in lime. B) Molecular model of GH1-nucleosome interaction (Brown et al., 2006). The nucleosomal DNA is colored in blue, the dyad in yellow, GH1 in red.

The mutagenesis mapping of GH1 and nucleosome interactions *in vivo* (Brown et al., 2006) identified two distinct nucleosome binding sites on the surface of mouse linker histone H1.0 globular domain. As shown in Figure 1.20 B, GH1 interacts with the major groove of nucleosomal DNA which is one helical turn away from the dyad via a cluster of basic residues from binding site 1; and binds to the minor groove of linker DNA through the other cluster of positively charged residues of binding site 2.

The structure of chromosome- complex of 197 bp nucleosome with *Xenopus laevis* full length linker histone H1.0- was solved by cryo-EM single particle reconstruction (Bednar et al., 2017). The structure reveals the on-dyad binding mode of H1, with the globular domain interacts with both the nucleosomal DNA at the dyad and the two linker DNAs. The orientation of GH1 is in agreement with that suggested by (Fan and Roberts, 2006) and (Zhou et al., 2015) that the “wing” motif (W1 in Figure 1.21 A) binds to the nucleosome dyad, while the HI-HII loop (L1 in Figure 1.21 A) and helix 3 ($\alpha 3$) each binds to one linker DNA. Also suggested from the structure is that the C-terminal domain interacts with one of the linker DNAs (Figure 1.21 B) which endows the chromosome complex with polarity.

Figure 1.21 Schematic representation of the 197 bp chromosome structure. A) Left panel is an overview of the chromosome structure, the H1 globular domain is shown in violet. Right panel is a close-up view of GH1 orientation, the 3 α -helices $\alpha 1$ -3 are in violet, β -hairpin in blue, the loop connecting α -helix 1 and 2 is labeled as L1, the “wing” motif-loop of the β -hairpin is labeled as W1, nucleosome dyad is shown in red, linker DNA bound with L1 is named “Linker-L1”, the other bound with α -helix 3 named “Linker- $\alpha 3$ ”. B)

Representation of H1 C-terminal domain (CTD) bound with one of the linker DNAs. The CTD density is shown in violet blobs. Figure obtained from (Bednar et al., 2017).

The cryo-EM reconstructed structure of the 30 nm chromatin fiber reveals an asymmetric localization of *Homo sapiens* linker histone H1.4 interacting with nucleosome dyad and both linker DNAs (Figure 1.22 A) (Song et al., 2014). It is suggested that, in the tetra-nucleosome structural unit of the chromatin fiber, linker DNA trajectories are determined by the specific localization and orientation of linker histone in such a way that sequentially neighboring nucleosomes are connected by straight linker DNA with every other nucleosome stacking on top of each other forming a two-start zigzag conformation (Figure 1.22 B). Furthermore, self-association of linker histones leads to a left-hand twist of the neighboring tetra-nucleosome units (Figure 1.22 C).

Figure 1.22 Cryo-EM structure of the 30 nm chromatin fiber. A) Representation of the asymmetric localization of linker histone in the nucleosome in two different views. H1 density is colored in green; the red arrow dotted line indicates the nucleosome dyad axis. B) A tetra-nucleosome structural unit of the chromatin fiber with straight linker DNAs connecting sequentially neighboring nucleosomes. C) Illustration of the left-handed twist of the chromatin fiber with three tetra-nucleosome units (in total 12 x nucleosome) each

coded in different colors; N1 to N12 indicate the sequential number of nucleosome along the chromatin fiber. Figure obtained from (Song et al., 2014).

The differences among various models of the linker histone–nucleosome complex might come from 1) different linker histone variants/subtypes use diverse binding modes; 2) different types of nucleosomal DNA and lengths of linker DNA, as well as nucleosome preparation methods; 3) different contexts, i.e. whether interaction is between linker histone and mono-nucleosome or within a long nucleosome array. All these and other not listed factors could induce different binding modes, which subsequently lead to distinct higher-order structures of chromatin (Figure 1.23) (Zhou et al., 2015).

Figure 1.23 Different linker histone binding modes lead to different chromatin compaction. A) Chromatin associated with linker histones adopting the on-dyad binding mode is folded into a more condensed higher-order structure compared with B) chromatin that associated with off-dyad positioned linker histones. Linker histones/globular domains (e.g. chicken GH5) adopting the on-dyad binding mode are in red color, those (e.g. fruit fly GH1) adopting off-dyad binding mode are in orange. Figure obtained from (Zhou et al., 2015).

It has been suggested that chicken globular H5 (GH5) takes on-dyad mode when binding with the nucleosome (Zhou et al., 2015), while fruit fly globular H1 (GH1) adopts off-dyad binding mode (Zhou, 2013). Besides, as also suggested by in Figure 1.23, the sedimentation experiments indicate that chromatin containing GH5 is more compacted than that with GH1. The determinants of either on- or off-dyad linker histone binding modes were recently investigated through mutagenesis and NMR (Zhou et al., 2016). Five critical residues of GH5 mutated to the corresponding residues of GH1 converted the binding mode of GH5 from on-

dyad to off-dyad. What's more, the mutated GH5 and GH1 share similar capability in chromatin compaction. All these combined together suggest that a small group of residues from the linker histone globular domain directs the on- or off-dyad binding mode with the nucleosome and controls the chromatin higher-order structure (Zhou et al., 2016; Zhou et al., 2015).

1.5.5 Linker histone and chromatin structure

Binding of linker histone to nucleosome leads to a stem-like conformation at the entry and exit site of linker DNA (Bednar et al., 1998). Electron microscopy shows a distinct difference in the curvature of entering and exiting DNA to and from the nucleosome core. In linker histone-containing chromatin, linker DNA enters and leaves the nucleosome almost at the same site, whereas in linker histone-depleted chromatin, the entry and exit sites are roughly at the opposite positions on the nucleosome core (Thoma et al., 1979).

In the absence of linker histone, DNA only wraps ~ 1.7 turns around the nucleosome core, with the two ends bending away from each other as a result of electrostatic repulsion. The central globular domain of linker histone alone is able to increase the wrapping to 1.8 to 1.9 turns, but fails to cross the entry and exit ends of DNA. The crossing of the two ends is achieved by the CTD with the formation of a four-stranded stem structure (Hamiche et al., 1996).

Figure 1.24 Schematic illustration of packing of chromatin into 30 nm fiber assisted by linker histones. (Figure from website (Buratovich, 2014)) The first-order of DNA organization involves the wrapping of 147 bp DNA around globular histone octamer core, forming the basic unit of chromatin named nucleosome; neighboring nucleosomes are connected by 10-80 bp of linker DNA, forming "beads-on-the-string" with 10 nm thickness;

additional 20 bp DNA is organized with the binding of linker histone on or near the nucleosome dyad; further compaction leads to the formation of “30 nm” fiber.

There are very different effects of salt on the morphology of chromatin higher-order structure between H1-containing chromatin and H1-depleted chromatin. With increasing ionic strength, H1-containing chromatin undergoes a progressive condensation into a well-organized superstructure (Figure 1.24), whereas H1-depleted chromatin becomes more condensed though, doesn't have a well-defined structure (Thoma et al., 1979).

1.6 Chromatin higher-order structure

Figure 1.25 illustrates the different compaction levels in the process of DNA packing into the condensed mitotic chromosome (Felsenfeld and Groudine, 2003). The first level of compaction is the formation of nucleosomes with interval of 10-80 bp of free DNA between each other. This extended “beads-on-a-string” fiber, with the diameter of 11nm, leads to a 5-10 folds compaction of DNA (Kornberg, 1974). Further compaction results in a more condensed fiber with a diameter at ~30 nm, because of which, chromatin at this level is also called 30 nm fiber.

Figure 1.25 Schematic illustration of chromatin compaction into condensed fiber. (Figure from website (Collaborative, 2014)) The DNA double strand first coils around the histone

octamers and sealed by linker histones forming the 10 nm “beads-on-the-string”; the chromosome string then coils into 30 nm fiber; followed by further condensation into 300 nm fiber; and finally the 700 nm thick fiber.

1.6.1 30 nm chromatin fiber, one-start or two-start?

Within the 30 nm fiber, individual nucleosomes cannot be readily resolved. Numerous works have been done on the organization of the nucleosomes inside the fibers, giving rise to various yet debating conclusions. The debates are mainly focused on whether the 30 nm fiber is formed from one-start or two-start nucleosome organization (Figure 1.26).

One-start model

In the one-start solenoid model, nucleosomes are in a sequential arrangement, with neighboring nucleosomes connected by bent linker DNA and packed side to side along the same helical path (Figure 1.26 A). Electron microscopy shows winding of the nucleosome array into a helical structure, with the diameter of at least 30 nm, and 6 nucleosome cores per turn on average (Finch and Klug, 1976; Robinson et al., 2006; Thoma et al., 1979). It is also implicated that linker histones are located on the interior of the solenoid (Thoma et al., 1979). Analysis of the X-ray diffraction pattern of 30 nm chromatin fiber also gives rise to the nucleosome arrangement in accord with the solenoid model (Finch and Klug, 1976; Widom and Klug, 1985).

Two-start model

In the two-start model (Figure 1.26 B and C), sequentially adjacent nucleosomes are positioned in different helical stacks. A study by disulfide cross-linking and Scal endonuclease digestion shows that the maximal length of the digested nucleosome array segment is half of the starting length, which can only be explained by the two-start model (Dorigo et al., 2004). There are two predominant sub-models for the two-start arrangement of nucleosomes. In the supercoiled model (Figure 1.26 B), the two stacks of nucleosomes, in the form of zigzag arrangement, are twisted into a supercoiled structure. The diameter of the supercoil is independent of the NRL (Woodcock et al., 1984; Worcel et al., 1981). In the crossed-linker model (Figure 1.26C), the two parallel nucleosome stacks are left-handed twisted around the long axis of the stacks and compressed along the fiber axis. Results from both electron microscopy and X-ray scattering

indicated a linear correlation between the fiber diameter and the linker DNA length (Michael F. Smith, 1990; Williams et al., 1986).

Figure 1.26 Different models of the 30-nm chromatin fiber. (A) one-start solenoidal model, (B) two-start supercoiled model, and (C) two-start crossed-linker model. Yellow color denotes linker DNA, purple color denotes DNA wrapping around nucleosome cores (Dorigo et al., 2004).

1.6.2 Do 30 nm fibers really exist *in vivo*?

Using both small-angle X-ray scattering and cryo-electron microscopy, the structure of human mitotic chromosomes were found to be rather irregular folded fibers than in the form of compacted 30 nm fibers (Nishino et al., 2012). The maximal regular structural dimension is 11 nm, which coincides with the diameter of “beads-on-a-string”. The hierarchical organization of chromatin and the existence of 30 nm fiber in living cells were argued by the *in vivo* imaging of the interphase nuclei with the super-resolution stochastic optical reconstruction microscopy (STORM) (Ricci et al., 2015). Nucleosomes are observed to be grouped into discrete domains which are connected with nucleosome-depleted regions along the chromatin fiber. 30 nm chromatin fiber structure was not found in neither differentiated nor embryonic stem cells.

1.7 Objectives

The structure of chromatin has been studied since decades ago. Both linker histone and H4K20Me3 are found to be related with chromatin compaction (Everetts et al., 2013; Kim et al.,

2015). The loss of H4K20Me3 is a hallmark of multiple cancers (Fraga et al., 2005). Linker histone binds to the nucleosome at or near the dyad and interacts with one or two linker DNAs (Cui and Zhurkin, 2009; Fan and Roberts, 2006; Zhou, 2013; Zhou et al., 2015). Structural studies of chromatosome came up with multiple debating models in terms of linker histone localization and interaction with the nucleosome (Bednar et al., 2017; Brown et al., 2006; Zhou et al., 2015). The decrease in linker histone level is closely related to multiple diseases (Ye et al., 2017). Besides, the effect of linker histone on chromatin compaction is influenced by NRLs (Routh et al., 2008). The actual molecular mechanisms of H4K20Me3, linker histone and NRLs mediating chromatin structure and dynamics are not clear. To address this question, the author will prepare nucleosome, chromatosome and nucleosome array *in vitro* reconstituted from precisely positioning DNA templates, histone octamers and linker histones. Cryo-EM single particle analysis will be conducted to reconstruct the chromatosome 3D structure to elucidate the molecular details of the interactions of linker histone with the nucleosome. Cation-induced precipitation assays will be applied to investigate the effects of linker histone and H4K20Me3 on self-association of nucleosome and chromatin. Solution SAXS and X-ray diffraction will be used to determine dimensions of chromatosome and nucleosome in solid and solution states, respectively. AUC velocity experiments will be conducted to study the influence of H4K20Me3 on chromatin folding in various salt conditions. With all the efforts, the author hopes to reveal new exciting details of the mechanisms of the dependences of chromatin structure on H4K20Me3, linker histone and NRLs.

2 Materials and methods

2.1 Materials

2.1.1 Chemicals

Chemical	Supplier
40 % Acrylamide solution	Bio-Rad
Ammonium persulfate, APS	Bio-Rad
Ampicillin sodium salt	USB
2 % Bis-acrylamide solution	Bio-Rad
β -mercaptoethanol	Sigma
Chloramphenicol	Sigma
Chloroform	Fisher Scientific
Coomassie Brilliant Blue R-250	Bio-Rad
Dimethyl sulfoxide, DMSO	Fisher Scientific
Dithiothreitol, DTT	Gold BioTechnology
Ethanol	Merck
Ethylenediaminetetraacetic acid, EDTA	USB
Glacial acetic acid	Merck
Glucose	Sigma
Glycerol	USB
Glycine	Bio-Rad
Guanidinium hydrochloride	Sigma
HEPES	Sigma
Imidazole	Sigma
Isoamyl alcohol/ isopentyl alcohol/ 3-methyl-1-butanol	Sigma
Isopropanol/ 2-propanol/ isopropyl alcohol	Merck
Isopropyl-thiogalactopyranoside, IPTG	Gold BioTechnology

Magnesium chloride hexahydrate	USB
4-Morpholinoethanesulfonic acid, MES monohydrate	Fluka
Phenylmethanesulfonyl fluoride, PMSF	Fluka
Phenol solution	Sigma
Polyethylene glycol, PEG-6,000	Sigma
Potassium acetate	Sigma
Potassium dihydrogen phosphate, KH_2PO_4	Merck
Potassium chloride	USB
Potassium hydrogen phosphate/ di-potassium hydrogen phosphate, K_2HPO_4	Merck
Sodium acetate	Sigma
Sodium chloride	Merck
Sodium dodecyl sulfate, SDS	Bio-Rad
Sodium hydroxide	Merck
Tetramethylethylenediamine, TEMED	Bio-Rad
Tris-base	Vivantis
Triton X-100	Bio-Rad
Trizma base	Sigma
Urea	USB

2.1.2 Buffers

Buffer	Composition
Lysis buffer 1	50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl pH 8.0
Lysis buffer 2	0.2 M NaCl, 1 % SDS
Lysis buffer 3	4 M potassium acetate, 2 M glacial acetic acid,
Histone octamer unfolding buffer	7 M guanidinium HCl, 10 mM Tris-HCl, 10 mM DTT

MES buffer	20 mM MES, 0.1 M sodium chloride, pH 5.0
Ni buffer A	100 mM Tris-HCl pH 8.0, 700 mM NaCl, 10 mM β -mercaptoethanol, 0.5 mM PMSF
Ni buffer B	100 mM Tris-HCl pH 8.0, 700 mM NaCl, 10 mM β -mercaptoethanol, 250 mM imidazole
Ni lysis buffer	100 mM Tris-HCl pH 8.0, 700 mM NaCl, 1 % Triton X-100, 10 mM β -mercaptoethanol, 0.5 mM PMSF
Phosphate buffer	50 mM K_2HPO_4 , 50 mM KH_2PO_4 , pH 7.4
Refolding buffer	2 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, 10 mM β -mercaptoethanol
S-200 unfolding buffer	7 M guanidinium HCl, 20 mM sodium acetate, 10 mM DTT
SAUDE 1000	7 M urea, 20 mM sodium acetate, 1 M NaCl, 5 mM β -mercaptoethanol, 1 mM EDTA
SAUDE 200	7 M urea, 20 mM sodium acetate, 200 M NaCl, 5 mM β -mercaptoethanol, 1 mM EDTA
S buffer A	50 mM Tris-HCl pH 8.8, 500 mM NaCl
S buffer B	50 mM Tris-HCl pH 8.8, 1 M NaCl
SDS-PAGE running buffer (10x)	250 mM Tris-base, 1.92 M glycine, 1 % SDS
SP Sepharose lysis buffer	50 mM Tris-HCl pH 8.0, 50 mM EDTA, 2 % Triton X-100, 0.5 mM PMSF, 1 mg/ml lysozyme
SP Sepharose wash buffer	50 mM Tris-HCl pH 8.0, 50 mM EDTA, 2 % Triton X-100, 0.5 mM PMSF, 1 M NaCl
TBE buffer (10x)	890 mM Tris-borate, 890 mM boric acid, 20 mM EDTA, pH 8.3
TE (10, 0.1)	10 mM Tris-HCl pH 8.0, 0.1 mM EDTA pH 8.0
TE (10, 50)	10 mM Tris-HCl pH 8.0, 50 mM EDTA pH 8.0
TEK 1.4 M	10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 1.4 M KCl
TEK 1.2 M	10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 1.2 M KCl
TEK 1.0 M	10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 1.0 M KCl
TEK 0.8 M	10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 0.8 M KCl

TEK 0.6 M	10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 0.6 M KCl
TEK 0.01 M	10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 0.01 M KCl
TES-9	100 mM NaCl, 20 mM Tris-HCl, 1 mM EDTA
Washing buffer (for core histone inclusion bodies)	50 mM Tris-HCl, 100 mM NaCl, 1 mM β -mercaptoethanol

2.1.3 *Escherichia coli* competent cells

Competent cell	Usage
BL21(DE3)pLysS	Protein expression
Top10	12-177-601 DNA amplification

2.1.4 Culture Media

Bacterial medium	Composition/1L
2x TY	16 g Bacto Tryptone, 10 g Bacto Yeast extract, 5 g NaCl, 0.1 % glucose
Terrific Broth (TB)	12 g Bacto Tryptone, 24 g Bacto Yeast extract, 17 mM KH ₂ PO ₄ , 72 mM K ₂ HPO ₄ (1 M), 0.4 % glycerol
TYE plate	10 g Bacto Tryptone, 5 g Bacto Yeast extract, 8 g NaCl, 15 g Bacto agar

2.1.5 Enzymes and reaction buffers

Enzyme	Supplier	Reaction buffer
EcoRV (20,000 U/mL)	New England BioLabs	NEBuffer 3: 100 mM sodium chloride, 50 mM Tris-HCl, 10 mM magnesium chloride, 1mM DTT pH 7.9
Scal-HF (20,000 U/mL)	New England BioLabs	CutSmart: 50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 μ g/mL BSA, pH 7.9 @ 25°C
Lysozyme ($\geq$ 40,000 U/mg)	Sigma	/

RNase A (10 mg/ml)	Fermentas	/
--------------------	-----------	---

2.1.6 Gel and running buffer

Gel	Recipe (for 1 gel)	Running buffer
10 % DNA-PAGE (5 mL/gel)	1000 µL 5xTBE, 1250 µL 40 % acrylamide, 425 µL 2 % Bis-acrylamide, 50 µL 10 % APS, 5 µL TEMED, 2270 µL water	1x TBE
18 % SDS-PAGE resolving gel (3 mL/gel)	750 µL 1.5 M Tris-HCl pH 8.8, 1350 µL 40 % acrylamide, 450 µL 2 % Bis-acrylamide, 30 µL 10 % SDS, 30 µL 10 % APS, 3 µL TEMED, 387 µL water	1x SDS-PAGE running buffer
5 % SDS-PAGE stacking gel (2 mL/gel)	500 µL 0.5 M Tris-HCl pH 6.8, 250 µL 40 % acrylamide, 250 µL 2 % Bis-acrylamide, 20 µL 10 % SDS, 20 µL 10 % APS, 2 µL TEMED, 958 µL water	
6 % EMSA (5 mL/gel)	250 µL 5xTBE, 750 µL 40 % acrylamide, 375 µL 2 % Bis-acrylamide, 50 µL 10 % APS, 5 µL TEMED, 3570 µL water	0.25x TBE

2.1.7 Gel ladders

Ladder	Supplier
PageRuler Plus Prestained protein ladder	Thermo Scientific
Gene Ruler 1 Kb DNA ladder	Thermo Scientific
Gene Ruler 100 bp DNA ladder	Thermo Scientific
Gene Ruler 100 bp plus DNA ladder	Thermo Scientific

2.1.8 Purification columns

Column/resin	Supplier	Chromatography	Technical specifications	Usage
--------------	----------	----------------	--------------------------	-------

Sephacryl SF-1000	GE	Gel filtration chromatography	Bed volume: 750 mL Pressure Max.: 1.5 bar	12-202-601 DNA template purification
Sephacryl S-200	GE	Gel filtration chromatography	Bed volume: 320 mL Bed dimensions: 26x 600 mm Pressure Max.: 1.5 bar	Core histone purification
Superdex 200 pg	GE	Gel filtration chromatography	Bed volume: 120 mL Bed dimensions: 16x 600 mm Pressure Max.: 3 bar	Histone octamer purification
Resource S	GE	Cation exchange chromatography	Strong cation exchanger Bed volume: 6 mL Bed dimensions: 16x 30 mm Pressure Max.: 6 bar Ligand: sulphonate group	Core histone, linker histone H1 ⁰ purification
SP Sepharose FF	GE	Cation exchange chromatography	Strong cation exchanger Bed volume: 25 mL Pressure Max. 1.5 bar Ligand: sulfopropyl	Linker histone H1 ⁰ purification
HisPur Ni-NTA spin column	Thermo Scientific	Nickel Affinity chromatography	Resin bed volume: 3 mL Binding capacity: ≤60 mg of a 28 kDa 6xHis-tagged protein Ligand: nickel-charged nitrilotriacetic acid (NTA)	Linker histone H1 ⁰ purification

2.2 Methods

2.2.1 Preparation of 601-DNA

The procedures of DNA preparation described below referred to published work (Dorigo et al., 2003). The tandem DNA array was constructed from the Widom 601 DNA sequence which is selected from the SELEX experiment. Widom 601 DNA sequence binds to the histone octamer with highest affinity than any natural or non-natural DNA sequences. Furthermore, the Widom 601 sequence has a strong nucleosome positioning ability, producing nucleosome arrays with very high regularity (Lowary and Widom, 1998).

