

Structural studies of the nucleosome, assembled with different histones, DNA fragments and HMGN proteins, and derivatized with ruthenium anticancer agents

Reyhan Muhammad

2013

Reyhan Muhammad. (2013). Structural studies of the nucleosome, assembled with different histones, DNA fragments and HMGN proteins, and derivatized with ruthenium anticancer agents. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/53656>

<https://doi.org/10.32657/10356/53656>

**STRUCTURAL STUDIES OF THE NUCLEOSOME,
ASSEMBLED WITH DIFFERENT HISTONES, DNA
FRAGMENTS AND HMGN PROTEINS, AND
DERIVATIZED WITH RUTHENIUM ANTICANCER
AGENTS**

REYHAN MUHAMMAD

SCHOOL OF BIOLOGICAL SCIENCES

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

2013

Acknowledgement

Graduate study is truly a massive battle between enlightening passion and motivating dream against discouraging reality. At the beginning of this battle, a graduate student joined the fight with the dream of shaping the reality according to student's dream. As the student would immediately realize, the reality most of the time did not obey the student's dream. The student tried to fight back so many times, yet the reality still prevailed. Only till the moment of desperation the student realized that reality is not the enemy. Reality is a track that keeps a train fueled by student's passion and dream on the right path. Graduate study is not meant to show which side is stronger, it is to show that passion, dream, and reality are meant to be harmoniously together in life.

I would like to thank my supervisor Dr. Curt Alexander Davey for the opportunities to work in his lab. I am very grateful for all the guidance, motivational supports, and degree of freedom he gave me during my doctoral study. His extraordinary passion for his nucleosome research has inspired me the most important characteristics a scientist should possess.

I would like to thank all members of Davey's structural genomic laboratory. I am indebted to Wu Bin, Dileep, and Kareem for showing me doing experiment is an art. Their help and friendship since the very beginning of my research in Davey group, back since I was an undergraduate student, mean a lot to me. I am also grateful to Hong Tian, Michelle, Zenita, Eugene, Gabriella, and Joey for all their experimental help and advice, and on top of all that for their companion in the laboratory.

I would like to thank people from Prof. Grüber's, Prof. Susana's, Prof. Lars', and Prof. Dröge's laboratories for their help with experiment and enjoyable moment sharing the lab life with them. I am also grateful to all friends and colleagues in School of Biological Sciences. This school has been fun and memorable.

Last but not least, my eternal gratefulness for my family. I will never be able to thank them enough in my entire life. My father and mother, who always love, support, advice, and be there for me in all situations, and my sister for her unselfishness to me, her ignorant brother. My wife, Agustina Kadaristiana who is always cheering me and showing me her love is always unconditioned. They are irreplaceable for me. This thesis and the struggle are dedicated to them. I hope this will bring them happiness.

Table of Contents

Acknowledgement	2
Table of Contents	3
List of Figures	7
List of Tables	11
Abbreviations.....	12
Abstract	14
Introduction.....	16
1.1. Chromatin Architecture	17
1.1.1. Nucleosome Core Particle.....	17
1.1.2. Centromeric Nucleosome Core Particle.....	22
1.1.3. Telomeres.....	24
1.1.4. Higher Order Chromatin	25
1.2. Chromatin Regulation.....	26
1.2.1. Nucleosome Organization and Dynamics	26
1.2.2. Histone Variants	28
1.2.3. Histone Post-translational Modification	30
1.2.4. Nucleosome Remodelling Complex.....	32
1.3. High Mobility Group Protein Family	32
1.3.1. High Mobility Group A Protein	33
1.3.2. High Mobility Group B Protein	33
1.3.3. High Mobility Group N Protein.....	34
1.4. Chromatin as anticancer drug target.....	37
1.5. Scope of this thesis	38
Materials and Methods	40
2.1. Palindromic DNA Production	40

2.1.1. DNA constructs and Plasmids.....	41
2.1.2. Cell transformation.....	41
2.1.3. Plasmid extraction	42
2.1.4. EcoRV Digestion and Purification of DNA Insert.....	43
2.1.5. Large Scale Dephosphorylation	43
2.1.6. HinfI/EcoRI Digestion and Purification of Half-palindromic Repeats by FPLC.....	44
2.1.7. Ligation and Purification of palindromic DNA.....	44
2.2. Histone Octamer Preparation	45
2.2.1. Cell transformation.....	45
2.2.2. Histone protein Overexpression.....	45
2.2.3. Inclusion Body Preparation	46
2.2.4. FPLC Purification of Histones	46
2.2.5. Histone Octamer Refolding.....	47
2.2.6. FPLC Purification of Histone Octamer.....	47
2.3. Nucleosome Core Particle Reconstitution	47
2.3.1. Off-centered NCP Reconstitution	47
2.3.2. Centered NCP Reconstitution.....	48
2.3.3. NCP Purification.....	48
2.4. HMGN Production	48
2.5. Crystal Preparation.....	49
2.6. Crystal Processing.....	52
2.6.1. Crystallographic Data Collection	52
2.6.2. Crystallographic Data Analysis	52
2.7. Electrophoretic Mobility Shift Assay (EMSA)	52
2.8. Western Blot	53
2.9. Crosslinking	53
2.10. Fluorescence Assay.....	53

Nucleosome Core Particle Production and Crystallization.....	55
3.1. Palindromic DNA Production	55
3.2. Histone Octamer Production	56
3.3. Nucleosome Core Particle Reconstitution	59
3.4. Nucleosome Core Particle Crystallization	62
Nucleosome Core Particle Structure.....	63
4.1. Overall NCP Structure	63
4.1.1. <i>Xenopus laevis</i> NCP145 and Human NCP145 structure	68
4.1.2. Human NCP146b structure	74
Ruthenium Agent Interaction with Nucleosome Core Particle	77
5.1. Result.....	77
HMGN Interaction with Nucleosome Core Particle	87
6.1. Result.....	87
6.1.1. HMGN triggers variant-specific conformational changes in nucleosome structure without apparent loss of histone proteins	87
6.1.2. Higher stoichiometry of HMGN is observed on S2 species	89
6.1.3. Binding assessment of HMGN NBD truncation series to Nucleosome Core Particle	91
6.1.4. Identification of the NBD's histone binding partner	94
6.1.5. Crystallization of HMGN-NCP Complex.....	98
6.1.6. Crystallization of an HMGN NBD-NCP Complex	100
6.1.7. Searching for NCP's divalent ion-free crystallization buffers	104
Discussion.....	105
7.1. Human and <i>Xenopus laevis</i> Nucleosomes adopt nearly identical structures	105
7.2. Steric hindrance governs RAPTA compound-NCP interactions.....	106
7.3. Arene ring and PTA ligand confer site selectivity for RAPTA binding.....	107
7.4. RAPTA compound association to <i>Xenopus laevis</i> NCP and human NCP	111
7.5. RAPTA binding sites overlap with a nuclear protein binding site on the NCP surface	113

7.6. HMGN binding to Nucleosome	114
7.7. NBD Binding Characterization.....	116
7.8. Crystallization Strategy	119
Appendix	124
References.....	128

List of Figures

Figure 1. Hierarchical organization of chromatin.....	17
Figure 2. Structure of NCP	19
Figure 3. Translational and rotational parameters for DNA bp steps.	20
Figure 4. Major groove and minor groove bending for nucleosomal DNA	21
Figure 5. Proposed models for centromeric NCP structure.....	23
Figure 6. Centromeric NCP crystal structure.....	24
Figure 7. Chromatin higher order structure.....	26
Figure 8. Histone post-translational modifications..	31
Figure 9. HMGN binding to chromatin structure.....	34
Figure 10. Domain organization of HMGN proteins.	35
Figure 11. HMGN-NCP interaction.....	37
Figure 12. Scheme for Palindromic DNA production	40
Figure 13. Gel electrophoresis result for each step of palindromic DNA production... ..	55
Figure 14. FPLC-MonoQ (ionic exchange chromatography) purification profile of half-palindromic repeats.....	56
Figure 15. FPLC-MonoQ (ionic exchange chromatography) purification profile of final palindromic DNA	56
Figure 16. HiLoad 16/60 Superdex 200 (size exclusion chromatography) – Histone purification profile.	57
Figure 17. ResourceS (ionic exchange chromatography) – Histone purification profile.	58
Figure 18. HiLoad 16/60 Superdex 200 (size exclusion chromatography) – HO purification profile.....	59
Figure 19. NCP products reconstituted with different DNA-to-HO ratios.	60
Figure 20. Two NCP positioning species after reconstitution.....	61
Figure 21. Difference between off-centered and centered NCP	61

Figure 22. NCPs reconstituted from different histones (human and frog histones) and DNAs (145 bp, 147 bp, 146b bp, 601L-145 bp).....	62
Figure 23. Standard NCP crystal morphology from $MnCl_2$ -KCl crystallization buffers.....	62
Figure 24. NCP structures with one half of the particle shown.	65
Figure 25. Positions of amino acid substitutions between <i>Xenopus laevis</i> and Human NCP.	66
Figure 26. Amino acid differences on both <i>Xenopus laevis</i> and human histones are discernible in the crystal structures..	67
Figure 27. Two non-conserved amino acids of H2B make direct contact with DNA as shown by NCP crystal structures.....	68
Figure 28. Nucleosomal DNA bp parameters for human NCP145 and <i>Xenopus laevis</i> NCP145.....	70
Figure 29. Location of DNA stretching in <i>Xenopus laevis</i> NCP145 (SHL +1.5) and human NCP145 (SHL +5)..	71
Figure 30. Local DNA conformational difference between <i>Xenopus laevis</i> NCP145 and human NCP145 at SHL -5..	72
Figure 31. Difference in manganese localization in human NCP145 and <i>Xenopus laevis</i> NCP145.....	73
Figure 32. Nucleosomal DNA bp step parameters for human NCP146b and <i>Xenopus laevis</i> NCP146b.....	75
Figure 33. B-factor distribution for DNA of human NCP146b and <i>Xenopus laevis</i> NCP146b.....	76
Figure 34. Structure of RAPTA compounds.....	77
Figure 35. Two consensus binding sites for RAPTA compounds on the surface of <i>Xenopus laevis</i> NCP145.....	79
Figure 36. RAPTA-B-treated NCP shows anomalous signal on nucleosomal DNA at nucleotide -15 (chain J)	80
Figure 37. Interaction of RAPTA compounds with H2A Glu 61 and H2A Glu 64 (Site A).....	84
Figure 38. Interaction of RAPTA compounds with H2B His 106 and H2B Glu 102 (Site B).	85
Figure 39. RAPTA-B treatment yields an anomalous signal on the symmetry-related location of Site B (Site B').....	86

Figure 40. Formation of distinct slow-migrating species (S2) triggered by truncated HMGN1 and HMGN2, but not HMGN1 in the presence of divalent ion.....	87
Figure 41. Western blot analysis of HMGN - NCP interactions.	88
Figure 42. Optimization of fluorescence assay for HMGN2 – NCP studies.....	89
Figure 43. Fluorescence analysis of the S2 species.....	91
Figure 44. EMSA results of NCP interaction with full length NBD (XN11 and XN21) and C-terminal truncated NBD (XN11,XN12,XN13).....	
Figure 45. SDS PAGE gel of glutaraldehyde-crosslinked NCP – NBD samples.....	94
Figure 46. SDS PAGE of glutaraldehyde crosslinking of H2A-H2B dimer and H3-H4 tetramer with NBD peptides..	95
Figure 47. Western Blot analysis of glutaraldehyde-treated NCP-NBD complexes....	97
Figure 48. HMGN1- NCP Crystals produced in 0.1 M Sodium citrate tribasic dihydrate pH 5.5, 18% w/v Polyethylene glycol 3,350 (PEGRx Tm 1 No.29, Hampton Research)	99
Figure 49. Fusion NCP after reconstitution.....	100
Figure 50. Fusion N1-H2A and N1-H3 NCP crystals	100
Figure 51. XN11-NCP co-crystals produced in low-ionic strength crystallization buffer and EMSA result	102
Figure 52. NCP packing in low-salt crystals.....	103
Figure 53. NCP crystals produced in divalent ion-free crystallization buffers.....	104
Figure 54. Different accessibility of RAPTA binding sites between the two NCP faces.	107
Figure 55. Important amino acids that shape the unique binding site for RAPTA compounds at Site A and Site B	109
Figure 56. Displacement of the arene ring of RAPTA-H at Site A from the consensus position.....	111
Figure 57. Superimposed structures of <i>Xenopus laevis</i> NCP and human NCP at Site B. Both structures are from NCP crystals treated with 0.75 M RAPTA-I for two days.	113
Figure 58. H2A glutamic acid 61 and 64, and H2B glutamic acid 102 involved in RAPTA binding also interact with other nuclear regulatory proteins such as Sir3 and RCC1. These amino acids are also recognized by Kaposi Sarcoma LANA antigen, and mediate chromatin condensation through interaction with the H4 tail.....	114

Figure 59. Hypothesis on S2 formation.....	116
Figure 60. Sequence alignment of the HMGN1 NBD and HMGN2 NBD peptide truncation series.	118
Figure 61. Prediction of HMGN1 NBD binding site on the NCP based on an HMGN1 NBD-NCP crosslinking - western blot profile	119
Figure 62. Location of potential NBD binding sites inside NCP crystals	123

List of Tables

Table 1. Divalent salt - KCl Crystallization Screens	50
Table 2. Spermine - KCl Crystallization Screen	51
Table 3. Data collection and refinement statistics for <i>Xenopus laevis</i> NCP145 and human NCP145 and NCP146b.....	64
Table 4. RAPTA compound identity	77
Table 5. Data collection and refinement statistics for six RAPTA-treated NCP crystals.	81
Table 6. Positions for anomalous signals of Ruthenium atoms of RAPTA compounds.	83
Table 7. Intensity measurements of 1:1 complexes and S1 complexes.	90
Table 8. Western blot densitometry measurements for quantification of degree of histone-NBD peptide crosslinking.....	98
Table 9. Data collection and refinement statistics for low-salt NCP-XN1-2 crystals .	103

Abbreviations

72bp+3	72 base pairs with 3 base overhangs
9bp+3	9 base pairs with 3 base overhangs
ACR complex	ATP-dependent chromatin remodelling complex
Amp	ampicillin
bp	base pairs
CAF1	chromatin assembly factor 1
Camp	chloramphenicol
CenNuc	centromeric nucleosome
CenNCP	centromeric nucleosome core particle
CHUD	chromatin unfolding domain
CIA	chloroform-Isoamyl alcohol
CIAP	calf intestine alkaline phosphatase
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
H1, H2A, H2B, H3, H4	histone proteins
HAT	histone acetyl transferase
HDAC	histone deacetylase
HIRA	histone regulator A
HMG	high mobility group protein
HMGA	high mobility group A protein
HMGB	high mobility group B protein
HMGN	high mobility group N protein
HO	histone octamer

FPLC	fast protein liquid chromatography
IPTG	isopropyl-thiogalactopyranosid
KCaco	Potassium Cacodylate
MME	monomethyl ether
MPD	2-Methyl-2,4-pentanediol
MWCO	molecular weight cutoff
NBD	nucleosome binding domain
NCP	nucleosome core particle
NFR	nucleosome-free region
NLS1, NLS2	nuclear localization signal 1 and 2
NMR	nuclear magnetic resonance
NRC	nucleosome remodelling complex
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethanesulphonyl fluoride
PTA	1,3,5-triaza-7-phosphaadamantane
RAPTA	Ruthenium - Arene - PTA
RD	regulatory domain
RNA	ribonucleic acid
RNaseI	ribonuclease I
S-200	SAUDE-200
SDS	sodium dodecyl sulfate
SHL	superhelical location
TB	terrific broth
TBE	tris-Borate-EDTA
TE	tris-EDTA
TEMED	tetramethylethylenediamine
TY	tryptone-yeast extract

Structural Studies of the Nucleosome, Assembled with Different Histones, DNA Fragments and HMGN Proteins, and Derivatized with Ruthenium Anticancer Agents

Abstract

The nucleosome forms the basic unit of chromatin in eukaryotic cells. This compact structure made of four histone proteins and DNA serves dual interrelated functions for storage and regulation of genomic activities. Nucleosome dynamics, histone variants, histone post-translational modification, nucleosome positioning and occupancy are integral in determining the status of the cell. Structural and biochemical studies of the nucleosome have granted a detailed mechanistic view for understanding the activity of this interesting structure inside the cell. We are interested in studying how nucleosomes interact with other nuclear proteins inside the cell and how we can exploit the nucleosome for anticancer drug development.

Xenopus laevis nucleosome core particle has been widely used as a model for nucleosome studies, benefiting from an atomic understanding of its structure. Even though histones and thus nucleosome structure are commonly conserved throughout evolution, we aim to develop a nucleosome platform more relevant for biomedical studies— human nucleosome core structures. We showed that human nucleosome core particle assembled from different DNA fragments can be crystallized to give high resolution structures. Furthermore, we did preliminary studies on the interaction of ruthenium-based compounds with *Xenopus laevis* and human nucleosomes, and showed that both gave satisfying results.

We conducted crystallographic analysis for several derivatives of novel ruthenium-based anticancer drugs, termed RAPTA compounds, with regard to their interaction with the nucleosome. We worked on the principle of optimizing a lead RAPTA-nucleosome structure in order to improve binding affinity and specificity towards the nucleosome. The balance between providing new groups for enhancing RAPTA interaction and avoiding steric clashes with the nucleosome structure is an important determinant in designing new RAPTA compounds. We found that the smallest RAPTA compound (RAPTA-B), with an unsubstituted arene ring, still maintains similar binding selectivity compared to larger RAPTA compounds. Interestingly, the small

arene group allows RAPTA-B to access a site on the nucleosome that is sterically restricted due to internucleosome interactions.

Nuclear proteins ranging from ATP-dependent chromatin remodelers, histone chaperones, histone modifying enzymes, and structural proteins assist the regulation of chromatin activities. We are interested in studying high mobility group N (HMGN) proteins due to their abundance inside the nucleus, their importance for genomic activities, and their specific interaction with the nucleosome core. Our group showed distinct structural changes on nucleosome structure elicited by different types of HMGN proteins. Several biochemical assays were conducted to investigate the nature of HMGN binding to the nucleosome with the goal of explaining how HMGN proteins may work inside the nucleus. Crystallizations of nucleosome with HMGN proteins have also been conducted, and novel potential crystallization buffers for nucleosomal studies have been discovered.

Chapter I

Introduction

Life depends on the expression, regulation, maintenance, and transmission of genetic material. The history of genetic material dates back to 1865 when Gregor Johann Mendel described mathematically, for the first time, the pattern of inheritance. He introduced the concept of gene variation (genotype) and the observed characteristics of organisms (phenotype). Subsequent experiments, notably by Oswald T. Avery and colleagues, identified DNA as the responsible inheritance factor. They showed DNA not only passable from one generation to the next, but also its transforming capability across the organism (Avery *et al.*, 1944). In 1953, James D. Watson and Francis Crick, with help from Rosalind Franklin and Maurice Wilkins, determined the general structure of DNA. Their double helix structure opened the world of molecular genetics. Their DNA basepairing concept proposed many breakthrough ideas, including how DNA is replicated and how the DNA message is translated by cells (Watson and Crick, 1953).

For proper cell activity, there has to be a system whereby genome expression is regulated. This regulation system is tightly associated with a maintenance system as the cell needs to store its huge genome into the size of a cell or nucleus, yet easily access its genome whenever required for appropriate genome transactions. One example is in *Escherichia coli* where the circular DNA is associated with several protein factors, such as histone-like HU proteins, DNA topoisomerase and DNA gyrase, which turn the DNA into a compact supercoiled structure (Sinden and Pettijohn, 1981). When specific genes need to be expressed, *Escherichia coli* DNA sequences will project out from the supercoiled structure into the cytoplasm and become accessible for transcription (Durrenberger *et al.*, 1988).

The eukaryotic genome is divided into several different chromosomes; each is a single, enormously long linear DNA. Each chromosome undergoes ~500-fold compaction in size during interphase, raised to ~10000-fold compaction during mitosis. In order to achieve this, DNA associates with several protein factors, and this interaction successively coils and folds the DNA into higher levels of chromatin organization (Figure 1). Each layer in the compaction hierarchy adds extra levels of chromatin regulatory control (Felsenfeld and Groudine, 2003).

To begin with, eukaryotic genome compaction starts with the coiling of ~147 basepairs (bp) of DNA around a protein complex, the histone octamer (HO), to form the nucleosome core (Davey *et al.*, 2002). Linker DNA of variable length (from 10-80 bp) connects neighbouring nucleosome core regions to form a “beads-on-a-string” structure, a hallmark of chromatin in an open state, euchromatin (Oudet *et al.*, 1975). Higher compaction states of chromatin, meanwhile, require additional factors. Linker histones (H1), modifications on histone protein tails (Eberharter and Becker, 2002) and the nucleosome surface (Zhou *et al.*, 2007), and the specific length of nucleosome spacing (Routh *et al.*, 2008) are involved in higher order structure formation. At the moment, compaction structure up to the level of 30-nm chromatin fibre has been elucidated (Robinson and Rhodes, 2006; Schalch *et al.*, 2005). Further experiments are required to explain higher compaction states.

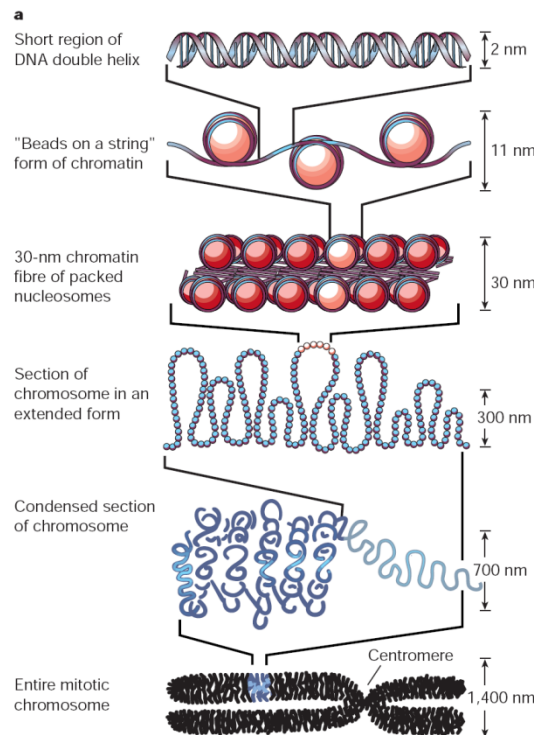


Figure 1. Hierarchical organization of chromatin (adapted from Felsenfeld and Groudine, 2003)

1.1. Chromatin Architecture

1.1.1. Nucleosome Core Particle

The fundamental repeating unit of chromatin is called the nucleosome (Figure 2). An individual nucleosome encompasses both free linker DNA and the nucleosome core

particle (NCP). While linker DNA may vary between 10-80 bp, the NCP is associated with ~147 bp DNA, which is wrapped in 1.67 left-handed superhelical turn around the HO. HO itself consists of 2 copies of each core histone protein; H2A, H2B, H3, and H4 (Richmond and Davey, 2003).

There are 14 binding sites that mediate histone protein-DNA direct interactions, occurring when the DNA minor grooves faces inward toward the HO core. To help designation of location along nucleosomal DNA, a term called Superhelical Location (SHL) is used. SHL is numbered as integer when the DNA major groove faces toward the HO, with SHL-0 being located at nucleosome dyad (Figure 2). Direct DNA-histone interactions are designated as SHL ± 0.5 - ± 6.5 (non-integer number) on the nucleosome structure. H2A-H2B dimers are engaged in direct interaction with DNA at SHL ± 3.5 to ± 6.5 , while the H3-H4 tetramer contacts nucleosomal DNA at SHL -2.5 to SHL +2.5. The binding involves both electrostatic and van der Waals interactions. Electrostatic interactions are formed by DNA phosphate groups with histone amide groups and side chains of serine, threonine, arginine, or lysine. This electrostatic interaction is supported by the following van der Waals interactions between the histones and the hydrocarbon portion of the DNA phosphodiester backbone. There are about 116 direct interactions and 121 water-mediated interactions made between histone and DNA (Davey *et al.*, 2002). Around 83% of water-mediated interactions allow distant histone side chains to reach the DNA, while the rest of them support the already formed histone-DNA direct interactions (Davey *et al.*, 2002). Both direct and indirect interactions give important contributions to nucleosome stability.

Direct interactions involving the phosphodiester group of DNA at SHL positions allow histones to hold DNA tightly and at the same time give histones freedom to bind DNA in a sequence-independent manner. Although the NCP crystal structure suggests that ten arginine side chains (Figure 2b) make direct contact with cytosine O2 and adenine and guanine N3 atoms (Davey *et al.*, 2002), it does not dictate strictly what nucleotide should be incorporated on that position. Different physical properties (for example, DNA bendability) of different DNA base steps, however, cause a certain DNA sequence preference for nucleosomes formation, thus giving a certain degree of sequence-dependent nucleosome positioning to chromatin (Satchwell *et al.*, 1986; Segal *et al.*, 2006).

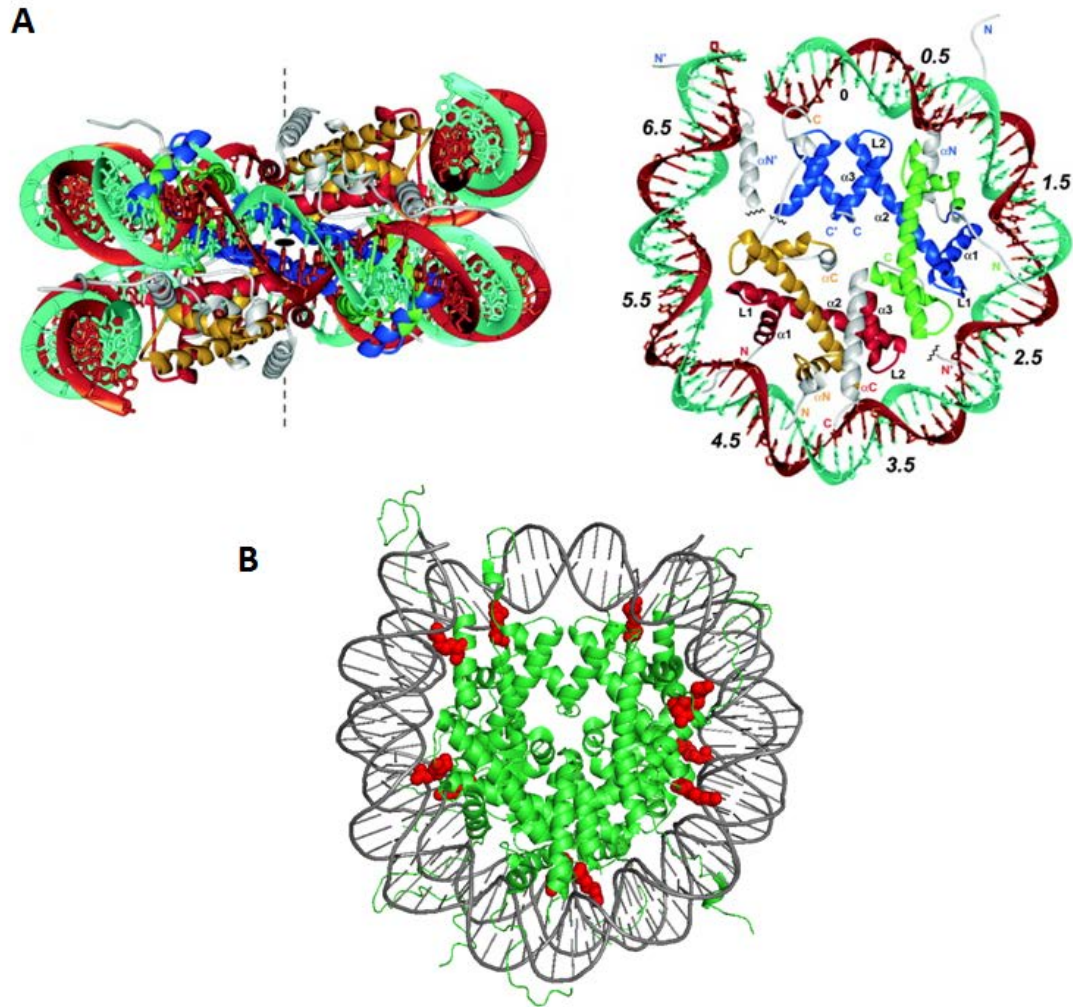


Figure 2. (A) Structure of NCP as shown from its 2-fold pseudo-symmetry dyad axis (left), and along its superhelical axis (right). For clarity only a half particle is shown for image on the right. Seven interaction sites (SHL 0.5 - 6.5) between DNA and HO are numbered accordingly (adapted from Davey *et al.*, 2002). (B) At 10 out of 14 DNA-histone direct interaction sites, arginine (red) is found to penetrate into DNA minor groove and establish direct interactions with the base nucleotide.

1.1.1.1. Nucleosomal DNA Geometry

DNA coiling around HO noticeably results in the distortion of DNA conformation from the ideal B-form DNA. Nucleosomal DNA, for example, has different average twist (~10.2 bp/turn) compared to free B-form DNA (~10.5 bp/turn) (Luger *et al.*, 1997). When DNA bends, both DNA major and minor grooves facing the histone core must narrow due to associated compression, while the grooves facing outside become wider. DNA curvature on the superhelix is also non-uniform. This deviation from ideal superhelix structure is mainly due to (1) the anisotropic flexibility of DNA, (2) local

structural features intrinsic to the DNA sequence, and (3) irregularities dictated by the underlying HO. Several conformational parameters are introduced to describe DNA geometry on NCP (Figure 3).

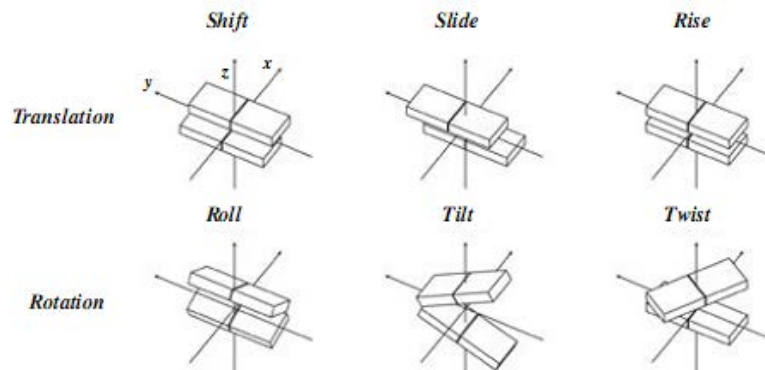


Figure 3. Translational and rotational parameters for DNA bp steps. Shift, slide, and rise are translations along the x, y, z axes, respectively. Roll, tilt, and twist are rotations about the x, y, z axes respectively (adapted from Richmond and Davey, 2003).

Blocks of DNA major groove and minor groove, each consisting of five bp, alternately face the histone core structure. DNA roll (compression at the groove edge) and DNA tilt (compression at the phosphodiester edge) are the factors contributing to DNA curvature on the histone surface. Compression at the major groove edge (positive roll) is more favourable than compression at the minor groove edge (negative roll) due to the tighter sterical hindrance of the latter. This is evident from the uniformity of curvature in the bent major groove blocks as compared to one in the minor groove (figure 4a).

In order to accommodate less favourable minor groove compression, there are at least two modes of minor groove bending observed in the nucleosome core. Highly flexible bp steps (CA=TG bp step) in the centers of minor groove blocks give rise to localized kinks at this particular bp step, and its pronounced negative roll makes a major contribution to wrapping (Figure 4c). Alternatively, in the absence of highly flexible DNA basepair step, negative roll is distributed over three to four basepair steps, and is accompanied by alternating shift of bp into the major and minor groove in order to prevent steric clash at the narrowed minor groove edge (figure 4b) (Richmond and Davey, 2003). In one special case, for short 145 bp nucleosomal DNA, however, extreme kinking in the minor groove can also be coupled with DNA stretching (Ong *et al.*, 2007).

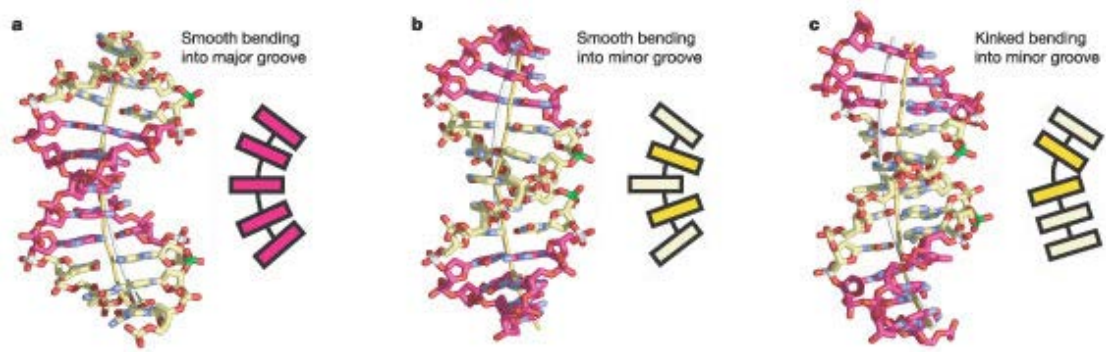


Figure 4. Major groove and minor groove bending for nucleosomal DNA (adapted from Richmond and Davey, 2003)

1.1.1.2. Histone Octamer

Histone core proteins are amongst the most highly conserved eukaryotic proteins known. This strong restriction indicates that almost all of their amino acids are involved in their function, and mutation of any residue may impose abnormalities to the organism. Histone core proteins possess highly charged tails enriched in lysine and arginine residues, which help charge neutralization of DNA and facilitate DNA compaction. Under physiological conditions in the absence of DNA, H2A and H2B form a heterodimer whereas H3 and H4 form a tetramer. Only under high salt conditions or in the presence of DNA do the dimers and tetramer assemble together to form HO.

Structurally, each core histone can be separated into four domains; the central histone-fold domain, N-terminal tail, C-terminal tail, and histone-fold extension domain. The central histone-fold domain is constructed from three α -helices that are connected by two loops, denoted as $\alpha 1$ -L1- $\alpha 2$ -L2- $\alpha 3$. This domain is thought to contain the primary site responsible for specific dimerization of H2A-H2B and H3-H4. Moreover, the C-halves of $\alpha 2$ and $\alpha 3$ from two H3 proteins are responsible for H3-H3 association, thus bringing two H3-H4 dimers to form the whole $(\text{H3-H4})_2$ tetramer. Finally, to complete HO structure H2A and H4 associate to form two four-helix bundles on each side of the $(\text{H3-H4})_2$ tetramer. Further interactions occur within an octamer to stabilize the final structure, such as H3-H2A', H4-H2A', H4-H2B', H2A-H2A', and H2A-H2B' (Luger *et al.*, 1997).

Assembly of NCP *in vivo* begins with the association of (H3-H4)₂ tetramer with DNA, followed by the association of two molecules of H2A-H2B dimer to the (H3-H4)₂-DNA complex in order to complete NCP formation. Histone chaperones and assembly factors modulate this process *in vivo*.

1.1.1.3. Linker Histone

Unlike core histone proteins, linker histone H1 is found in a ratio of one copy for about 200 bp of genomic DNA, suggesting the stoichiometry of 1:1 for H1:nucleosome. This linker histone interacts with the terminal DNA of the nucleosome core through its central globular domain, and with linker DNA through its C-terminal domain. The interaction condenses chromatin and limits regulatory protein access to nucleosomal DNA.

Among histone proteins, H1 has the highest number of subfamilies. Although different subtypes are differentially expressed in different cells and at different developmental stages, they still show some redundancy in function as deletion of one subtype can be compensated by overexpression of other types. Deletion of three or more H1 subtypes, however, is lethal for the cell. This may suggest that basal structural changes mediated by the H1 family may suffice to support a dynamic network of chromatin-binding proteins, as long as the H1 amount is adequate inside the cell (Thomas, 1999).

1.1.2. Centromeric Nucleosome Core Particle

The centromere region on chromatin is responsible for faithful chromosomal segregation during cell division. This region serves as an attachment site of the newly-replicated chromosome (sister chromatid) to the kinetochore complex, which eventually will pull sister chromatids apart from each other to two opposite spindle poles. Centromeric DNA is usually a long tandem repeat of a 160-180 bp sequence, which in human is called alpha-satellite DNA. This sequence was initially thought to be the determining factor for centromere localization, signalling where centromeric proteins will eventually accumulate and perform their functions. However, subsequent observations indicated that alpha-satellite DNA is not necessarily associated with centromeres, and in some case, centromeres can even form on DNA sequences

lacking the alpha-satellite sequence. Hence the role of DNA in centromere localization is still unclear (Verdaasdonk and Bloom, 2011).

The main mark for centromere is the incorporation of a different H3 variant into the centromeric nucleosome (CenNuc). This H3 variant is called CENP-A in mammals, CID in flies, and Cse4 in budding yeast. CenNuc is hypothesized to adopt a different structure than the ubiquitous nucleosome, which subsequently underlies the CenNuc ability to perform centromeric functions. Several findings reported possible structures of CenNuc, and six models have been proposed for it (figure 5) (Furuyama and Henikoff, 2009; Lavelle *et al.*, 2009; Mizuguchi *et al.*, 2007; Sekulic *et al.*, 2010; Williams *et al.*, 2009)

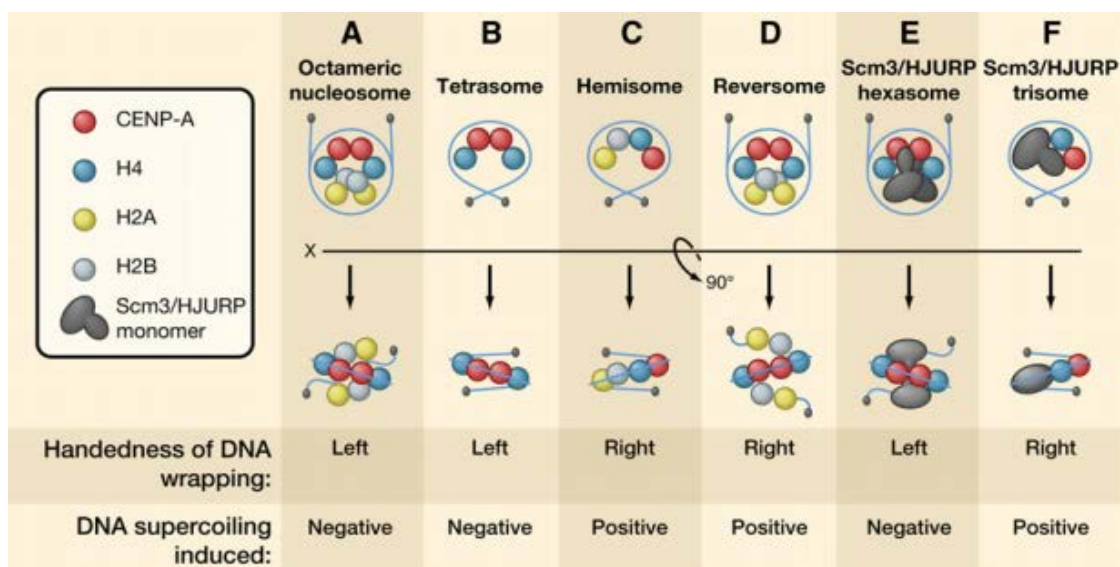


Figure 5. Proposed models for centromeric NCP structure (adapted from Black and Cleveland, 2011)

The most direct explanation for CenNuc structure perhaps comes from the crystal structure of centromeric nucleosome core particle (CenNCP) (Figure 6). In this model, CenNCP contains two copies of each histone H2A, H2B, CENP-A, and H4. DNA wraps in a left-handed superhelix around the HO. The shorter α N helix of CENP-A, in comparison to the one of H3, fails to stabilize the last 13 bp of DNA termini, hence CenNCP maintains only the central 121 bp DNA on its surface. Moreover, two extra amino acid residues, Arg 80 and Gly 81, on the loop 1 region of CENP-A are responsible for the protrusion of the loop out from the core structure (figure 6b). This protruding loop is important for CENP-A retention on centromeres, and perhaps also

provides a binding platform for other factors to recognize CenNCP (Tachiwana *et al.*, 2011).

CenNucs are organized into numbers of interspersed blocks in the centric heterochromatin. Each block contains about one thousand CenNucs, and these blocks are separated from each other by blocks of ubiquitous nucleosomes. This organization allows CenNucs to be positioned on the surface of centric heterochromatin, where they can interact with other proteins required for kinetochore assembly on chromatin (Verdaasdonk and Bloom, 2011).

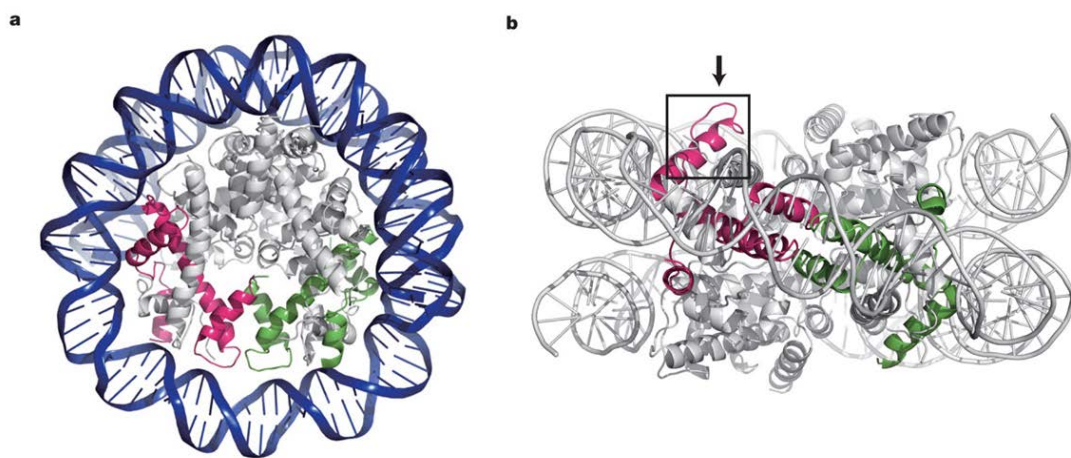


Figure 6. Centromeric NCP as shown looking down the superhelical axis (a). Only 121 bp DNA (blue) is found in the crystal structure. Loop 1 region of CENP-A stick out out from the NCP structure (b) (adapted from Tachiwana *et al.*, 2011)

1.1.3. Telomeres

Eukaryotic linear chromosomes contain a special feature at their termini. These termini, termed telomeres, contain long tandem repeats of short sequences (10-15 kbp of TTAGGG repeat in human). The strand that is G-rich runs 5'-3' towards the tail, and becomes single stranded at the end to form a so-called 3'-end tail. The presence of 3'-end tails poses two main challenges for chromatin maintenance; the protection of the tail from being recognized as a DNA break by the chromatin surveillance system, and the 3'-end replication problem which results in 3'-end tail shortening (Zakian, 1995).

Shelterin, a complex of six protein components, helps to maintain telomeric DNA regions. One of its components, Pot1, specifically binds to single stranded 3' tails (Lei *et al.*, 2004). In addition to shelterin, 3' tails can also penetrate double strand telomeric regions, and force base pairing with the corresponding C-rich strand, thus creating a structure called t-loop (Griffith *et al.*, 1999). Another form of secondary structure, which consists of stacked layers of four guanines basepairing with each other (G-quadruplex), may also form at the 3' tail (Lipps and Rhodes, 2009). Altogether they serve as mechanisms to protect 3'-ends from being recognized as DNA break sites.

In order to overcome chromosome shortening after every cycle of replication, an enzyme is employed to replenish DNA at the end of telomeres. This enzyme, called telomerase, is a reverse transcriptase that uses its own RNA template for the synthesis of new DNA on 3' tails (Gillis *et al.*, 2008). Telomerase is active mostly during the embryonic stage, where in normal mature cells only a minute amount of telomerase is still found active. Increased activity of this enzyme is found in cancer cells (Shay and Bacchetti, 1997).

Nucleosomes are also found in telomeric regions. The nucleosomes are regularly and closely positioned with linker DNA only about 10 bp long. Due to the peculiar nature of telomeric DNA sequences, telomeric nucleosomes are intrinsically more mobile than the ubiquitous ones. TRF1, one of the shelterin components, has been found to induce nucleosome sliding on telomeric sequence *in vitro* (Pisano *et al.*, 2010).

1.1.4. Higher Order Chromatin

Early observations show that *in vivo* nucleosomes undergo further compaction into a much thicker chromatin fiber. The diameter of this more compact fiber is ~30nm, which is about 3-fold the diameter of the nucleosome. Several factors, including length of the linker DNA, the presence of linker histone, and modifications on the histone core, are responsible for 30-nm chromatin fiber regulation.

Depending on the length of linker DNA between adjacent nucleosome cores, chromatin fiber could come in two different conformations; zigzag two-start helix type or solenoid one-start helix type (Figure 7). Short linker DNA (~20 bp) favours formation of zigzag two-start helix as observed from both the crystal structure of

tetranucleosome and electron microscopy images of longer nucleosome arrays. On the other hand, longer linker DNA favours the solenoid one-start helix. Judging from the distribution of linker DNA length inside the cell, the solenoid one-start helix is perhaps the most common structure. There is higher incidence of 40-50 bp linker DNA *in vivo*, while short 20 bp linker DNA is relatively rarer (Robinson and Rhodes, 2006; Schalch *et al.*, 2005).

Tight stacking of nucleosomes in chromatin fiber is mediated by histone tails, notably the tail of H4. Deletion of H4 tail, for example, inhibits salt-dependent compaction of chromatin fiber (Robinson *et al.*, 2008). In addition to that, specific acetylation on H4 Lysine 16, which is associated with active gene transcription, abolishes chromatin compaction (Shogren-Knaak *et al.*, 2006). Another factor, such as linker histone, plays roles in compaction by dictating the path of DNA as it exits from the nucleosome core (Robinson *et al.*, 2008).

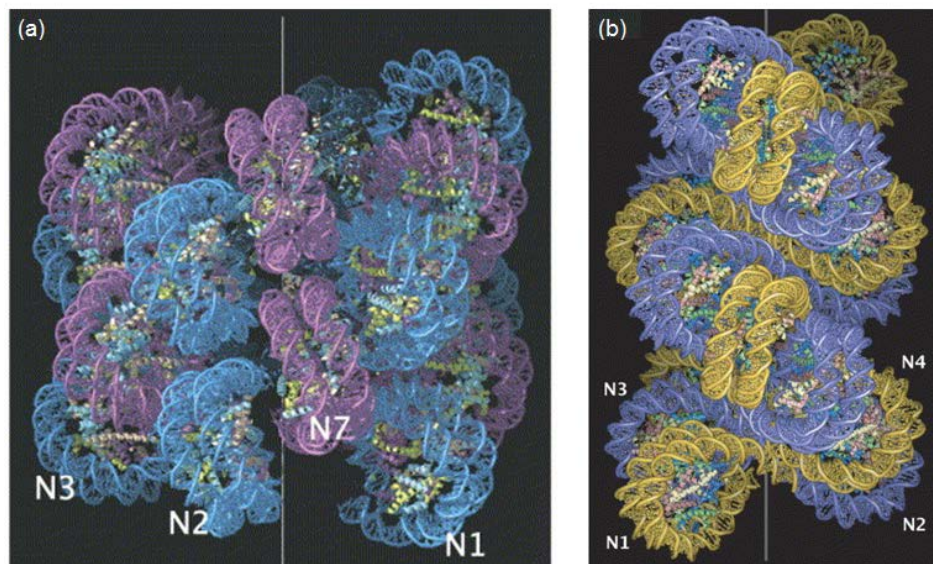


Figure 7. Chromatin higher order structure. Solenoid one-start helix model (a) and zig-zag two-start helix model (b) are depicted (adapted from Robinson and Rhodes, 2006).

1.2. Chromatin Regulation

1.2.1. Nucleosome Organization and Dynamics

The repressive nature of nucleosome for DNA-templated processes suggests that nucleosome organization on chromatin is important for appropriate chromatin function. On the very basic level, the DNA sequence itself can dictate to a certain degree the allocation of nucleosomes. Analysis of several DNA sequences on the nucleosome,

for example, demonstrates a periodicity of certain dinucleotide sequences of about 10 bp, which coincides with the periodicity of alternating minor and major grooves facing the histone core. A|T sequences which prefer narrow minor grooves tend to be positioned at sites where the minor groove undergoes compression on nucleosome (i.e. when this groove faces the core histone), hence the position of nucleosome along DNA is dictated (Satchwell *et al.*, 1986). Moreover, rigid sequences such as poly dA:dT tracks disfavour nucleosome formation, and this particular sequence can be found on many yeast promoters which are devoid of nucleosomes (Anderson and Widom, 2001).

The DNA sequence alone, however, is not sufficient to fully explain nucleosome positioning (Jiang and Pugh, 2009; Zhang *et al.*, 2009). DNA sequence for example cannot explain why nucleosome position is precisely assigned for the first nucleosome on the gene, but not for the subsequent nucleosomes. To explain this observation, the concept of statistical positioning is then introduced (Mavrich *et al.*, 2008). In this concept, DNA sequence on one particular site may also affect how nucleosomes are positioned on other sites. Sequences that disfavour nucleosome formation, such as Poly dA:dT tracts, can promote highly positioned nucleosomes over neighbouring locations. This sequence type is thought to serve as a nucleosome barrier, and for the sequence just next to this barrier the nucleosome is highly positioned. Using this model, a nucleosome free region (NFR) prior to one gene is thought to impose a highly positioned nucleosome at the start of the gene. Subsequent nucleosomes along the gene show reduced positioning as their positions are getting further away from NFR. However, even with this model, *in vitro* nucleosome reconstitution on genomic DNA failed to reproduce *in vivo* nucleosome organization. Perhaps new factors other than DNA sequences need to be considered in addition to statistical positioning models.