Transformation

The construct of 12-177-601 DNA template is a gift from Prof. Timothy J. Richmond, Department of Biology, ETH Zürich. 100 ng of plasmid vector carrying the 12-177-601 sequence was added into 50 µL Top10 *E. coli* competent cells (pre-thawed on ice), gently mixed and incubated on ice for 30 min. Heat-shock was then performed at 42°C water bath for exactly 45s, followed by immediate chilling on ice for 3 min. With 700 µL of 2x TY medium added into the transformation tube, cells were incubated in the shaking incubator set at 37°C and 225 rpm for 1 h. 20 µL and 120 µL of the cells were each plated on a TYE plate containing 100 µg/mL ampicillin. The plates were incubated at 37°C overnight.

***E. coli* glycerol stock**

3 large and isolated colonies were picked from the TYE plate and each to inoculate 30 mL of 2x TY medium containing 100 µg/mL ampicillin. Cells were cultured in 37°C shaking (225 rpm) incubator for 6-7 h or until very turbid. Cells were pelleted by 4,000 g centrifugation at 4 °C for 5 min. Excess medium was removed leaving only 5 mL to re-suspend cell pellet. 100 % sterile glycerol was added to the cells with final glycerol concentration at 20-30 %. The glycerol stock was distributed into 500 µL aliquots and stored in -80 °C freezer.

Cell culture

100 mL of 2x TY medium containing 100 µg/mL ampicillin was inoculated with one aliquot of glycerol stock and incubated at 37°C with shaking (225 rpm) for 4-5 h. The starter culture was

distributed equally to 12 flasks each with 500 mL of TB medium (containing 100 µg/mL ampicillin). Cells were amplified in the 37°C incubator shaking at 225 rpm for at least 18 h.

Plasmid isolation

Cells were harvested equally into six 500 mL centrifuge bottles (Beckman Coulter) by centrifugation at 7000 rpm, 4 °C for 7 min each time. The cells were processed following the alkaline lysis method. 60 mL of lysis buffer I was added into each centrifuge bottle to re-suspend and homogenize cell pellets. 120 mL of lysis buffer II was then added into each bottle and mixed by vigorous shaking for 2 min. The cell lysate was incubated on ice for a total of 20 min, during which vigorous shaking was repeated every 5 min. 210 mL of lysis buffer III was added into each bottle and mixed immediately and gently. The cell lysate was incubated on ice for a total of 20 min, with gentle mixing every 5 min. The cell debris was pelleted by centrifugation at 10,000 g, 4 °C for 20 min. The supernatant containing the desired plasmids was filtered through an 8-fold sterilized gauze. Into the supernatant, 0.52 volume of isopropanol was added. The mixture was incubated for at room temperature for 30 min, followed by centrifugation at 16,000 g, 20°C for 30 min. The pellets which contained the desired plasmids were air-dried and distributed equally into two 50 mL Falcon tubes. After being dissolved each in 20 mL TE (10, 50) buffer, 120 µL RNase A (10 mg/ml) was added into each tube and incubated at 37°C with shaking (225 rpm) overnight.

Plasmid purification

The insoluble matter was separated by centrifugation at 10,000 g, 20°C for 10 min. The supernatant was transferred to two TEFLON tube (20 mL in each tube). 10 mL phenol was added into each 20 mL supernatant and mixed very well by 1 min vortex, followed by centrifugation at 27,000 g, 20°C for 20 min. The aqueous upper part was carefully transferred to a new TEFLON tube. The phenol extraction was repeated for two more times or until the interface was clear. Residual phenol was extracted by adding 10 mL CIA into each 20 mL sample followed by centrifugation at 12,000 g, 20°C for 10 min.

The aqueous upper part was transferred into a PPCO tube, into which 0.2 volume of 4 M NaCl and 0.4 volume of 40 % PEG-6000 were added to achieve 0.5 M NaCl and 10 % PEG-6000 in the

final mixture. The sample was mixed for 5 min in 37°C shaker (225 rpm) and seated on ice for 30 min, followed by centrifugation at 27,000 g, 4 °C for 10 min. The supernatant was discarded. The pellet was air-dried and dissolved by a total of 10 mL TE (10, 0.1). Residual PEG in the sample was extracted twice by 10 mL CIA each time and centrifugation at 10,000 g, 20°C for 10 min.

The aqueous phase was equally transferred into two PPCO tubes. Into each tube, 2.5 volumes of 100 % cold ethanol and 0.1 volume of 3 M sodium acetate were added and incubated on ice for 30 min, followed by centrifugation at 27,000 g, 4 °C for 20 min. The supernatant was discarded. The pellet was air-dried and dissolved in a total of 10 mL TE (10, 0.1).

Enzymatic digestion

The 12-177-601 DNA was released from the plasmid vector (pWM530) by adding 400U of EcoRV per 1mg DNA. There are multiply EcoRV restriction sites on the pWM530 vector backbone, so the vector would be chopped into smaller fragments with 300-500 bp size range. DNA concentration in the digestion system was adjusted to 2 mg/ml. The reaction was incubated overnight at 37°C with shaking (225 rpm). To check completeness of the digestion, sample was loaded onto a 10 % DNA-PAGE gel with 1xTBE buffer at constant 160 V for 50 min.

2.2.1.1 Purification of the 12mer tandem array of 601-DNA

PEG fractionation

PEG (Polyethylene glycol) fractionation is a method to size-selectively precipitate larger DNA fragments, thus separate them from the smaller ones. The size range of precipitated DNA fragments can be shifted by adjusting the PEG concentration.

Into the digested plasmid, 40 % PEG-6000 and 4 M NaCl were added to achieve starting concentration of 5 % PEG and 0.5 M NaCl in the mixture. A solution of 35 % PEG-6000 and 0.5 M NaCl was used to titrate the sample until it became slightly turbid. The sample was incubated on ice for 30 min, followed by centrifugation at 10,000 g, 4 °C for 30 min. Supernatant was immediately separated from the pellet right after the centrifugation. The titration and centrifugation was repeated until no more precipitation was observed. The pellets were

dissolved in TE (10, 0.1) and together with the supernatant analyzed by gel electrophoresis. If the pellet fractions were pure enough for downstream application, the purification was completed. Otherwise, size exclusion chromatography was carried out for further purification.

Size exclusion chromatography

The gel filtration column (90*2.5 cm) was packed with Sephacryl S-1000 Superfine resin following the GE Healthcare packing procedure. Before loading DNA sample, the column was equilibrated with 1.5-2 column volume (750 mL) of TES-9 buffer. Around 30 mg of DNA sample with volume no more than 10 mL was loaded to the column and eluted by TES-9 buffer with flow rate at 2 mL/min. The peak fractions were analyzed by gel electrophoresis. Fractions with pure 601 DNA were pooled for concentration and stored at -20 °C.

2.2.1.2 Preparation of mono-repeat of 177bp 601-DNA

The single repeat of 177 bp Widom 601 DNA was obtained from the restriction enzymatic digestion of the 12mer tandem repeats of 601-177 DNA by Scal (Takahashi et al., 1985), followed by phenol extraction to remove the enzyme and ethanol precipitation to separate out extra salts.

The aforementioned 12mer 177 bp tandem array of 601-DNA was constructed in such a way that each neighboring 177 bp repeat was interconnected by Scal restriction site (Figure 2.1). The mono-repeat of 177 bp 601-DNA was prepared simply by Scal digestion followed by phenol purification. 400U of Scal-HF was added per 1 mg of 12-177 DNA. Digestion was incubated at 37°C with shaking (225 rpm) for 24 h. The 12-177 DNA should be completely cut into solely monomers. The completeness could be checked by 10 % PAGE with electrophoresis setting as previously mentioned.

Scal recognition sites:

Figure 2.1 Scal restriction endonuclease recognition sequence. The arrow heads indicate the cutting sites of the enzyme. The digestion cleavage produces DNA fragments with

blunt ends. Figure from New England Biolabs website (<http://www.neb.sg/products/r3122-scai-hf>).

Once the digestion was confirmed to be complete, one volume of phenol/chloroform/isoamyl alcohol (24:24:1) was added into the reaction and mixed vigorously by vortex until the emulsion was formed. The sample was then centrifuged at 14,000 g for 5 mins at room temperature which would generate two separate layers- the upper aqueous phase with enriched DNA and the lower organic phase, while the enzyme was denatured and trapped in between. The aqueous phase was transferred to a fresh tube and extracted by phenol/CIA mixture one more time or until no protein is visible at the interface. The aqueous phase was then mixed by one volume of CIA followed by the same centrifugation setting to extract any trace of phenol.

The aqueous phase was then mixed with 0.1 volume of sodium acetate (3 M, pH 5.2) and 2.8 volume of cold absolute ethanol. The mixture was incubated on ice for 30 mins, followed by centrifugation at 15,000 g for 30 mins at 4 °C. After removing the supernatant, the DNA pellet was rinsed with 70 % ethanol and centrifuged again for 15 mins. The DNA pellet was air dried and dissolved in TE (10, 0.1) buffer. The DNA concentration was determined based on the absorbance at 260nm by NanoDrop spectrophotometer. The ratio of A_{260}/A_{280} was used to examine any residual protein contaminants. The sample was stored at -20 °C.

2.2.2 Preparation of core histone H2A, H2B, H3 and H4

Preparation procedures of the *Xenopus laevis* and *Homo sapiens* recombinant core histones described below referred to published work (Luger et al., 1999).

Transformation

100 ng of plasmid vector carrying one type of the core histones was transformed into 50 µL of *E. coli* BL21(DE3)pLysS *E. coli* competent cells (pre-thawed on ice), following the aforementioned transformation details. The transformed cells plated on TYE plates containing 100 µg/mL ampicillin and 25 µg/mL chloramphenicol to select for both pET-3a and pLysS. The plates were incubated overnight at 37°C.

Histone expression

50 mL 2x TY medium containing 100 µg/mL ampicillin and 25 µg/mL chloramphenicol was inoculated with one colony from the plate and cultured at 37°C with shaking (225 rpm) for 4-5 h or until turbid. The starter culture was distributed equally to 15 flasks each with 500 mL of 2x TY (containing 100 µg/mL ampicillin and 25 µg/mL chloramphenicol). Cells were cultured at 37°C incubator shaking at 225 rpm until the OD₆₀₀ reached 0.6. IPTG was added to a final concentration of 0.4 mM to induce the histone expression. Cells were harvested 3 h after the induction by centrifugation at 7,000 rpm, 20°C for 7 min.

Separation of inclusion bodies

Cell pellets were re-suspended in a total of 200 mL wash buffer. Cells were broken down by sonication with 25 % amplitude, 1 sec on and 1 sec off for 1 h or until homogeneous. To avoid over-heating, cell sample was kept on ice throughout sonication. The inclusion bodies containing desired histone were harvested by centrifugation at 18,000 g, 4 °C for 15 min. Pellets were washed three times in 150 mL washing buffer (first two times with 1 % Triton X-100), followed by centrifugation at 18,000 g, 4 °C for 10 min after each wash.

Core histone purification

Size exclusion chromatography

1 mL of DMSO was added to homogenize the inclusion bodies. 40 mL of S-200 unfolding buffer was then added to solubilize histones as much as possible. Insoluble matter was removed by triple centrifugation each time at 18,000 g, 20°C for 30 min. Hiprep 26/60 Sephacryl S-200 HR column was washed with de-gassed water and equilibrated with SAUDE 1000 buffer. Sample with no more than 20 mL was loaded to the column and eluted by SAUDE 1000 buffer. Peak fractions were analyzed by 18 % SDS-PAGE. Fractions with relatively pure histone were pooled and dialyzed overnight against 4L of 5 mM β-mercaptoethanol. The histone sample was lyophilized and stored at -20 °C before being further purified.

Cation exchange chromatography

The dry histone was dissolved in SAUDE 200 buffer and centrifuged at 10,000 g for 10 min to remove any insoluble matter. Resource S column was pre-equilibrated with SAUDE 200 buffer.

Sample with no more than 50 mg of histone was loaded to the column and eluted by SAUDE 1000. Peak fractions were analyzed by 18 % SDS-PAGE. Pure histone fractions were pooled and dialyzed overnight against 4L of 5 mM β -mercaptoethanol. The pure histone sample was lyophilized and stored at -80 °C.

The four types of histones from *X. laevis* and *H. sapien* were each purified following the procedures described as above.

Preparation of H4K20Me3

The histone H4K20Me3 was prepared by our collaborator Dr. Yang Renliang.

The recombinant histone H4 with site-specific mutation K20C was over-expressed and purified following the above protocol. The site-specific installation of the tri-methyl group followed the published approach (Simon et al., 2007) which allows rapid production of site- and degree-specific methyl-lysine installation into recombinant histones. Detailed procedure is as below: 40 mg of H4K20C protein was dissolved in 4.9 mL of alkylation buffer (6 M Gdn·HCl, 1 M HEPES, 10 mM D/L methionine, pH 7.8) and mixed with 100 μ L of 1 M DTT. After 10 min incubation, the alkylation reaction was initiated by adding 300 mg of (2-bromoethyl) trimethylammonium bromide into the solution. The alkylation reaction was carried out at 50°C for 5 h, followed by adding 50 μ L of 1 M DTT and 100 mg of (2-bromoethyl) trimethylammonium bromide. The reaction was continued for another 3 h at 50°C. The reaction mixture was dialyzed against water (containing 0.1 % TFA) overnight. The end product was obtained after lyophilization. The molecular weight of the end product was measured by MALDI-TOF MS.

2.2.3 Histone octamer refolding and purification

Procedures described below referred to published work (Rogge, 2013).

Histones H2A, H2B, H3, and H4 were first dissolved separately in the unfolding buffer to concentration at around 2-3 mg/ml. Accurate concentration was calculated based on the absorbance at 276 nm and extinction coefficient of each type of histone. H2A, H2B, H3 and H4 were mixed with equimolar ratio of 1:1:1:1. Final protein concentration of the histone mixture was adjusted by unfolding buffer to 1-2 mg/ml. Refolding of histone octamers was achieved by

triple dialysis against 1L of refolding buffer for 5 hrs each time and preferentially in cold room to avoid aggregations. Insoluble matter was removed by centrifugation at 18,000 g for 10 min. The supernatant with the desired histone octamers was concentrated by the Amicon Ultra centrifugal filter units with 3,000 g to a final concentration at around 10 mg/ml. The sample was then ready for purification by size exclusion chromatography. The Hiload 16/600 Superdex 200 pg gel filtration column was equilibrated with Refolding buffer. For each run, HO sample with no more than 5 mL was loaded to the column and eluted by Refolding buffer. Peak fractions were analyzed on 18 % SDS-PAGE gel. Pure histone octamer fractions were pooled and concentrated by Amicon Ultra centrifugal filter unit to around 10 mg/ml. For storage, equal volume of 100 % glycerol was added to the concentrated HO to give a 50 % glycerol concentration in the HO sample. Concentration of HO was determined based on the absorbance at 276nm and extinction coefficient. HO sample was stored at -20 °C.

Preparation of all types of histone octamers (*X. laevis* wild type and tri-methylation on H4K20, *H. sapien* wild type) used in this thesis followed the same protocol as described above. *X. laevis* wild type HO and HO with tri-methylation on H4K20 were used in the reconstruction of 12-177 wild type and H4K20 tri-methylated nucleosomal array, respectively. *H. sapien* wild type HO was used in the reconstruction of 177bp nucleosome and chromatosome.

2.2.4 Preparation of linker histone H1⁰

Transformation

The *Homo sapien* H1⁰ construct histone is a generous gift from Dr. Curt A. Davey, School of Biological Sciences, NTU. The plasmid was transformed into *E. coli* BL21(DE3)pLysS competent cells. The cells were plated on TYE plates containing 100 µg/mL ampicillin and 25 µg/mL chloramphenicol and incubated at 37°C overnight.

Large-scale expression

The starter culture of 50 mL 2x TY medium containing 100 µg/mL ampicillin and 25 µg/mL chloramphenicol was inoculated with three large colonies from the transformation plate and grown at 37°C with shaking (225 rpm) for 4-5 h or until turbid. The starter culture was then

distributed equally to 8 flasks of 500 mL 2x TY medium containing 100 µg/mL ampicillin, 25 µg/mL chloramphenicol and 0.1 % glucose. Cells were cultured at 37°C with shaking (225 rpm) until OD₂₆₀ reached 0.6. IPTG was added to a final concentration of 0.4mM to induce the expression of linker histone. Cells were harvested 2 h after the induction by centrifugation at 7,000 g, 4 °C for 7 min. Cell pellets were stored at -20 °C.

Linker histone purification

The purification strategy was adapted from Thomas Schalch's thesis (Schalch, 2004).

Cell lysis

Cell pellets were thawed on ice and re-suspended in S-buffer A. Cell lysis was done by sonication with 1s on and 1s off at 25 % amplitude for a total of 1 h. The homogenized cell lysate was double centrifuged at 35,000 g, 4 °C, for 30 min. After the second centrifugation, the supernatant which contained desired linker histone was separated from the pellet and used for downstream purification.

Two-step cation exchange chromatography

The SP Sepharose FF column was equilibrated with S-buffer A. Supernatant from the cell lysis step was loaded to the column, followed by S-buffer A washing until the junk peak completely came out and the UV absorbance dropped to the baseline. Linker histone was eluted by salt gradient increasing from S-buffer A to 100 % S-buffer B. Peak fractions were analyzed by 18 % SDS-PAGE gel electrophoresis. Fractions with relatively pure H1 were pooled and dialyzed twice against 2L of S-buffer A at 4 °C, each time for at least 6 h. After being recovered from the dialysis, the sample was loaded to Resource S column for a second cation exchange chromatography to further improve the purity. The column was washed with S-buffer A until the junk peak completely came out and the UV absorbance dropped to the baseline. In the elution step, salt concentration was gradually raised to 100 % S-buffer B within 2 min forming a salt gradient by which the linker histone was eluted from the column. Peak fractions were analyzed by 18 % SDS-PAGE gel electrophoresis. Fractions containing pure linker histone were

pooled and dialyzed twice against phosphate buffer (pH 7.4) in the cold room. For storage, linker histone sample was flash-frozen in liquid nitrogen, lyophilized and stored at -80 °C.

2.2.5 *In vitro* reconstitution of nucleosome arrays

Procedures described below referred to published work (Huynh et al., 2005; Schalch, 2004) (Rogge, 2013).

To prepare saturated 12mer nucleosome arrays in which each 177 bp repeat was occupied with one histone octamer, a set of small scale test reconstitutions were set up with different molar ratio amount of *Xenopus* HO to 12mer DNA template at around 12:1 (e.g. 11:1, 11.5:1, 12:1, 12.5:1 and 13:1). The 177 bp DNA repeat concentration was fixed at 5µM. Calf thymus genomic DNA (ctDNA for short) with length of ~150 bp was added as DNA buffer to absorb possible excess HO, amount of ctDNA was set at 0.6 molecule per 12mer DNA template. In addition, each reconstitution mixture contained 2 M KCl and 10 mM DTT. The reconstitution size was 200 µl. The mixture was transferred to the dialysis button (Hampton Research) covered by membrane tubing with 6-8,000 MWCO which was fixed in place by a rubber ring. To ensure proper nucleosome array formation, the reconstitution went through gradual change from 2 M to lower salt conditions by stepwise dialysis against 1.4 M, 1.2 M, 1.0 M, 0.8 M, 0.6 M, 0.3 M and 10 mM TEK buffer (with 10 mM Tris pH 7.5, 0.1 mM EDTA, and 1 mM DTT, in addition to different concentrations of KCl as aforementioned) at room temperature. Dialysis in each KCl concentration took 1.5-2 h. At the end of the dialysis, each reconstitution was recovered from the button and centrifuged at 20,000 g for 10 min to remove any possible precipitate. Samples were stored at 4 °C.