Nucleosome active remodelling occurs *in vivo*. This ATP-dependent process modulates nucleosome dynamics along DNA, hence suggesting its importance for several genetic transactions. However, only recently it is found that the Nucleosome Remodelling Complex (NRC) also plays a significant role in nucleosome positioning. *In vitro* nucleosome reconstitution successfully reproduced *in vivo* nucleosome arrangements only when the remodelling complex and ATP were introduced into the system. It is hypothesized that the remodelling process overrides nucleosome sequence preferences and allows several strongly locked nucleosomes (i.e. due to the underlying DNA sequence) to slide along the DNA and re-arrange themselves (Zhang

et al., 2011). In support to that, another recent finding reports the absence of previously-observed dinucleotide periodicity on highly positioned nucleosomes *in vivo*, suggesting that DNA sequence is not the main determining factor for strongly positioned nucleosome (Valouev *et al.*, 2011).

Even though NRC is largely responsible for nucleosome positioning inside the cell, other factors can still contribute towards nucleosome organization. DNA methylation can determine nucleosome occupancy for different DNA sequences. Combinations of histone post-translational modifications and histone variants can recruit different NRCs, which subsequently lead to different nucleosome organizations. Histone variants may also have different DNA preferences compared to ubiquitous histone proteins for nucleosome positioning. The binding of certain DNA factors or stably positioned nucleosomes can serve as a barrier which then promotes statistical positioning of nucleosomes. And finally higher order chromatin structure may also favour certain nucleosome organizations in order to promote compaction (Segal and Widom, 2009).

External factors are not the sole determinant that grants dynamics to nucleosomes. Even the nucleosome by itself can undergo spontaneous conformational changes (Poirier *et al.*, 2009). These changes on the surface of the nucleosome facilitate transient access to the terminal DNA. This site exposure, which happens on the range of seconds to microseconds, is not restricted to single nucleosomes but extends to chromatin arrays, even when they are in a compact state (Poirier *et al.*, 2009). Binding of one factor to the terminal DNA of a transiently disassembled nucleosome could stabilize the open state and allow other regulatory factors to further access the inner DNA which was previously buried inside the core structure (Tims *et al.*, 2011).

1.2.2. Histone Variants

In addition to canonical core histone proteins, cells also employ a range of histone core variants. H2A variants include H2A.Z, H2A.X (Redon *et al.*, 2002), macroH2A, and H2A-Bbd (Bao *et al.*, 2004). H3 variants include H3.3 and the already-mentioned CENP-A. Differences in their sequences from their ubiquitous histone counterparts endow the histone variants with different properties in chromatin factor binding and nucleosome stability.

The H3.3 variant in most animals only differs by four amino acid substitutions from native H3 proteins. However this minor difference, especially three amino acids on the

$\alpha 2$ helix that are responsible to contact H2A-H2B in native NCP, seems to be important as H3.3 NCP exhibit reduced salt-dependent stability compared to ubiquitous NCP (Jin and Felsenfeld, 2007). Inside the cell of higher eukaryotes, deposition of H3.3 into nucleosome is replication-independent and is mediated by different histone chaperones (histone regulator A, HIRA) from one that deposits H3 into nucleosome (chromatin assembly factor 1, CAF1) during DNA synthesis. H3.3 NCP is associated with transcribed genes, promoters, and gene regulatory elements (Tagami *et al.*, 2004).

Both H2A.Z and H2A.X are reported to be involved in several processes, such as gene activation and silencing, DNA repair, and chromatin fiber formation. In most of their functions, their identities are not the only determining factor, but also their post-translational states inside the cell. Notably H2A.Z has been identified on most promoter regions in yeast and human cells, where H2A.Z-containing nucleosomes are involved in RNA polymerase II recruitment. H2A.X on the other hand is best known for its involvement in DNA repair. Phosphorylation of H2A.X in response to double-strand breaks helps to recruit DNA repair proteins, histone modifying enzymes, and chromatin remodelling complexes (Adam *et al.*, 2001; Altaf *et al.*, 2009).

Macro H2A (mH2A) and H2A Barr body-deficient (H2A.Bbd) have distinct features on their C-terminus with regard to canonical H2A. mH2A has a C-terminus macrodomain (200 residues), while H2A.Bbd has a truncated C-terminal domain. The presence of a macrodomain, perhaps due to the steric hindrance imposed, reduces transcription factor and H1 binding to chromatin. The enrichment of mH2A on Xi female chromosome and inhibition of polyADP-ribose polymerase by mH2A further highlight a mH2A role as transcriptional silencer inside the cell (Abbott *et al.*, 2005; Costanzi and Pehrson, 2001).

In contrast to mH2A, H2A.Bbd is associated with active chromatin. The protein shares only 48% identity with the ubiquitous H2A. H2A.Bbd nucleosomes only protect 110-130 bp of nucleosomal DNA from micrococcal nuclease digestion due to truncation of its C-terminus domain (Eirin-Lopez *et al.*, 2008). Moreover H2A.Bbd NCP has a smaller acidic patch than native H2A, which negatively affects chromatin compaction (Zhou *et al.*, 2007).

1.2.3. Histone Post-translational Modification

Another factor that contributes to chromatin regulation is histone post-translational modification. Several modifications ranging from acetylation, methylation, phosphorylation, ubiquitination, sumoylation, deimination, ADP ribosylation, crotonylation, β -N-acetylglucosamine modification, histone tail clipping, and histone proline isomerisation can influence chromatin regulation both directly and indirectly (Bannister and Kouzarides, 2011; Tan *et al.*, 2011). These modifications occur both on histone tails and histone globular domains. Several major modification sites on histones are summarized in Figure 8.

Direct alteration of chromatin structure due to histone post-translational modification can be achieved in many ways. Acetylation and phosphorylation of highly basic histone tails remove positive charge, hence altering their interaction with negatively-charged DNA. Acetylation of H3K56, for example, increases nucleosomal DNA breathing by seven-fold around the termini due to the disruption of histone-DNA contacts (breathing is used to describe spontaneous nucleosomal DNA unwrapping-rewrapping) (Neumann *et al.*, 2009). In addition, acetylation can also affect inter-nucleosome interaction as shown by inhibition of chromatin fiber formation due to H4K16 acetylation (Robinson *et al.*, 2008; Shogren-Knaak *et al.*, 2006). Furthermore, a direct role of histone modification on nucleosome structure also includes the effect of large molecule attachment to nucleosomes by ubiquitination and sumoylation. Ubiquitin and SUMO proteins presumably work like a wedge that can reshape nucleosome conformation (Sun and Allis, 2002). Ultimately, change on nucleosome structure induced by the presence of histone modification may both facilitate and inhibit access of different DNA binding factors to DNA, hence affecting genomic activities (Seet *et al.*, 2006).

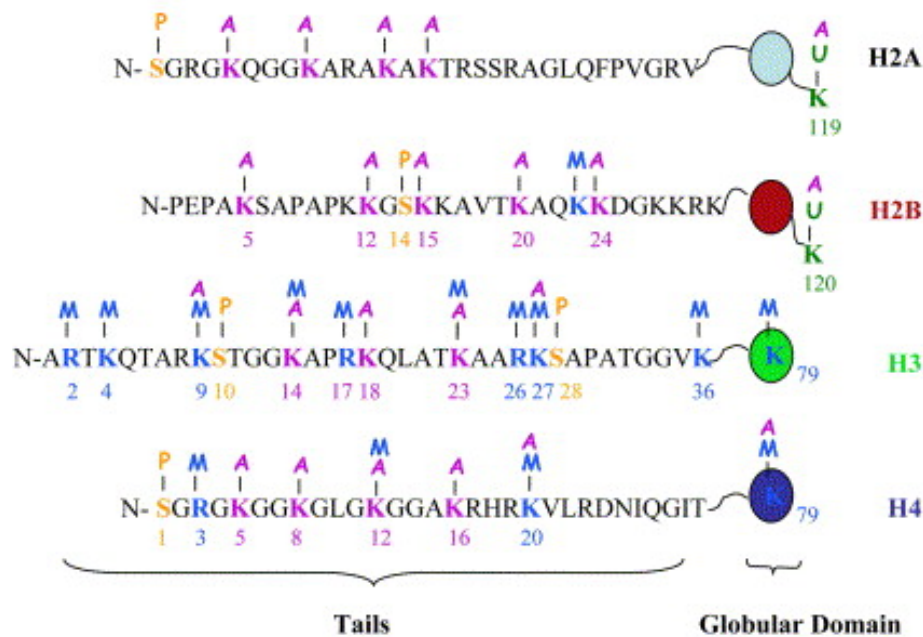


Figure 8. Histone post-translational modifications. Several major modifications on histone (Phosphorylation, P; Acetylation, A; Methylation, M; Ubiquitination, U) and their site of occurrence are shown (adapted from Santos-Rosa and Caldas, 2005).

Post-translational modifications on histones may generate specific recognition sites for many other protein factors, including nucleosome remodelling complexes. Several domains have been identified as modified-histone binders, such as bromodomains, chromodomains, PHD domains, and Tudor domains. Histone modification-bridged interactions between nucleosome and protein factors subsequently initiate distinct biological processes (Bannister and Kouzarides, 2011).

The reversible nature of histone post-translational modifications grants huge control over chromatin regulation. Specific biological processes can be triggered by the level of histone modifications present on the chromatin. Trimethylation of H3K4, for example, is associated with active transcription (Santos-Rosa *et al.*, 2002). Moreover, similar modification on the same amino acid on the histones may give different outcomes depending on the position of that particular nucleosome in the genome. Set2-mediated methylation of H3K36 is present within open reading frames of actively transcribed genes. Its mislocalization to promoter regions, however, inhibits gene transcription (Landry *et al.*, 2003). Histone modifications may also crosstalk with each other, promoting or inhibiting subsequent modifications on histones. In addition to that, multivalency of histone modifications on one particular amino acid creates a dynamic

competition for protein factors binding, hence again allowing diversity of possible biological outcomes (Bannister and Kouzarides, 2011).

1.2.4. Nucleosome Remodelling Complex

Internal nucleosomal DNA requires major nucleosome remodelling in order to facilitate chromatin factor binding. This remodelling process is facilitated by several ATP-dependent chromatin remodelling complexes (ACR complexes). On the basis of their sequence and structure, there are four major families of ACR complexes; SWI/SNF, ISWI, CHD, and INO80 complexes. The SWI/SNF complex may transiently expose DNA regulatory sites by creating DNA loops on the nucleosome surface. Isw2, a member of the ISWI family, and Ino80 help nucleosome sliding processes, which in turn expose or hide critical DNA regulatory sites. SWR1, another member of INO80 families, and Chd1 from the CHD family assist replacement of H2A and H3 by H2A.Z and H3.3, respectively. Altogether these mechanisms may control DNA accessibility to different regulatory factors and mark certain DNA locations for specific activities (Hargreaves and Crabtree, 2011; Jiang and Pugh, 2009).

Greater access to nucleosomal DNA can be achieved by complete eviction of nucleosomes from the underlying DNA. Conversely, nucleosome deposition can help turn down access to nucleosomal DNA. The RSC complex, homologs of the SWI/SNF complex, helps the eviction of nucleosomes in addition to other remodelling capabilities (Schwabish and Struhl, 2007).

1.3. High Mobility Group Protein Family

The High Mobility Group (HMG) protein family is abundantly and ubiquitously expressed in eukaryotic cells. They are members of nuclear structural proteins that bind to chromatin fiber and induce conformational change in chromatin structure, hence affecting several DNA-dependent processes. *In vivo* observations reveal HMG protein dynamics. They continuously move among chromatin-binding sites where they associate only transiently. However, since the travel time is shorter than the binding time, most of HMG proteins are usually found to be associated with chromatin at any given instance.

HMG proteins and other nuclear proteins form dynamic interactions inside the nucleus. They may compete or cooperate with each other for a particular binding site on

chromatin, thus influencing which chromatin activity is going to occur. This interaction is reversible and can be shifted to a new equilibrium by changing the state of the proteins and their binding sites, such as acetylation of histone proteins, as well as changing the concentration of the HMG proteins. The versatility of these kinds of events confers a broad control of chromatin activity.

There are three groups of protein under the HMG protein family, named HMGA, HMGB, and HMGN. These proteins have different structural motifs and target different sites on chromatin. Mutations in these proteins have been known to cause several diseases, signifying their importance in cell function (Gerlitz *et al.*, 2009).

1.3.1. High Mobility Group A Protein

The High Mobility Group A (HMGA) protein family consists of four proteins; HMGA1a, HMGA1b, HMGA1c, and HMGA2, where the first three result from alternative splicing of the HMGA1 gene (Reeves and Beckerbauer, 2001). Each protein contains three basic DNA-binding domains, called AT hooks, and an acidic C-terminal region with unknown function. HMGA proteins bind preferably to the minor groove of AT-rich DNA sequences, a process that is facilitated by two minor-groove-penetrating arginine residues on 'AT hook' domains (Huth *et al.*, 1997). This DNA-HMGA interaction induces DNA bending, which is important for several DNA-dependent activities. Indeed, *in vivo*, HMGA serves as one of the transcription factors that regulate transcription processes both positively and negatively. HMGA proteins are mainly expressed during embryonic development, and their abnormal expression has been associated with some tumor formations (Reeves *et al.*, 2001).

1.3.2. High Mobility Group B Protein

The High Mobility Group B (HMGB) protein family is by far the most abundant family of HMG proteins. There are three canonical HMGB variants in metazoan cells, namely HMGB1, HMGB2, and HMGB3, and several other HMGB-related proteins having some truncations in the domain architecture. This protein family contains two box domains, designated as A and B domains, followed by an acidic extension on the C-terminal region. Both A and B domains have similar overall structure but with some substitutions of amino acids at important sites that result in different bending abilities. The B domain is able to bend linear DNA, while the A domain fails to do so effectively. Nonetheless, both domains show binding preferences towards distorted DNA and bind to the DNA minor groove. The acidic patch at the C-terminus interacts with both

box domains and reduces the HMGB binding affinity for double-stranded DNA, whereas it enhances the affinity for distorted DNA (Thomas and Travers, 2001).

In vivo, HMGB proteins seem to interact with linker DNA and subsequently produce a kink that affects the global chromatin structure (Nightingale *et al.*, 1996). In this way, HMGB may influence several chromatin-related activities, such as transcription and recombination (Thomas and Travers, 2001). In addition, HMGB proteins may also interact with several nuclear factors and help their assembly and binding to DNA (Agresti *et al.*, 2005). Apart from their nuclear activity, recently it has been discovered that HMGB proteins also function as extracellular signaling molecules (Lotze and Tracey, 2005).

1.3.3. High Mobility Group N Protein

High Mobility Group N (HMGN) proteins are the only family among HMG proteins known to have a preference toward binding the nucleosome core relative to naked DNA. HMGN binding results in an open conformation of chromatin, thus facilitating access of other regulatory factors to the DNA. The HMGN-induced open state of chromatin has been demonstrated clearly using electron microscopy; circular minichromosomes became larger in diameter upon HMGN introduction, which is an indication of partial release of DNA from the nucleosome core without a change in nucleosome number (Figure 9) (Catez *et al.*, 2003). This open structure, however, does not necessarily mean free access of nucleosomal DNA to any of the regulatory factors. *In vitro* remodeling processes by a number of ATP-dependent remodeling factors have been shown to be inhibited in the presence of HMGN (Rattner *et al.*, 2009). In addition, it seems that HMGN-NCP complexes also have higher stability than the NCP alone, as observed from thermal denaturation experiments (Yau *et al.*, 1983).

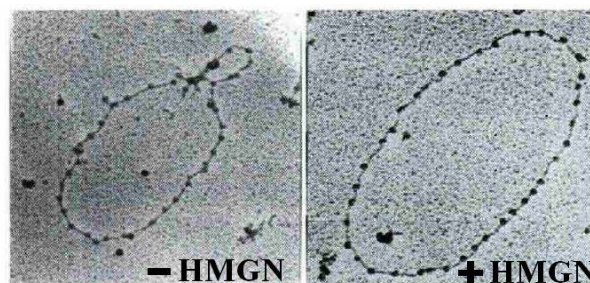


Figure 9. HMGN binding to chromatin structure. In the absence of HMGN, the minichromosome diameter is smaller than one in the presence of HMGN. The increase in diameter, but not the number of nucleosomes in the minichromosomes, indicates decompaction of nucleosome structure (adapted from Catez *et al.*, 2003).

In vivo, only two HMGN molecules of the same variant can bind to an individual nucleosome. Clustering of HMGN-loaded nucleosomes is variant-specific, giving rise to an average of ~six HMGN-nucleosomes, responsible for decompaction of the chromatin structure in that particular region. Decompaction could result from direct action of HMGN on nucleosome structure, from post translational modification of the core histones promoted by HMGN binding, or by counteracting H1-nucleosome binding (Postnikov *et al.*, 1995).

To date, there are five HMGN proteins that have been identified; HMGN1, HMGN2, HMGN3, HMGN4, and NSBP1. The HMGN proteins contain four conserved domains: starting from the N-terminus is the nuclear localization signal (NLS-1) domain, followed by the nucleosome-binding domain (NBD), a second nuclear localization signal (NLS-2), and a regulatory domain (RD, previously named as CHUD) at the C-terminus. Mutation in any of these domains could render the protein inactive. NBD mutations decrease binding affinity and specificity of HMGN for the nucleosome, while NLS-2 and RD mutations affect several downstream processes, such as histone post translational modifications. In addition, several studies show RD-deficient HMGN is unable to unfold chromatin and to stimulate transcription (West *et al.*, 2004).

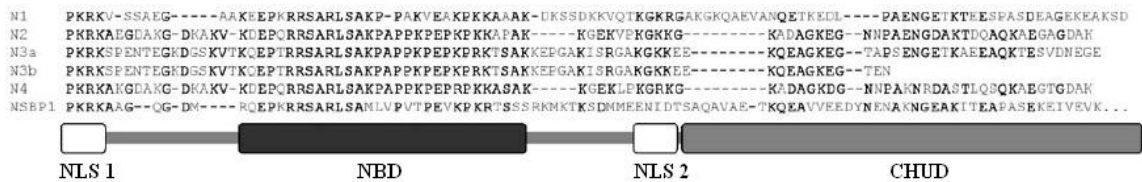


Figure 10. Domain organization of HMGN proteins (adapted from West and Bustin, 2004). The CHUD domain was renamed to the Regulatory Domain (RD) in recent published works.

Being the most conserved domain in HMGNs, the NBD is responsible for two important properties of HMGN binding. This 30-amino acid-long component is the minimum motif required for nucleosome binding (Crippa *et al.*, 1992), and its interaction with nucleosomes is responsible for HMGN variant-specific association with chromatin *in vitro* (Postnikov *et al.*, 1995). Furthermore, mutation studies of the NBD successfully identified residues critical for nucleosome-specific association. There are eight amino acids, RRSARLSA, which are absolutely conserved in all HMGNs, and embedded in this sequence are four residues, R-S-RL, which are absolutely essential for nucleosome binding. Mutation in any of these four residues abolishes nucleosome association both *in vivo* and *in vitro* (Ueda *et al.*, 2008).

In vivo, HMGN is involved in several chromatin-related activities. HMGN proteins enhance transcription of nucleosomal DNA but not naked DNA substrates. HMGN-deficient mutants are viable but suffer from increased sensitivity to DNA stress damage, consistent with the recent findings on HMGN involvement in global genome repair mechanisms (Subramanian *et al.*, 2009). A comprehensive review of HMGN mutation phenotypes has been summarized elsewhere (Hock *et al.*, 2007).

Lack of specificity in binding to any particular DNA sequence makes HMGN a suitable candidate as a general chromatin-opening factor inside the nucleus. Indeed about 50% of HMGN is found to be associated with chromatin-processing megaprotein complexes. Association of HMGN with such assemblies helps target HMGN to particular chromatin sites, where it can exert an influence on chromatin structure. Alternatively, free HMGN, upon binding to a target site, could be stabilized by association with other proteins, thus prolonging HMGN occupancy and its influence on that site (Lim *et al.*, 2002).

Several biochemical studies reveal possible binding mechanisms of HMGN to the nucleosome (Figure 11). DNAase I footprinting, hydroxyl radical footprinting and protein-DNA crosslinking uncovered HMGN-DNA interactions at positions 25 bp from the DNA termini. These results also expose potential binding of HMGN near the NCP dyad axis, as shown by the protection at 65 and 75 bp from the DNA termini. In addition to the DNA, HMGN proteins make apparent contacts with histone proteins, in particular with the N-terminal tail of H3 (Ueda *et al.*, 2006; West *et al.*, 2004). Details regarding the specific mode and mechanism of nucleosome recognition for HMGN binding, however, are still lacking.

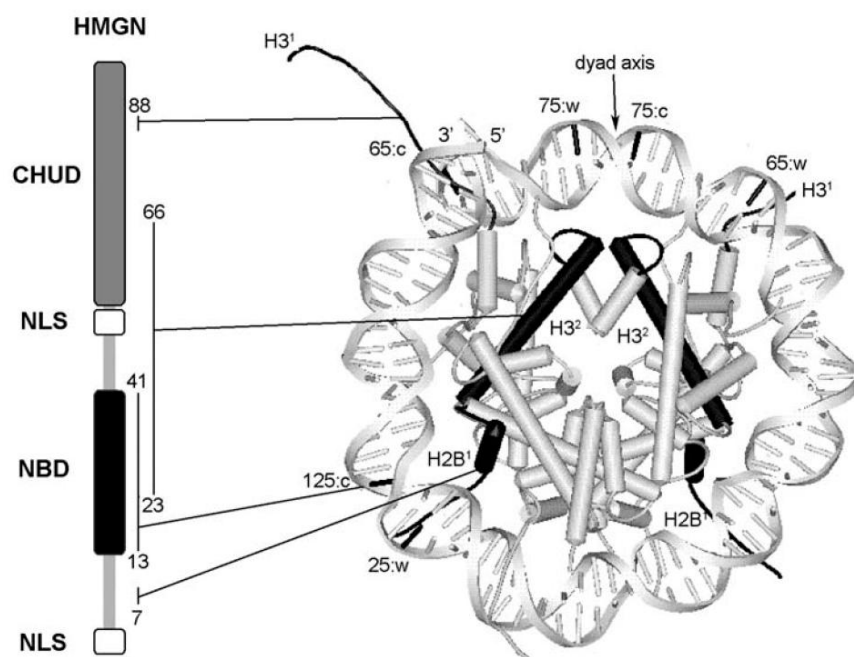


Figure 11. HMGN-NCP interaction. DNA basepairs that make contact with HMGN are numbered. Crick (c) and Watson (w) correspond to different DNA strands. Histone subunits that make HMGN contacts are colored black (adapted from West and Bustin, 2004).

1.4. Chromatin as anticancer drug target

Modulation of genomic transactions by various nucleosome activities suggests a therapeutic potential for the nucleosome. Nucleosomes encompass about 83% of genomic DNA (Segal *et al.*, 2006), thus signifying their importance for consideration in drug design (Davey and Davey, 2008). Several nucleosomal activities have been linked with cancer occurrence, and their importance in cancer progression can be exploited for cancer therapy.

In general, there are three chromatin-related modifications involved in cancer development; DNA methylation, histone modification, and nucleosome remodelling. These factors work closely with each other in promoting cancer development. On the DNA level, cytosine methylation may attract methylated DNA binding proteins, like MBD2 and MeCP2, which indirectly promote nucleosome remodelling due to their association with several nucleosomal remodelling complexes, such as NuRD and SWI/SNF chromatin remodelling complex (Jones and Baylin, 2007). Deacetylation of H4 lysine 16 and trimethylation of H4 lysine 20 are also some examples of human cancer hallmarks. Furthermore, on the chromatin level, nucleosome occupancy on certain promoter regions has also been linked with cancer occurrence. One remarkable example is the MLH1 promoter region, which is almost devoid of

nucleosomes in normal cells. Cancer cells place three nucleosomes on this promoter region, which subsequently silence the corresponding gene expression (Lin *et al.*, 2007).

Platinum-based chemotherapeutic agents elicit their effect on cancer cells mainly through DNA lesion formation. 1-2 intrastrand crosslinking of GG or AG dinucleotides, and to a smaller extent 1-3 GXG intrastrand crosslinks interrupt normal genomic activities, such as DNA transcription, and eventually trigger apoptosis (Jung and Lippard, 2007). Some of the compounds, such as cisplatin, oxaliplatin and carboplatin have been widely used in cancer treatment. In addition to their effect on DNA-templated genetic activities such as transcription, *in vitro* experiments showed that cisplatin-treated DNA adopts new conformational restraints that change nucleosomal positioning (Danford *et al.*, 2005). Moreover, in the case of nucleosomal DNA, lesions on DNA formed by platinum drugs inhibit nucleosome mobility (Wu and Davey, 2008). Taken together, targeting DNA in anticancer treatment may lead to different aspects of chromatin regulation, thus broadening the scope of the effect.

Histone proteins are also promising targets for anticancer therapeutic agents. Drug association to histone proteins, which may interrupt histone interaction with oncogenic proteins, may halt tumor progression. RAPTA-C, one of the ruthenium-based anticancer drugs, shows significant accumulation on cancer cell chromatin upon treatment. Its binding on histone indicates some degree of site selectivity as shown by crystallographic studies (Wu *et al.*, 2011). Moreover, RAPTA-C binding sites coincide with important sites for chromatin compaction (Chodaparambil *et al.*, 2007), nucleosome recognition (Makde *et al.*, 2010), and histone post-translational modifications, thus implying potential function of RAPTA-C in interrupting several chromatin regulations at once.

1.5. Scope of this thesis

We conducted crystallographic studies on NCP containing human histone proteins associated with different DNA fragments. Albeit being highly conserved throughout evolution, employing human NCP is more relevant for studying nucleosome function and examining nucleosome interaction with potential anticancer drugs. Having high resolution structures of human NCPs will certainly promote the use of human NCP for future study.

RAPTA-C interaction with the nucleosome has been studied in great detail using crystallographic analysis before (Wu *et al.*, 2011). We aim to improve this ruthenium-based compound by examining other derivatives of RAPTA-C with regard to their binding to nucleosomes. Our crystallographic study is focused on RAPTA-C derivatives with modified arene rings.

Finally, we are also interested to investigate the nature of HMGN binding to NCP. This protein plays several important roles inside the cell, such as nucleosome decompaction, gene repair and transcription. Knowledge gained may improve our understanding of the dynamics of genomic regulation.

Chapter II

Materials and Methods

Nucleosome core particle production was done with reference to established protocols (Dyer *et al.*, 2004). All chemical reagents without specific annotation were purchased from Sigma (St. Louis, MO, USA). All enzymes without specific annotation were purchased from New England Biolabs (Beverly, MA, USA). Competent cells used for histone and DNA overexpression were purchased from Promega (USA). RAPTA drugs were kindly provided by Christian Hartinger (University of Vienna) and Paul Dyson (EPFL). HMGN peptides were purchased from 1st BASE (Singapore). Histone, DNA, and HMGN proteins and peptides sequences are given in the Appendix.

2.1. Palindromic DNA Production

The scheme for palindromic DNA production is shown in figure 12. Palindromic DNA plasmids contain repeats of different halves of the final palindromic DNA (termed half-palindromic repeats for subsequent explanation). Palindromicity of the final DNA is ensured by the nature of the secondary restriction site.

Plasmids for 145 bp and 147 bp DNA production utilize two different sequences of *Hin*I restriction site for the final ligation. Plasmid for 146b bp DNA utilizes one *Eco*RI site for final ligation.

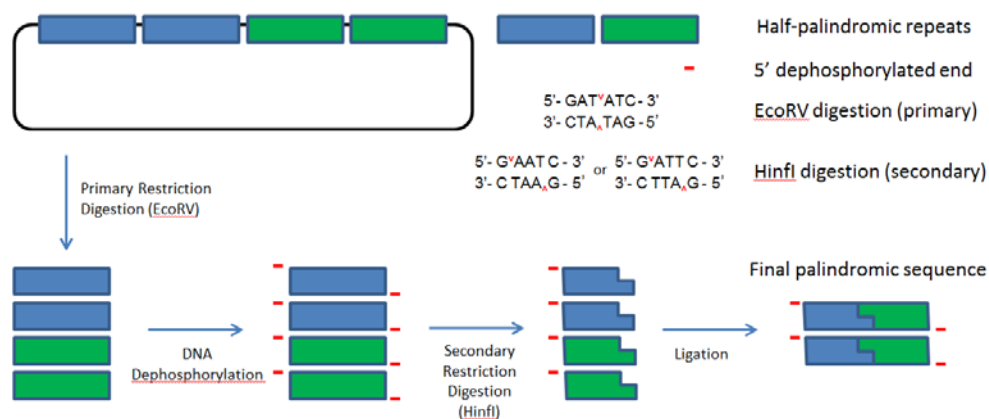


Figure 12. Scheme for Palindromic DNA production. There are two different types of repeats present on the plasmid (blue and green boxes). Each of them is the different half of the final palindromic DNA and has different *Hin*I restriction site. These *Hin*I sites ensure ligation of different repeat types for final palindromic sequence production.

2.1.1. DNA constructs and Plasmids

Four different palindromic DNA constructs were used in this study. Three of them (145 bp, 146b bp and 147 bp DNA) originate from human α -satellite DNA sequence, while 601L-145 bp DNA is palindromic DNA from left-half sequence of Widom 601 DNA. Widom 601 DNA was identified by SELEX experiment as DNA with highest known affinity to histone octamer (Lowary and Widom, 1998).

P24- α 10s plasmid was overexpressed for palindromic 145 bp DNA production. The plasmid contains 24 half-palindromic repeats, consisting of 12 repeats of each different half of the final palindromic 145 bp DNA. Each repeat is flanked by EcoRV restriction sites. The appropriate final palindromic ligation is ensured by two different sequences of HinfI restriction sites.

P32- α 8s plasmid was overexpressed for palindromic 147 bp DNA production. The plasmid contains 32 half-palindromic repeats, consisting of 16 repeats of each different half of the final palindromic 147 bp DNA. Each repeat is flanked by EcoRV restriction sites. The appropriate final palindromic ligation is ensured by two different sequences of HinfI restriction sites.

P16- α 6s plasmid was overexpressed for palindromic 146b bp DNA production. The plasmid contains 16 half-palindromic repeats. Each repeat is flanked by EcoRV restriction site. The appropriate final palindromic ligation is ensured by an EcoRI restriction site.

601L-145 plasmid was overexpressed for 601L-145 bp DNA production. The final 145 bp DNA is palindromic and its half-palindromic repeats are derived from the left-half portion of 601 high-positioning Widom DNA sequence. Each repeat is flanked by EcoRV restriction sites. The appropriate final palindromic ligation is ensured by HinfI restriction site.

2.1.2. Cell transformation

Palindromic DNA plasmid was transformed into HB101 competent cells. Fifty microliters of cells were incubated with 10 ng of plasmid on ice for ten minutes before being transferred to 42°C waterbath and incubated for 45 seconds. Cells were cooled down on ice for two minutes and subsequently transferred to a 37°C shaking incubator. After one hour, cells were plated to ampicillin-containing agar media (100 μ g/ml ampicillin) and incubated at 37°C overnight.

2.1.3. Plasmid extraction

Five round-shaped and nicely separated colonies formed on the ampicillin agar plate were picked and subsequently grown in ampicillin-containing terrific broth (TB) media [1.2% (w/v) Bacto Tryptone, 2.4% (w/v) yeast extract, 0.4% (v/v) glycerol, 17 mM KH_2PO_4 , and 72 mM K_2HPO_4 + 100 $\mu\text{g/ml}$ ampicillin] inside a 37°C shaking incubator. Cells were initially inoculated into 5 ml TB media, and after reaching turbidity, were transferred to 250 ml TB media, and finally into 7.5 L TB media. Cells were allowed to grow for 18 hours after being transferred to 7.5 L TB media.

Following 18 hours of growth, cells were pelleted down by centrifugation at 7000xg, 4°C. Pelleted cells were subsequently resuspended in 360 ml of Alkaline Lysis Solution I [50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0)]. After complete resuspension of the cells, 720 ml of Alkaline Lysis Solution II [0.2 M NaOH, 1% (w/v) sodium dodecyl sulfate (SDS)] were added, and cell suspension was vigorously mixed for 20 minutes. At the end of lysis step, 1260 ml of Alkaline Lysis III (4 M potassium acetate, 2 M acetic acid) were added into the cell suspension. Cells were mixed gently for another 20 minutes.

Cell debris generated during lysis step was removed by centrifugation at 10000xg, 4°C, for 20 minutes. Supernatant was filtered through stacks of sterile gauze pads to further remove the debris. Isopropanol (0.52x volume) was added to the filtrate, and the suspension was allowed to stand at room temperature for 30 minutes. Precipitates were collected by centrifugation at 10000xg, 20°C, for 30 minutes.

Pellets obtained after centrifugation were air-dried and later resuspended in 40 ml TE(10,50) [10 mM Tris-HCl (pH 8.0), 50 mM EDTA]. 240 μl of 10 mg/ml RNase A (BioRad Laboratories, USA) was added to the suspension, and RNA digestion was allowed to run at 37°C in a shaking incubator overnight.

The suspension was centrifuged at 10000xg, 4°C, for 30 minutes on the following day in order to remove any insoluble matter. The supernatant was subjected to three rounds of phenol extraction (10 ml phenol for every 20 ml of solution) to remove protein part from the solution. Plasmid DNA in the extracted aqueous phase was subsequently precipitated in a solution with final concentration of 0.5 M NaCl and 10% (w/v) PEG 6000. Precipitated plasmid DNA was collected by centrifugation (12000xg, 4°C, 20 minutes), resuspended in TE(10,50), and subjected to CIA extraction. Following CIA extraction, plasmid DNA was ethanol precipitated and finally dissolved

in TE(10,0.1) [10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA] . Plasmid DNA concentration was measured by NanoDrop spectrophotometer (Thermo Scientific, USA).

2.1.4. EcoRV Digestion and Purification of DNA Insert

Plasmid DNA was diluted in 50 mM Tris (pH 7.6), 100 mM NaCl, 10 mM MgCl₂ (Merck, Germany), 1 mM DL-1,4 dithiothreitol (DTT; Acros Organics, USA) reaction buffer, which gave the final DNA concentration of about 2 mg/ml. EcoRV was added accordingly (300 units per mg of DNA). The digestion was carried out for 24 hours at 37°C.

The vector was separated from the insert by PEG fractionation. 40% (w/v) PEG 6000 and 4 M NaCl were added to the digestion product to give a final concentration of 9.5% (w/v) PEG and 0.5 M NaCl. The solution was incubated on ice for one hour and subsequently centrifuged at 27,000xg, for 20 minutes at 4°C. The supernatant was then subjected to ethanol precipitation and the pellet was resuspended in TE(10,0.1). DNA concentration was measured by UV reading at 260 nm and the purity was determined by analyzing the samples by 10% DNA PAGE [10% monoacrylamide (Bio-Rad, California, USA), 0.25% bis (Bio-Rad, California, USA), 0.25X TBE, 0.1 % APS, TEMED (Acros Organics, USA)]. DNA PAGE was captured by Chemi Genius Bio Imaging System device (Syngene, USA) using GeneSnap v6.03 software (Syngene, USA).

2.1.5. Large Scale Dephosphorylation

DNA insert was diluted in 50 mM Tris (pH 8.0), 10 mM MgCl₂, 0.1 mM EDTA reaction buffer. The final concentration of DNA was 2 mg/ml. Calf Intestine Alkaline Phosphatase (CIAP) was added at a concentration of 36 units per mg of DNA. The reaction was carried out for 24 hours at 37°C.

Dephosphorylation reaction was subjected to phenol-CIA extraction (1:1, 0.5x volume) in order to remove the enzyme and other contaminants. The solution was centrifuged at 22,000xg, for 20 minutes, at 20°C. The upper phase was ethanol precipitated (0.3 M Na-Acetate, 2.5x volume of 100% cold ethanol) and subsequently dissolved in TE(10,0.1). The completion of dephosphorylation was verified by T4 DNA ligation test of dephosphorylated versus un-dephosphorylated DNA samples. Control and samples were analyzed by 6% DNA PAGE (6% monoacrylamide, 0.15% bis, 0.25X TBE, 0.1 % APS, TEMED).

2.1.6. HinfI/EcoRI Digestion and Purification of Half-palindromic Repeats by FPLC

Second restriction endonuclease digestion is required to generate the ligation site for the final step in palindromic DNA production. The site is HinfI for both P24- α 10s and P32- α 8s repeats. On the other hand, P16- α 6s repeats use EcoRI site for this purpose.

Dephosphorylated DNA was diluted to 2 mg/ml concentration in a reaction buffer containing 50 mM Tris (pH 8.0), 50 mM NaCl, 10 mM MgCl₂, and 1 mM DTT for HinfI digestion, or 100 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM MgCl₂ for EcoRI digestion. HinfI was added at a ratio of 276 units per mg of DNA, while EcoRI was added at a ratio of 400 units per mg of DNA. The reaction was carried out for 24 hours at 37°C. Completion of the digestion was verified by 10% DNA PAGE.

Half-palindromic repeats were separated from shorter DNA fragments (generated after second restriction endonuclease digestion) by using a MonoQ ion exchange column (Amersham Biosciences, Uppsala, Sweden) on an ÄKTA FPLC system (Amersham Biosciences, Uppsala, Sweden) at room temperature. About 4 mg of DNA was injected in each round of purification, and a flow rate of 2 ml/min was maintained. The column was equilibrated with buffer A [20 mM Tris (pH 7.5), 0.1 mM EDTA, 300 mM NaCl] and then a gradient up to 40% buffer B [20mM Tris (pH 7.5), 0.1 mM EDTA, 1 M NaCl] was set over a period of 2 minutes. Shorter DNA fragment eluted at around 40% buffer B as a sharp peak. Then the gradient was set to run from 40% to 70% of buffer B over a period of 15 minutes. Half-palindromic repeat DNA eluted out at about 47% of buffer B. The column was then washed with 100% buffer B in order to remove any DNA remaining on the column. Purified half-palindromic repeat DNA from several purification batches were combined together and ethanol precipitated. The DNA was redissolved in TE(10,0.1) and purity was confirmed by 10% DNA PAGE.

2.1.7. Ligation and Purification of palindromic DNA

Half-palindromic repeat DNA was subjected to ligation reaction in a buffer containing 100 mM Tris (pH 7.6), 20 mM MgCl₂, 10 mM DTT, 2 mM rATP. Final concentration of DNA was maintained at 2 mg/ml. T4 DNA ligase was added into the buffer at ratio of 500 units per mg of DNA. The reaction mixture was incubated for 24 hours at room temperature. In order to maximize the yield of ligation, the sample was kept at 4°C after a 1-day room temperature incubation. The result of ligation reaction was verified by 10% DNA PAGE.

Palindromic 147/146b/145 bp DNA was separated from unligated DNA sample by using a MonoQ ion exchange column connected to FPLC system at room temperature. About 2.5 mg of DNA was injected in each round of purification to achieve clear separation of ligated and non-ligated product. The buffer flow rate was maintained at 2 ml/min rate. A gradient was set to run from 0% to 50% buffer B in 1 minute. The buffer was allowed to run at 50% buffer B for a few minutes to elute any short DNA fragment. A gradient to 70% buffer B was subsequently set over a period of 10 minutes. Unligated DNA eluted at about 60% buffer B, while 147/146b/145 bp palindromic DNA eluted from the column at around 62% buffer B. The purity of final palindromic DNA was assessed by 10% DNA PAGE. Pure fractions were combined together, ethanol-precipitated, and then redissolved in TE(10,0.1) to obtain a suitable concentration.

2.2. Histone Octamer Preparation

2.2.1. Cell transformation

Histones genes, H2A, H2B, and H4, were inserted to the pET3a vector while H3 was cloned into the pET3d vector (Luger *et al.*, 1999). Histone plasmid was transformed into BL21(DE3)pLysS competent cells. Fifty microliters of cells were incubated with 10 ng of plasmid in ice for ten minutes before being transferred to a 42°C waterbath and incubated for 45 seconds. Cells were cooled down on ice for two minutes and subsequently transferred to a 37°C shaking incubator. After one hour, cells were plated to ampicillin and chloramphenicol-containing agar media (100 µg/ml ampicillin, 25 µg/ml chloramphenicol) and incubated at 37°C overnight.

2.2.2. Histone protein Overexpression

Five round-shaped and nicely separated colonies formed on the ampicillin-chloramphenicol agar plate were picked and subsequently grown in ampicillin and chloramphenicol-containing 2xTY media [1.6% (w/v) Bacto Tryptone, 1.0% (w/v) yeast extract, 0.5% NaCl, 0.1% glucose + 100 µg/ml ampicillin + 25 µg/ml chloramphenicol] inside a 37°C shaking incubator. Protein expression was induced by adding 0.4 mM Isopropyl-beta-D-thiogalactopyranoside (IPTG) into cell culture once OD₆₀₀ reached 0.55. After three hours, cells were pelleted by centrifugation at 7000xg, 4°C.

2.2.3. Inclusion Body Preparation

Pelleted cells were fully resuspended in wash buffer [50 mM Tris (pH 7.5), 100 mM NaCl, 1 mM beta-mercaptoethanol], and subsequently flash frozen in liquid nitrogen in order to break their plasma membrane. Cells were re-thawed on the following day, and sonicated until the suspension became homogenous. Sample was then washed four times in wash buffer, with 1% (v/v) Triton X-100 included on the third wash. Relatively clean inclusion body (crude histone) obtained after the last wash was collected by centrifugation.

2.2.4. FPLC Purification of Histones

Approximately 6 grams of crude histone were denatured in 40 ml S-200 unfolding buffer [7 M Guanidium HCl, 20 mM Na-acetate (pH 5.2), 10mM DTT] and 1 ml dimethylsulfoxide (Fisher Scientific, UK). This suspension was homogenized as extensively as possible before spinning down at 18000xg, for 30 minutes at 20°C. Centrifugation was carried out three times to remove traces of precipitating contaminants.

Unfolded histone was injected into a HiLoad 16/60 Superdex 200 column (GE Healthcare, UK) connected to an AKTA FPLC system at 0.8 ml/min flow rate of SAUDE-1000 buffer [7 M Urea (USB, Cleveland, USA), 20 mM Na-acetate (pH 5.2), 1 M NaCl, 5 mM beta-mercaptoethanol, 1 mM EDTA]. The size exclusion was conducted at room temperature. Histones eluted after about 80 ml volume. Purity of all fractions was assessed by 18% SDS PAGE [resolving gel: 0.375 M Tris HCl (pH 8.8), 18% Acrylamide, 0.3% Bis, 0.1% SDS, 0.1% APS, TEMED, stacking gel: 0.125 M Tris HCl (pH 6.8), 5% Acrylamide, 0.25% Bis, 0.1% SDS, 0.1% APS, TEMED]. Pure fractions were pooled together, while less pure ones were further purified by a Resource-S ion exchange column (Amersham Biosciences, UK). Purification using a Resource-S column was done at room temperature. Lyophilized histones (less pure fractions) were dissolved in SAUDE-0 [7 M Urea (USB, Cleveland, USA), 20 mM Na-acetate (pH 5.2), 5 mM beta-mercaptoethanol, 1 mM EDTA] unfolding buffer. The histones were then injected into a Resource-S ion exchange column connected to an AKTA FPLC system at 5 ml/min flow rate of SAUDE-0 unfolding buffer. The buffer was gradually exchanged from SAUDE-0 to 60% SAUDE-1000 over a period of 6 minutes. Each fractionation was verified by 18% SDS PAGE.

Purified histones (still in denaturing condition) were dialyzed (dialysis membrane purchased from Laboratories Inc., USA) against double-distilled water containing 5 mM beta-mercaptoethanol for 24 hours, and the concentration was estimated based on absorbance at 276 nm. Individual histones were then aliquoted, lyophilized, and stored at -80°C.

2.2.5. Histone Octamer Refolding

Lyophilized histones were dissolved and denatured individually in unfolding buffer [7 M Guanidinium HCl, 10 mM Tris (pH 7.5), 10 mM DTT]. The histones were mixed together in an accurate equimolar stoichiometry. Unfolding buffer was added to the mixture to bring the final concentration to about 1 mg/ml. The solution was subsequently dialyzed at room temperature against refolding buffer [2 M NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 10 mM beta-mercaptoethanol] for about 20 hours. Refolding buffer was replaced with fresh buffer every 5 hours. Dialyzed HO solution was centrifuged thrice to remove any precipitant before concentrating down to about 4-5 ml by using Amicon concentrator device (MWCO 10 kDa; Millipore, USA). Final concentration of HO was about 5-10 mg/ml.

2.2.6. FPLC Purification of Histone Octamer

HO from the previously described refolding was further separated from histone dimers and tetramers by passing through a HiLoad 16/60 Superdex 200 column equilibrated with refolding buffer. The purification is run at 4°C with a flow rate of 1.2 ml/min. HO eluted at about 60 ml, followed by H2A/H2B dimer at around 80 ml. The purity of the HO was determined by 18% SDS PAGE. Pure HO fractions were pooled together and concentrated using an Amicon concentrator. HO was stored in 50% glycerol - 50% refolding buffer at -20°C. Absorbance reading at 276 nm gave only an approximate concentration of the HO.

2.3. Nucleosome Core Particle Reconstitution

2.3.1. Off-centered NCP Reconstitution

Six micromolar palindromic DNA was mixed with 5.4 μ M HO and incubated for one hour at 4°C in a buffer of 10 mM DTT and 2 M KCl. The mixture was dialyzed against TCS buffer [20 mM Tris (pH 7.5), 1 mM EDTA, 1 mM DTT] with gradual decrement of KCl concentration from 1.5 M to 0 M over 3 days. NCP was stored at 4°C.

2.3.2. Centered NCP Reconstitution

Reconstitution was done in the same way as that for off-centered NCP preparation. The dialysis, however, started with TCS containing 0.85 M KCl. KCl concentration is gradually decreased to 0.2 M KCl at room temperature over 8 hours. The NCP was subsequently incubated for 1 hour at 37°C. Finally the NCP was dialyzed against TCS buffer for about four hours. Centered NCP can also be prepared by one-hour 37°C-heat treatment of the off-centered NCP. The efficiency of latter method, however, is lower than the former one. Reconstituted NCP was finally dialyzed against 20 mM Potassium Cacodylate (KCaco) (Acros Organics, USA) and assessed by a 5% DNA PAGE.

2.3.3. NCP Purification

In some cases, such as production of protein-fused NCP, further purification is required. For this purpose, newly reconstituted NCP was injected into a HiLoad 16/60 Superdex 200 column equilibrated with TCS containing 10 mM KCl. Pure NCP fraction was dialyzed against 20 mM KCaco.

2.4. HMGN Production

HMGN proteins production was done with reference to established protocol (Ong *et al.*, 2010). HMGN1 and HMGN2 were both overexpressed separately in BL21(DE3)pLysS competent cell. Protein expression was induced by adding 0.4 mM Isopropyl-beta-D-thiogalactopyranoside (IPTG) into cell culture once OD₆₀₀ reached 0.7. After three hours, cells were pelleted by centrifugation at 7000xg, 4°C. Cell lysate was homogenized in lysis buffer containing 50 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol, and 1 mM phenylmethylsulphonyl fluoride (PMSF). Insoluble aggregate was clarified and supernatant was subjected to size-exclusion chromatography using a 26/60 Sephacryl S-200 column (GE Healthcare, Uppsala, Sweden) pre-equilibrated with a buffer containing 20 mM Tris-HCl (pH 7.5), 0.2 M NaCl, 1 mM EDTA, and 1 mM PMSF. HMGN proteins were further purified using Resource S cation-exchange column (GE Healthcare, Uppsala, Sweden). The purification was done over a gradient of 0.2 to 0.4 M NaCl in buffer containing 20 mM Tris-HCl (pH 7.5) and 1 mM EDTA. Pure HMGN proteins were dialyzed into a buffer containing 20 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 1 mM EDTA, and stored at -80°C

2.5. Crystal Preparation

NCP crystals were grown by vapour diffusion at 17°C over the course of one to four weeks in droplets containing 4 mg/ml NCP, 70-90 mM MnCl₂, 50-80 mM KCl, 20 mM KCaco (pH 6.0). Crystallization droplets were equilibrated against 35-45 mM MnCl₂, 25-40 mM KCl, 10 mM Kcaco (pH 6.0) on hanging drop crystallization plates.

For HMGN-NBD:NCP complex crystallizations, screening was performed at varying NCP concentrations (4-8 mg/ml of NCP147), HMGN NBD:NCP ratios (2:1 to 50:1), pH (5.5; 6 or 6.5), salt precipitants (MnCl₂-KCl, MgCl₂-KCl, CoCl₂-KCl) (Table 1), Spermine-KCl (Table 2), and reservoir concentration (NaCl; 150 mM – 70 mM). The effect of organic precipitants, 1.5%-3% Polyethylene glycol monomethyl ether 2000 (Hampton Research, USA)], and other crystal screens [PEG/Ion Screen, PEG/Ion Screen 2 (Hampton Research, USA)] were also tested.

Fully grown native NCP crystals were transferred to a harvesting buffer of 40 mM MnCl₂, 35 mM KCl, 20 mM KCaco. MPD (2-methyl-2,4-pentanediol) and trehalose were slowly introduced to the harvesting buffer by step-wise buffer exchange until MPD and trehalose concentration reached 24% (v/v) and 2% (w/v), respectively. For low ionic strength condition, the final crystals were in 12.5mM MnCl₂, 10mM KCl, 10mM KCaco, 24% (v/v) MPD, and 2% (w/v) trehalose.

For drug binding studies, harvested NCP crystals were buffer-exchanged to 10 mM MgSO₄, 20 mM Kcaco (pH 6.0), 24% (v/v) MPD, and 2% (w/v) trehalose. Drugs were dissolved in the same buffer to give a final concentration of 250 µM or 750 µM. NCP crystals and drug were mixed together and incubated at 17°C.