The degree of saturation of the reconstituted nucleosome array was verified by both agarose gel analysis and electrophoretic mobility shift assay (EMSA for short). 20 ng of nucleosome array of each titration sample was loaded to the 0.8 % agarose running with 0.2x TBE buffer at 120 V for 1 h, followed by 0.5 mg/ml Ethidium Bromide staining and imaging.

Before the EMSA, the nucleosome array was first digested by Scal, since as aforementioned, each 177 bp repeat was tandemly connected by Scal restriction sequence. 20U of Scal-HF was used for 2µg of 12-177 nucleosome array in 10 µL reaction. Digestion was incubated at 37°C for

3 h. 200 ng of each digested sample was loaded to 6 % native PAGE, with 0.25x TBE running buffer, at constant 100 V for 1 h. The gel was stained by 0.5 µg/mL Ethidium Bromide solution for 5 mins, and visualized by the UV transilluminator (Syngene G:BOX gel documentation).

2.2.6 *In vitro* reconstitution of nucleosomes

Nucleosomes were prepared in a similar way as nucleosome arrays that all the ingredients were initially mixed at 2 M salt and went through dialysis against step-wise decreasing concentration of salt. Ideally, nucleosome could be reconstituted by simply mixing equal molar amount of histone octamer and 601-177 bp DNA. However, due to the error in the measurements of the DNA and histone octamer concentrations, unavoidable aggregations during salt dialysis and other unknown reasons, different HO to DNA molar ratios around 1:1 were screened in order to obtain optimal nucleosome sample. The *human* HO titration was done in 50 µL scale, in which DNA concentration was fixed at 6 µM (0.65 µg/µL) or higher, whereas HO amount varies corresponding to different HO/DNA molar ratios. Other components included in the reconstitution mixture were 2 M NaCl and 10 mM DTT. Unlike in nucleosome array reconstitution, no competitor DNA was added. Similar as reconstitution of nucleosome array, the nucleosome reconstitution was achieved by stepwise dialysis against lower salt buffer with 0.85 M (3 h), 0.5 M (3 h), 10 mM NaCl (twice, 5 h each time) at room temperature. Titration samples were recovered and centrifuged at 10,000 rpm for 5 mins to remove any aggregates. Titration samples were analyzed by EMSA on 6 % PAGE, with freshly made 0.25x TBE running buffer, at constant 120 V for 50 min.

2.2.7 *In vitro* reconstitution of chromatosomes

The chromatosome studied was composed of 3 parts: single repeat of 177 bp Widom 601 nucleosome positioning sequence, human histone octamer which was prepared as stated above, and full-length human linker histone H1⁰. The *in vitro* reconstitution of chromatosome was divided into two parts- reconstitution of nucleosome and addition of linker histone. The nucleosome reconstitution mixture was prepared with optimal HO/DNA relative amount defined by the *human* HO titration as described in previous section, followed by dialysis against 0.85 M and 0.6 M NaCl buffer (20 mM HEPES pH 8.0, 1 mM DTT, 1 mM EDTA). The nucleosome

sample was recovered at this point and measured concentration based on absorbance at 260nm with NanoDrop spectrophotometer. For the 50 μ L linker histone titration scale, nucleosome sample was divided into 40 μ L aliquots. Linker histone powder was dissolved in 0.6 M dialysis buffer. A broad linker histone titration was first applied with H1 to DNA molar ratio ranging from 0.5:1 to 3:1. Linker histone with amounts referred to different H1/DNA molar ratios were added into the nucleosome aliquot. Each titration sample was topped up with 0.6 M dialysis buffer to 50 μ L and transferred to the 50 μ L dialysis button which was set up as previously mentioned. For proper binding of linker histone to nucleosome, the linker histone titration mixture went through a slow dialysis against 10 mM NaCl buffer (20 mM HEPES pH 8.0, 1 mM DTT, 1 mM EDTA). As demonstrated in Figure 2.2, dialysis buttons were bagged in the 3,000 MWCO membrane tubing containing 20-30 mL of 0.6 M buffer, which was then put in the 2L reservoir of 10 mM buffer. Dialysis took place in the cold room overnight. Samples were recovered and centrifuged at 9,000 g, 4 $^{\circ}$ C for 5 min to remove any precipitates. The supernatant was analyzed by EMSA on 5 % PAGE, the result of which was used as a guidance to set up a narrow linker histone titration.

Figure 2.2 Schematic illustration of the dialysis set-up.

2.2.8 Cryo-electron microscopy (Cryo-EM) single particle analysis

Cryo-grid preparation

Preparation of cryo-grids were done with FEI vitrobot. The process chamber was maintained at 4 °C and 100 % humidity. Quantifoil copper grids with R2/2 holey carbon films were hydrophilized by the plasma cleaning system with 1 min glow discharge. 4-5 μ L chromatosome sample was loaded onto the grid with excess solution blotted by two layers of filter paper on each side of the grid. Blot setting was optimized at 2 second blot time and 0 blot force. The grid was then plunge-frozen in liquid ethane which was maintained at liquid nitrogen temperature. Sample grid was clipped in the cartridge and transferred to the cassette and ready to be loaded onto the microscope or stored in liquid nitrogen.

EM imaging and data collection

The cryo-grids were imaged by FEI Tecnai Arctica cryo transmission electron microscope which was equipped with a 200kV field emission gun and Falcon III direct electron detection camera. The C2 condenser lens was excited to 39-40 %, with 100 μ m aperture size. Micrographs were collected at 78,000x nominal magnification in movie mode with 20-25 $e^-/\text{\AA}^2 \cdot s$ electron dose at various defocus values (-1.5, -1.75, -2.0, -2.25, -2.5, -2.75, -3.0 μ m).

Data preprocessing

Movie frames were built into stacks and went through beam-induced motion correction (Bai et al., 2013; Li et al., 2013). The initial evaluation of the micrographs was done in EMAN2 to discard those with astigmatism or with very few or very crowded particles. Estimation of the CTF parameters were done by using Niko Grigorieff's CTFFIND3 (Mindell and Grigorieff, 2003). Manual particles picking was done in EMAN2 by Steven Ludtke's E2boxer.py with box size of 168 pixels. The coordinates of selected particles were saved in the .box files which were used as input in RELION-1.4 to extract particles.

Data processing

The following data processing was all done by RELION-1.4 (Scheres, 2012). In the reference-free 2D class averaging, particles were classified into 150 classes in 25 iterations, the mask diameter was set at 210 $\text{\AA}$, full phase and amplitude CTF correction was performed inside the Bayesian framework. Particles from good classes were selected (re-grouped if necessary) and saved.

The selected particles from the 2D classification were used as input for the unsupervised 3D classification. The 167 bp-chromatosome crystal structure (Zhou et al., 2015) (PDB ID: 4QLC) low-pass filtered to 80 Å was used as the reference model. More particles were excluded to further reduce the heterogeneity of the dataset. Only particles from the good class(es) were selected and saved. The subset was refined to high resolution by the 3D auto-refine procedure in a fully automated manner. Over-fitting was avoided by applying the gold standard Fourier Shell Correlation (FSC) calculation to estimate the resolution (Bai et al., 2013; Scheres and Chen, 2012). The initial model could either be the same one used in the 3D classification or one of the refined maps from the 3D classification. After the 3D auto-refinement was done, the map was sharpened and masked in the RELION post-processing task. The gold-standard FSC curves were calculated from the masked map. Resolution was estimated from the gold-standard FSC curves.

2.2.9 Precipitation assay (PA)

In the field of chemistry, precipitation is a process to generate insoluble matters out from a solution. Here in this thesis, precipitation assay is applied to study the precipitation progress of nucleosome, chromatosome or nucleosome array, for the purpose of investigating the biophysical properties of the cationic counterions induced particle aggregation.

The nucleosome, chromatosome or 12-177 nucleosome array samples are mixed with different amount of cations (Mg^{2+} or $CoHex^{3+}$), the precipitation (or particle aggregation) is measured indirectly by recording the amount of soluble material in the solution calculated from the absorbance measured at 260 nm. Below is a detailed description of the experiment procedure.

10 µl sample in 10 mM Tris-HCl (pH 7.5) buffer with OD_{260} at 2 were mixed with 10 µl of salt (Mg^{2+} or $CoHex^{3+}$) buffer. 20 to 30 mixtures were prepared with different salt concentrations. Incubation took place at room temperature for 15 mins followed by centrifugation at 20,000 g for 15 mins. OD_{260} of the supernatant was measured and plotted against cation concentration. The plot was fitted by Boltzmann sigmoidal function:

$$y = \frac{A_1 - A_2}{1 + e^{(x-x_0)/dx}} + A_2$$

The midpoint- x_0 in the Boltzmann sigmoidal function- of the fitted curve was set as EC50 value, which denotes the salt concentration that induced 50 % sample precipitation (Figure 2.3).

Figure 2.3 Sample curve of the Boltzmann function. $(x_0, (A_1 + A_2)/2)$ is the center of the curve, A_1 and A_2 are the minimum and maximum value of variable y , when applied to the precipitation curve are the complete precipitation (OD_{260} close to 0) of the sample and the initial OD_{260} value of the sample, respectively. Figure obtained from OriginPro.

2.2.10 Small-angle X-ray Scattering (SAXS) and X-Ray diffraction

X-ray diffraction is proved to be a suitable method for the detection and characterization of the phase properties of nucleosome aggregates (Berezhnoy et al., 2016; Mangenot et al., 2003b). The 601-177 bp chromosome sample was prepared at 8 mg/ml concentration and precipitated by different concentrations of multivalent cations (Mg^{2+} and $CoHex^{3+}$). Samples were packed into transparent thin (1.5 mm in diameter) quartz capillary tubes and sealed with wax. The X-ray diffraction experiments in this work were carried out using the SAXS beamline at 23A workstation in National Synchrotron Radiation Research Center (NSRCC), Taiwan. The phase properties of the nucleosome/ chromosome aggregates were identified from the periodicity of the Bragg peaks of the diffraction spectrum.

Figure 2.4 Schematic illustration of the principle of SAXS experiment. Figure from website: http://coms-nano.jp/coms-nano3/?page_id=622

SAXS is a technique used for the determination of the microscale or nanoscale structural properties of particles in a solution system. X-ray shoots the experimental material and is scattered by the molecules. The elastic scattering carrying structural information of macromolecules is recorded at very low (small) angles (0.1-10°) (Figure 2.4).

The 177 bp nucleosome and chromatosome samples were prepared at 3 mg/ml concentration in 10 mM KCl and 10 mM Tris (pH 7.5). SAXS data collection was performed at the beamline 23A workstation in National Synchrotron Radiation Research Center (NSRCC), Taiwan.

Our collected SAXS spectra provide description of the shape and structure of the 177 bp nucleosome and chromatosome. The SAXS data were analyzed by Primus (Konarev et al., 2003) from the ATSAS software package (version 2.8.1) (Franke et al., 2017). Radius of gyration (R_g) of the 177 bp nucleosome and chromatosome were calculated from Guinier approximation (Guinier, 1939; Konarev et al., 2003) at low q value ($sR_g < 1.3$) region where the particle interactions are negligible.

$$\ln I(q) = I_0 + \left(\frac{-R_g^2}{3} \right) * q^2$$

In the above equation, $I(q)$ is the scattering intensity, I_0 is the forward scattering, q is the scattering vector.

In the history of SAXS application, R_g is the first structural parameter determined from the SAXS spectrum. R_g is defined as the mass-weighted geometric mean of the distance of each atom of the particle from the center of mass. Therefore, it provides direct information on the shape and size of the particle.

$$R_g = \sqrt{\frac{\sum m_i r_i^2}{\sum m_i}}$$

In the equation above, m_i is mass of atom i , r_i is the distance of atom i from the structure's center of mass.

2.2.11 Analytical ultracentrifugation (AUC)

Analytical ultracentrifugation is a powerful technique for quantitative analysis of molecules in solution. In this thesis, AUC was used for the investigation of folding behaviors of nucleosome array. The sedimentation coefficient of nucleosome arrays in different salt conditions were calculated from the sedimentation velocity experiments. Velocity experiments were carried out on the Beckman XL-I analytical ultracentrifuge equipped with Beckman monochromator and the eight-channel AN-50 Ti rotor. The nucleosome array sample was diluted by TEK 0.01 buffer to OD₂₅₉ of 0.6-0.8. The AUC cell was assembled with a 2-channel epon centerpiece with 12 mm path length. 400 μ L each of nucleosome array sample in different cationic salt concentrations was injected into the sample channel of the cell. 420 μ L (different volume from that of the sample for the ease of meniscus determination) of buffer solution was injected into the reference channel.

The transportation of chromatin fibers was recorded by an optical system throughout the sedimentation process. Sample absorbance at 259nm wavelength was collected. The radial scan range was defined according to the absorbance curve from the single scan run at 3,000 rpm. 40 scans of absorbance data were collected with 10 min interval at 12,000 rpm and 20°C.

Figure 2.5 Schematic illustration of the AUC double-sector centerpiece. Sample and buffer solutions are injected into the corresponding sectors. Volume of reference buffer is

slightly higher than sample volume to avoid the overlapping of the two menisci at the same radial position. Figure obtained from (Ralston, 1993).

Figure 2.6 A collection of 38 scans with 10 mins interval between each scan. The scan boundary moves towards bottom (left side of X-axis) of the cell as the sedimentation progresses. Decreasing of the plateaus is due to the sector-shaped cell channels which lead to radial dilution of sample towards the bottom of the cell.

Absorbance data were edited in UltraScan II software (Demeler, 2005) where sample meniscus and data range were defined, radially invariant noise was subtracted from the data, and absorbance plateau and baseline were determined. The edited data were analyzed by Enhanced van Holde-Weischet method (Vanholde and Weischet, 1978) with the UltraScan package II version 9.9, from which the distribution of sedimentation coefficients at different boundary positions were generated. Sedimentation coefficient is used to characterize the behavior of a particle during sedimentation process. It is defined as the ratio of sedimentation velocity to acceleration.

$$s = \frac{v_t}{a}$$

S-value describes how fast the particle is travelling towards the bottom of the centrifugal vessel. It is in the dimension of time, and one unit is defined as 10^{-13} s, called one Svedberg, named after the Swedish chemist, winner of the 1926 Nobel Prize in Chemistry, Theodor Svedberg.

The partial specific volume ($\bar{v}$) of nucleosome array is calculated based on 0.499 ml/g for Na-DNA in water (Cohen and Eisenberg, 1968) and 0.75 ml/g for histone octamer (Eisenberg and Felsenfeld, 1981). $\bar{v}$ of 12-177 nucleosome array is 0.622 ml/g.

The sedimentation coefficients were corrected to $S_{20,w}$ by using a partial specific volume of 0.622 mL/g for saturated 12-177 nucleosome array, density and viscosity specific to the buffer solution (Dorigo et al., 2003). The average of $S_{20,w}$ from 20 % to 80 % of the boundary was reported as the S-value.

Complementarily, the sedimentation velocity data were independently analyzed by Dr. Nikolay Korolev using Sedfit, which is another powerful software package with a variety of models and extensive set of rigorous statistical tools, allowing for efficient analysis of samples with discrete species and size distributions. When applied here in the study of nucleosome arrays, sedimentation coefficient distribution, $c(s)$ was calculated, from which one could get a better knowledge of the heterogeneity of the sample, which is a great advantage over other AUC software packages.

3 Results and Discussions

3.1 Large scale production and purification of 601 DNA

3.1.1 12-177-601 DNA preparation

The plasmid containing 12-177-601 DNA was amplified in *E. coli* TOP10 cells and isolated with the alkaline lysis method. Protein precipitation were removed by phenol extraction. RNA contaminants were separated away by PEG-6000 precipitation. Yield of plasmid DNA was ~20 mg per 1 L of cell culture. The purified plasmid was digested by EcoRV overnight to release the 12-177-601 DNA and also to chop the vector backbone (with multiple EcoRV sites) into smaller fragments. The sample was analyzed by 1.2 % agarose gel () at the end of the digestion.

Figure 3.1 Agarose gel analysis of the EcoRV digestion of 12-177-601 plasmid. Sizes of DNA ladder are shown on the left. Pure 12-177 DNA was loaded to the “control” lane. The 12-

177-601 DNA cut-out and the plasmid vector fragments are indicated.

The 12-177 DNA was separated from the vector fragments by PEG precipitation. PEG precipitation was carried out in several rounds. In each round, most of the 12-177-601 DNA was precipitated, while majority of the vector fragments remained soluble in the supernatant. The pellets containing the 12-177-601 DNA were re-solubilized in TE (10, 0.1) buffer and analyzed by 1.5 % agarose gel (Figure 3.2).

High purity of the 12-177 DNA was observed from the agarose gel (Figure 3.2). No further purification procedures were applied. The yield of purified 12-177 DNA was ~300 µg per 1 mg of plasmid DNA. Purified DNA sample was stored in the -20 °C freezer.

Figure 3.2 Agarose gel analysis of the pellets from PEG precipitation. Sizes of DNA ladder are shown on the left. The 12-177-601 DNA is indicated.

3.1.2 601-177 monomer DNA preparation

Instead of starting from the monomer 601-177 plasmid amplification, a short cut was taken for preparation of the 601-177 DNA- the purified 12-177-601 DNA array was digested into monomers by *ScaI* restriction enzyme. As shown in Figure 2.1, *ScaI* restriction cleavage produces DNA fragments with blunt ends, which makes the 601-177 monomer DNA ideal for downstream experiments. The 601-177 DNA was purified from the *ScaI* overnight digestion by ethanol precipitation. The purification efficiency is ~85-90 %. Purity of the 601-177 DNA was analyzed by 10 % PAGE (Figure 3.3). The pure DNA sample was stored in -20 °C freezer.

Figure 3.3 10 % DNA-PAGE analysis of the purified 601-177 DNA. The DNA ladder refers to the left panel. The position of the monomer 177-601 DNA is indicated by arrow head.

3.2 Preparation of core histones

All the four types of recombinant core histones H2A, H2B, H3 and H4, whether originated from *X. laevis* or *H. sapiens*, were expressed and purified following the same procedure. Purification results of *X. laevis* histone H2A were shown below as an example.

The purification was achieved by two steps of chromatography. The solubilized inclusion body was first applied to the Sephacryl S-200 size exclusion column. Fractions of the chromatography (Figure 3.4) were analyzed by 18 % SDS-PAGE (Figure 3.5).

Figure 3.4 Size exclusion chromatography of histone H2A. Sample injection point is marked by the pink dotted line; the UV absorbance curve at OD280 is shown in blue; the elution fractions are shown in red.

Figure 3.5 18 % SDS-PAGE analysis of fractions from the Sephacryl S-200 size exclusion chromatography. Histone octamer (HO) sample from previous prep is loaded as the marker, H2A/H2B (sizes of which are too close to get resolved on the gel) position is shown as the middle band. Lane 1-6 are fractions A1, A3, A5, A7, A9 and A11 from the chromatography as in Figure 3.4.

Fractions shown in Figure 3.5 lane 1-6 all contained histone H2A and with very little amount of contaminants. Resource S cation exchange chromatography was applied for further polishing, the purification profile shown as Figure 3.6.

Figure 3.6 Resource S cation exchange chromatography of histone H2A purification. Sample injection point is marked by the pink dotted line; the UV absorbance curve at OD280 is shown in blue; relative percentage of the elution buffer is shown in lime; buffer conductivity in brown; the elution fractions are shown in red. The H2A elution peak is indicated.

Purity of the fractions of the chromatography were analyzed by 18 % SDS-PAGE (Figure 3.7).

The histone yield was ~30 mg per 1 L of cell culture.

Figure 3.7 18 % SDS-PAGE analysis of the H2A peak fraction (Figure 3.6) from the Resource S cation exchange chromatography. Histone octamer (HO) sample from previous prep is loaded as the marker, H2A/ H2B (sizes of which are too close to get resolved on the gel) position is shown as the middle band.

3.3 Histone octamer preparation

3.3.1 Purification of *Xenopus* wild-type histone octamer

The equimolar amount of H2A, H2B, H3, and H4 were mixed and dialyzed against the refolding buffer. The mixture was then concentrated and purified by the size exclusion chromatography with the Hiload 16/600 superdex 200pg column (Figure 3.8).

The collected fractions were analyzed by 18 % SDS-PAGE, as shown in Figure 3.9.

Figure 3.8 Size exclusion chromatography profile of the *Xenopus* HO. The UV absorbance curve at OD280 is shown in blue; the shadow under the peaks are collected fractions; the retention volume is indicated on top of each peak.

Figure 3.9 18 % SDS-PAGE analysis of the fractions of the size exclusion chromatography. The first lane on the left is protein marker, size of which refers to the left panel; lane 2-9 are fractions labeled with numbers corresponding to the fractions in Figure 3.8.

Fractions 13-17 from the second peak (Figure 3.8) were regarded as pure *Xenopus* HO.

3.3.2 Purification of *Xenopus* HO-H4K20Me3

Xenopus histone octamer with tri-methylation on histone H4 lysine 20 (HO-H4K20Me3) was re-folded and purified following the same protocol as for wild-type HO. Figure 3.10 shows the profile of the size exclusion chromatography.

Figure 3.10 Size exclusion chromatography profile of the *Xenopus* HO-H4K20Me3. The UV absorbance curve at OD280 is shown in blue; the elution peaks were labelled as P1, P2, P3 and P4, respectively.

Figure 3.11 18 % SDS-PAGE analysis of the fractions of the size exclusion chromatography. The first lane on the left is protein marker, size of which refers to the left panel; lane 2-5 are fractions labeled with numbers corresponding to the fractions in Figure 3.10.

The collected fractions were analyzed by 18 % SDS-PAGE, as shown in Figure 3.11. Fraction 4 from the second peak (Figure 3.10, P2) were regarded as pure *Xenopus* HO-H4K20Me3.

3.3.3 Purification of *human* HO

Human wild-type histone octamer was re-folded and purified following the same protocol as for *Xenopus* HO. Figure 3.12 shows the profile of the size exclusion chromatography.

Figure 3.12 Size exclusion chromatography profile of the *human* HO. Sample injection point is marked by the pink dotted line; the UV absorbance curve at OD280 is shown in cyan; elution fractions are shown in red.

Figure 3.13 18 % SDS-PAGE analysis of the fractions of the size exclusion chromatography. The first lane on the left is protein marker, size of which refers to the left panel; lane 2-7 are fractions labeled with numbers corresponding to the fraction numbers in Figure 3.12.

Fractions were collected and analyzed by 18 % SDS-PAGE, as shown in Figure 3.13. Fraction 4 and 5 from the second peak (Figure 3.12, P2) were regarded as pure *human* HO.

Generally, for each of the three types of HO, the yield was 50-60 % of the total mass of the four histones.

3.4 Linker histone H1⁰ large scale expression and purification

The linker histone was over-expressed in *E. coli* BL21(DE3)pLysS cells, harvested and pelleted, followed by sonication. Purification of linker histone was carried out by two-step cation exchange chromatography. Both SP Sepharose and Resource S are strong cation exchanger. SP Sepharose is stronger than Resource S and was applied in the first step of purification to remove majority of the contaminants, meanwhile avoid losing H1⁰ protein (as shown in Figure 3.15, H1⁰ is completely absent in the wash fraction X1), whereas the expense is some contaminants with similar pI cannot be effectively separated from H1⁰. In the second step of purification, the relatively weak Resource S column was applied to get pure H1⁰, even at the cost of losing some of it (as shown in Figure 3.17, the wash fraction X1 contains small amount of H1⁰).

Figure 3.14 SP Sepharose cation exchange chromatography profile of linker histone H1⁰. Two complete runs are shown. The two sample injection points are marked by pink lines; the UV absorbance curve at OD₂₈₀ is shown in blue; relative percentage of the elution buffer is shown in lime; buffer conductivity in brown; the elution fractions are shown in red. The H1⁰ elution peaks are indicated.

A very high UV peak was observed in the washing step (Figure 3.14). The $H1^0$ peak came out during the elution step. Fractions from both peaks were analyzed on 18 % SDS-PAGE gel (Figure 3.15).

Figure 3.15 18 % SDS-PAGE analysis of the fractions from the SP Sepharose cation exchange chromatography of $H1^0$. The first lane on the right is protein marker, size of which refers to the right panel; “LS” is the total loaded sample. X1 is fraction of the washing step which came out shortly after sample injection. A1 and A1’ are salt buffer elution fractions from the first and second run, respectively.

The gel showed that even though the peak that came out at the washing step was large, it did not contain any $H1^0$ protein at all. The $H1^0$ peak fractions were pooled and further purified by cation exchange chromatography with the Resource S column (Figure 3.16).

Figure 3.16 Resource S chromatography profile of linker histone $H1^0$. Two complete runs are shown. The two sample injection points are marked by pink lines; the UV absorbance

curve at OD₂₈₀ is shown in blue; relative percentage of the elution buffer is shown in lime; buffer conductivity in brown; the elution fractions are shown in red. The H1⁰ elution peaks are indicated.

Figure 3.17 18 % SDS-PAGE analysis of fractions from the Resource S cation exchange chromatography of H1⁰. The first lane on the right is protein marker, size of which refers to the right panel; “LS” is the total loaded sample. X1 is fraction of the washing step which came out shortly after sample injection. A1 and A1’ are salt buffer elution fractions from the first and second run, respectively.

Fractions of both the washing and elution steps were analyzed by 18 % SDS-PAGE (Figure 3.17). The yield of histone H1⁰ was ~5 mg per 1 L of cell culture.

3.5 Nucleosome/ chromatosome reconstitution

3.5.1 177 bp-nucleosome reconstitution

Multiple reconstitutions of 177 bp nucleosome with different HO to 177-601 monomer DNA molar ratios were set up in parallel. The reconstitution mixture went through gradual buffer exchange from high salt to low/zero salt condition. At the end of dialysis, the nucleosome samples were recovered from the dialysis vessel and analyzed by EMSA with 6 % native PAGE ().

Figure 3.18).

Figure 3.18 EMSA analysis of the 177 bp-nucleosome samples on 6 % native PAGE. From left to right, lane 1 is naked 177 bp DNA as a marker; lane 2 is DNA ladder, size of which is indicated as in the left panel; lane 3 to 9 are the reconstituted 177 bp-nucleosome samples with apparent HO to 177 bp DNA molar ratios from 0.4:1 to 1.4:1 (0.4, 0.8, 0.9, 1.0, 1.1, 1.2, 1.4, respectively). From top to bottom, as labeled to the right of the gel picture, are 177 bp nucleosome composed of one molecule of 177 bp DNA and one HO; incomplete nucleosome which is formed from 177 bp DNA and incomplete HO, and naked 177 bp DNA.

The gel ().

Figure 3.18) demonstrates that different relative amounts of HOs lead to the reconstitution with different compositions. At lower HO amount, as shown in lanes 3 to 5, besides 177 bp-nucleosome, free 177 bp DNA and incomplete nucleosome are present, indicating HO deficiency. With increase of HO amount to and above the 177 bp DNA saturation level, all the 177 bp DNA is bound with HO, indicated by the absence of 177 bp DNA band on the gel (lane 6 to 9). However, the similar quality of NCP band shown in lane 6 to 9 have dramatic difference in nucleosome concentration: as HO amount exceeds that sufficient for 177 bp DNA saturation, the 177 bp-nucleosome aggregation happens. The aggregates could be observed as white precipitates upon sample recovery at the end of dialysis. There is a large drop in nucleosome concentration for samples loaded in lane 8 and 9. From gel analyses, it could be concluded that the optimal HO to 177 bp DNA molar ratio is observed around lane 6 and 7. In summary, the optimal 177 bp nucleosome sample is determined from the EMSA results and the nucleosome solubility. The yield of 177 bp nucleosome reconstitution was ~50-60 %.

3.5.2 177 bp-chromatosome reconstitution

Multiple reconstitutions of 177 bp chromatosome with different linker histone H1⁰ to 177-601 DNA molar ratios were set up in parallel. The reconstitution mixture went through gradual buffer exchange from high salt to low/zero salt condition. At the end of dialysis, the chromatosome samples were recovered from the dialysis vessel and analyzed by EMSA with 5 % native PAGE (Figure 3.19).

Figure 3.19 EMSA analysis of the 177 bp chromatosome samples on 5 % native PAGE. From left to right, lane 1 and 2 are naked 177 bp DNA and 177 bp nucleosome, respectively, loaded as markers; lane 3 to 7 are the reconstituted 177 bp-chromatosome samples with apparent H1⁰ to 177 bp DNA molar ratios from 0.4:1 to 1.5:1 (0.4, 0.8, 1.2, 1.4, 1.5, respectively). From top to bottom, as labeled to the right of the gel picture, are 177 bp chromatosome compose of one molecule of 177 bp DNA, one HO, and one molecule of linker histone H1⁰; 177 bp nucleosome without linker histone bound with, and naked 177 bp DNA.

The intensity of 177 bp-chromatosome and -nucleosome bands in Figure 3.19 displays different relative amounts of these two species. At the lowest apparent H1⁰ to 177 bp DNA molar ratio tested (lane 3, 0.4:1), the chromatosome band is faint compared with the nucleosome band. As linker histone amount increases, more nucleosomes are bound with the linker histone, the chromatosome band intensity gets dramatically higher, meanwhile the nucleosome band fades away and completely disappeared as shown in lane 5 to 7, which means the linker histone amount has reached (or exceed) the required amount for nucleosome saturation. However, excess linker histones would induce aggregation resulting in low amount of the chromatosome, as observed for sample loaded to lane 7. An optimal 177 bp chromatosome is determined by

both the homogeneity of the chromatosome and its solubility. The optimal apparent linker histone to 177 bp DNA molar ratio is observed around lane 5 and 6 (corresponding to 1.2:1 and 1.4:1 molar ratio of H1⁰ to 177 bp DNA). The 177 bp chromatosome reconstitution yield was ~40-50%.

3.6 Nucleosome array reconstitution

3.6.1 12-177-601 nucleosome array reconstitution

Multiple reconstitutions of 12-177 nucleosome arrays with different HO (wtHO or HO with H4K20 trimethylation) to 12-177-601 DNA molar ratios were set up in parallel. The reconstitution mixture went through gradual buffer exchange from high salt to low/zero salt condition. At the end of dialysis, the nucleosome array samples were recovered from the dialysis vessel and analyzed by 0.8 % agarose (Figure 3.20 and Figure 3.22) and ScaI restriction digestion plus the EMSA (electrophoretic mobility shift assay) with 6 % native PAGE (Figure 3.21 and Figure 3.24). The yield of reconstituted 12-177 nucleosome array was ~70%.

12-177-601 nucleosome array with wtHO

The bands highlighted in the yellow rectangle (in Figure 3.20) from lane 3 to lane 8 represent nucleosome arrays with different saturation amount of HO. As the relative amount of wtHO increases, more wtHO binding positions were occupied, the array became heavier and at the same time more negative charges of the DNA were neutralized by the positively-charged wtHOs, as a result of which, migration of the nucleoprotein complex became retarded, as shown on the gel (Figure 3.20) the shortening migration distance from lane 3 to 8. As the 12-177-601 DNA approaching saturation, the observed amount of competitor DNA decreased and eventually almost disappeared as shown in lane 8. The “loss” of competitor DNA was due to the excess input of wtHO, with which the competitor DNA reconstituted forming nucleosome, which could be visualized on the gel from lane 5 to 8 (squared by blue rectangle). Synchronized with the formation of nucleosome was the appearance of nucleosome array dimer, the formation of which was suggested to be induced by the excess wtHO as well. Empirically, an ideal nucleosome array sample saturated with wtHOs was the one somewhere close to that in

lane 5 or 6 (corresponding apparent molar ratio of wtHO to 12-177 DNA is 12:1 and 14.4:1, respectively), where all the 12 nucleosome positioning 601 sequences have been loaded with wtHOs and those slightly excess wtHOs were absorbed by competitor DNA.

Figure 3.20 0.8 % agarose gel analysis of nucleosome arrays with different wtHO to 12-177-601 DNA molar ratios. From left to right, lane 1 is DNA ladder, size of which is indicated as in the left panel; lane 2 is loaded with a mixture of 12-177-601 DNA and the calf thymus competitor DNA as control; lane 3 to 8 are the reconstituted nucleosome arrays with apparent molar ratio of wtHO to 12-177-601 DNA as 6, 9.6, 12, 14.4, 16.8, 19.2:1, respectively. From top to bottom are nucleosome array dimers (in pink rectangle) which are induced by the excess wtHO, 12-177 nucleosome array (in yellow rectangle), 12-177-601 DNA, nucleosomes (in blue rectangle) that are formed from competitor DNA and excess wtHO, and competitor DNA.

Figure 3.21 Analysis of the Scal digestion product of the 12-177-wtHO nucleosome array samples on 6 % native PAGE. From left to right, lane 1 is DNA ladder, size of which is indicated as in the left panel; lane 2 is loaded with 177-601 monomer DNA as an additional marker; lane 3 to 6 are Scal-digested nucleosome arrays with apparent wtHO to 12-177-

601 DNA molar ratios from 12.25:1 to 13:1. From top to bottom are incomplete- or undigested nucleosome array trapped in the wells, nucleosome dimer, 177 bp-nucleosome monomer, and 177 bp naked DNA.

The Scal digestion assay was applied, as an alternative way to verify the saturation state of the reconstituted 12-177-wtHO nucleosome arrays. The digested samples were analyzed by the 6 % native PAGE EMSA (Figure 3.21).

Similar to the agarose gel electrophoresis (Figure 3.20), Figure 3.21 demonstrated nucleosome arrays with different saturation amount of wtHO. The cutting sites of Scal are at either side of the nucleosome positioning sequences, digestion of saturated nucleosome arrays would yield 177 bp-nucleosome solely, whereas the under-saturated ones would produce both 177 bp-nucleosome and naked 177 bp DNA, and for the over-saturated ones, the digestion would be incomplete since some Scal cutting sites were blocked by excess wtHOs. Besides, the formation of nucleosome array dimers would also impede Scal digestion. In the EMSA gel (Figure 3.21), the observation of a single 177 bp-nucleosome band from lane 3 and 4 indicated optimal saturation of the 12-177-wtHO nucleosome arrays; whereas the “nucleosome array incomplete digest” and “nucleosome dimer” bands from lane 5 and 6 suggested that nucleosome arrays were over-loaded with wtHOs.

12-177-601 nucleosome array with trimethylation on H4K20

Selection of the optimal reconstituted 12-177-HO_H4K20M3 samples followed the same criteria as for 12-177-wtHO elaborated above. The agarose gel (Figure 3.22) suggested that the ideal nucleosome array sample saturated with HO_H4K20Me3 was the one somewhere close to that in lane 5 or 6 (corresponding apparent molar ratios of HO to 12-177 DNA were 12.5:1 and 13:1, respectively), where all the 12 nucleosome positioning 601 sequences have been loaded with HO_H4K20Me3 and those slightly excess HOs are absorbed by competitor DNA. In the EMSA gel (Figure 3.23), the optimal saturation of the 12-177-HO_H4K20M3 nucleosome array was suggested to be sample of lane 3 with apparent molar ratio of HO_H4K20M3 to 12-177 DNA as 12.5:1.

Figure 3.22 0.8 % agarose gel analysis of nucleosome arrays with different HO_H4K20Me3 to 12-177-601 DNA molar ratios. From left to right, lane 1 is DNA ladder, size of which is indicated as in the left panel; lane 2 is loaded with a mixture of 12-177-601 DNA and the calf thymus competitor DNA as control; lane 3 to 8 are the reconstituted nucleosome arrays with apparent molar ratio of HO_H4K20Me3 to 12-177-601 DNA as 11.5, 12, 12.5, 13, 13.5, 14:1, respectively. From top to bottom are nucleosome array dimers (in pink rectangle), 12-177 nucleosome array (in yellow rectangle), 12-177-601 DNA, nucleosomes (in blue rectangle) that are formed from competitor DNA and excess HO_H4K20Me3, and competitor DNA.

Figure 3.23 Analysis of the Scal digestion product of the 12-177-HO_H4K20Me3 nucleosome array samples on 6 % native PAGE. From left to right, lane 1 is DNA ladder, size of which is indicated as in the left panel; lane 2 is loaded with 177-601 monomer DNA as an additional marker; lane 3 to 7 are Scal-digested nucleosome arrays with apparent HO_H4K20Me3 to 12-177-601 DNA molar ratios as 12.25, 12.5, 12.75, 13, 13.25:1,

respectively. From top to bottom are incomplete- or un-digested nucleosome array trapped in the wells, nucleosome dimer, 177 bp-nucleosome monomer, and 177 bp naked DNA.

3.7 Cryo-Electron Microscopy single particle analysis of 177 bp-chromatosome

Cryo-EM is a powerful technique to determine structures of large macromolecular complexes at near atomic resolution. Multiple attempts have been undertaken to determine the structure of the chromatosome complex, with the best description of the chromatosome given for the 197 bp-chromatosome structure (Bednar et al., 2017) recently (as discussed in Introduction). The chromatosome structure (Figure 1.21) in couple with the full-length *Xenopus* linker histone H1.0 was solved to 11 Å resolution. The H1 globular domain is bound with the on-dyad mode, interacting with nucleosomal DNA and both linker DNA arms; the H1 CTD density is not well resolved, it is suggested to interact with one of the linker DNA.

In this thesis, the author works on the 177 bp-chromatosome with full-length human linker histone H1⁰. As discussed in the introduction, various factors- subtype/ variant or deletion/ mutation of linker histone, linker DNA length (or NRL in the context of nucleosome array) and possibly other unknown factors- could lead to divergent chromatosome structures with on- or off-dyad linker histone positions, different linker histone orientations, as well as different linker DNA flexibility and geometries. All of these would confer the chromatosome with distinct structure and symmetry, and eventually distinct chromatin higher-order structures. To better clarify these uncertainties, the author aims to solve the structure of 177 bp chromatosome by using the cryo-EM single particle analysis method.

The cryo-grids were prepared using freshly glow-discharged Quantifoil R2/2 holey carbon grids. 4 µl sample at 0.4 mg/ml concentration was used per grid. The sample was flash-frozen in liquid ethane with a Vitrobot (FEI) plunge-freezer with blot force 0, blot time 2 s. The samples were imaged on a 200 kV Tecnai Arctica (FEI) equipped with an autoloader for 12 grids and a Falcon III direct electron detector. Micrographs were collected using a nominal magnification of 78,000 x, corresponding to an object pixel size of 1.28 Å. In total, 2300 micrographs were collected at different defocus (ranging between -1.5 and -3.0 µm). Figure 3.24 shows a representative

micrograph of the 177 bp-chromatosome particles. The darker area at the lower left corner is the support carbon film (Figure 3.24), while the lighter part contains chromatosome particles in vitrified ice.

Two representative views are highlighted in Figure 3.24. Particles in red circles adopt a disk view, and those in yellow are shown as side view.

Figure 3.24 Cryo-EM micrograph of 177 bp-chromatosome. The micrograph was recorded on a 200 kV Arctica at 78,000 x magnification. The scale bar is 50 nm. Representative views are highlighted, disk view in red and side view in yellow.

The particle picking was done semi-automatically with “e2boxer” in the Eman software package (Tang et al., 2007). Box size was set at 168 pixels (1.28 Å/pix). In total, 124,000 particles were picked from the 2,300 micrographs. Image analysis was done using the RELION software

package (Scheres, 2012). The CTF estimations for the micrographs were done in CTFFIND3 (Mindell and Grigorieff, 2003). A soft spherical mask with diameter at 210 Å was applied to the particle images, the area outside the mask was set to 0 prior to the Fourier transformation, to minimize the noise around the particle and increase the sensitivity in the alignment and classification.

In the 2D class averaging, the particles were divided into 150 classes during the first iteration, with ~800 particles per class. The 2D class averaging yielded classes with high number of particles and good alignment, from which the one can see an outline of the DNA superhelix and partial densities of HO (Figure 3.25, outlined in red); and classes with low number of particles and smeared appearances (Figure 3.25, outlined in purple). The latter were outliers to be discarded. The 2D class averaging process was repeated two more times with particles from “good classes” of the previous run as the input dataset.

Figure 3.25 2D class averaging results. 150 classes from 124,000 particles sorted by particle distributions from high to low. Example of good classes with good alignment and clean background are highlighted by red outline; while outliers highlighted in purple were deposited into small classes with low particle distribution, poor alignment and noisy background.

2D averages with good alignment are selected as input for 3D classification (Figure 3.26).

Figure 3.26 Selection of “H1-bound” 2D class averages. 2D averages with compacted linker DNA arms were selected for 3D classification input, as highlighted in red.

3D classification was performed to further resolve structural heterogeneity. An initial model was built from the crystal structure of 197 bp chromosome (Figure 3.27 A) (Bednar et al., 2017). The electron density map was low-pass filtered to 80 Å (Figure 3.27 B) by Gaussian filter.

Figure 3.27 Preparation of the initial reference map for 3D classification. (A) Crystal structure of 197 bp-chromosome at 5.4 Å (PDB ID: 5NL0) (Bednar et al., 2017) (B) The electron density map of the chromosome, shown as a blue sphere.

density maps was low-pass filtered to 80 Å. Images were generated in Chimera (Pettersen et al., 2004).

The reference was refined against the entire dataset to achieve a consensus model and used for 3D classification. A soft spherical mask with diameter of 210 Å was applied to both the particles and the reference to reduce the background noise. The number of classes was set to 3. The initial density map was used in the 3D classification process to generate random seeds and classify the whole dataset into 3 classes. The refinement converged by the end of the 25 iterations.

Figure 3.28 The three classes from the 3D classification. The number of particles in each class was indicated. The scale bar is 50 Å.