High Ionic Strength Crystallization Buffer

Divalent Salt (mM)	60	70	80	90	100	110	120
KCl (mM)	60	60	60	60	60	60	60
KCaco (mM)	20	20	20	20	20	20	20

Divalent Salt (mM)	60	70	80	90	100	110	120
KCl (mM)	70	70	70	70	70	70	70
KCaco (mM)	20	20	20	20	20	20	20

Divalent Salt (mM)	60	70	80	90	100	110	120
KCl (mM)	80	80	80	80	80	80	80
KCaco (mM)	20	20	20	20	20	20	20

Divalent Salt (mM)	60	70	80	90	100	110	120
KCl (mM)	90	90	90	90	90	90	90
KCaco (mM)	20	20	20	20	20	20	20

Divalent Salt (mM)	60	70	80	90	100	110	120
KCl (mM)	100	100	100	100	100	100	100
KCaco (mM)	20	20	20	20	20	20	20

Divalent Salt (mM)	60	70	80	90	100	110	120
KCl (mM)	110	110	110	110	110	110	110
KCaco (mM)	20	20	20	20	20	20	20

Divalent Salt (mM)	60	70	80	90	100	110	120
KCl (mM)	120	120	120	120	120	120	120
KCaco (mM)	20	20	20	20	20	20	20

Low Ionic Strength Crystallization Buffer

0.75	0.8	0.85	0.9	0.95	1	1.05	1.1	1.15	1.2	1.25
0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
20	20	20	20	20	20	20	20	20	20	20

0.75	0.8	0.85	0.9	0.95	1	1.05	1.1	1.15	1.2	1.25
0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
20	20	20	20	20	20	20	20	20	20	20

0.75	0.8	0.85	0.9	0.95	1	1.05	1.1	1.15	1.2	1.25
0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
20	20	20	20	20	20	20	20	20	20	20

0.75	0.8	0.85	0.9	0.95	1	1.05	1.1	1.15	1.2	1.25
0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
20	20	20	20	20	20	20	20	20	20	20

0.75	0.8	0.85	0.9	0.95	1	1.05	1.1	1.15	1.2	1.25
0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
20	20	20	20	20	20	20	20	20	20	20

0.75	0.8	0.85	0.9	0.95	1	1.05	1.1	1.15	1.2	1.25
1	1	1	1	1	1	1	1	1	1	1
20	20	20	20	20	20	20	20	20	20	20

0.75	0.8	0.85	0.9	0.95	1	1.05	1.1	1.15	1.2	1.25
1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
20	20	20	20	20	20	20	20	20	20	20

*Three divalent salts are screened separately. They are $MnCl_2$, $MgCl_2$, and $CoCl_2$

Table 1. Divalent salt - KCl Crystallization Screens

Spermine (mM)	10	15	20	25	30	35	40	10	15	20	25	30	35	40
KCl (mM)	10	10	10	10	10	10	10	35	35	35	35	35	35	35
Kcaco (mM)	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Spermine (mM)	10	15	20	25	30	35	40	10	15	20	25	30	35	40
KCl (mM)	15	15	15	15	15	15	15	40	40	40	40	40	40	40
Kcaco (mM)	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Spermine (mM)	10	15	20	25	30	35	40	10	15	20	25	30	35	40
KCl (mM)	20	20	20	20	20	20	20	50	50	50	50	50	50	50
Kcaco (mM)	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Spermine (mM)	10	15	20	25	30	35	40	10	15	20	25	30	35	40
KCl (mM)	25	25	25	25	25	25	25	75	75	75	75	75	75	75
Kcaco (mM)	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Spermine (mM)	10	15	20	25	30	35	40	10	15	20	25	30	35	40
KCl (mM)	30	30	30	30	30	30	30	100	100	100	100	100	100	100
Kcaco (mM)	20	20	20	20	20	20	20	20	20	20	20	20	20	20

Table 2. Spermine - KCl Crystallization Screen

2.6. Crystal Processing

2.6.1. Crystallographic Data Collection

Preliminary data collection was performed using an in-house x-ray instrument (MicroMax-007HF rotating anode generator, Rigaku Raxis IV+ image plate detector). Crystals were mounted onto goniometer head in line with the nitrogen cryostream at -100°C. A total of 180 images, each corresponding to 1° crystal oscillation and 10 minutes x-ray exposure, were collected.

Final data collection was done at PXI-X06SA and PXIII-X06DA beamlines of Swiss Light Source (PSI Villigen, Switzerland). PXI-X06SA was equipped with a PILATUS detector, while PXIII-X06DA utilized a MarCCD detector. Crystals were mounted onto goniometer head in line with nitrogen cryostream at temperature of -180°C (PXI-X06SA) or -167°C (PXIII-X06DA). NCP data was collected at 1.0 Å wavelength, with 0.5° crystal oscillation and 0.5 second x-ray exposures at both PXI-X06SA and PXIII-X06DA facilities. Drug-NCP crystal datasets were collected at 1.5 Å wavelength, with 0.5° crystal oscillation and 0.5 second x-ray exposures at PXIII-X06DA beamline. A total of 360 images were collected for all datasets.

2.6.2. Crystallographic Data Analysis

Diffraction images were integrated by MOSFLM (Leslie, 2006) and scaled by Scala. Phases were obtained by molecular replacement using Molrep and Phaser, using various NCP structures (1KX5, 2CV5, 2NZD Protein Data Bank) as starting models. Model building was done with Coot (Emsley and Cowtan, 2004), and the structures were refined with Refmac5. Anomalous maps were created using Sfall and FFT. Scala, Molrep, Phaser, Refmac5, Sfall, and FFT are part of the CCP4 package (CCP4, 1994). Restraint parameters for RAPTA agents were derived from RAPTA-C small-molecule crystal structures (Wu *et al.*, 2011). Molecular visualizations were prepared with Pymol (DeLano Scientific LLC, San Carlos, CA, USA).

2.7. Electrophoretic Mobility Shift Assay (EMSA)

One µg of NCP was mixed with HMGN NBD peptide according to the molar ratio desired. The mixture was brought to a final 9 µl solution with 1X TBE buffer (89 mM Tris-HCl, 89 mM boric acid, 2.5 mM EDTA, pH 8.3) and incubated for 30 minutes at 4°C. One µl of 50% Sucrose was added before loading the sample on a 6%

polyacrylamide gel. The electrophoresis was carried out at 90 V for ~ 2 hours at 4°C. Ethidium bromide was used to visualize bands under UV illumination.

2.8. Western Blot

HMGN-nucleosome assemblies separated by gel electrophoresis in 1X TB + 1 mM MgCl₂ running buffer were blotted onto a polyvinylidene difluoride membrane (Pall Corporation, USA) by semi-dry transfer at 20 V for 40 min in a buffer of 0.5X TBE (for 12-array) or 1X TB with 1mM MgCl₂ (for NCP). Membranes were subsequently blocked with 1-2% BSA dissolved in TBST buffer [0.1 M Tris-HCl (pH 7.5), 0.15 M NaCl, and 0.05-0.1% Tween 20] at 4°C. Core histone protein-specific antibodies (H2A, ab13923; H2B, ab1790; H3, ab1791; H4, ab10158; Abcam, UK) were used for primary incubation of membranes, which was carried out for 4 hr at 4°C. Membranes were subsequently washed in TBST buffer, followed by secondary antibody incubation with HRP-conjugated antibodies (ab6721; Abcam, UK) for 1 hr at 4°C. Binding of histone antibodies was detected using Lumi-Light^{PLUS} Western blotting substrate (Roche, Germany).

2.9. Crosslinking

HMGN NBD peptide was mixed with NCP at 1.5:1 ratio in 100 mM KCl, 20 mM KCaco buffer and incubated for 30 minutes at 4°C. Glutaraldehyde stock in 100 mM KCl, 20 mM KCaco at 0.004% concentration was titrated to the mixture to give a final glutaraldehyde concentration of 0.001%. Reaction was allowed to proceed for 30 minutes. Tris buffer was used to stop the crosslinking.

2.10. Fluorescence Assay

Thiophosphate group are required at DNA 5'-end termini for labelling DNA with fluorescence dyes. For this purpose, DNA was incubated in solutions containing 1 U T4 PNK per µg DNA, 1 mM ATP-γS, and 1X PNK buffer for 2 hours at 37°C. Phosphorylated DNA was subsequently mixed with Cy5 maleimide-conjugated dye (GE Healthcare, UK) and the labelling process was allowed to run at 65°C for 2.5 hours. Unbound dye was removed from the solution by spin column, and DNA was reconstituted into NCP.

HMGN protein was labelled with ATTO488 (ATTO-TEC GmbH, Germany) at 1:1 or 2:1 dye-to-HMGN stoichiometry. Labelling was done in 0.1 M Na-bicarbonate (pH 8.3) buffer for an hour at room temperature. Reaction was stopped by adding 1 M Tris-HCl

(pH 7.5) to give a final Tris-HCl concentration of 100 mM in the labelling solution. Unbound dye was removed from the solution by spin column.

Chapter III

Nucleosome Core Particle Production and Crystallization

3.1. Palindromic DNA Production

DNA was expressed and processed through series of steps, and the products of each step were verified by gel electrophoresis (Figure 13). Purification profile of half-palindromic repeats DNA, and palindromic DNA after final ligation step are shown in Figure 14 and Figure 15, respectively.

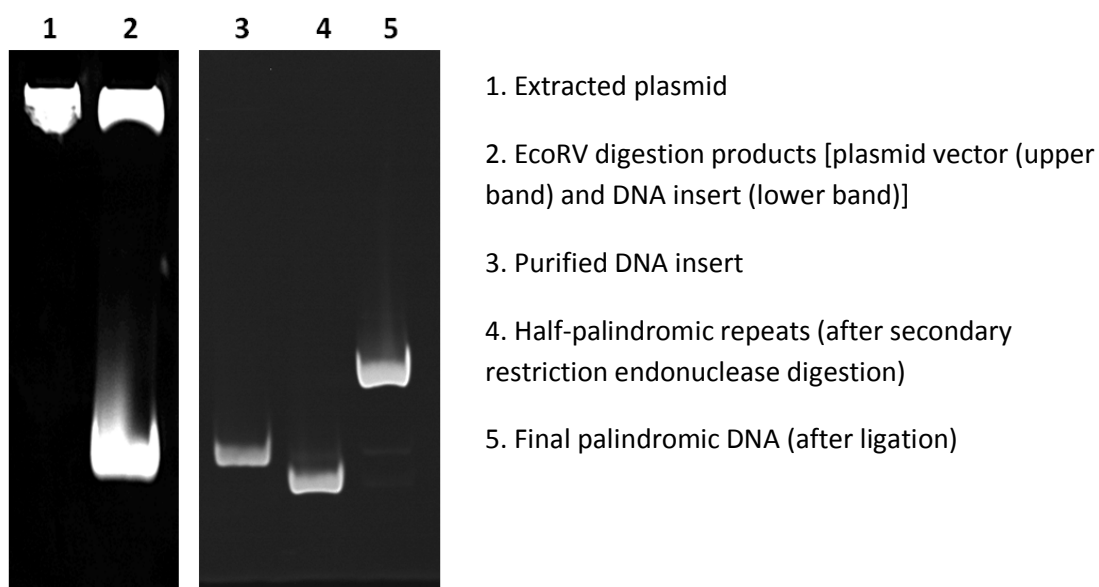


Figure 13. Gel electrophoresis result from each step of palindromic DNA production

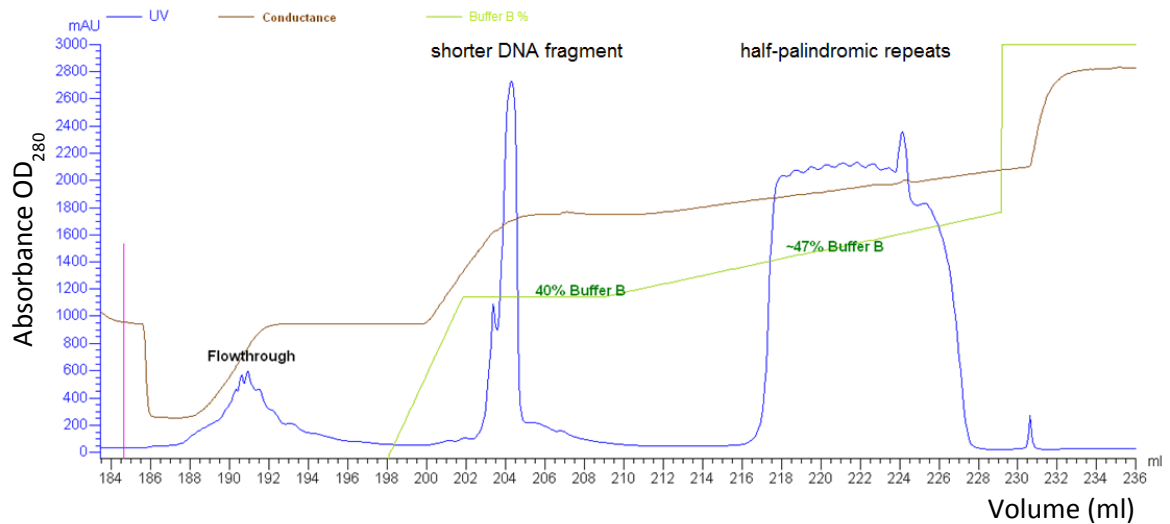


Figure 14. FPLC-MonoQ (ionic exchange chromatography) purification profile of half-palindromic repeats after secondary restriction endonuclease digestion (HinfI digestion for 145 bp, 147 bp, and 601L-145 bp DNA. EcoRI digestion for 146b bp DNA)

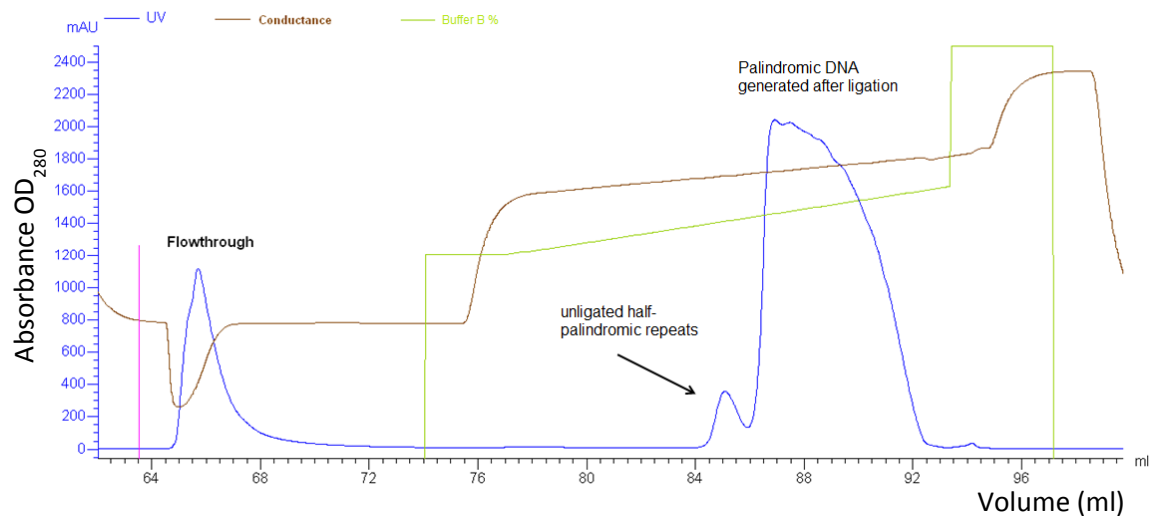


Figure 15. FPLC-MonoQ (ionic exchange chromatography) purification profile of final palindromic DNA (after ligation).

3.2. Histone Octamer Production

Crude histone extracts were purified by HiLoad 16/60 Superdex 200 column. Histone purification profiles look similar for each histone protein (Figure 16). There are three

peaks observed during purification; high molecular weight contaminant (1st peak), histone protein (2nd peak, blue box Figure 16), and guanidinium (3rd peak). Further purification of histone proteins requires passing the protein through Resource-S column (Figure 17).

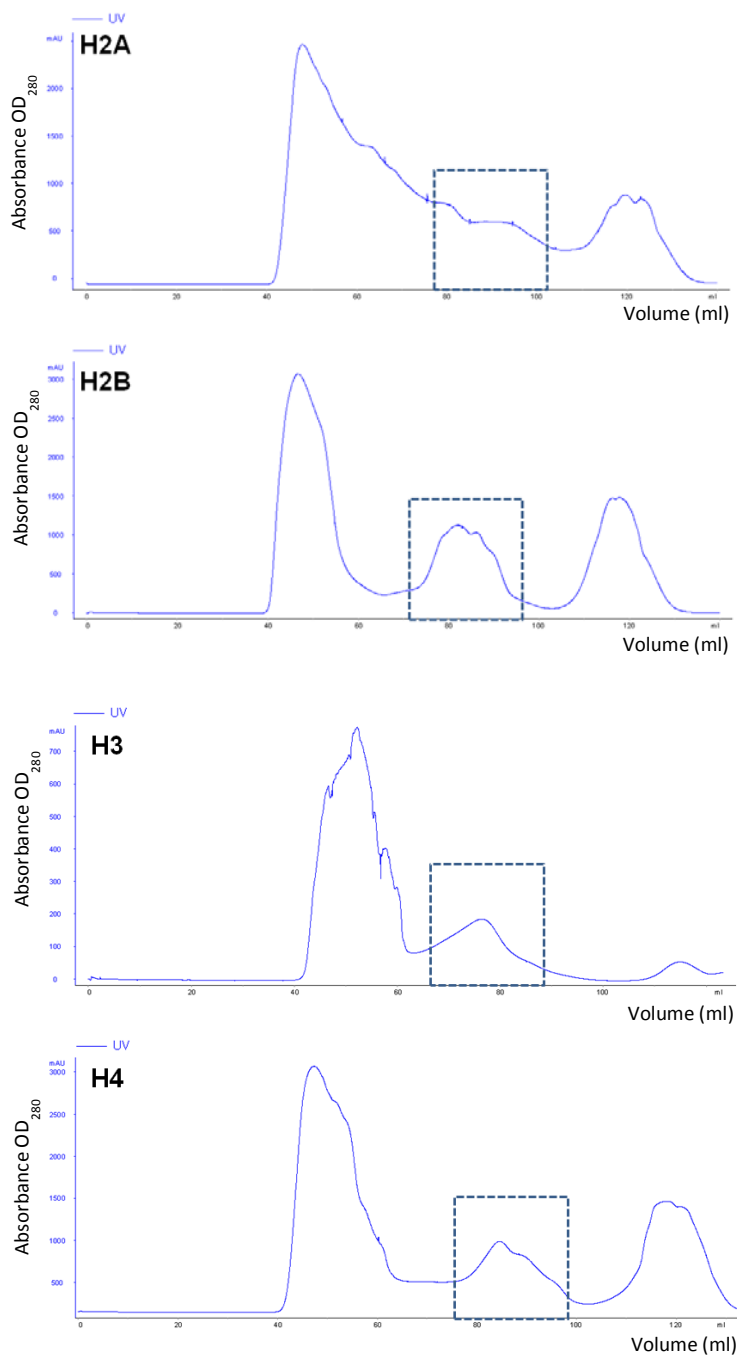


Figure 16. HiLoad 16/60 Superdex 200 (size exclusion chromatography) – Histone purification profile. Histones eluted from the column around 80 ml volume (dashed-line box).

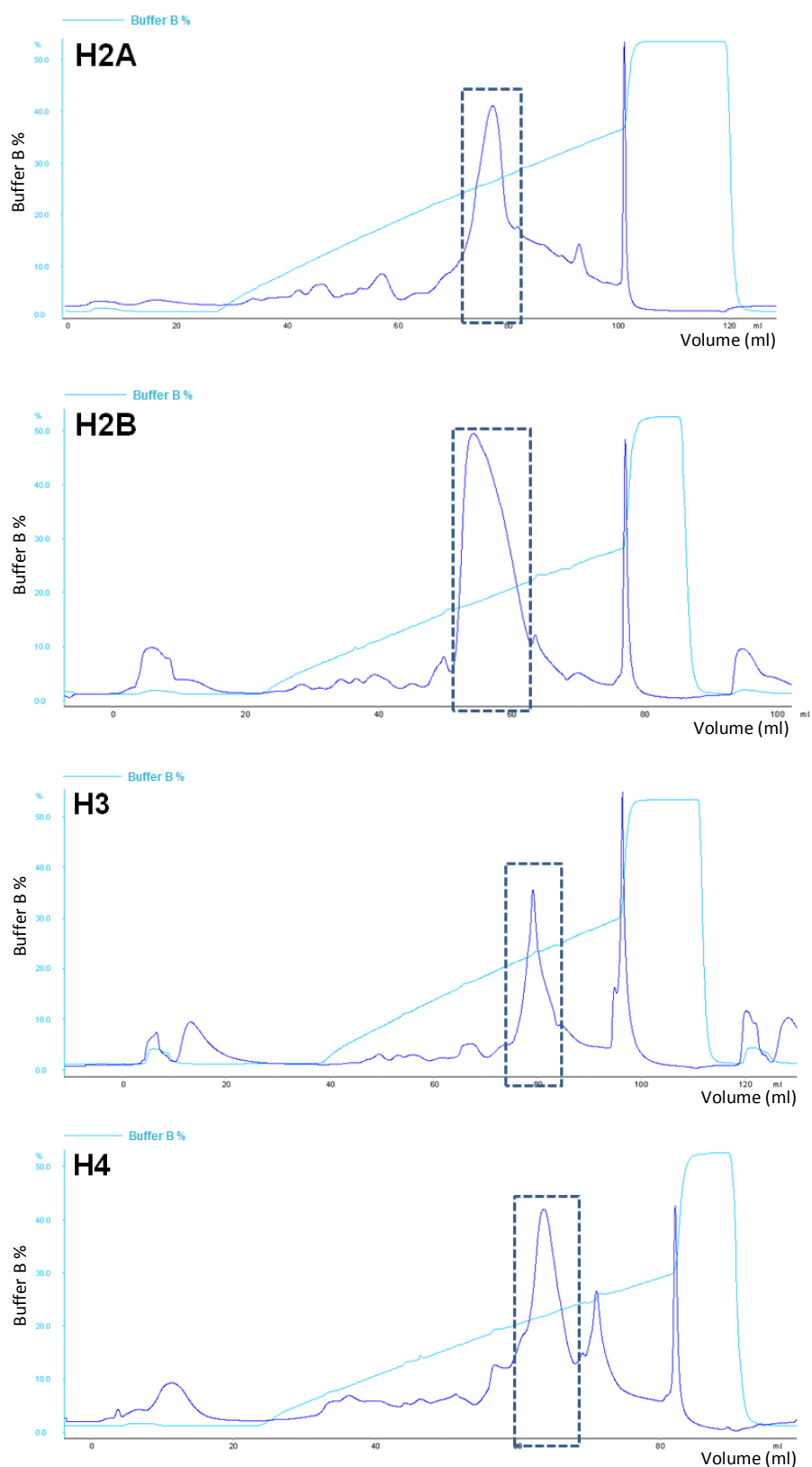


Figure 17. ResourceS (ionic exchange chromatography) – Histone purification profile. Peaks corresponding to pure histone elution are indicated by a dashed-line box. Blue line is Absorbance OD₂₈₀

HO refolding was performed with an equimolar ratio of individual histones. Histone precipitation was carefully monitored during dialysis to establish the optimum refolding condition. We found that refolding at room temperature yielded the highest amount of good HO product. Subsequent FPLC purification helped to separate HO from other histone complexes that may form during the refolding. The quality of HO was assessed by 18% SDS PAGE to confirm histone stoichiometry (Figure 18).

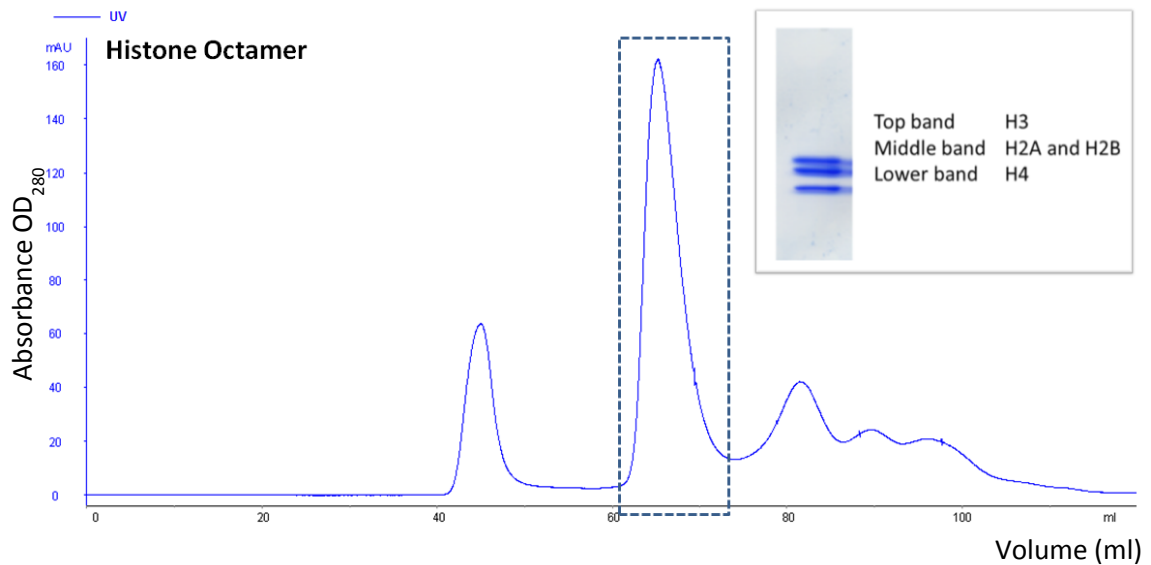


Figure 18. HiLoad 16/60 Superdex 200 (size exclusion chromatography) – HO purification profile. Peak corresponding to HO elution is boxed. HO quality was assessed by 18% SDS PAGE (insert). Histone dimer and monomer peaks appear after the HO peak.

3.3. Nucleosome Core Particle Reconstitution

HO concentration can not be measured precisely from absorbance readings at 276 nm. Therefore, several test reconstitutions with different estimations in HO concentration were carried out in order to pinpoint the optimal ratio. Reconstitution with overestimated histone concentration will produce a NCP solution with free DNA in the solution. This free DNA may perturb the NCP since it facilitates H2A-H2B dimer dissociation when the mixture is incubated at near physiological conditions (Figure 19, bottom panel). On the other hand, reconstitution with underestimated histone concentration will lead to the formation of non-homogenous NCP species (Figure 19, smeared NCP band at top panel), and in some cases NCP dimer and high molecular aggregates.

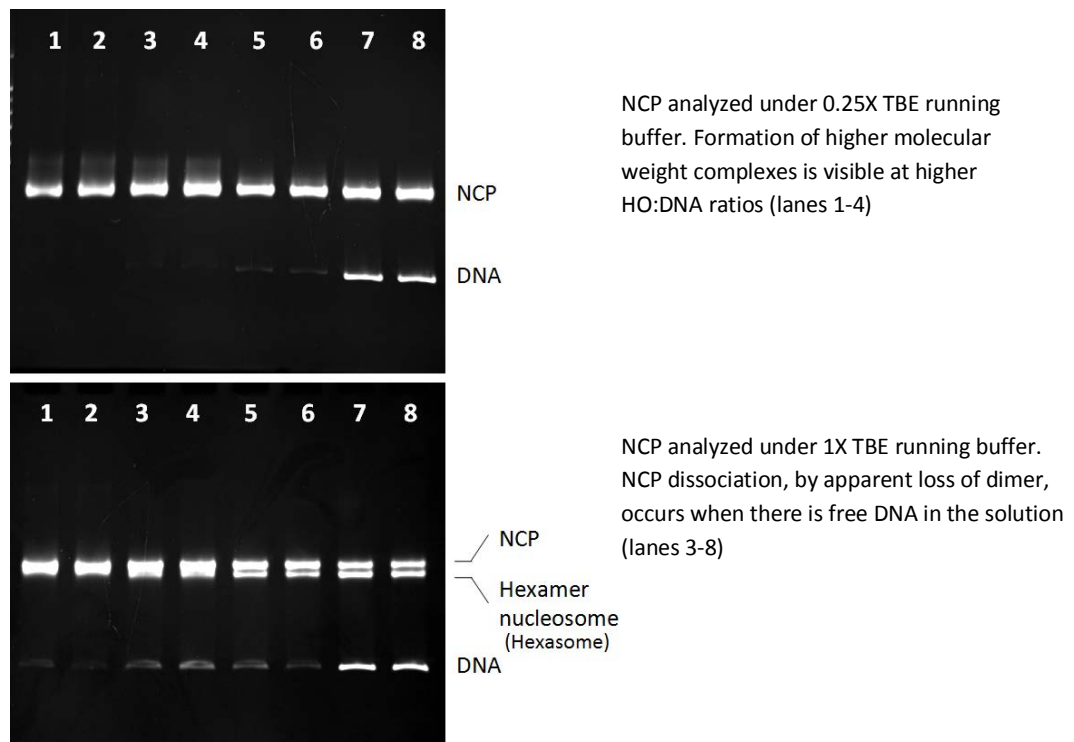


Figure 19. NCP products reconstituted with different DNA-to-HO ratios (increasing ratio from lane 1 to 8). NCP products were run on two different ionic conditions; 0.25X TBE (upper panel), and 1X TBE (lower panel), to assess the quality of the NCP reconstitution product.

Even with optimal DNA-to-histone ratios, NCP reconstitutions typically give rise to NCP in two distinct translational positioning states. These species differing in position by one turn of the double helix, namely off-centered and centered NCP, can be resolved on 5% DNA PAGE gels, with the centered form migrating faster than the off-centered one (Figure 20). The difference in structure between these two species is illustrated in Figure 21. Heat treatment of the newly reconstituted NCP is required to convert off-centered NCP to the centered one.

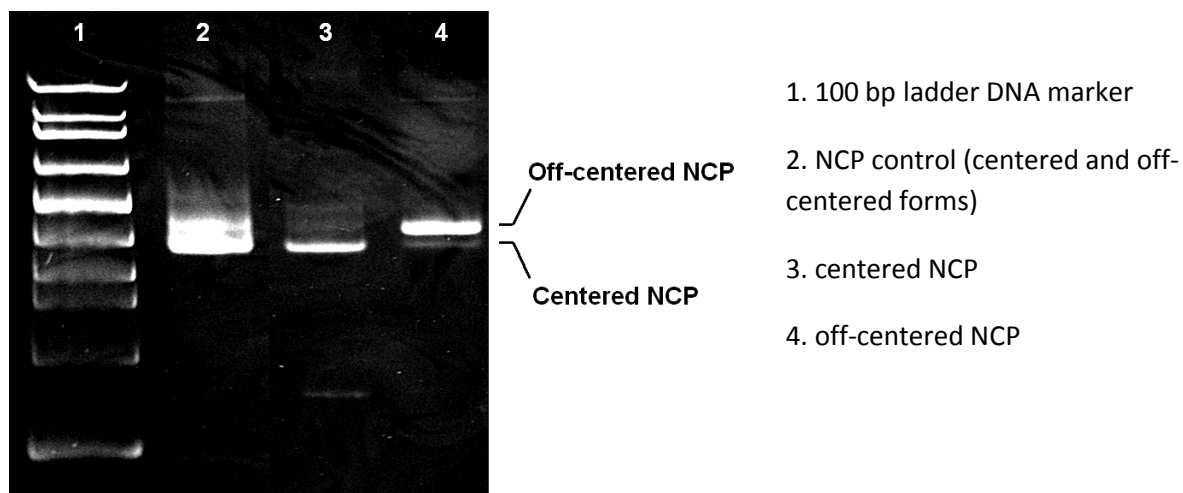


Figure 20. Two NCP positioning species after reconstitution

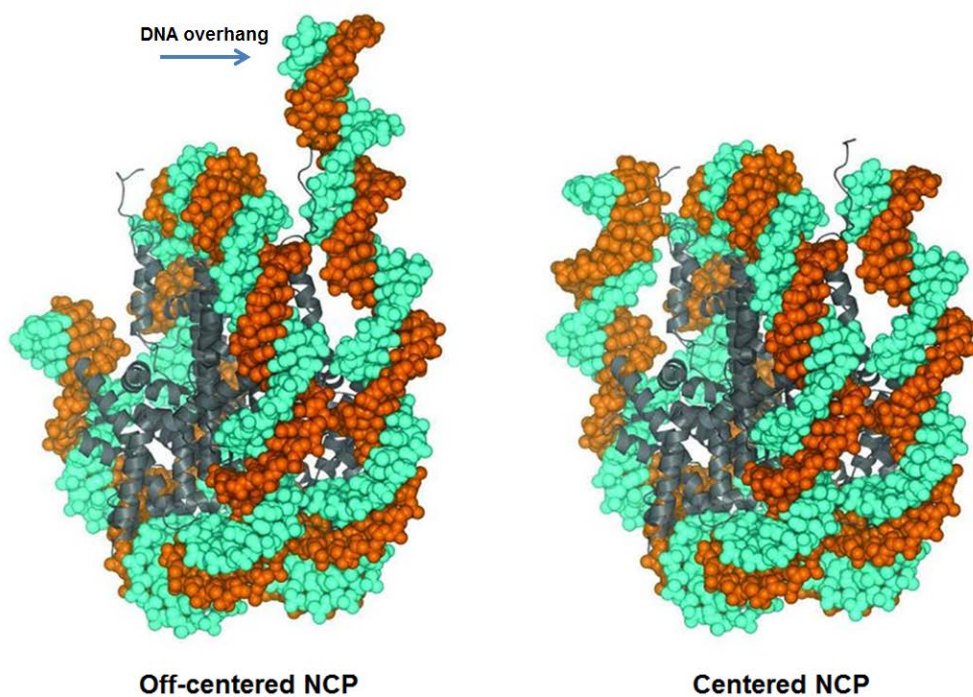


Figure 21. Difference between off-centered and centered NCP

Finally, the DNA-to-HO ratio that gave the best reconstitution product was scaled up. Figure 22 depicts centered NCPs reconstituted with four different DNAs; 145 bp, 147 bp, 146b bp, 601L-145 bp.

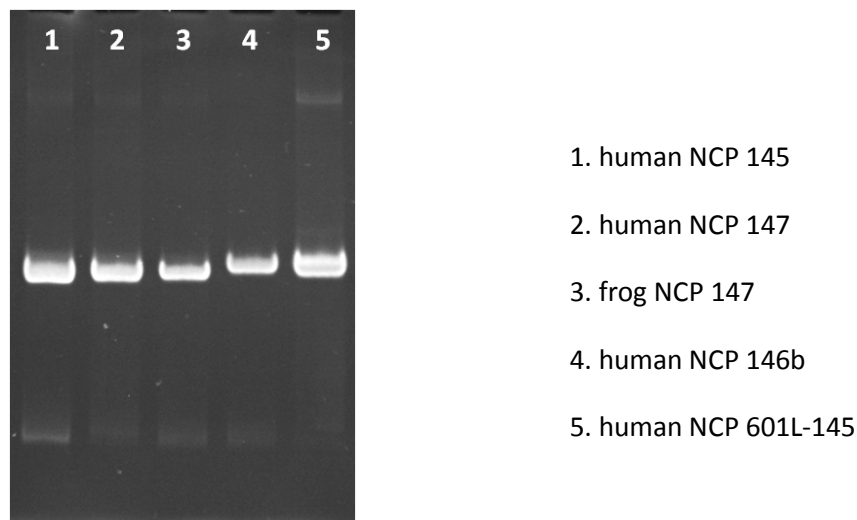


Figure 22. NCPs reconstituted from different histones (human and frog histones) and DNAs (145 bp, 147 bp, 146b bp, 601L-145 bp)

3.4. Nucleosome Core Particle Crystallization

For crystallographic analysis of nucleosome, centered NCPs were crystallized in standard NCP crystallization buffer [70-90 mM MnCl_2 , 50-80 mM KCl, 20 mM Kcac (pH 6.0)]. Crystal morphology produced in this buffer is depicted in Figure 23. The same crystal morphology and size was also produced in CoCl_2 -KCl buffer. Their diffraction quality, however, is significantly lower compared to crystals produced in MnCl_2 -KCl buffer.

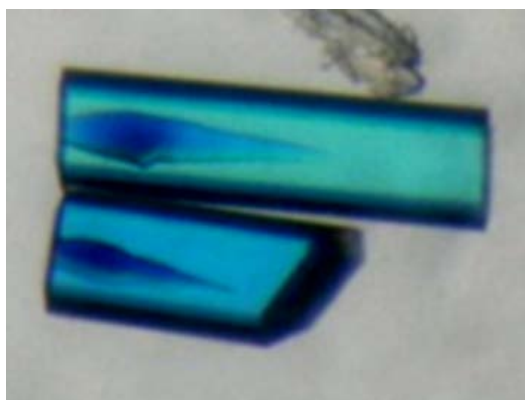


Figure 23. Standard NCP crystal morphology from MnCl_2 -KCl crystallization buffers. Crystals are about 1-2 mm length in their longest axis

Chapter IV

Nucleosome Core Particle Structure

4.1. Overall NCP Structure

We crystallized four human NCP structures reconstituted with 145 bp, 146b bp, 147 bp, and 601L-145 bp DNA sequences. Despite excellent crystal dimension and morphology, especially NCP601L145 being the largest and having the best morphology, only human NCP145, NCP146b, and NCP147 crystals gave diffraction qualities better than 3 Å. In addition we also obtained *Xenopus laevis* NCP145 crystals with much better diffraction data quality than the previous published structure (Ong *et al.*, 2007).

Crystals of human NCP145 and NCP146b have similar space group and unit cell parameters relative to their *Xenopus laevis* counterpart (Table 3). The human NCP146b unit cell is also smaller in size compared to Human NCP145. We did not manage to solve human NCP147 due to merohedral twinning within the crystals. In general, our human NCP structures (Figure 24) are quite comparable to *Xenopus laevis* NCP and to several other NCP structures from different organisms (Harp *et al.*, 2000; White *et al.*, 2001). This finding reiterates previous conclusions about conserved NCP structure across species.

There are in total 18 amino acid substitutions per nucleosome half between human and *Xenopus laevis* NCP. Eleven of them are apparent from the diffraction data (Figure 25). Our high data resolution allows amino acid substitutions to be clearly discerned based on 2Fo-Fc electron density maps (Figure 26). The absence of negative density around the substituted amino acids in Fo-Fc map was used to prove the amino acid identity differences between human and *Xenopus laevis* NCP. Out of eleven, only two of the amino acids are involved in histone-DNA interactions, while the rest are buried inside the histone core or facing away from nucleosomal DNA. Two amino acids that make contact with DNA are parts of the H2B N-terminal tail. They are Arg 31 and Ser 32 on human H2B, or Lys 31 and Thr 32 on *Xenopus laevis* NCP (Figure 27). These differences, however, do not cause substantial changes in DNA conformation around those amino acids.

Data Collection and Refinement Statistics for *Xenopus laevis* NCP145 and Human NCP145 and NCP146b

Data Collection			
NCP Crystals	NCP145	NCP145	NCP146b
Species	X.laevis	Human	Human
Space Group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
a, b, c (Å)	105.5, 109.9, 180.3	106.1, 109.7, 180.3	105.8, 109.6, 176.3
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	2.04-93.8 (2.04-2.15)	2.13-47.8 (2.13-2.25)	2.57-109.6 (2.57-2.71)
R-merge	4.6% (45.4%)	5.3% (52.4%)	11.4% (49.2%)
I/σI	14.4 (1.2)	17.4 (2.4)	9.2 (2.2)
Completeness (%)	90.9 (61.3)	97.9 (88.0)	92.1 (65.6)
Redundancy	4.6 (2.1)	6.9 (4.8)	5.2 (3.3)
Refinement			
Resolution (Å)	2.04-93.8	2.13-47.8	2.57-93.1
No. reflections	118697	113070	59493
R-work / R-free	24.2% / 26.0%	23.2% / 26.2%	22.4% / 26.9%
No. atoms	12119	12159	12027
Protein	6084	6084	6011
DNA	5939	5939	5980
H2O	96	136	36
B-factors (Å ²)	70.2	69.3	94.5
Protein	45.5	43.8	66.5
DNA	95.8	96	122.7
H2O	49.6	45.1	72.1
Rmsd			
Bond lengths (Å)	0.008	0.009	0.008
Bond angles (°)	1.46	1.42	1.49

Table 3. Data collection and refinement statistics for *Xenopus laevis* NCP145 and human NCP145 and NCP146b. Data sets are based on single crystal diffraction, and values in parentheses are for the highest-resolution shell.

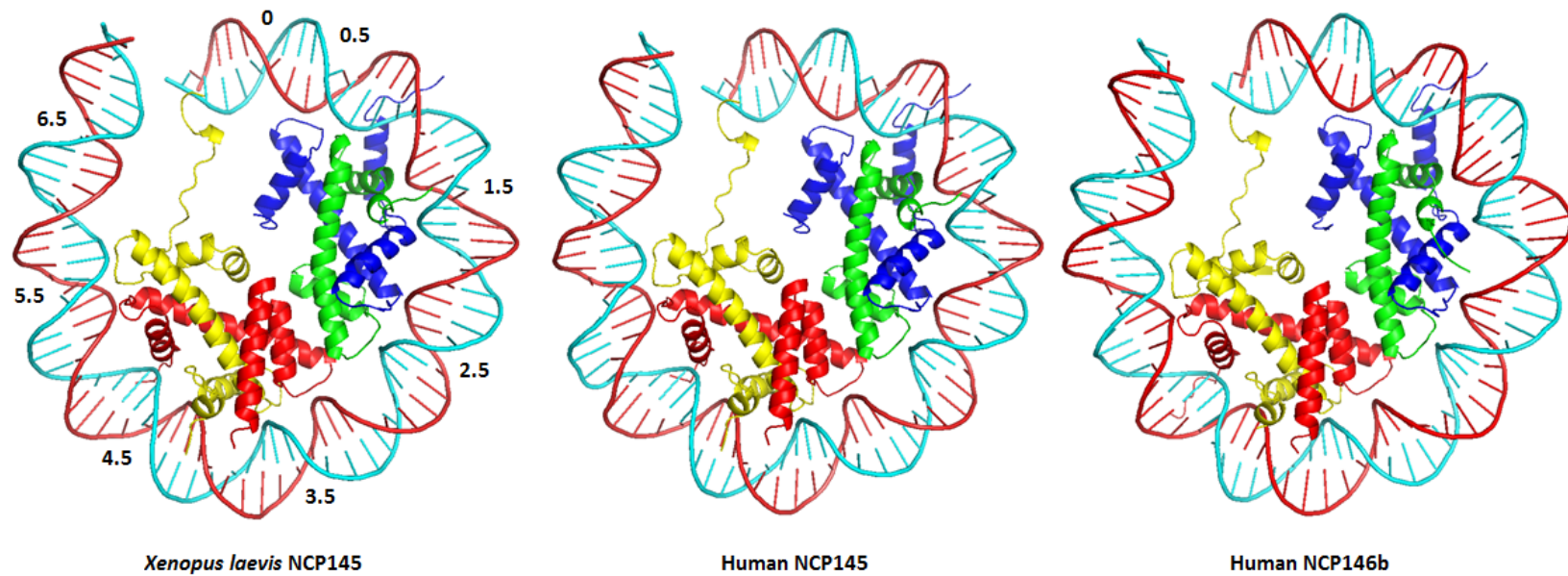


Figure 24. NCP structures with one half of the particle shown. Histones are color-coded; H2A (yellow), H2B (red), H3 (blue), H4 (green). SHL positions are shown on the *Xenopus laevis* NCP145.

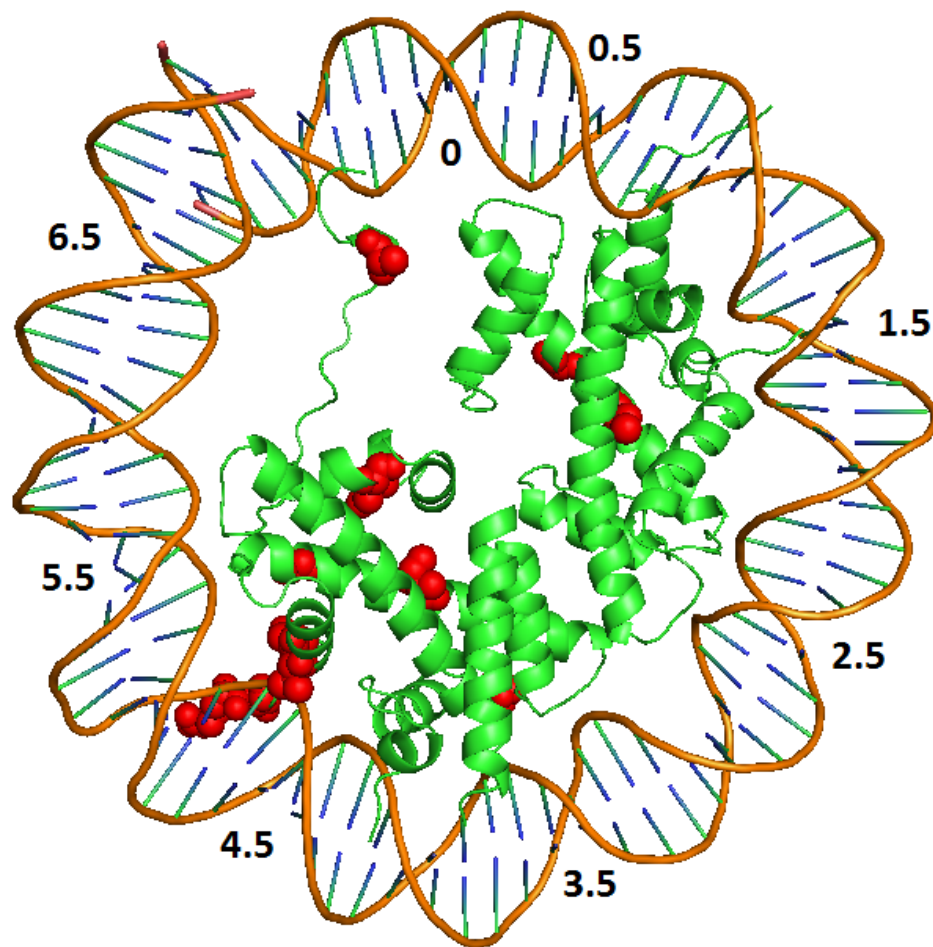


Figure 25. Positions of amino acid substitutions between *Xenopus laevis* and Human NCP. For clarity only half-NCP structure is shown.

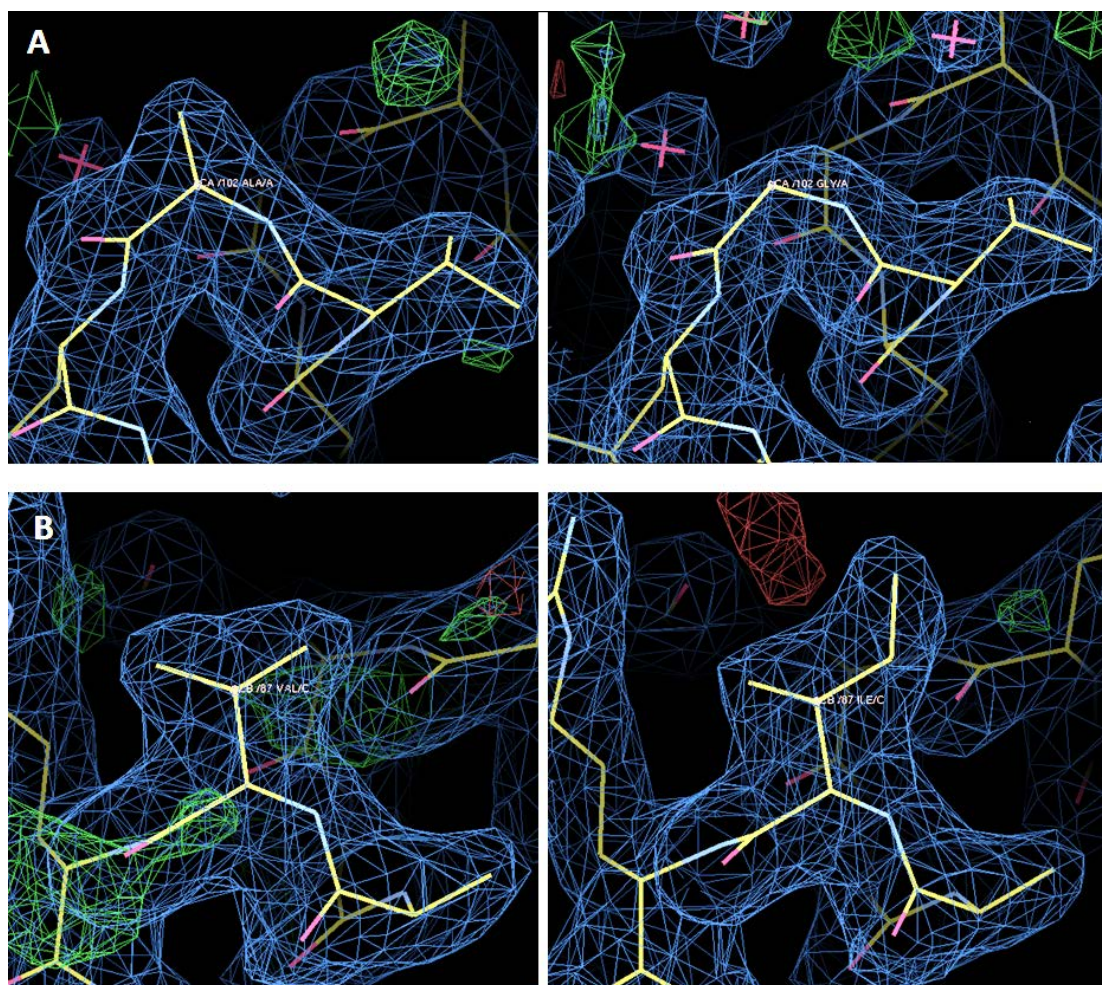


Figure 26. Amino acid differences on both *Xenopus laevis* and human histones are discernible in the crystal structures. (A) Alanine 102 on *Xenopus laevis* H3 (left) is replaced by glycine 102 on human H3 (right). (B) Valine 87 on *Xenopus laevis* H2A (left) is replaced by isoleucine 87 on human H2A (right). A 2Fo-Fc map (blue) is contoured at 1.5σ and a Fo-Fc map (green and red) is contoured at 3σ.