The 3 classes of 3D reconstructions are shown in Figure 3.28. The class distribution is 0.29, 0.27 and 0.44 for class 1, 2 and 3, respectively. Number of input particles was 52,600, corresponding to 15,000, 14,200 and 23,200 particles per class. The structure of class 1 (Figure 3.28 A) shows as a nucleosome. No extra densities found on/around the nucleosome dyad region and/or close to the linker DNAs indicating the binding of linker histone.

Binding of linker histone is suggested to bring the two linker DNA arms closer to each other forming a more compact structure. The Class 2 and 3 show more compacted linker DNAs compared to class 1. Particles from these two classes were applied for separate auto-refinement and post-process masking.

In auto-refinement, unmasked maps were used for the FSC calculation to avoid potential overfitting. However, the resolutions may be under-estimated due to noises in the solvent. To achieve accurate resolution, in post-process, a soft mask was applied to reduce solvent noise. Resolution was calculated based on masked FSC curves.

Figure 3.29 Refined density maps of class 2 and 3 from the 3D classification. The reconstructions are shown in disk, side and tilted view, from top to bottom. The disk and side views are related by a 90° rotation about the longer axis; the side and tilted views are related by a 30° rotation about the longer axis. (A) Refined density map of class 2, (B) refined density map of class 3, (C) 197 bp chromosome (Bednar et al., 2017).

The 3D auto-refinement and post-processing results (Figure 3.29) suggest the structure of class 2 to be 177-nucleosome rather than chromosome, with the absence of linker histone densities, as well as relatively less compact linker DNAs orientation. The agreement of class 3 with the 197 bp-chromosome cryo-EM structure in both linker histone densities and orientation of

linker DNAs suggests that structure of class 3 might contain linker histone density in the triangle area formed by the nucleosome dyad and linker DNAs.

Figure 3.30 Fourier Shell Correlation (FSC) plots for the EM density map. Resolution was estimate at FSC=0.143. Class 2 (A) is 10 Å, class 3 (B) is 11 Å. A comparison of the FSC corrected (red), masked (orange), unmasked (green) and phase-randomized (purple) map is shown.

The masking effect was assessed by the high-resolution noise substitution method (Chen et al., 2013) with the output shown as the “FSC_corrected” curve in Figure 3.30. Phases of the two unfiltered maps upon convergence of the 3D auto-refinement were randomized before masking, the Fourier shell correlation was then calculated from the masked maps. The resolution-dependent differences between the FSCs of the phase-randomized and masked maps were analyzed for the assessment of masking effects. The resolution of the class 2 and 3 reconstruction was estimated at 10 and 11 Å from the “FSC_masked” curve at FSC= 0.143 following the gold-standard.

The atomic structure of 167 bp-chromatosome (bound with globular H5) (PDB ID: 4QLC) (Zhou et al., 2015) was fitted into the cryo-EM density map of the 177 bp-chromatosome (Figure 3.31) with Chimera using rigid-body fitting function. The atomic coordinates of 167 bp-chromatosome was converted into a density map at 3.5 Å. Each atom was described as a Gaussian distribution of width proportional to the resolution and amplitude proportional to the atomic number. The simulated map of 167 bp-chromatosome was fitted in the experimental

density map of 177 bp-chromatosome using the “map-in-map” function in Chimera. The correlation coefficient of the fitting is 0.76, average map value is 0.09, with 2226 of 13355 atoms outside the contour surface.

Both the nucleosome core and the linker DNAs of 167 bp chromatosome are fitted well into the density map. The “extra” densities in the end of the 177 bp chromatosome linker DNA region is likely due to the 5 bp difference in linker DNA lengths. Notably, the densities in between the two linker DNAs might indicate presence of linker histone CTD, which is absent in this crystal structure and that is highly enriched in positively charged amino acids- we hypothesize that the CTD is acting as an electrostatic bridge interacting with both linker DNAs making it possible for them to come close to each other, although this density may also partly originated from flexibility in the linker DNA itself or the long core histone H3 N-terminal tail (Stützer et al., 2016). The close-up view (Figure 3.32, right panel) of the linker histone orientation within the chromatosome adds more confidences to this assignment (in Figure 3.32, the triangle region is zoomed in to show structural details of the linker histone globular domain). The β -strand in the C-terminus of the globular domain points right towards the assumptive position of the CTD. The fitting also shows that small segment of linker DNA (Figure 3.32, red circled) of the 177 bp chromatosome falls outside of the density map, which means that the 177 bp chromatosome in this study is slightly less compacted at the linker DNA region. The other explanation is that binding of the full-length linker histone at the triangle region generate steric hindrance preventing linker DNA from coming closer to the nucleosome dyad.

Figure 3.31 The atomic structure of 167 bp-chromatosome fitted into the class 3 map of 177 bp chromatosome after post-processing. Two different views are shown - disk view (A) and titled disk view (B) are related by 30° rotation about the longer axis. The 177 bp chromatosome is shown in purple and blue, histones are represented by secondary structure features; the reconstruction map is shown in surface rendering (grey).

Figure 3.32 Orientation of linker histone globular domain within the chromatosome structure. The right panel is a close-up view of the squared triangle region where the linker histone globular domain interacts with the two linker DNAs. The potential linker histone CTD density is circled in yellow. The end of the β -strand at the C-terminus of the globular domain is marked by a yellow star. The assumptive CTD is sketched in green dashed line. The DNA segment outside of the contour is circled in red.

The crystal structure of the chromatosome with linker histone including the CTD was recently published (Bednar et al., 2017) at 5.5 Å resolution. The chromatosome is composed of 197 bp 601 DNA with two linkers each 25 bp long entering and exiting the nucleosome core particle (Figure 3.33, left panel). For the purpose of better fitting and comparison, the 197 bp was shortened to 177 bp by excising 10 bp from each linker DNA, leading to 15 bp each of the linker DNA (Figure 3.33, right panel). This modified complex is called “model-177 bp chromatosome”. The atomic structure and the simulated density map of “model-177 bp chromatosome” were fitted into the reconstructed 177 bp-chromatosome density map (Figure 3.34). From the atoms-in-map fitting, 1199 of 13691 atoms lay outside the contour surface. The correlation coefficient of the map-in-map fitting is 0.84, the average map value is 0.09.

Figure 3.33 Modification of the 197 bp chromatosome crystal structure to the “model-177 bp chromatosome”. The atoms of 10 bp were removed from each of the two linker DNAs from the pdb files, so that the original 25 bp linker DNA was shortened to 15 bp, and the 197 bp chromatosome turned to “model-177 bp chromatosome”. The cutting sites are denoted by the lightning bolts.

Figure 3.34 The atomic structure of the “model-177 bp chromatosome” fitted into the class 3 map of 177 bp chromatosome after post-processing. Two different views are shown, disk view (A) and tilted disk view (B) are related by a 30° rotation about the longer axis. The 197 bp chromatosome is shown in camel and blue, histones are represented by secondary structures; the reconstruction map is shown in surface style with 70 % transparency.

Both the nucleosome core and two linker DNAs of the “model-177 bp chromosome” structure are fitted very well into the density map. The linker histone globular domain achieved generally a good fitting, a close-up view is discussed (Figure 3.35). In the absence of linker histone, the 601-177 bp-nucleosome has a two-fold symmetric structure (if the exact DNA sequence not considered). Binding of linker histone confers asymmetry to the nucleosome. As shown in Figure 3.35, the “model-177 bp chromosome” atomic structure was fitted into the 177 bp-chromosome density map by Chimera rigid-body fitting. The “model-177 bp chromosome” atomic structures in (A) and (B) are related by 180° rotation about the longer axis. The atoms-in-map fitting indicated that out of 13691 atoms there were 1407 (A) and 1199 (B) atoms lying outside the contour surface. The correlation coefficient of the map-in-map fitting was 0.83 and 0.84, respectively. These fitting output suggested a better match in (B) as compared with that in (A). The close-up view of the two fittings revealed that the Figure 3.35 (A) position of the linker histone in the crystal structure does not match the electron densities observed in the space between two linker DNAs. In contrast, 180° rotation of the crystal structure about the longer axis shown in Figure 3.35 (B) results in a better agreement with the electron density map, only a small segment of the C-terminal β -hairpin and α -helix fall outside of the density map.

Figure 3.35 Binding of linker histone confers asymmetry to the nucleosome. The atomic structure of the “model-177 bp chromatosome” was fitted into the 177 bp chromatosome density map by Chimera with the “model-177 bp chromatosome” positioned in two different orientations 180° relative to each other about the longer axis. The middle panel is an enlarged view of the triangle region, the left panel is 90° clockwise rotation of the middle panel view.

Figure 3.36 177 bp-chromatosome map contains densities that are suggested to be assigned to the linker histone C-terminal domain. Rigid-body fitting of 167 bp-chromatosome (A) and 197 bp-chromatosome (B) crystal structure into the 177 bp-chromatosome density map. The linker histone globular domain is indicated by red squares. Densities circled in yellow are suggested to originate from the linker histone C-terminal domain.

There are indications in literature (Lu and Hansen, 2004; Roque et al., 2005; Vila et al., 2001) that binding of linker histone to either DNA or nucleosome restricts the flexibility of the CTD and induces it to form ordered secondary structures. However, it seems that the flexibility is still a key issue in getting the complete picture when linker histone binds to the nucleosome. The latest result of the 197 bp-chromatosome structure from cryo-EM single particle analysis (Bednar et al., 2017) demonstrated that one of the linker DNAs with extra densities might be assigned to the CTD. The cross-linking experiments (Bednar et al., 2017) also suggested association of CTD with one of the linker DNAs. Our result does not show differences in densities between the two linker DNAs in the 3D map of the 177 bp chromatosome. It is rather the triangle region formed by the nucleosome dyad and two linker DNAs, which is more voluminous than the linker histone globular domain alone which may indicate the CTD localization (Figure 3.36).

3.8 Self-association of the 177 bp nucleosome and chromosome

To investigate differences in biophysical properties of 177 bp nucleosome and chromosome, precipitation assay (PA) study has been carried out. Solutions of the 177 bp nucleosome and chromosome in 10 mM Tris-HCl (pH 7.5) were titrated by Mg^{2+} and $CoHex^{3+}$.

3.8.1 The polyelectrolyte property of 177 bp-nucleosome and chromosome

Figure 3.37 Precipitation assays of 177 bp-nucleosome and chromosome induced by Mg^{2+} (A) or $Co(HeX)^{3+}$ (B). The percentage of soluble sample measured by absorbance at OD_{260} was plotted against corresponding cationic concentrations. Three independent

assays were performed for each PA experiments. The raw data points shown as different symbols (square, circle or triangle) were fitted by Boltzmann sigmoidal function (solid lines).

For one complete set of precipitation assay experiment, the 177 bp nucleosome or chromatosome sample in 10 mM Tris-HCl (pH 7.5) buffer with $OD_{260} \sim 1.0$ were mixed with different concentrations of Mg^{2+} or $CoHex^{3+}$ salts. After centrifugation the amount of soluble sample was measured and plotted against corresponding cation concentrations (Figure 3.37). Three sets of PA experiments were done independently for each type of cations. EC50 value was calculated for evaluation and comparison of salt-dependency of the nucleosome/ chromatosome aggregation. The three PA repeats are in a good agreement. The average of EC50 values calculated from each of the 3 fitted curves is reported in Table 3.1. The Mg^{2+} induced self-association of the 177 bp-nucleosome achieved a maximum 90 % precipitation. The other PAs all reached nearly 100 % precipitation.

Table 3.1 Summary of EC50 values

Sample \ Cation	Mg^{2+} (mM)	$CoHex^{3+}$ (μM)
177 bp-nucleosome	3.0+/-0.1	31.3+/-0.8
177 bp-chromatosome	0.6+/-0.0	19.1+/-0.2

Cation induced self-association of the DNA, nucleosome and chromatin arrays is governed by electrostatic forces (Korolev et al., 2012) and references cited therein. The nucleosome and the chromatosome are complexes of DNA and histones, both of which are charged molecules. The 177 bp DNA carries 354 e negative charge (-354 e) from the phosphate groups, human HO carries net positive charge +146 e on the lysine and arginine residues. As a result, the 177 bp-nucleosome carries an overall negative charge of **-208 e**. In the case of 177 bp-chromatosome, the DNA negative charge is further neutralized by the positive charge of linker histone located mainly on the C-terminal tail region (Figure 3.38). Binding one molecule of full-length linker histone to the 177 bp nucleosome ends up with **-156 e** net negative charge of the resulting chromatosome. As has been discussed in the previous section '3.7 Cryo-Electron Microscopy

single particle analysis of 177 bp-chromatosome' and other's work (Bednar et al., 2017), the linker histone CTD is suggested to interact with the linker DNAs.

Figure 3.38 Cartoon illustration of the charge distribution on human linker histone hH1⁰. The amino acid sequence of the hH1⁰ is listed in the “Appendix”. The NTD, GD and CTD are coded in different colors, the net charges are +2 e, +11 e, and +39 e, respectively.

The trivalent CoHex³⁺ is more potent in inducing the nucleosome and chromosome aggregation than the divalent Mg²⁺ - there are ~100 x and 30 x differences ($[Mg^{2+}] / [CoHex^{3+}]$) in cation concentration required for the precipitation of nucleosome and chromosome, respectively (Table 3.1). The higher negative charge of nucleosome makes it more sensitive to the valency difference of the interacting cations. In addition, chromosome more readily self-associates compared to the nucleosome, with five times lower Mg²⁺ and ~1.5 times lower CoHex³⁺ in terms of the EC₅₀ values. These observations are in agreement with the electrostatic mechanism of nucleosome and chromosome condensation, that the less negative chromosome is more sensitive to screening by cations than the more negatively-charged nucleosome, and that this difference is less obvious when aggregation is induced by higher valent cations.

Mg²⁺ induced nucleosome and chromosome precipitation is reversible

Both nucleosome and chromosome precipitation induced by CoHex³⁺ is completed below 40 μM (Figure 3.37 B). No redissolution was observed at higher CoHex³⁺ concentration (up to 1000 μM for the chromosome, Figure 3.37 B). However, it is interesting to observe the resolubilization of the precipitated samples at higher concentration of Mg²⁺. The plots demonstrated that the resolubilization was a complete reversion of precipitation, the samples returned back to the original soluble state at around 100 mM Mg²⁺ (Figure 3.39). Similar result was also reported earlier (de Frutos et al., 2001) in which the resolubilization of NCPs was obtained at high concentrations of divalent cations Mg²⁺ or Ca²⁺. In this work, the

resolubilization process was observed for both 177-nucleosome and chromosome. Figure 3.39 demonstrates the combined process of precipitation followed by complete resolubilization with increasing amount of Mg^{2+} . The curves in Figure 3.39 show that precipitation transition proceeds in a broader range of Mg^{2+} concentrations than the reverse process of resolubilization both for the 177 bp nucleosome and chromosome. The precipitation of 177 bp-nucleosome and chromosome took place within a broad range of Mg^{2+} concentration from 0.1 to 10 mM; in comparison, the resolubilization happened in a narrow range between 10 mM and 100 mM.

Figure 3.39 Resolubilization of Mg^{2+} induced precipitation assay. The precipitation assay of 177 bp-nucleosome (A) and 177 bp-chromosome (B) were followed by complete resolubilization with higher amount of Mg^{2+} .

The resolubilization data points were fitted by Boltzmann sigmoidal function (Figure 3.40), from which EC_{50} values were obtained and summarized in Table 3.2. The resolubilization of the 177 bp nucleosome happens right after the maximal precipitation, while a delay was observed for the chromosome: maximal precipitation of the 177 bp-chromosome is reached at 4 mM Mg^{2+} , and resolubilization only begins at ~ 35 mM, at which point the 177 bp-nucleosome has already reached its resolubilization EC_{50} . The reason could be that binding of linker histone led to increased stability of condensed phase of the 177 bp chromosome, due to addition of positively charged linker histone. It might be that stability of the chromosome condensed

phase to resolubilization is due to ability of the linker histone to crosslink chromatosomes, particularly by extension of long flexible, disordered and positively charged CTDs.

Figure 3.40 Resolubilization of 177 bp-nucleosome and chromatosome induced by Mg^{2+} . Experimental data points were fitted by Boltzmann sigmoidal function.

Table 3.2 EC50 of the Mg^{2+} induced 177 bp-nucleosome and chromatosome resolubilization

Sample	EC50 Mg^{2+} (mM)
177 bp-nucleosome	34.2+/-0.3
177 bp-chromatosome	51.2+/-0.6

The divalent cation Mg^{2+} is ubiquitous component of cytoplasm and plays important roles in genomic activities such as gene replication and transcription, in which cases chromatin de-compaction is required for the access of relevant enzymes or molecular machines.

Furthermore, the combined capability of Mg^{2+} in inducing nucleosome/ chromatosome precipitation and resolubilization might be critical for regulating different states of chromatin compaction or relaxation for various purposes of cell processes.

3.9 X-Ray diffraction investigations of chromosome phase behavior

Small angle X-ray scattering is a technique where the X-rays are scattered by sample and the elastic part of the scattered beam carrying structural information of macromolecules is recorded at very low (small) angles ($0.1\text{-}10^\circ$). The 601-177 bp chromosome sample was prepared at 8 mg/ml concentration and precipitated by different concentrations of multivalent cations (Mg^{2+} and CoHex^{3+}). Samples were packed into transparent thin (1.5 mm in diameter) quartz capillary tubes and sealed with wax. Solution scattering experiments were done with 3 mg/ml nucleosome/ chromosome in 10 mM KCl buffer. SAXS data collection was performed at the beamline 23A workstation in National Synchrotron Radiation Research Center (NSRCC), Taiwan.

The integrated spectral intensity $I(q)$ of the precipitated samples were plotted as a function of the scattering vector q in OriginPro software (Figure 3.41).

Figure 3.41 The SAXS spectrum of 177 bp chromosome in the presence of 25 mM Mg^{2+} (left panel). The 2D diffraction pattern is shown on the right.

Phases of the particle aggregates could be identified from characteristic Bragg peaks. As for the nucleosome core particle (NCP), various phases have been identified (Livolant et al., 2006) (Berezhnoy et al., 2016) as shown in Figure 3.42. Isotropic phase (Figure 3.42 A) is mainly observed when NCPs are in monovalent salt conditions. Multivalent salts could lead to the NCPs

phase separation as columnar phases with different degrees of order: (1) columnar phase without any inter-columnar order (Figure 3.42 B), (2) the more ordered 2D columnar hexagonal phase (Figure 3.42 C), and (3) the most ordered 3D orthorhombic quasi-hexagonal phase (Figure 3.42 D) (Berezhtnoy et al., 2016).

Figure 3.42 Schematic presentation of different NCP phases. (A) Soluble NCP sample in solution state has isotropic phase; (B) columnar phase without inter-columnar order, the stacking of NCP is described by the stacking distance within a single column; both the 2D hexagonal columnar phase (C) and 3D orthorhombic quasi-hexagonal phase (D) are characterized by the particle stacking distance h and the inter-columnar distance a_H , as indicated. (E) The lamella-columnar stacking phase, the red line indicates a single layer of particles as part of the bilayer structure. The distance between bilayers is described as d_L . The yellow dots indicate the facing direction of nucleosome dyad. The figures are adopted from (Berezhtnoy et al., 2016; Mangenot et al., 2003a).

All the three columnar phases are described by the stacking distance h within a single column (Figure 3.42 B, C, and D), and the latter two are also characterized by the inter-columnar distance a_H (Figure 3.42 C and D). The two parameters h and a_H are calculated from the first peaks of two peak sets, (q_1 , q_2 , q_3 , q_4 , q_5 , etc.) and (q_{1h} , q_{2h} , q_{3h} , q_{4h} , etc.), respectively (Figure 3.43).

Figure 3.43 SAXS spectra of 177 bp chromosome samples in the presence of Mg^{2+} . Concentrations of Mg^{2+} are indicated in figure. Three sets of characteristic peaks were identified as shown on the 25 mM Mg^{2+} curve. The stacking parameters were calculated from the peak positions and presented on top of the peaks.

$$h = 2\pi/q_{1h}$$

$$a_H = 2/\sqrt{3} * 2\pi/q_1$$

The characteristic peak periodicity is $q_1: q_2: q_3: q_4: q_5 = 1: \sqrt{3}: \sqrt{4}: \sqrt{7}: \sqrt{9}$ and $q_{1h}: q_{2h}: q_{3h}: q_{4h} = 1: 2: 3: 4$. The lamella-columnar phase was observed from the native calf thymus NCPs in the presence of high concentration of monovalent salt under certain osmotic pressure (Livolant et al., 2006; Mangenot et al., 2003a; Mangenot et al., 2003b). The lamellar phase is described by the distance between each single bilayer d_L , which is calculated from the position of the first peak of the peak set (q_{1L}, q_{2L}, q_{3L} , etc.) (Figure 3.43). The peak periodicity is $q_{1L}: q_{2L}: q_{3L} = 1:2:3$.

$$d_L = 2\pi/q_{1L}$$

3.9.1 177 bp chromosome forms lamella-columnar phase

The 177 bp chromosome samples were measured at three different Mg^{2+} concentrations. The spectrum at 25 mM Mg^{2+} was well reproduced at 28 mM Mg^{2+} (Figure 3.41 and Figure 3.43).