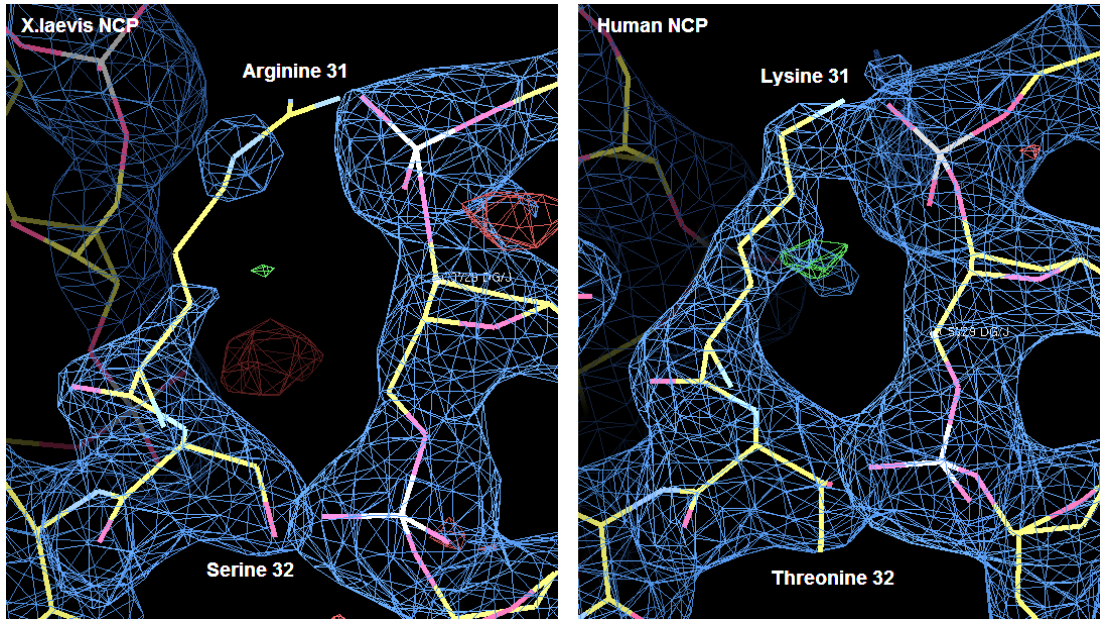


Figure 27. Two non-conserved amino acids of H2B make direct contact with DNA as shown by NCP crystal structures. Out of 18 non-conserved amino acids between *Xenopus laevis* and human histones, only two are observed to interact with the DNA in the crystal structures. A 2Fo-Fc map (blue) is contoured at 1.5σ and a Fo-Fc map (green and red) is contoured at 3σ .

4.1.1. *Xenopus laevis* NCP145 and Human NCP145 Structure

Although the overall NCP structures are quite similar, the DNA conformations of the human NCP145 and *Xenopus laevis* NCP145 crystal structures, particularly around the SHL +2 and +5 positions, seem to adopt different configurations (Figure 28). At the SHL +1.5 CC=GG basepair step, extreme DNA kinking into minor groove associated with large negative roll value (-50.8°) on *Xenopus laevis* NCP145 is replaced by a more subtle kink into the minor groove (-25.8°) in human NCP145. Moreover, this extreme DNA kink is accompanied by higher twist values in *Xenopus laevis* NCP145 as compared to human NCP145. This increased twist value in *Xenopus laevis* NCP145 allows DNA on this region to stretch over, thus advancing *Xenopus laevis* nucleosomal DNA by 1 bp as compared to human NCP (Figure 29, right box).

Nucleosomal DNA around SHL +5 of human NCP145 seems to replicate the same stretching strategy employed by nucleosomal DNA of *Xenopus laevis* NCP145 at SHL +2. Kinking into minor groove at SHL +4.5 TG=CA basepair step is slightly more pronounced in human NCP145 than in *Xenopus laevis* NCP145. Roll value at this particular basepair step is -18.7° and -6.1° in human NCP145 and *Xenopus laevis*

NCP145 respectively. The following DNA sequences on human NCP145 also adopt higher twist value than *Xenopus laevis* NCP145, thus advancing nucleosomal DNA of human NCP145 by 1 bp relative to *Xenopus laevis* NCP145 (Figure 29, left box).

Subtle differences in the DNA structure is also observed around the SHL -5 position (Figure 30). The outer DNA strand backbone of human NCP145 (Figure 30, blue) is slightly distorted in this region, resulting in rougher DNA curvature as compared to *Xenopus laevis* nucleosomal DNA (Figure 30, green) at the same position. Human nucleosomal DNA seems to utilize two separate kinks at TA=TA and AC=GT basepair steps to bend the DNA at this position, whereas *Xenopus laevis* nucleosomal DNA only utilize one kink, yet with higher roll value, at the central basepair step (CT=AG). The human CT=AG basepair step, in contrast, shows a much lower roll value. This single curving location allows *Xenopus laevis* DNA at SHL -5 of NCP145 to bend more smoothly than that of human NCP145.

Both *Xenopus laevis* and human NCP145 share similar extreme kinking and DNA stretching around the SHL -2 positions. Unlike stretching at +2 or +5 SHL, the roll value of TG=CA at -1 SHL is positive, indicating kinking towards major groove (Figure 28, overlapped green and blue bars). DNA stretching associated with this kinking is observed along SHL -1.0 until SHL -2.5. Since both NCP145s have the same number of stretches, the effective nucleosomal DNA lengths for both *Xenopus laevis* and human NCP145 are about the same. In summary, the main difference between the NCP145 crystal structures is that DNA stretching around SHL +2 in *Xenopus laevis* NCP145 has repositioned to SHL +5 in human NCP145, while stretching at SHL -1 to SHL -2.5 is conserved in both *Xenopus laevis* and human NCP145.

With respect to ion binding sites, both NCPs generally share similar Mn^{2+} and Cl^- ion localization. There is, however, one ion site that is unique to each of the NCPs. Manganese ion associated with N7 atom of both guanines at -33 J and -34 J is only observed in *Xenopus laevis* NCP145 (Figure 31, top panel). On the other hand, the opposite symmetry-related positions (guanines -33 I and -34 I) bind to Mn^{2+} in both human and *Xenopus laevis* NCP145 crystals (Figure 31, bottom panel, left). Different DNA register at chain J between human and *Xenopus laevis* NCP145 due to aforementioned stretching between SHL +1.5 and +2.5 in *Xenopus laevis* NCP145 is thought to be the reason for differential manganese localization in human and *Xenopus laevis* NCP145 (Figure 31, bottom panel, right).

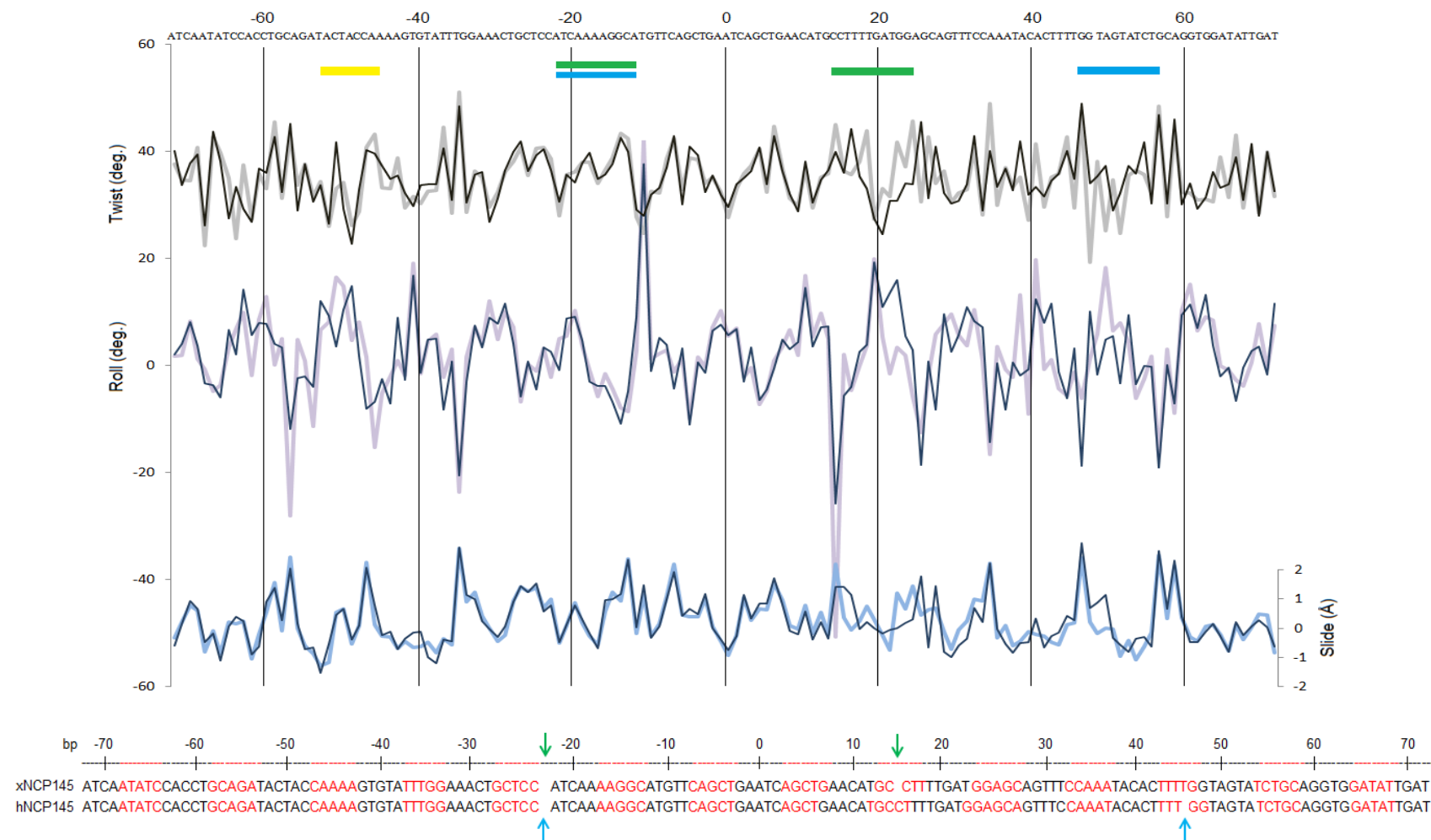


Figure 28. Nucleosomal DNA bp parameters for human NCP145 (thin line) and *Xenopus laevis* NCP145 (thick line). Positions of stretching for *Xenopus laevis* NCP145 are marked by green bars and green arrows, while positions of stretching for human NCP145 are marked by blue bars and blue arrows. A yellow bar marks local structural difference between both NCP (Figure 30). DNA bases with major groove facing the HO are colored black, while ones with minor groove facing the HO are colored red.

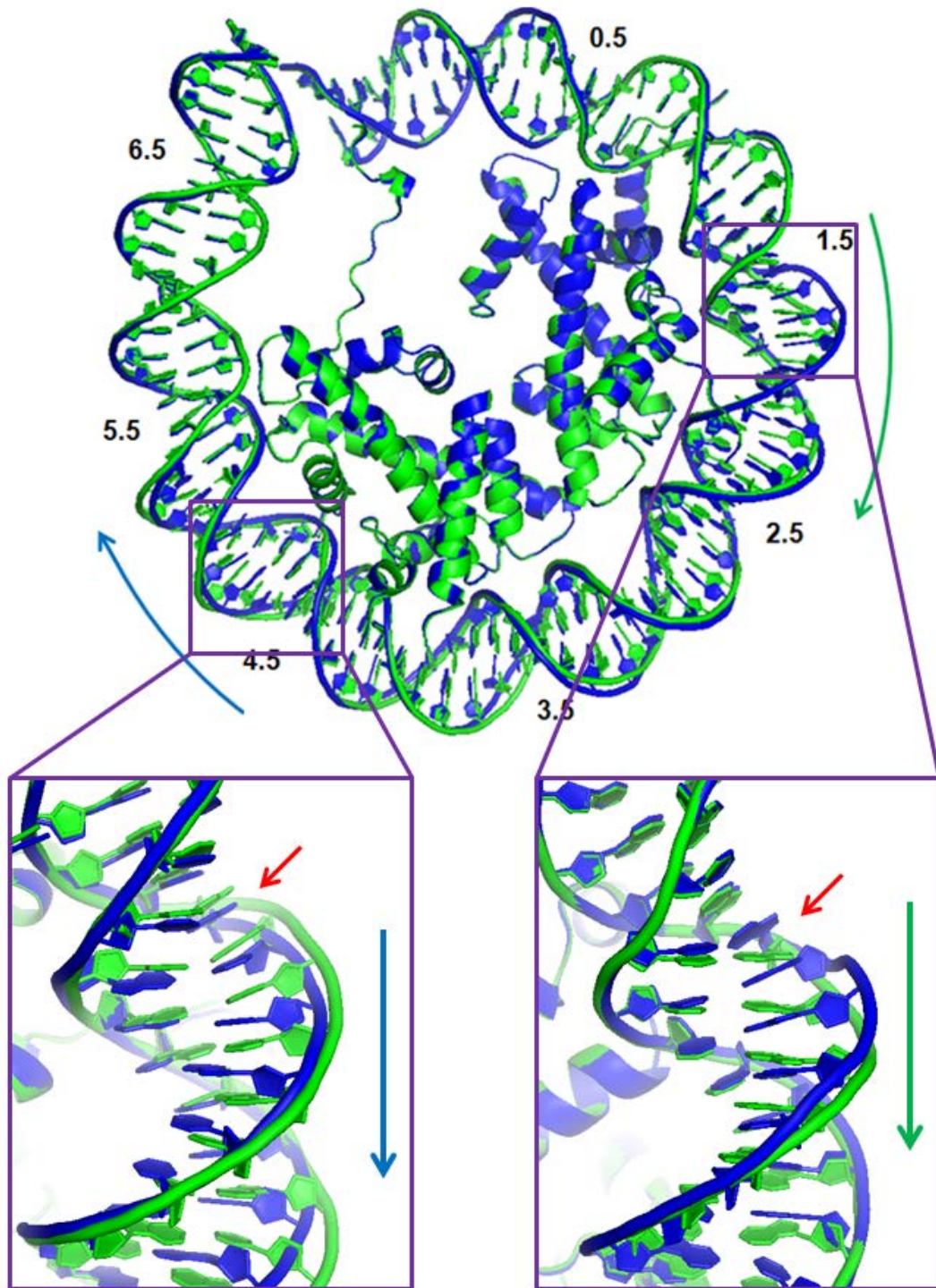


Figure 29. Location of DNA stretching in *Xenopus laevis* NCP145 (green, SHL +1.5) and human NCP145 (blue, SHL +5). Green and blue arrows show the direction of stretching on *Xenopus laevis* NCP145 and human NCP 145, respectively. Red arrows point to the start of stretching; at a CC=GG bp step in *Xenopus laevis* NCP145 (right box), and at a TG=CA bp step in human NCP145 (left box).

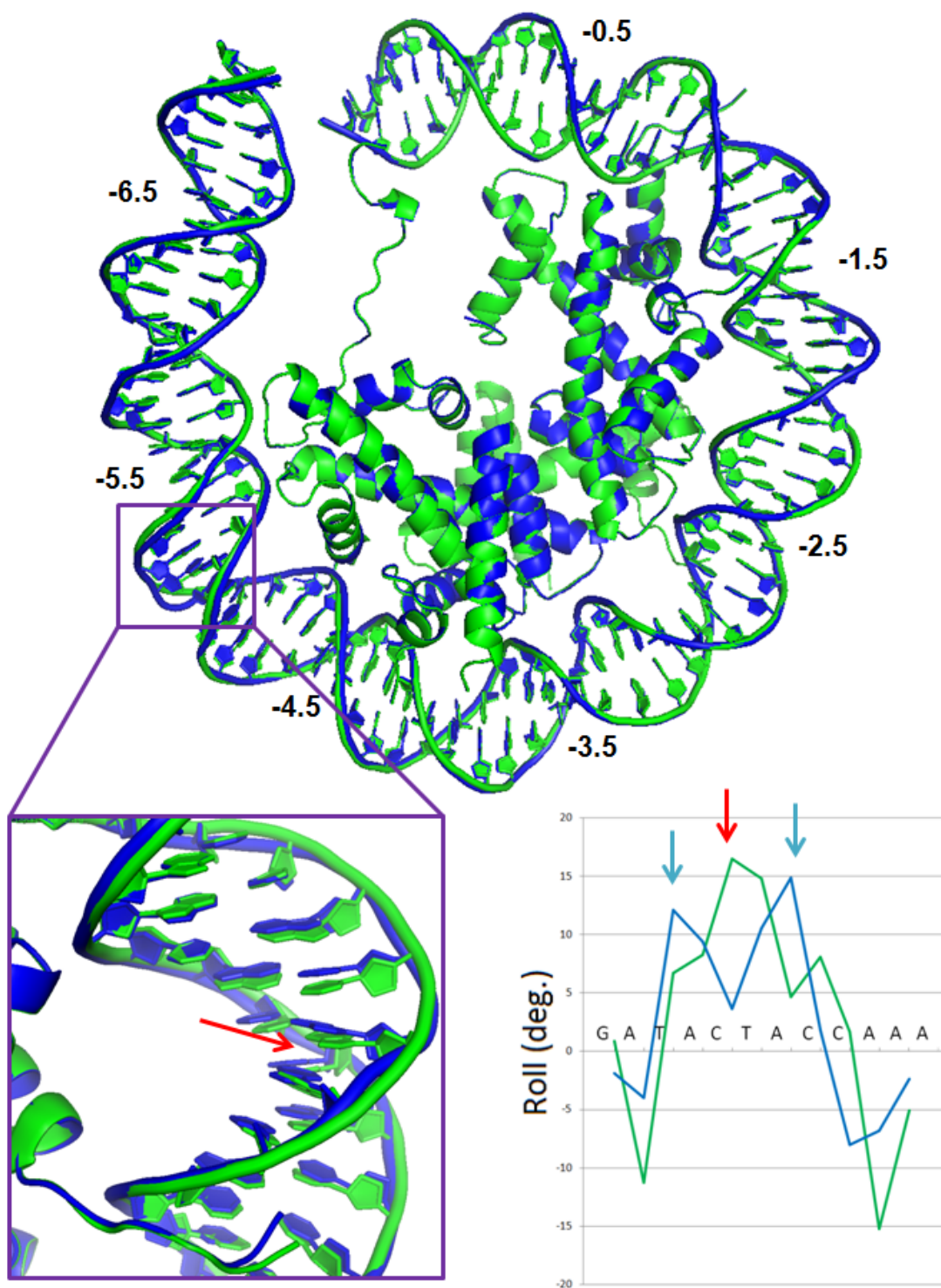


Figure 30. Local DNA conformational difference between *Xenopus laevis* NCP145 (green) and human NCP145 (blue) at SHL -5. A red arrow points to the location of a positive roll peak that mediates smooth bending of DNA in *Xenopus laevis* NCP145 at that region. Blue arrows point to two positive roll peaks in human NCP145.

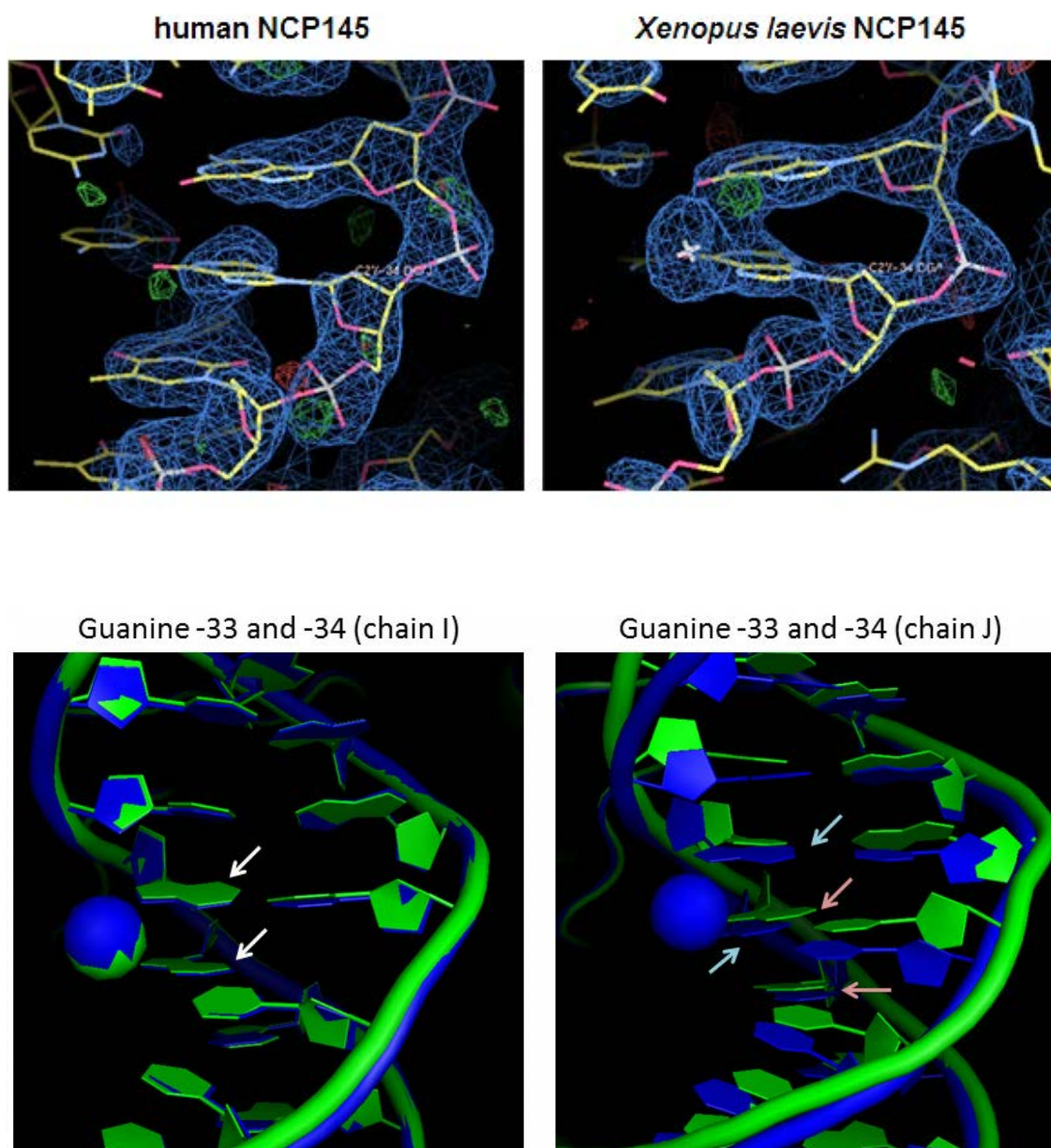


Figure 31.

Top Panel: Difference in manganese localization between human NCP145 (left panel) and *Xenopus laevis* NCP145 (right panel). Guanine -33 and -34 (chain J on *Xenopus laevis* NCP145) coordinate one Mn^{2+} ion at the major groove, while none is observed at the same position on human NCP145. A 2Fo-Fc map (blue) is contoured at 1.5σ and a Fo-Fc map (green and red) is contoured at 3σ .

Bottom Panel: Difference in manganese localization between symmetry-related positions. Human NCP145 (green) and *Xenopus laevis* NCP145 (blue) has similar DNA register at guanine -33 and -34 in chain I (left panel, white arrows). On the other hand, guanine -33 and guanine -34 in chain J have different DNA register (right panel). Cyan arrows point at guanine -33 and -34 in *Xenopus laevis* NCP145 chain J (blue). Brown arrows point at guanine -33 and -34 in human NCP145 chain J (green). Manganese atoms are shown in blue (*Xenopus laevis*) and green (human)

4.1.2. Human NCP146b Structure

Both direct structural inspection and DNA parameter calculation show that *Xenopus laevis* and human NCP146b have very similar structures (Figure 32). In spite of this, we found that the electron density for the terminal DNA (SHL +5 to +7) of human NCP146b is poorly defined, perhaps indicating high DNA flexibility around this area. B-factor plot of the human NCP146b DNA also indicates that this particular DNA terminus is quite disorganized in the crystal structure (Figure 33, plot on the right side). On the contrary, the opposing DNA terminus (SHL -5 to SHL -7; Figure 33, plot on the left side) is relatively more ordered. We did not observe any significant differences between the human and *Xenopus laevis* NCP146b structures.