Three sets of peaks were identified from the SAXS spectrum of 177 bp chromosome (Figure 3.43). The stacking phase of 177 bp chromosome aggregate was identified from the periodicity of the Bragg peaks, which suggested the formation of a lamella-columnar phase. Parameters of the lamella-columnar phase were calculated as stacking distance within the column $h = 58.7 \text{ \AA}$, inter-columnar distance $a_H = 111.2 \text{ \AA}$, bilayer-bilayer distance $dL = 226 \text{ \AA}$.

3.9.2 Binding of linker histone changed the nucleosome stacking phase

The reported stacking distance h and inter-columnar distance a_H of NCP is 55-57 $\text{\AA}$ and 110-112 $\text{\AA}$, respectively (Berezhnoy et al., 2016). Both the stacking distance ($h = 58.7 \text{ \AA}$) and inter-columnar distance ($a_H = 111.2 \text{ \AA}$) are comparable with those found in condensed phases of NCPs. In comparison, our preliminary data of 177 bp nucleosome indicates formation of columns that appear to be disordered suggested by the SAXS spectra (Figure 3.44) with the lack of characteristic peaks. It is proposed here that the lack of long-range ordered structure is originated from the flexibility of the two linker DNA arms. Binding of linker histone leads to a more rigid 177 bp chromosome structure. Furthermore, formation of the bilayer structure might be due to the asymmetric shape of the 177 bp chromosome. When tightly stacked onto each other, the stacking distance of nucleosome/ chromosome is defined by the thickness of the wedge-shaped particle (viewed perpendicularly to the axis of the DNA superhelix) and suggested to be mediated by core histone interactions (Allahverdi et al., 2010). Without any core histone modifications, the coincident stacking distance means that binding of linker histone creates no apparent effect on the columnar stacking.

The inter-columnar distance is related to the diameter of the disk-shaped particle (when looking down the axis of the DNA superhelix), the similar inter-columnar distance suggested that within a single bilayer, the 177 bp chromosome particles are laterally aligned back-to-back (the nucleosome dyad side is defined to be the 'face', while the opposite site the 'back'), as shown in Figure 3.42 E, the dyads (designated by yellow dots) are all facing towards the external side of the bilayer.

Figure 3.44 SAXS spectra of 177 bp nucleosome samples in the presence of Mg^{2+} . Concentrations of Mg^{2+} are indicated in figure.

However, the bilayer-bilayer distance determined in this work ($dL = 226 \text{ \AA}$) is considerably different from the value $dL = 358\text{-}376 \text{ \AA}$ for NCP reported by (Mangenot et al., 2003a). This disagreement may reflect the different conditions that induced the ordered phase in the two systems. This difference in conditions inducing the ordered phase may also result in different supermolecular structures. Furthermore, the difference could be explained by other factors: the differences between NCP and 177 bp chromatosome in the intrinsic structural features and charge distribution. The two 15 bp linker DNA arms, together with the bound linker histone, add additional volume to the NCP (Figure 3.45). As a result, the chromatosome has larger surface areas which would be more favorable for a tighter compaction. There are only small differences in the overall charges ($-156 e$ for 177 bp chromatosome, $-148 e$ for NCP), however, the charge distribution is significantly different. The long C-terminal domain of linker histone is rich in the positive residues. It is suggested in the cryo-EM work from this thesis and other's work (Bednar et al., 2017) that the CTD binds to one or both of the linker DNAs in the context of a single chromatosome. Here in the case of particle stacking, there is reason to believe that the positively charged CTD could act as a potential charge mediator to counter-balance (to a certain degree) the repulsive interactions among the negatively charged DNA of neighboring particles.

Figure 3.45 Structural comparison of 177 bp chromatosome with nucleosome core particle. The 177 bp chromatosome structure shown in grey semi-transparent volumes is reconstructed by cryo-EM single particle analysis as part of this thesis work. The crystal structure (PDB ID- 3LZ0) (Vasudevan et al., 2010) of nucleosome core particle with 145 bp DNA, shown in light and dark blue, is fitted into the 177 bp chromatosome density map by rigid-body fitting in Chimera. From the atoms-in-map fitting, 900 of 11908 atoms lay outside the contour surface. The correlation coefficient of the map-in-map fitting is 0.7, the average map value is 0.09. The extra densities of the 177 bp chromatosome compared with the NCP structure is the 'triangle' region formed by two linker DNA arms together with the linker histone.

3.10 Solution SAXS studies of 177 bp nucleosome and chromatosome

The 177 bp nucleosome and chromatosome samples were prepared at 3 mg/ml concentration in 10 mM KCl and 10 mM Tris (pH 7.5). Solution SAXS spectra of these samples are as shown in Figure 3.46.

Intensity $I(q)$ of the diffraction spectrum originates from two factors: form factor $P(q)$ represents scattering by individual nucleosome or chromatosome; structure factor $S(q)$ refers to scattering by interacting nucleosomes or chromatosomes. An ideal sample for solution SAXS experiment requires a dilute monodisperse solution with no particle interactions. Nucleosomes at low monovalent cation concentrations (e.g. 10 mM KCl used in this work) are suggested to have weak repulsive interactions (Mangenot et al., 2002) and contribution of the structure

factor $S(q)$ to the SAXS spectrum is negligible. Therefore, intensity of scattering spectrum is defined by form factor $P(q)$. Consequently, our recorded solution SAXS spectra provide description of the shape and structure of the 177 bp nucleosome and chromosome.

Figure 3.46 Solution SAXS spectra and Guinier plots of 177 bp nucleosomes (orange curve) and chromosome (green curve). The black arrow in (A) indicates the signature dip at $q=0.14 \text{ \AA}^{-1}$. (B) Guinier fitting plot of SAXS spectra at low q region, marked by the blue square in (A). Data points used for fitting were shown in solid circles.

The SAXS data were analyzed by Primus (Konarev et al., 2003) from the ATSAS software package (version 2.8.1) (Franke et al., 2017). Radius of gyration (R_g) of the 177 bp nucleosome and chromosome were calculated from the slope of the Guinier (liner) region of the spectrum. The R_g , determined in this way from SAXS, of a 146 bp “601” DNA NCP was previously reported to be $41.6 \text{ \AA}$ by (Mangenot et al., 2002; Yang et al., 2011) in low salt conditions. Guinier approximation (Guinier, 1939; Konarev et al., 2003) is valid at low q value ($sR_g < 1.3$) region where inter-particle distance is big and particle interactions are negligible.

$$\ln I(q) = I_0 + \left(\frac{-R_g^2}{3} \right) * q^2$$

$I(q)$ is the scattering intensity, I_0 is the forward scattering, q is the scattering vector. R_g is directly related to the shape and mass distribution of the particle.

Table 3.3 Summary of nucleosome/ chromatosome geometry

Samples	R_g (Å)
177 Nucleosome	46.7
177 Chromatosome	49.0

The linear region extends well beyond the empirical $q_{\max}$ cutoff of $1.3/R_g$ for globular structures. The R_g values (shown in Table 3.3) indicate that the 177 bp chromatosome has larger size than 177 bp nucleosome. Increase of the R_g value of chromatosome relative to the nucleosome corresponds to 15.5 % increase of the volume of the particle (calculated as $(49/46.7)^3$). This roughly corresponds to the increase of the chromatosome size caused by binding of linker histone.

A dip is observed at $q = 0.14 \text{ Å}^{-1}$, as indicated by the black arrow in Figure 3.46. This dip is characteristic of the degree of unwrapping of the DNA that winds around the HO. Unwrapping results in a reduction in this dip as shown for example by (Bertin et al., 2007) and (Andresen et al., 2013). A deeper dip indicates less unwrapped DNA. The 177 nucleosome has a reduction in this dip compared to 145 bp NCP, which shows that longer linker DNA results in unwrapping of the core DNA that wraps the HO (Aghil Soman et al, unpublished data from our laboratory).

One can see in Figure 3.46 that the chromatosome spectrum shows deeper dip than the one of the nucleosome indicating that DNA is fully wrapped in the chromatosome. Smaller dip observed in the nucleosome spectrum indicates some degree of DNA unwrapping from the HO. Similar unwrapping of the DNA was observed in solutions of NCP with tailless H3 and H4 (Howell et al., 2013). Our data indicates that repulsion of the stretches of the linker DNA at the entry and exit of the nucleosome might cause partial DNA unwrapping. Presence of the linker histone reduces DNA-DNA repulsion and increases DNA wrapping. Consequently, the result in Figure 3.46 shows that a more compacted structure of the 177 bp chromatosome compared to

177 bp nucleosome due to the binding of linker histone, which is in agreement with the cryo-EM results.

Taken together, comparisons of R_g and the signature unwrapping dip suggest the 177 bp chromatosome has a more compacted structure than the 177 bp nucleosome. The results indicate a key influence of linker histone as a structural factor in DNA compaction.

3.11 Analytical Ultracentrifugation (AUC) investigation of effects of H4K20Me3 on chromatin folding

AUC sedimentation velocity experiment is a biophysical method for the study of hydrodynamic properties of macromolecules in solution. The size, shape and compaction of the molecules are defined from the interpretation of the movement of the solutes in the centrifugal field. The movement of molecules is recorded with an optical system. A series of sedimentation curves are collected throughout the course of sedimentation process. The sedimentation coefficient (S-value) of the particles is deduced by the vHW analysis of the sedimentation curves. The S-value is an intrinsic property of a particle, it is used to characterize how fast the particles are moving in the direction of the centrifugal force. For a certain type of nucleosome array, different S-values reflect different degrees of compaction.

AUC sedimentation velocity experiments were carried out for wild-type and modified (H4K20Me3) 12-177 nucleosome arrays. The potential effects of H4K20Me3 on the compaction of chromatin driven by intra-molecular interactions were investigated.

The sedimentation process was analyzed by two software packages UltraScan and Sedfit independently. The S-values of chromatin fibers were calculated and used for evaluation of the compaction of chromatin fibers.

At low salt conditions (e.g. TEK buffer), the nucleosome array is extended as “beads-on-a-string” conformation. In the electrostatic view, this extended form is induced by the repulsive interactions among the negatively charged linker DNA (Korolev et al., 2012). In this work, monovalent and divalent counterions (Na^+ and Mg^{2+}) are added to the nucleosome array sample to screen the repulsions and induce attractive nucleosome-nucleosome interactions.

Therefore, nucleosome array folding is driven by the salt-induced inter-nucleosome interactions. Extensive studies have indicated that histone tails play important roles in the regulation of nucleosome array folding (Howell et al., 2013) (Berezhnoy et al., 2016; Luque et al., 2016). Histone H4 tail is a hotspot for post-translational modifications. H4K20Me3 is reported to be a heterochromatin marker (Jorgensen et al., 2013; Schotta et al., 2004). The AUC experiments in this thesis is aimed to elucidate the effect and mechanisms of H4K20Me3 on the nucleosome array folding.

3.11.1 12-177-wtHO nucleosome array

The 12-177 nucleosome array reconstituted with wild-type *Xenopus* histone octamers were prepared in TEK buffer (10 mM Tris pH 7.5, 0.1 mM EDTA, 10 mM KCl). Sedimentation velocity experiments were carried out for nucleosome samples in TEK buffer, and with addition of different concentrations of Mg^{2+} (0.2, 0.4, 0.6, 0.8 and 1.0 mM) or Na^+ (20, 40, 60, 80, 100 mM). Characteristic S-values determined by vHW and Sedfit methods are in good agreement both for Mg^{2+} salt (Table 3.4 and Table 3.5) and Na^+ salt (Table 3.6 and Table 3.7).

Figure 3.47 vHW curves of 12-177-wtHO nucleosome array with different Mg^{2+} concentrations. The S-values increase with the step-wise increases of Mg^{2+} from 0.2 to 1.0 mM. Mg^{2+} concentrations were indicated by different colors.

Table 3.4 Summary of S-values of 12-177-wtHO nucleosome array folded in different Mg^{2+}

Mg^{2+} conc./ mM	S-value
0	34.6+/-0.1
0.2	38.9+/-0.4
0.4	42.0+/-0.9
0.6	45.9+/-0.8
0.8	49.9+/-2.1
1.0	49.7+/-1.6

The average of S-values from the 20 % to 80 % boundary positions is reported as the S-value of the nucleosome array. Besides, the homogeneity of the sample is estimated from the straightness of the vHW curves. For a monodisperse system, the vHW curve is vertical with a distinct S-value of the single species. The deviation of the vHW curve from the vertical line could be due to different compaction of the single nucleosome fiber or the formation of multimeric fiber complex, or in most situations both, as will be discussed below.

Figure 3.48 Sedfit curves of 12-177-wtHO nucleosome array with different Mg^{2+} concentrations. The peaks move towards higher S-values with the step-wise increases of Mg^{2+} from 0.2 to 1.0 mM. Mg^{2+} concentrations are marked by different colors.

Table 3.5 Summary of Sedfit calculation of S-values of 12-177-wtHO nucleosome array folded in different Mg^{2+}

Mg²⁺ conc./ mM	Main peak	Size of the peak	Other peaks	Size of the peak
0	35.0± 0.9	100.0 %		
0.2	37.8± 0.8	100.0 %		
0.4	42.6 ± 0.7	98.3 %	12.0 ± 1.1	1.3 %
0.6	45.8 ± 1.2	98.9 %	7.4 ± 1.6	0.8 %
0.8	49.0 ± 2.5	89.6 %	71.1 ± 5.8	9.2 %
1.0	49.7 ± 3.4	88.6 %	70.7 ± 4.7	9.5 %

The height and width of the S-value distribution peaks directly reflect the heterogeneity of the folded nucleosome array sample. The broadening of the main peak indicates that the nucleosome array is not uniformly folded. The appearance of a second or multiple peaks suggests the formation of nucleosome array dimer and multimers.

The maximal S-value calculated by Sedfit analysis is 49-50, in the buffer condition with 0.8 and 1.0 mM Mg²⁺. From 0 to 0.6 mM Mg²⁺, the 12-177 nucleosome array sample is quite homogeneous, as shown by the relatively high and narrow single peak (Figure 3.48). A second peak came out with higher Mg²⁺ amount (0.8 and 1.0 mM), it accounts for 9-10 % of the population, the S-value of which is 70-71. The larger species is proposed to be nucleosome array dimer and/or other oligomeric species formed through fiber-fiber interactions.

Higher Na⁺ concentration leads to more compacted nucleosome fibers, as shown in Figure 3.49 and Table 3.6. The highest S-value is 50S achieved with 100 mM Na⁺. However, the “bending” of the vHW curves with higher Na⁺ concentrations (60-100 mM) suggests non-uniformed nucleosome array folding and/ or the formation of multimeric nucleosome arrays.

Figure 3.49 vHW curves of 12-177-wtHO nucleosome array with different Na^+ concentrations. The S-values increase with the step-wise increases of Na^+ from 20 to 100 mM. Na^+ concentrations were indicated with different colors.

Table 3.6 Summary of S-values of 12-177-wtHO nucleosome array folded in different Na^+

Na^+ conc./mM	S-value
0	34.6+/-0.1
20	39.6+/-0.7
40	42.3+/-0.6
60	45.4+/-1.6
80	49.0+/-2.4
100	49.6+/-1.6

The maximal nucleosome array compaction was obtained at ~49S in buffer conditions with 80-100 mM Na^+ . There is a pronounced drop and broadening of the S-value distribution peaks with 60 mM and higher Na^+ concentration, which suggests a polydisperse system, as discussed for the UltraScan analysis (Figure 3.49). From Table 3.7, the appearance of a second peak at larger S-values (75-80S) is consistently observed for Na^+ concentrations of 60-100 mM. This second

species accounts for ~10 % of the population. The 75-80S is slightly bigger than the 70S obtained for nucleosome array folded in buffers with 0.8 and 1.0 mM Mg^{2+} .

Figure 3.50 Sedfit curves of 12-177-wtHO nucleosome array with different Na^+ concentrations. The peaks move towards higher S-values with the step-wise increases of Na^+ from 20 to 100 mM. Na^+ concentrations were indicated with different colors.

Table 3.7 Summary of Sedfit calculation of S-values of 12-177-wtHO nucleosome array folded in different Na^+

Na^+ conc./ mM	Main peak	Size of the peak	Other peaks	Size of the peak
20	41.0 ± 0.7	99.5 %	11.6 ± 0.8	0.3 %
40	44.1 ± 1.0	94.4 %	10.3 ± 0.9 74.0 ± 4.3	3.5 % 1.7 %
60	46.2 ± 1.7	89.6 %	17.8 ± 1.4 75.6 ± 6.5	2.4 % 7.1 %
80	48.5 ± 2.6	80.9 %	79.1 ± 8.9	13.1 %
100	49.2 ± 3.1	85.3 %	22.9 ± 2.7 78.1 ± 9.2	2.1 % 11.2 %

In a summary, the AUC results show that the S-values of the 12-177 nucleosome array in TEK buffer condition calculated independently from UltraScan (34.64) and Sedfit (34.97) are in perfect agreement. Besides, it is in consistent with the 34-36S calculated by others (Allahverdi et al., 2010; Korolev et al., 2012). Larger S-values were obtained with higher amount of cations.

The maximal compaction with S-value at 49-50 was achieved at 0.8-1.0 mM Mg^{2+} or 80-100 mM Na^+ . Results from the two software agreed well with each other in general. However, with higher cation concentrations (0.8-1.0 mM Mg^{2+} or 60-100 mM Na^+), the nucleosome array sample lost its monodispersity, the isolated nucleosome fibers display the formation of a small amount (~10 %) of nucleosome array dimers/oligomers with 70-80 S-values.

3.11.2 12-177-HO_H4K20Me3 nucleosome array

To investigate the effects of H4K20 tri-methylation on nucleosome array compaction (intra-molecular interactions), AUC velocity experiments were performed with the methylated 12-177 nucleosome array in buffer conditions with 1.0 mM Mg^{2+} or 100 mM Na^+ , at which the wild type 12-177 nucleosome array achieved maximal compaction.

Multiple nucleosome array samples with various amount of HOs in TEK buffer (10 mM Tris pH 7.5, 0.1 mM EDTA, 10 mM KCl) were set up for the AUC experiments. The sedimentation velocity data were analyzed in the same way as for the wild-type nucleosome array.

Figure 3.51 vHW curves of 12-177-HO_H4K20Me3 nucleosome array with different amount of histone octamers in TEK buffer. The apparent molar ratio of HO to 12-177 DNA are marked by different colors.

Table 3.8 Summary of S-values of 12-177-HO_H4K20Me3 nucleosome array with different amount of histone octamers

HO to 12-177 DNA app molar ratio	S-value
11:1	33.4+/-0.2
11.5	33.5+/-0.3
12.25:1	34.7+/-0.2
12.5:1	35.6+/-0.1
12.75:1	36.4+/-0.2
13:1	36.7+/-0.2

Figure 3.52 Sedfit curves of 12-177-HO_H4K20Me3 nucleosome array with different amount of histone octamers at 0 salt condition. The apparent molar ratio of HO to 12-177 DNA are indicated with different colors.

Table 3.9 Summary of Sedfit calculation of S-values of 12-177-HO_H4K20Me3 nucleosome array with different amount of histone octamers at 0 salt condition

HO to 12-177 DNA app molar ratio	Main peak	Size of the peak	Other peaks	Size of the peak
11	29.6± 1.0	98.3 %	0.4 ± 0.2	1.4 %
11.5	33.5 ± 0.8	99.7 %		
12.25	33.3 ± 0.9	100 %		
12.5	36.3 ± 0.8	100 %		
12.75	35.1 ± 0.8	100 %		
13	37.6 ± 0.7	100 %		

As shown in Table 3.8 and Table 3.9, the S-values vary from 30 to 38. Variation in S-values reflects different degrees of HO saturation. It is suggested that under the TEK buffer condition, the nucleosome fibers adopt the elongated “beads-on-a-string” structure. The 12-177 nucleosome array with “beads-on-a-string” structure has the S-value at ~35 (Allahverdi et al., 2010; Kornberg, 1974), as has also been confirmed in section “3.11.1 12-177-wtHO nucleosome array” of this thesis. Whether with slightly higher or lower amount of HOs, the compactness of the nucleosome arrays is supposed to remain constant. Thus the variation in S-values here is not a direct reflection of different nucleosome array compactness, but rather a result of various abundance of HOs among the six tested samples. As shown in the formula below, with similar sample composition ($\bar{v}$) and buffer condition (ρ) ($\bar{v}$ is the solute partial specific volume, ρ is the buffer density, N is for Avogadro constant, f is frictional coefficient). The S-value (s) and molecular weight (M) are positively correlated:

$$s = \frac{M(1 - \bar{v}\rho)}{Nf}$$

As suggested by gel electrophoretic analyses (Figure 3.22 and Figure 3.23), nucleosome array samples with HO to 12-177 DNA apparent molar ratio at 12.5:1 and 12.75:1 correspond to saturated amount of HOs. S-values from UltraScan and Sedfit (Table 3.8 and Table 3.9, respectively), which are generally consistent with each other, have further verified that these two samples are desirable for AUC experiments in cationic conditions with either 1.0 mM Mg^{2+} or 100 mM Na^+ .

Figure 3.53 vHW curves of 12-177-HO_H4K20Me3 nucleosome array in the presence of 1 mM Mg^{2+} . The apparent molar ratio of HO to 12-177 DNA are indicated with different colors.

Table 3.10 S-values of 12-177-HO_H4K20Me3 nucleosome array with 1 mM Mg^{2+}

HO to 12-177 DNA app molar ratio	S-value
12.5:1	50.4+/-1.3
12.75:1	51.1+/-1.7

Figure 3.54 Sedfit curves of 12-177-HO_H4K20Me3 nucleosome array in the presence of 1 mM Mg^{2+} . The apparent molar ratio of HO to 12-177 DNA are indicated with different colors.

Table 3.11 Sedfit calculation of S-values of 12-177-HO_H4K20Me3 nucleosome array in 1 mM Mg^{2+}

HO to 12-177 DNA app molar ratio	Main peak	Size of the peak	Other peaks	Size of the peak
12.5	50.6 ± 3.4	79.7 %	8.8 ± 2.9 80.5 ± 5.0	2.2 % 14.1 %
12.75	51.4 ± 3.7	95.1 %	16.5 ± 2.8 70.3 ± 4.8	0.45 % 3.7 %

Figure 3.55 vHW curves of 12-177-HO_H4K20Me3 nucleosome array in the presence of 100 mM Na^+ . The apparent molar ratio of HO to 12-177 DNA are indicated with different colors.

Table 3.12 S-values of 12-177-HO_H4K20Me3 nucleosome array in 100 mM Na^+

HO to 12-177 DNA app molar ratio	S-value
12.5:1	48.2 ± 1.4
12.75:1	50.8 ± 1.5

Figure 3.56 Sedfit curves of 12-177-HO_H4K20Me3 nucleosome array in the presence of 100 mM Na⁺. The apparent molar ratio of HO to 12-177 DNA are indicated with different colors.

Table 3.13 Sedfit calculation of S-values of 12-177-HO_H4K20Me3 nucleosome array in 100 mM Na⁺

HO to 12-177 DNA app molar ratio	Main peak	Size of the peak	Other peaks	Size of the peak
12.5	48.3 ± 2.3	88.5 %	12.0 ± 1.3 75.3 ± 5.1	2.7 % 8.0 %
12.75	49.6 ± 1.6	89.0 %	12.0 ± 1.3 78.7 ± 5.0	2.5 % 7.8 %

Results of UltraScan and Sedfit are in good agreement. The 12.5:1 nucleosome array sample has slightly lower S-values than the 12.75:1 sample with ~1-2S differences for the two cationic conditions. S-values obtained from the 12.75:1 sample were reported as the characteristic S-values of 12-177-HO_H4K20Me3 nucleosome array.

In summary, the S-value of modified nucleosome array in cation conditions of 1 mM Mg²⁺ or 100 mM Na⁺ are in the range of 50-51S. A small fraction (<10 %) of a larger species (70-80S) were observed in both cationic conditions, as discussed for the wild-type nucleosome array, these are suggested to be the dimeric/multimeric nucleosome array complexes.

H4K20Me3 alone makes no distinct effect on nucleosome array folding

Figure 3.57 Comparison of the compaction of wild-type and modified nucleosome array. The Sedimentation coefficient distribution curves were generated by Sedfit. Cationic conditions and nature of the nucleosome arrays were indicated in the figure.

Despite numerous literature data showing positive correlation between the formation of heterochromatin and presence of H4K20Me3 histone modification (Fraga et al., 2005; Jorgensen et al., 2013; Mikkelsen et al., 2007; Pesavento et al., 2008; Schotta et al., 2004), the results of this work indicate that this conserved heterochromatin marker does not make noticeable influence on nucleosome array compaction when compared with the wild-type. As shown in Figure 3.57, the modified and wild-type nucleosome arrays are folded to almost the same degree in either 100 mM Na^+ or 1 mM Mg^{2+} .

Similarly, the effects of H4K20Me3 on chromatin folding was studied (Lu et al., 2008) with nucleosome array with longer NRLs (12-207). In the presence of 1 or 1.5 mM Mg^{2+} , the wild-type 12-207 nucleosome array was only able to fold into moderate condensation with average S-value at 35S. This observation was in agreement with other literature (Correll et al., 2012), where a negative correlation was observed between the NRL and the degree of salt-induced nucleosome array compaction. While nucleosome arrays with shorter NRLs are fully compacted

by Na^+ or Mg^{2+} , those with longer NRLs only achieve moderate folding. Salts have less influence in the compaction of nucleosome arrays with longer NRLs than shorter NRLs. Linker histone is critical for the complete compaction of nucleosome array with longer NRLs (Carruthers et al., 1998; Correll et al., 2012; Robinson et al., 2008).

However, the folding of modified (H4K20Me3) 12-207 nucleosome array was promoted to a certain degree yet not full compaction in favorable salt conditions. The degree of folding is not uniform and majority of the population is under full compaction (Lu et al., 2008).

In contrast, in this thesis, the 12-177 nucleosome array, whether wild-type or modified (H4K20Me3), both achieve full compaction in the absence of linker histone. This result is in agreement with other literature (Dorigo et al., 2003; Routh et al., 2008; Shogren-Knaak et al., 2006) that nucleosome arrays with shorter NRLs (177 bp or shorter) are able to fully compact with only the presence of favorable salts. Therefore, the positive effect of H4K20Me3 on the compaction of 12-207 nucleosome array observed in (Lu et al., 2008) was not manifested in 12-177 nucleosome array.

Together, my results and others suggest that for nucleosome array with longer NRLs, while H4K20Me3 promotes folding to some extent, linker histone is critical in achieving the full compaction; for nucleosome array with shorter NRLs, the presence of H4K20Me3 has no additive effects on the folding in favorable salt conditions.

3.12 Self-association of the 12-177 nucleosome array with unmodified HO or H4K20Me3 modification

To investigate effects of heterochromatin specific H4K20Me3 modification on the chromatin self-association, unmodified 12-177 nucleosome array and 12-177 nucleosome array with H4K20Me3 were prepared. Precipitation assay study was carried out with Mg^{2+} and $CoHex^{3+}$ as precipitating cations. The nucleosome array sample in 10 mM Tris-HCl (pH 7.5) buffer with $OD_{260} \sim 1.0$ were mixed with different concentrations of Mg^{2+} or $CoHex^{3+}$ salts. After centrifugation the amount of soluble sample was measured and plotted against corresponding cation concentrations (Figure 3.58). Three sets of PA experiments were done independently for each type of cations. EC50 value was calculated for evaluation and comparison of salt-dependency of the chromatin aggregation. The average of EC50 values calculated from each of the 3 fitted curves is reported in Table 3.14.

Figure 3.58 Precipitation assays of unmodified and modified (H4K20Me3) 12-177 nucleosome array induced by Mg^{2+} (A) or $CoHex^{3+}$ (B). The percentage of soluble sample measured by absorbance at OD260 was plotted against corresponding cationic concentrations. Three independent assays were performed for each PA experiments. The raw data points were fitted by Boltzmann sigmoidal function (solid lines for modified, and dashed lines for unmodified 12-177 nucleosome array).

Table 3.14 Summary of EC50 values

Sample \ Cation	Cation	
	Mg^{2+} (mM)	$CoHex^{3+}$ (μM)
12-177-wtHO	0.99+/-0.05	24.62+/-0.11
12-177-H4K20Me3	0.99+/-0.02	26.86+/-1.78

The nucleosome array is complex of negatively charged DNA and positively charged histones. 177 bp nucleosome is the building block of 12-177 nucleosome array, it carries -206 e net negative charge (-354 e from DNA, +148 e from unmodified *Xenopus* recombinant HO). Methylation of the amine group in the lysine residues does not change positive charge, it increases hydrophobicity of the ϵ -amine. Both unmodified and H4K20Me3 containing arrays have the same charge (overall negative). Higher hydrophobicity of the modified 12-177 nucleosome arrays does not influence their self-association properties. As shown in Figure 3.58,

the overlapped PA curves of modified and unmodified nucleosome array indicate same cation (Mg^{2+} and CoHex^{3+})-dependent self-association, with same/close EC_{50} values as shown in Table 3.14. Besides, the higher valent CoHex^{3+} is more efficient in inducing nucleosome array self-association compared to divalent Mg^{2+} , with $\sim 40\times$ differences ($[\text{Mg}^{2+}]/[\text{CoHex}^{3+}]$) in EC_{50} . These observations are in agreement with the electrostatic mechanism of cation induced nucleosome, chromosome (discussed in this thesis: section 3.8 Self-association of the 177 bp nucleosome and chromosome) and chromatin aggregation (Korolev et al., 2012).

Mg^{2+} induced resolubilization of nucleosome array

As shown in Figure 3.59, complete precipitation was achieved at 2 mM Mg^{2+} , the nucleosome array remained in aggregated state up to 80 mM Mg^{2+} . Resolubilization of 12-177 nucleosome array was observed at higher Mg^{2+} concentration (Figure 3.59 and Figure 3.60). EC_{50} of resolubilization is ~ 100 mM (Table 3.15) for both modified and unmodified nucleosome array.

Figure 3.59 Resolubilization of Mg^{2+} induced precipitation assay. The precipitation assay of unmodified (wt) and modified (H4K20Me3) 12-177 nucleosome array were followed by complete resolubilization with higher amount of Mg^{2+} .

The coincidence of Mg^{2+} induced precipitation-resolubilization curves further emphasize the similar polyelectrolyte properties of the modified and unmodified 12-177 nucleosome array. In contrast, resolubilization was not observed at higher $CoHex^{3+}$ concentration (up to 500 μM) (Figure 3.58 B). The divalent cation-induced resolubilization was observed in nucleosome and chromatosome from this work (section 3.8 Self-association of the 177 bp nucleosome and chromatosome) and other literature (de Frutos et al., 2001). The coincidence of ubiquitous distribution of Mg^{2+} in cytoplasm with the Mg^{2+} -specific two-way regulation of nucleosome array precipitation and resolubilization suggests a critical influence of Mg^{2+} in control of chromatin structures to cope with various genomic activities.

Figure 3.60 Resolubilization of 12-177 nucleosome array induced by Mg^{2+} . Unmodified (wt) array is marked in open circle, modified (H4K20Me3) array is shown in solid triangle. Experimental data points were fitted by Boltzmann sigmoidal function.

Table 3.15 EC50 of the Mg^{2+} induced 12-177 modified and unmodified nucleosome array resolubilization

Sample	EC50	Mg^{2+} (mM)
12-177-wtHO		114.54+/-9.57
12-177-H4K20Me3		98.33+/-3.58

Our data supports an explanation that H4K20Me3 exerts its influence on chromatin structure and dynamics indirectly, most probably serving as binding site for the proteins involved in formation of heterochromatin.

4 Conclusion

The dynamic structures of chromatin play critical roles in genomic activities within living cells. The chromatin fibers adopt with different levels of compaction to cope with various cellular activities. The structure of chromatin could be either such as in loose euchromatin to expose DNA for the access of molecular machines during replication and gene transcription, or such as in condensed heterochromatin non-coding regions, and condensed chromatosomes during cell mitosis. Various factors contribute to the formation and maintenance of the diversity of chromatin structures.

In this thesis, the author combined methods of molecular biology to construct precisely defined elementary units of eukaryotic chromatin: 177 bp nucleosome and chromatosome; and 12-177 (unmodified and modified with H4K20Me3) nucleosome arrays. Various biophysical methods: cryo-EM single particle analysis, SAXS, PA, and AUC were used to study the structure of these well-defined model systems.

Images of the 177 bp chromatosome were obtained with electron microscopy under cryogenic condition. The 3D structure of 177 bp chromatosome at its natural state was reconstructed by single particle analysis with resolution of the 3D map at 11 Å based on the gold-standard criteria. The two 15 bp linker DNA arms are positioned close to each other, together with the nucleosome dyad region, forming a triangle area, the densities in the middle are suggested to be from the bound linker histone. Fitting the 167 bp- or 197 bp-chromatosome atomic structure into the 177 bp chromatosome density map resulted in desirable overlap of the triangle density with the atomic structure of the linker histone globular domain. Besides, the extra density of the triangle area in the vicinity of the two linker DNA ends were proposed to be from the linker histone C-terminal domain. The hypothesis is that the highly positively charged CTD was localized in thus a way to act as an “electrostatic bridge” to screen the negative charges of the two linker DNA leading to a more rigid and compacted overall structure.

The linker histone is rich in positively charged lysine and arginine residues and overall it has 52 e net positive charge. The 177 bp nucleosome contains 208 e negative charge (-352 e from DNA and + 144 e from the HO), with the binding of one molecule of linker histone, the 177 bp

chromatosome has 25 % less negative charges compared with the 177 bp nucleosome. Therefore, both complexes can be considered as polyelectrolytes, the less negative chromatosome is more sensitive and reactive to the multivalent cations, the aggregation of 177 bp chromatosome requires five times lower Mg^{2+} and ~ 1.5 times lower $CoHex^{3+}$ concentrations compared with the 177 bp nucleosome.

The aggregation induced by the divalent cation Mg^{2+} is reversible with further addition of Mg^{2+} . The resolubilization was observed for both 177 bp nucleosome and chromatosome. The binding of linker histone not only conferred electrostatic changes to the nucleosome, but also strengthened the chromatosome self-association within the aggregates leading to a more stable state. Thus, it required more (1.5x) Mg^{2+} to invert the chromatosome aggregates compared with the nucleosome. This observation delivers possible biological significance in the control of chromatin accessibility by the linker histone- Mg^{2+} coordinated regulation.

The 177 bp chromatosome forms a lamella-columnar stacking phase in the aggregates induced by 25-28 mM Mg^{2+} . Characteristic stacking parameters were calculated from the SAXS X-ray diffraction spectrum- the columnar stacking distance $h = 58.7 \text{ \AA}$, inter-columnar distance $a_H = 111.2 \text{ \AA}$, bilayer-bilayer distance $dL = 226 \text{ \AA}$. The former two are in agreement with those reported for nucleosome core particles (with $h = 55-57 \text{ \AA}$ and $a_H = 110-112 \text{ \AA}$) (Berezhnoy et al., 2016). Besides, the h and a_H reveal the overall dimensions of the 177 bp-chromatosome, which has an average thickness of $58.7 \text{ \AA}$ and diameter of $111.2 \text{ \AA}$. The large difference in the bilayer-bilayer distance between 177 bp chromatosome ($dL = 226 \text{ \AA}$) and the $358-376 \text{ \AA}$ (Mangenot et al., 2003a) reported for NCP may reflect the distinct features of the chromatosome obtained from the interactions with linker histone as well as different conditions that induced the ordered phase in the two systems.

The R_g values of 177 bp chromatosome ($49 \text{ \AA}$) and nucleosome ($46.7 \text{ \AA}$) calculated from the solution SAXS spectra suggest 15 % increase of the volume of chromatosome relative to the nucleosome. This roughly corresponds to the increase of the chromatosome size caused by binding of linker histone. Presence of the dip at $q = 0.14 \text{ \AA}^{-1}$ is characteristic of the DNA wrapping on the HO. Decrease of the dip is attributed to the unwrapping of DNA from the

nucleosome core (Andresen et al., 2013; Bertin et al., 2007; Howell et al., 2013). The depth of the dip is related to the length of unwrapped DNA. The deeper dip in the chromatosome spectrum compared to the nucleosome spectrum suggests that DNA is fully wrapped in chromatosome, and there is some degree of DNA unwrapping from the HO in nucleosome. The results indicate that repulsion of the stretches of the linker DNA at the entry and exit of the nucleosome might cause partial DNA unwrapping. Presence of the linker histone reduces DNA-DNA repulsion and increases DNA wrapping.

From the AUC experiments, the sedimentation coefficient of 12-177 nucleosome array in TEK buffer (10 mM Tris pH 7.5, 0.1 mM EDTA, 10 mM KCl) was obtained at 35S, which is in consistent with the 34-36S calculated by others (Allahverdi et al., 2010; Korolev et al., 2012; Schalch, 2004). At 0.8-1.0 mM Mg^{2+} or 80-100 mM Na^+ , the maximal compaction of the wild-type 12-177 nucleosome array was achieved, with S-value at 49-50. The S-value of 12-177 nucleosome array with H4K20Me3 in 1 mM Mg^{2+} or 100 mM Na^+ was measured at ~50-51. The modified and wild-type nucleosome arrays are folded to almost the same degree in either 100 mM Na^+ or 1 mM Mg^{2+} . At this relatively high cation concentrations, both nucleosome arrays showed tendency of self-association. In literature (Correll et al., 2012; Lu et al., 2008), nucleosome arrays with longer NRLs are not able to fully fold in favorable cationic conditions. The presence of H4K20Me3 promote the compaction to moderate degree, linker histone seems indispensable for complete compaction. In comparison, nucleosome arrays with shorter NRLs (177 bp or shorter) are able to fully compact with only the presence of favorable salts (Routh et al., 2008). Therefore, the positive effect of H4K20Me3 on folding of the nucleosome arrays with long NRLs was not manifested.

Self-association of the modified (H4K20Me3) and unmodified 12-177 nucleosome array in the presence of Mg^{2+} and CoHex³⁺ was studied by precipitation assay. No difference was observed in the salt-induced precipitation between these two arrays. In coincidence with 177 bp nucleosome and chromatosome, Mg^{2+} induced resolubilization was observed for both modified and unmodified nucleosome arrays. The coincidence of ubiquitous distribution of Mg^{2+} in cytoplasm with the Mg^{2+} -specific two-way regulation of nucleosome array precipitation and resolubilization suggest a critical influence of Mg^{2+} in control of chromatin structures to

cope with various genomic activities. The resolubilization process for both types of the arrays proceeded under similar ionic conditions. H4K20Me3 exerts no change in charge properties of the nucleosome array. In the electrostatic view, the similar polyelectrolyte properties of the modified and unmodified 12-177 nucleosome arrays provide explanations to the conclusion that H4K20Me3 has no influence in the 12-177 nucleosome array self-association.

Similar properties of the wild-type and modified (H4K20Me3) arrays observed in AUC and PA studies support conclusion that trimethylation of the H4K20 exerts its role in biological processes by being binding site for specific proteins, rather than being able to cause changes in chromatin structure and dynamics by itself.

5 Perspectives

The author will continue with the cryo-EM project aiming to achieve the 3D structure of 177 bp chromosome with higher resolution. Focused local refinement (Bai et al., 2015) will be done on the triangle region with the RELION refinement program. The Volta phase plate will be used for in-focus data acquisition (Danev et al., 2014).

To elucidate linker histone H1⁰ positioning within the chromosome as well as specific interactions with the nucleosome, especially the location of CTD, site-specific cross-linking and hydroxyl radical foot printing experiments (Bednar et al., 2017) will be done.

The close spatial involvement as well as functional link (Luger et al., 1997; Stützer et al., 2016) between linker histone and histone H3 N-terminal tail suggested an alternative interpretation of the origin of the extra density in between the two linker DNAs. To verify this, H3 mutant with NTD deletion will be used for chromosome reconstitution in the cryo-EM single particle analysis.

It is suggested that long linker DNA is preferable for linker histone binding and more biologically relevant than shorter linker DNAs (Routh et al., 2008). 207 bp chromosome sample has been prepared. The solution SAXS experiments is in progress. Dimensions of the 207 bp nucleosome and chromosome will be compared to nucleosomes with shorter NRLs (145, 157, 177 bp). Cryo-EM single particle analysis will be carried out to reconstruct the 3D structures of 207 bp chromosome. To locate the binding position of linker histone CTD, the structures of 207 bp chromosome with full-length and CTD-deleted H1⁰ molecules will be reconstructed separately and compared with each other.

To further investigate the NRL-dependent effect of H4K20Me3 on chromatin fiber folding, nucleosome arrays with longer NRLs (e.g. 188, 197, 202, 207 bp) will be reconstituted and set-up for AUC velocity experiments. It would be interesting to determine if H4K20Me3 affects nucleosome array folding for various NRLs.

Findings of this thesis advance our understanding of the dependences of chromatin structure and dynamics on linker histone, trimethylation of H4K20, and nucleosome repeat lengths. The

reconstructed 3D structure of 177 bp chromatosome provides confidence and guidance in using cryo-EM to achieve atomic-resolution structure of nucleosome bound with full-length linker histone in the near future. The characteristic stacking of 177bp chromatosome under cationic conditions will enlighten future efforts in elucidating the mechanism of H1-mediated chromatin compaction. In sum, the linker histone related studies lead us one small step further in getting a complete picture of the associations of linker histone with several diseases, e.g. cancer, Alzheimer's disease, infertility and autoimmune diseases (Ye et al., 2017). The H4K20Me3 studies come up with new hypotheses - (A) NRL-dependent effect of H4K20Me3 on chromatin compaction; (B) H4K20Me3 acts as binding platform for specific protein factors regulating chromatin architecture. Furthermore, the studies on H4K20 trimethylation shed light on the cure for cancers which H4K20Me3 is closely relevant to, since the loss of H4K20Me3 is a characteristic marker of multiple human cancers (Fraga et al., 2005). Overall, this thesis will inspire deeper thought about the complexity of life and innocence of human beings and evoke one to re-position oneself in nature, society and the universe.

References

- Allahverdi, A., Yang, R., Korolev, N., Fan, Y., Davey, C.A., Liu, C.-F., and Nordenskiöld, L. (2010). The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Research*.
- Allan, J., Hartman, P.G., Crane-Robinson, C., and Aviles, F.X. (1980). The structure of histone H1 and its location in chromatin. *Nature* 288, 675-679.
- Allan, J., Mitchell, T., Harborne, N., Bohm, L., and Crane-Robinson, C. (1986). Roles of H1 domains in determining higher order chromatin structure and H1 location. *J Mol Biol* 187, 591-601.
- Andresen, K., Jimenez-Useche, I., Howell, S.C., Yuan, C., and Qiu, X. (2013). Solution Scattering and FRET Studies on Nucleosomes Reveal DNA Unwrapping Effects of H3 and H4 Tail Removal. *PLoS ONE* 8, e78587.
- Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A* 88, 10148-10152.
- Arents, G., and Moudrianakis, E.N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A* 92, 11170-11174.
- Bai, X.-c., Fernandez, I.S., McMullan, G., and Scheres, S.H.W. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* 2, e00461.
- Bai, X.-c., Rajendra, E., Yang, G., Shi, Y., and Scheres, S.H.W. (2015). Sampling the conformational space of the catalytic subunit of human γ -secretase. *eLife* 4, e11182.
- Beck, D.B., Burton, A., Oda, H., Ziegler-Birling, C., Torres-Padilla, M.E., and Reinberg, D. (2012). The role of PR-Set7 in replication licensing depends on Suv4-20h. *Genes Dev* 26, 2580-2589.
- Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., *et al.* (2017). Structure and Dynamics of a 197bp Nucleosome in Complex with Linker Histone H1. *Molecular Cell* 66, 384-397.e388.
- Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J., and Woodcock, C.L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proceedings of the National Academy of Sciences* 95, 14173-14178.
- Berezhnoy, N.V., Liu, Y., Allahverdi, A., Yang, R., Su, C.J., Liu, C.F., Korolev, N., and Nordenskiöld, L. (2016). The Influence of Ionic Environment and Histone Tails on Columnar Order of Nucleosome Core Particles. *Biophys J* 110, 1720-1731.
- Berger, S.L. (2002). Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* 12, 142-148.
- Bertin, A., Renouard, M., Pedersen, J.S., Livolant, F., and Durand, D. (2007). H3 and H4 Histone Tails Play a Central Role in the Interactions of Recombinant NCPs. *Biophysical Journal* 92, 2633-2645.
- Brown, D.T., Izard, T., and Misteli, T. (2006). Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nat Struct Mol Biol* 13, 250-255.
- Buratovich, M. (2014). Yale Scientists Find Marker for High-Quality Induced Pluripotent Stem Cells.

- Burlingame, R.W., Love, W.E., Wang, B.C., Hamlin, R., Nguyen, H.X., and Moudrianakis, E.N. (1985). Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3 Å. *Science* 228, 546-553.
- Carruthers, L.M., Bednar, J., Woodcock, C.L., and Hansen, J.C. (1998). Linker Histones Stabilize the Intrinsic Salt-Dependent Folding of Nucleosomal Arrays: Mechanistic Ramifications for Higher-Order Chromatin Folding†. *Biochemistry* 37, 14776-14787.
- Caterino, T.L., Fang, H., and Hayes, J.J. (2011). Nucleosome linker DNA contacts and induces specific folding of the intrinsically disordered H1 carboxyl-terminal domain. *Molecular and cellular biology* 31, 2341-2348.
- Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H.W., and Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* 135, 24-35.
- Cohen, G., and Eisenberg, H. (1968). Deoxyribonucleate solutions: Sedimentation in a density gradient, partial specific volumes, density and refractive index increments, and preferential interactions. *Biopolymers* 6, 1077-1100.
- Collaborative, B.-S.M. (2014). *Genome Structure*.
- Correll, S.J., Schubert, M.H., and Grigoryev, S.A. (2012). Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J* 31, 2416-2426.
- Cui, F., and Zhurkin, V.B. (2009). Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to AT-rich and methylated DNA. *Nucleic Acids Research* 37, 2818-2829.
- Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M., and Baumeister, W. (2014). Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the National Academy of Sciences* 111, 15635-15640.
- Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* 319, 1097-1113.
- de Frutos, M., Raspaud, E., Leforestier, A., and Livolant, F. (2001). Aggregation of Nucleosomes by Divalent Cations. *Biophysical Journal* 81, 1127-1132.
- Demeler, B. (2005). UltraScan: a comprehensive data analysis software package for analytical ultracentrifugation experiments. *Modern analytical ultracentrifugation: techniques and methods*, 210-229.
- Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T.J. (2003). Chromatin fiber folding: Requirement for the histone H4N-terminal tail. *Journal of Molecular Biology* 327, 85-96.
- Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. *Science* 306, 1571-1573.
- Eisenberg, H., and Felsenfeld, G. (1981). Hydrodynamic studies of the interaction between nucleosome core particles and core histones. *Journal of Molecular Biology* 150, 537-552.
- Evertts, A.G., Manning, A.L., Wang, X., Dyson, N.J., Garcia, B.A., and Collier, H.A. (2013). H4K20 methylation regulates quiescence and chromatin compaction. *Mol Biol Cell* 24, 3025-3037.

Fan, L., and Roberts, V.A. (2006). Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. *Proceedings of the National Academy of Sciences* 103, 8384-8389.

Fan, Y., Nikitina, T., Morin-Kensicki, E.M., Zhao, J., Magnuson, T.R., Woodcock, C.L., and Skoultchi, A.I. (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Molecular and cellular biology* 23, 4559-4572.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation. *Cell* 123, 1199-1212.

Fang, J., Feng, Q., Ketel, C.S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Simon, J.A., and Zhang, Y. (2002). Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Curr Biol* 12, 1086-1099.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. *Nature* 421, 448-453.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A* 73, 1897-1901.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. *Nature* 269, 29-36.

Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., *et al.* (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* 37, 391-400.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., *et al.* (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography* 50, 1212-1225.

Goll, M.G., and Bestor, T.H. (2002). Histone modification and replacement in chromatin activation. *Genes Dev* 16, 1739-1742.

Gozani (2017). RESEARCH | Gozani Lab.

Guinier, A. (1939). La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Ann Phys* 11, 161-237.

Hamiche, A., Schultz, P., Ramakrishnan, V., Oudet, P., and Prunell, A. (1996). Linker histone-dependent DNA structure in linear mononucleosomes. *J Mol Biol* 257, 30-42.

Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* 431, 1-12.

Hendzel, M.J., Lever, M.A., Crawford, E., and Th'ng, J.P. (2004). The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *J Biol Chem* 279, 20028-20034.

Howell, Steven C., Andresen, K., Jimenez-Useche, I., Yuan, C., and Qiu, X. (2013). Elucidating Internucleosome Interactions and the Roles of Histone Tails. *Biophysical Journal* 105, 194-199.

Huynh, V.A., Robinson, P.J., and Rhodes, D. (2005). A method for the in vitro reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone. *J Mol Biol* 345, 957-968.

- Izzo, A., Kamieniarz, K., and Schneider, R. (2008). The histone H1 family: specific members, specific functions ? . *Biological Chemistry* 389, 333-343.
- Jorgensen, S., Elvers, I., Trelle, M.B., Menzel, T., Eskildsen, M., Jensen, O.N., Helleday, T., Helin, K., and Sorensen, C.S. (2007). The histone methyltransferase SET8 is required for S-phase progression. *J Cell Biol* 179, 1337-1345.
- Jorgensen, S., Schotta, G., and Sorensen, C.S. (2013). Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res* 41, 2797-2806.
- Kasinsky, H.E., Lewis, J.D., Dacks, J.B., and Ausio, J. (2001). Origin of H1 linker histones. *Faseb J* 15, 34-42.
- Kim, J.-M., Kim, K., Punj, V., Liang, G., Ulmer, T.S., Lu, W., and An, W. (2015). Linker histone H1.2 establishes chromatin compaction and gene silencing through recognition of H3K27me3. *Scientific Reports* 5, 16714.
- Kinkade, J.M., Jr., and Cole, R.D. (1966). A structural comparison of different lysine-rich histones of calf thymus. *J Biol Chem* 241, 5798-5805.
- Klug, A., Rhodes, D., Smith, J., Finch, J.T., and Thomas, J.O. (1980). A low resolution structure for the histone core of the nucleosome. *Nature* 287, 509-516.
- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* 36, 1277-1282.
- Kornberg, R.D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* 184, 868-871.
- Korolev, N., Allahverdi, A., Lyubartsev, A.P., and Nordenskiold, L. (2012). The polyelectrolyte properties of chromatin. *Soft Matter* 8, 9322-9333.
- Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). ORC1 BAH domain links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 484, 115-119.
- Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Meth* 10, 584-590.
- Livolant, F., Mangenot, S., Leforestier, A., Bertin, A., Frutos, M.d., Raspaud, E., and Durand, D. (2006). Are liquid crystalline properties of nucleosomes involved in chromosome structure and dynamics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 364, 2615.
- Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* 276, 19-42.
- Lu, X., and Hansen, J.C. (2004). Identification of specific functional subdomains within the linker histone H10 C-terminal domain. *J Biol Chem* 279, 8701-8707.
- Lu, X., Simon, M.D., Chodaparambil, J.V., Hansen, J.C., Shokat, K.M., and Luger, K. (2008). The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nature structural & molecular biology* 15, 1122-1124.

- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251-260.
- Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. In *Methods in Enzymology* (Academic Press), pp. 3-19.
- Luque, A., Ozer, G., and Schlick, T. (2016). Correlation among DNA Linker Length, Linker Histone Concentration, and Histone Tails in Chromatin. *Biophysical Journal* 110, 2309-2319.
- Lyubitelev, A., Nikitin, D., Shaytan, A., Studitsky, V., and Kirpichnikov, M. (2016). Structure and functions of linker histones. *Biochemistry (Moscow)* 81, 213-223.
- Mangenot, S., Leforestier, A., Durand, D., and Livolant, F. (2003a). Phase Diagram of Nucleosome Core Particles. *Journal of Molecular Biology* 333, 907-916.
- Mangenot, S., Leforestier, A., Durand, D., and Livolant, F. (2003b). X-Ray Diffraction Characterization of the Dense Phases Formed by Nucleosome Core Particles. *Biophysical Journal* 84, 2570-2584.
- Mangenot, S., Leforestier, A., Vachette, P., Durand, D., and Livolant, F. (2002). Salt-Induced Conformation and Interaction Changes of Nucleosome Core Particles. *Biophysical Journal* 82, 345-356.
- Michael F. Smith, B.D.A., Shawn P. Williams, and John P. Langmore (1990). Radial density distribution of chromatin: evidence that chromatin fibers have solid centers. *The Journal of Cell Biology* 110, 245-254.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553-560.
- Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142, 334-347.
- Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A.S., Imamoto, N., Ishikawa, T., *et al.* (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO Journal* 31, 1644-1653.
- Pennings, S., Meersseman, G., and Bradbury, E.M. (1994). Linker histones H1 and H5 prevent the mobility of positioned nucleosomes. *Proc Natl Acad Sci U S A* 91, 10275-10279.
- Pesavento, J.J., Yang, H., Kelleher, N.L., and Mizzen, C.A. (2008). Certain and Progressive Methylation of Histone H4 at Lysine 20 during the Cell Cycle. *Molecular and cellular biology* 28, 468-486.
- Peterson, C.L., and Laniel, M.A. (2004). Histones and histone modifications. *Curr Biol* 14, R546-551.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605-1612.
- Ralston, G. (1993). *Introduction to analytical ultracentrifugation*, Vol 1 (Beckman California:).
- Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* 362, 219-223.
- Regha, K., Sloane, M.A., Huang, R., Pauler, F.M., Warczok, K.E., Melikant, B., Radolf, M., Martens, J.H., Schotta, G., Jenuwein, T., *et al.* (2007). Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. *Mol Cell* 27, 353-366.

Ricci, M.A., Manzo, C., Garcia-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145-1158.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**, 532-537.

Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008). 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. *Journal of Molecular Biology* **381**, 816-825.

Robinson, P.J.J., Fairall, L., Huynh, V.A.T., and Rhodes, D. (2006). EM measurements define the dimensions of the “30-nm” chromatin fiber: Evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences* **103**, 6506-6511.

Rogge, R.A. (2013). Assembly of nucleosomal arrays from recombinant core histones and nucleosome positioning DNA. *Journal Of Visualized Experiments: Jo Ve*.

Roque, A., Iloro, I., Ponte, I., Arrondo, J.L., and Suau, P. (2005). DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. *J Biol Chem* **280**, 32141-32147.

Routh, A., Sandin, S., and Rhodes, D. (2008). Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci U S A* **105**, 8872-8877.

Scaffidi, P. (2016). Histone H1 alterations in cancer. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1859**, 533-539.

Schalch, T. (2004). The 30-nm chromatin fiber: in vitro reconstitution and structural analysis. (Zurich: Swiss Federal Institute of Technology).

Scheres, S.H., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nature methods* **9**, 853-854.

Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180**, 519-530.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* **18**, 1251-1262.

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., *et al.* (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev* **22**, 2048-2061.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* **311**, 844-847.

Simon, M.D., Chu, F., Racki, L.R., de la Cruz, C.C., Burlingame, A.L., Panning, B., Narlikar, G.J., and Shokat, K.M. (2007). The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128**, 1003-1012.

Simpson, R.T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* **17**, 5524-5531.

- Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.-M., Zhu, P., and Li, G. (2014). Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science* **344**, 376-380.
- Stützer, A., Liokatis, S., Kiesel, A., Schwarzer, D., Sprangers, R., Söding, J., Selenko, P., and Fischle, W. (2016). Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Molecular Cell* **61**, 247-259.
- Syed, S.H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M.S., Hayes, J.J., Everaers, R., Angelov, D., Bednar, J., and Dimitrov, S. (2010). Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. *Proc Natl Acad Sci U S A* **107**, 9620-9625.
- Takahashi, H., Kojima, H., and Saito, H. (1985). A new site-specific endonuclease, Scal, from *Streptomyces caespitosus*. *Biochemical Journal* **231**, 229.
- Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46.
- Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol* **83**, 403-427.
- Vanholde, K.E., and Weischet, W.O. (1978). Boundary analysis of sedimentation-velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers* **17**, 1387-1403.
- Vasudevan, D., Chua, E.Y., and Davey, C.A. (2010). Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol* **403**, 1-10.
- Vila, R., Ponte, I., Collado, M., Arrondo, J.L., and Suau, P. (2001). Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA: an infrared spectroscopy study. *J Biol Chem* **276**, 30898-30903.
- Vila, R., Ponte, I., Jimenez, M.A., Rico, M., and Suau, P. (2000). A helix-turn motif in the C-terminal domain of histone H1. *Protein Sci* **9**, 627-636.
- White, A.E., Hieb, A.R., and Luger, K. (2016). A quantitative investigation of linker histone interactions with nucleosomes and chromatin. *Scientific Reports* **6**, 19122.
- Widom, J. (1998). Chromatin structure: linking structure to function with histone H1. *Curr Biol* **8**, R788-791.
- Widom, J., and Klug, A. (1985). Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. *Cell* **43**, 207-213.
- Wikipedia (2017). Linker histone H1 variants - Wikipedia.
- Williams, S.P., Athey, B.D., Muglia, L.J., Schappe, R.S., Gough, A.H., and Langmore, J.P. (1986). Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophys J* **49**, 233-248.
- Wolffe, A.P., and Hayes, J.J. (1999). Chromatin disruption and modification. *Nucleic Acids Res* **27**, 711-720.
- Woodcock, C.L., Frado, L.L., and Rattner, J.B. (1984). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J Cell Biol* **99**, 42-52.

- Woodcock, C.L., Skoultchi, A.I., and Fan, Y. (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* *14*, 17-25.
- Worcel, A., Strogatz, S., and Riley, D. (1981). Structure of chromatin and the linking number of DNA. *Proc Natl Acad Sci U S A* *78*, 1461-1465.
- Yang, C., van der Woerd, M.J., Muthurajan, U.M., Hansen, J.C., and Luger, K. (2011). Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. *Nucleic Acids Res* *39*, 4122-4135.
- Ye, X., Feng, C., Gao, T., Mu, G., Zhu, W., and Yang, Y. (2017). Linker Histone in Diseases. *International Journal of Biological Sciences* *13*, 1008-1018.
- Yue, H., Fang, H., Wei, S., Hayes, J.J., and Lee, T.-H. (2016). Single-Molecule Studies of the Linker Histone H1 Binding to DNA and the Nucleosome. *Biochemistry* *55*, 2069-2077.
- Zhou, B.-R., Feng, H., Ghirlando, R., Li, S., Schwieters, C.D., and Bai, Y. (2016). A Small Number of Residues Can Determine if Linker Histones Are Bound On or Off Dyad in the Chromatosome. *Journal of Molecular Biology* *428*, 3948-3959.
- Zhou, B.R. (2013). Structural insights into the histone H1-nucleosome complex. *Pnas* *110*, 19390-19395.
- Zhou, B.R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T.S., and Bai, Y. (2015). Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell* *59*, 628-638.
- Zlatanova, J., and Doenecke, D. (1994). Histone H1 zero: a major player in cell differentiation? *Faseb J* *8*, 1260-1268.

Appendix

- 177-601 DNA sequence (The 147-601 sequence is highlighted in blue)

acttacgcgccgccCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCT
TAAACGCACGTACGCGCTGTCCCCGCGTTTTAAACGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGT
GTCAGATATATACATCCTGTgcatgtattgaaagt

- 202-601 DNA sequence (The 147-601 sequence is highlighted in blue)

tatgtgatggaccctatacgcgccgccCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCT
AGCACCGCTTAAACGCACGTACGCGCTGTCCCCGCGTTTTAAACGCCAAGGGGATTACTCCCTAGTCTC
CAGGCACGTGTCAGATATATACATCCTGTgcatgtattgaacagcgaccttgccgg

- hH2A (130 a.a. MW=14136)

MSGRGKQGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAHVLEYLTAEILELAGNAA
RDNKKTRIIPRHLQLAIRNDEELNKLGRVTIAQGGVLPNIQAVLLPKKTESHHKAKGK*

- hH2B (126 a.a. MW=13890)

MPEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRSRKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFE
RIAGEASRLAHYNKRSTITSREIQTAVRLLPGELAKHAVSEGTKAVTKYTSK*

- hH3 (136 a.a. MW=15404)

MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLFPQRLVR
EIAQDFKTDLRFQSSAVMALQEACEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA*

- hH4 (103 a.a. MW=11368)

MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTY
TEHAKRKTVTAMDVVYALKRQGRTLYGFGG*

- hH1⁰ (NTD is in red, CTD is in orange)

MHHHHHHSSGLEVLFGPTENSTSAPAAKPKRAKASKSTDHPKYSDMIVAAIQAEKNRAGSSSRQSIQKYI
KSHYKVGENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEPKKSVAFKKTKEIKKVATPKKASKPKK
AASKAPTKKPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVKPAKSSAKRAGKKK

- hH1⁰ gene sequence

ccatgggc ATG CAT CAT CAC CAC CAT CAC AGC AGC GGT CTG GAA GTG CTG TTT CAG GGC CCA
ACC GAG AAT AGC ACG AGC GCG CCA GCC GCA AAG CCG AAA CGT GCA AAG GCC AGC AAG AAG
AGC ACC GAT CAT CCG AAA TAT AGC GAT ATG ATT GTG GCC GCC ATT CAG GCG GAA AAG AAC
CGT GCG GGC AGC AGC CGC CAG AGC ATC CAG AAG TAC ATT AAG AGC CAC TAC AAG GTG GGC
GAG AAC GCG GAT AGC CAG ATT AAA CTG AGC ATC AAG CGT CTG GTG ACG ACC GGT GTG CTG
AAG CAG ACG AAA GGC GTT GGC GCG AGC GGC AGC TTT CGC CTG GCG AAA AGC GAC GAA CCA
AAA AAG AGC GTG GCG TTC AAG AAG ACG AAG AAG GAG ATC AAG AAG GTG GCG ACG CCG
AAA AAG GCG AGC AAG CCA AAG AAG GCG GCA AGC AAG GCA CCA ACC AAG AAG CCG AAA
GCC ACG CCG GTT AAG AAG GCC AAG AAA AAA CTG GCA GCC ACC CCG AAG AAA GCG AAG AAA
CCA AAG ACC GTG AAG GCG AAG CCA GTT AAA GCG AGC AAA CCG AAA AAA GCC AAG CCG GTG
AAA CCG AAG GCA AAA AGC AGC GCG AAG CGT GCA GGT AAA AAG AAA taagctgagc