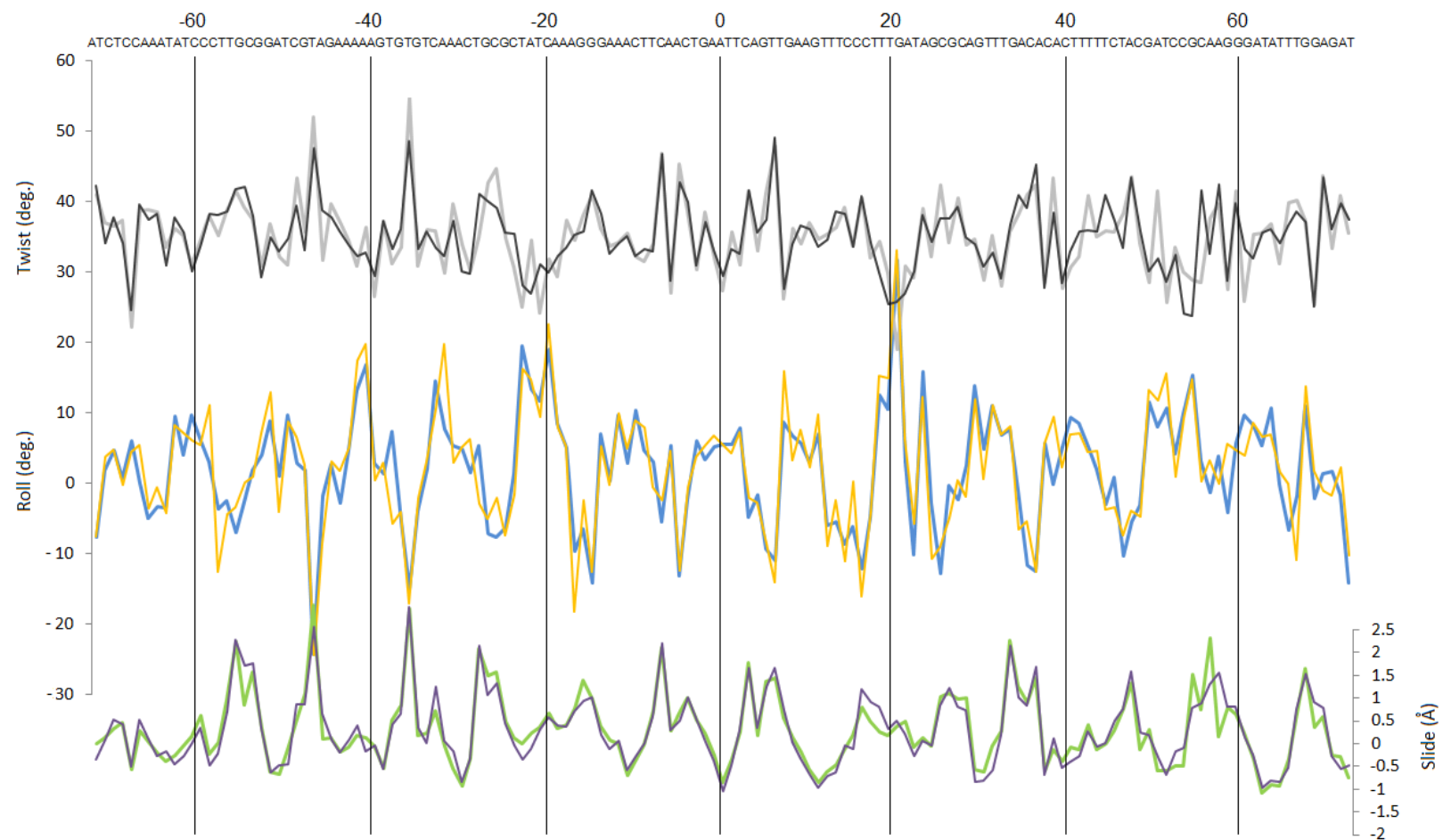


Figure 32. Nucleosomal DNA bp step parameters for human NCP146b (thin front line) and *Xenopus laevis* NCP 146b (thick background line).

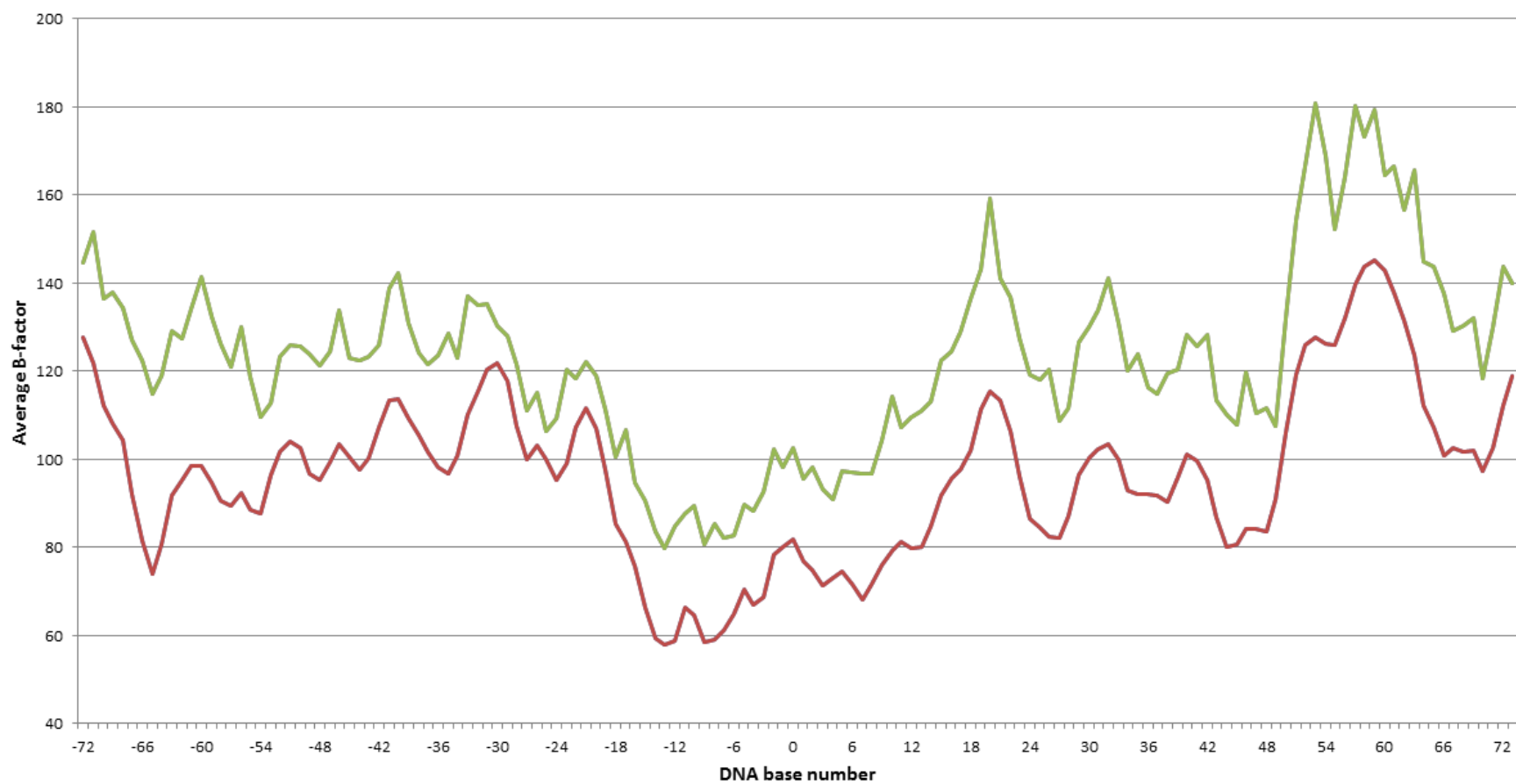


Figure 33. B-factor distribution for DNA of human NCP146b (green line) and *Xenopus laevis* NCP146b (brown line). B-factor data is based on the average B-factor of DNA chain I and J. Numbers on X-axis correspond to chain I numbering scheme.

Chapter V

Ruthenium Agent Interaction with Nucleosome Core Particle

5.1. Result

We inspected the interaction of five RAPTA derivatives (Figure 34) with the NCP by using X-ray crystallography. Native NCP crystallized under standard manganese condition was buffer exchanged to magnesium sulphate buffer before incubation with different RAPTA compounds at two different RAPTA concentrations. RAPTA-treated NCP crystals were subjected to diffraction analysis after one or two days of incubation.

RAPTA compounds consist of four different groups; a ruthenium (II) centre, coordinated to two chloride atoms, a 1,3,5-triaza-7-phosphaadamantane (PTA) ligand, and a η^6 -arene ring. The designation of each RAPTA compound is based on the modification present on the arene ring (Table 4). The initial NCP crystallographic study on a RAPTA drug was done with RAPTA-C (Wu *et al.*, 2011).

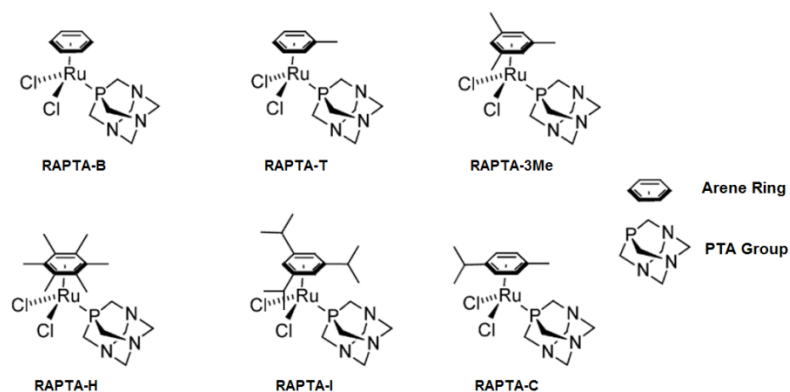


Figure 34. Structure of RAPTA compounds

Name	Arene identity	Molecular Formula
RAPTA-B	benzene	$\text{RuCl}_2(\eta^6\text{-C}_6\text{H}_6)(\text{PTA})$
RAPTA-T	toluene	$\text{RuCl}_2(\eta^6\text{-C}_7\text{H}_8)(\text{PTA})$
RAPTA-3Me	1,3,5-trimethylbenzene	$\text{RuCl}_2(\eta^6\text{-C}_9\text{H}_{12})(\text{PTA})$
RAPTA-H	hexamethylbenzene	$\text{RuCl}_2(\eta^6\text{-C}_6(\text{CH}_3)_6)(\text{PTA})$
RAPTA-I	1,3,5-triisopropylbenzene	$\text{RuCl}_2(\eta^6\text{-C}_{15}\text{H}_{24})(\text{PTA})$
RAPTA-C	4-Isopropyltoluene	$\text{RuCl}_2(\eta^6\text{-C}_{10}\text{H}_{14})(\text{PTA})$

Table 4. RAPTA compound identity

We replicated the previous RAPTA-C $[\text{RuCl}_2(\eta^6\text{-}p\text{-cymene})(\text{PTA})]$ - NCP study (Wu *et al.*, 2011) by treating crystals of *Xenopus laevis* and human NCP145, buffered in $\text{MgSO}_4 + \text{KCaco}$ (pH 6.0) solution, with 0.25 mM and 0.75 mM of RAPTA compounds. In contrast to the previous RAPTA-C treatment (0.75 mM RAPTA-C concentration, two days treatment), which still yielded 2.5 Å crystals data resolution, similar treatment of NCP145 with four of the compounds (RAPTA-B, RAPTA-T, RAPTA-3Me, RAPTA-H) severely reduced diffraction data quality. This result prompted us to use lower drug concentration (0.25 mM) for subsequent treatments. Diffraction images were collected after one or two days of incubation.

Crystallographic analysis of RAPTA-treated NCP shows no major crystal parameter alteration from native NCP crystals (Table 5). Several strong anomalous signals (anomalous peak $> 3\sigma$) corresponding to the ruthenium (II) atom of RAPTA compounds, were observed on the surface of the HO core (Table 6), but none on the DNA surface, with exception of RAPTA-B-treated NCP. Among the ruthenium signals on the surface of histones, two of them consistently appear near Glu 61 of H2A (chain G; Site A; Figure 35 and Figure 37) and His 106 of H2B (chain H; Site B; Figure 35 and Figure 38), which are the same binding sites observed for RAPTA-C as shown in the previous study. The ruthenium (II) atom of RAPTA compounds at Site A is engaged in coordination with H2A Glu 61 and H2A Glu 64. On the other hand, Ruthenium (II) atom at Site B is held in coordination with H2B His 106 and H2B Glu 102. NCP crystals treated with RAPTA-B at 0.25 mM concentration for two days are the only crystals showing a ruthenium anomalous signal on the DNA (Figure 36). Milder RAPTA-B treatment (one day treatment, 2.4 Å data resolution) did not give any sign of ruthenium presence on that site, while harsher treatment (0.75 mM RAPTA-B concentration, two days treatment) damaged the crystals too severely to see any ruthenium anomalous signal (3.7 Å data resolution). The relatively weak ruthenium signal on the DNA does not allow us to model RAPTA-B on that particular site.

Interestingly, human NCP treated with RAPTA-I only showed one RAPTA compound on the NCP surface. Site A gives a very strong anomalous signal ($>9.5\sigma$), while Site B is relatively empty. Comparison of anomalous density with the same compounds on *Xenopus laevis* NCP (site A 6.4σ , Site B 4.7σ) indicates a potential difference in the RAPTA-I binding profiles between the two NCP constructs (Figure 38).

Among the RAPTA compounds, RAPTA-H is the only one that does not give any interpretable anomalous signal after one day treatment. Data analysis of the two-day

drug treatment gives a very high B-factor for two RAPTA-H compounds observed on the surface of the NCP, which may suggest low occupancy of RAPTA-H on both site A and B. The effective size of arene ring on the top of RAPTA-H is extended by incorporation of six methyl groups, which seemingly inhibits the co-existence of two neighbouring RAPTA-H adducts on the same NCP particle in the crystals.

Despite the 'pseudo' two-fold symmetry of NCP, drug binding is mostly observed only on one side of the nucleosome. Anomalous signals are observed near H2A and H2B (chain G and H, respectively) on one side of the nucleosome, but not on the other H2A and H2B (chain C and D, respectively) symmetry-related sites. Both RAPTA-B and RAPTA-T, however, are the exceptions. RAPTA-B and RAPTA-T show anomalous signals on the symmetry opposed histones. Anomalous signals for RAPTA-B are present on both new Site A' (4.6σ) and new Site B' (4.7σ , Figure 39), while RAPTA-T is only present on new Site B', with apparently lower occupancy (3.1σ). The previous study of the RAPTA-C compound did not indicate any adduct formation at these sites on the other face of the NCP.

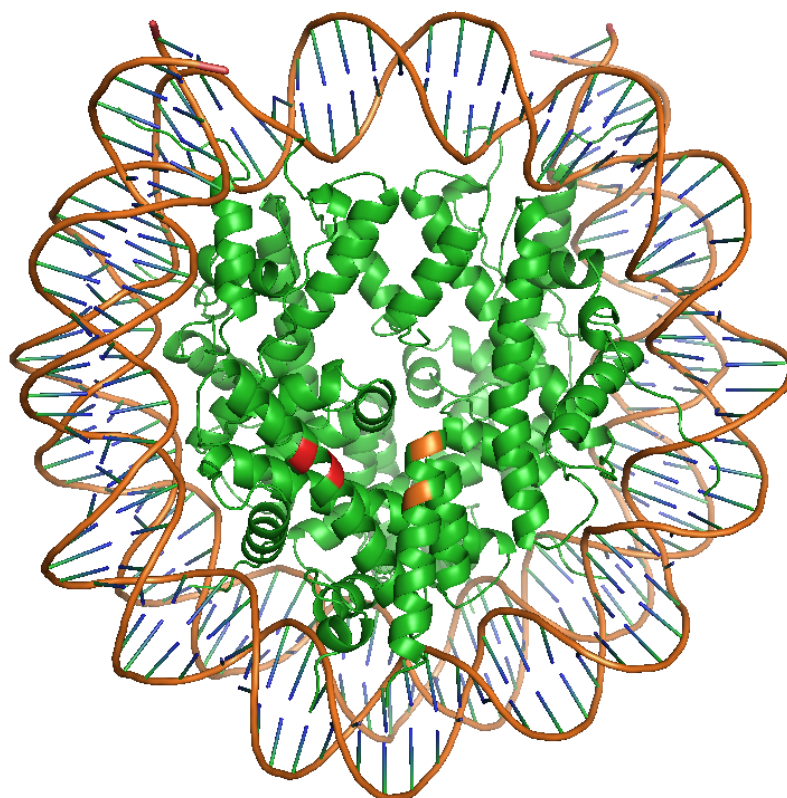


Figure 35. Two consensus binding sites for RAPTA compounds on the surface of *Xenopus laevis* NCP145. Site A on histone H2A (red patch) and Site B on histone H2B (orange patch) are coloured differently.

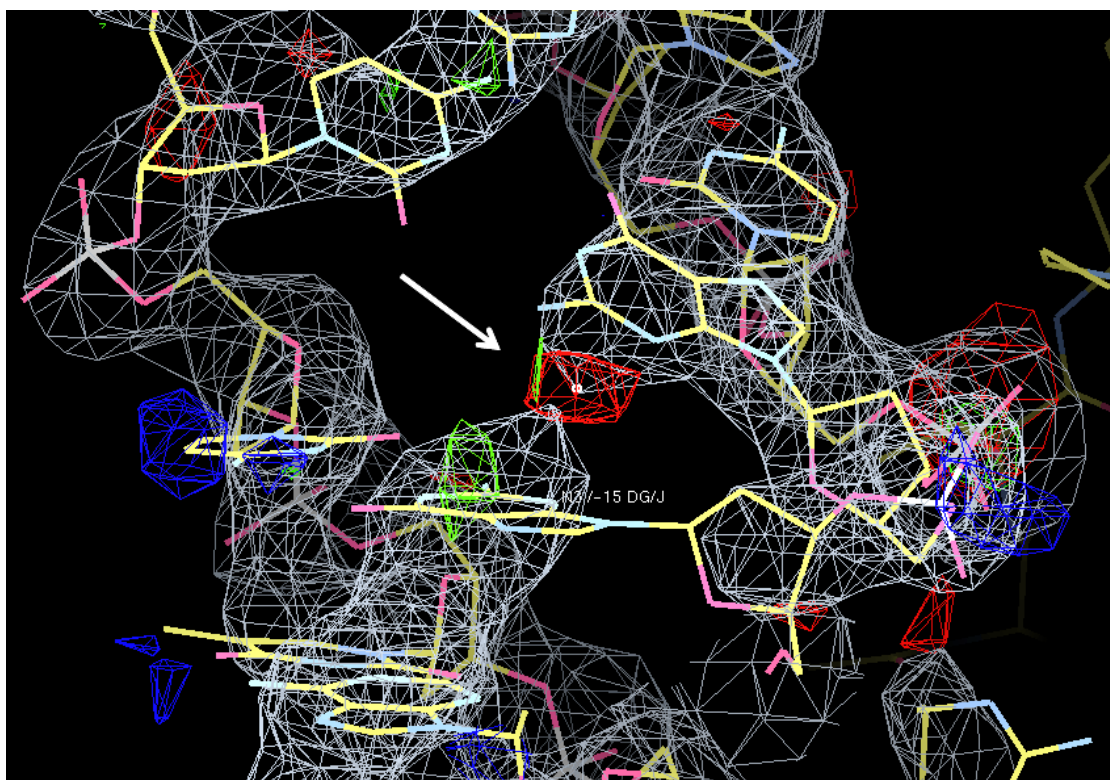


Figure 36. RAPTA-B-treated NCP shows anomalous signal on nucleosomal DNA at nucleotide -15 (chain J). Electron density maps are contoured at 1.5σ (2Fo-Fc, white mesh), 3σ (Fo-Fc, red mesh), and 3σ (anomalous map, green mesh).

Data Collection Statistics for Six RAPTA-treated NCP Crystals

Data Collection						
NCP Source	<i>X.laevis</i>	<i>X.laevis</i>	<i>X.laevis</i>	<i>X.laevis</i>	<i>X.laevis</i>	Human
Compounds	RAPTA-B	RAPTA-T	RAPTA-3Me	RAPTA-H	RAPTA-I	RAPTA-I
Soak duration (days)	2	1	2	2	2	2
Space Group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions						
a, b, c (Å)	106.4, 109.7, 181.2	106.7 109.8 182.3	106.3 109.7 181.8	106.7 109.8 181.4	106.9 110.1 182.9	107.2 109.6 181.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	2.79-60.4 (2.79-2.94)	2.81-60.8 (2.81-2.96)	2.85-60.6 (2.85-3.00)	2.58-60.5 (2.58-2.72)	3.12-61.0 (3.12-3.29)	2.81-60.6 (2.81-2.96)
R-merge	7.8% (48.9%)	5.9% (50.1%)	7.4% (50%)	4.6% (47.3%)	6.5% (49.2%)	5.8% (49.7%)
I/σI	14.5 (2.5)	15.5 (2.3)	13.4 (2.3)	18.6 (1.7)	13.5 (3.3)	14.3 (1.5)
Completeness (%)	91.1 (61.5)	96.2 (78.2)	98.9 (94.0)	89.6 (52.7)	99.3 (98.7)	95.0 (74.3)
Redundancy	6.6 (4.3)	6.6 (4.3)	6.7 (4.3)	5.8 (2.3)	6.6 (6.5)	6.1 (2.8)

Table 5. Data collection and refinement statistics for six RAPTA-treated NCP crystals. Data sets are based on single crystal diffraction, and values in parentheses are for the highest-resolution shell. (continued to next page)

Refinement Statistics for Six RAPTA-treated NCP Crystals

Refinement						
NCP Crystals	<i>X.laevis</i>	<i>X.laevis</i>	<i>X.laevis</i>	<i>X.laevis</i>	<i>X.laevis</i>	Human
Compounds	RAPTA-B	RAPTA-T	RAPTA-3Me	RAPTA-H	RAPTA-I	RAPTA-I
Resolution (Å)	2.79-60.4	2.81-60.8	2.85-60.6	2.58-60.5	3.12-61.0	2.81-60.6
No. reflections	47616	49694	48757	59398	38016	49261
R-work / R-free	21.0 % / 26.5%	21.1% / 24.8%	21.7% / 25.2%	22.7% / 27.0%	19.2% / 24.5%	21.5% / 27.9%
No. atoms	12080	12080	12085	12091	12093	12067
Protein	6084	6084	6084	6084	6084	6084
DNA	5939	5939	5939	5939	5939	5939
H2O	7	5	6	6	2	2
Ion	16	16	16	16	16	16
RAPTA compounds	34	36	40	46	52	26
B-factors (Å²)	113	102	99	105	109	112
Protein	76	69	69	70	77	78
DNA	149	135	130	140	141	148
H2O	58	54	53	54	55	59
Mg	58	57	57	59	61	58
SO₄	135	122	120	124	144	145
Rapta compounds	191	169	169	220	185	137
Rmsd						
Bond lengths (Å)	0.008	0.008	0.007	0.007	0.007	0.008
Bond angles (°)	1.48	1.5	1.48	1.5	1.5	1.5

Anomalous Signal Positions of Ruthenium Atoms of RAPTA compounds in NCP Crystals

RAPTA-B

Concentration	Days	Resolution	Sites (anomalous signal strength)
0.25 mM	1	2.5 Å	H2A Glu 61 (4 σ), H2B His 106 (3.3 σ)
0.25 mM	2	2.8 Å	H2A Glu 61 (6.56 σ), H2B His 106 (4.9 σ)
0.75 mM	2	3.7 Å	H2A Glu 61 (6.8 σ), H2B His 106 (4.9 σ), H2A' Glu 61 (4.6 σ), H2B' His 106 (4.7 σ), H4 Arg 17 (3.4 σ)

RAPTA-T

Concentration	Days	Resolution	Sites (anomalous signal strength)
0.25 mM	1	2.8 Å	H2A Glu 61 (7.2 σ), H2B His 106 (6.2 σ)
0.25 mM	2	3.3 Å	H2A Glu 61 (5 σ), H2B His 106 (6.9 σ), H2B' His 106 (3.1 σ), H2B Asp 48 (4.3 σ)

RAPTA-3Me

Concentration	Days	Resolution	Sites (anomalous signal strength)
0.25 mM	1	2.5 Å	H2A Glu 61 (5.1 σ), H2B His 106 (3.5 σ)
0.25 mM	2	2.85 Å	H2A Glu 61 (5.9 σ), H2B His 106 (4.4 σ), H2A' Glu 91 (3.8 σ), H3' Asp 77 (3.2 σ)

RAPTA-H

Concentration	Days	Resolution	Sites (anomalous signal strength)
0.25 mM	1	2.5 Å	-
0.25 mM	2	2.6 Å	H2A Glu 61 (5.7 σ), H2B His 106 (3.8 σ)

RAPTA-I

Concentration	Days	Resolution	Sites (anomalous signal strength)
0.75 mM	2	3.1 Å	H2A Glu 61 (6.4 σ), H2B His 106 (4.7 σ), H3' Asp 77 (3.8 σ)
1 mM	1	3.1 Å	H2A Glu 61 (8.7 σ), H2B His 106 (3.5 σ), H3' Asp 77 (4.6 σ)

RAPTA-I - human NCP

Concentration	Days	Resolution	Sites (anomalous signal strength)
0.75 mM	1	2.8 Å	H2A Glu 61 (11.1 σ)
0.75 mM	2	2.8 Å	H2A Glu 61 (9.6 σ)

Table 6. Positions for anomalous signals of Ruthenium atoms of RAPTA compounds. H2A', H2B', H3', H4' are symmetry counterparts of H2A, H2B, H3, H4 on the same NCP.

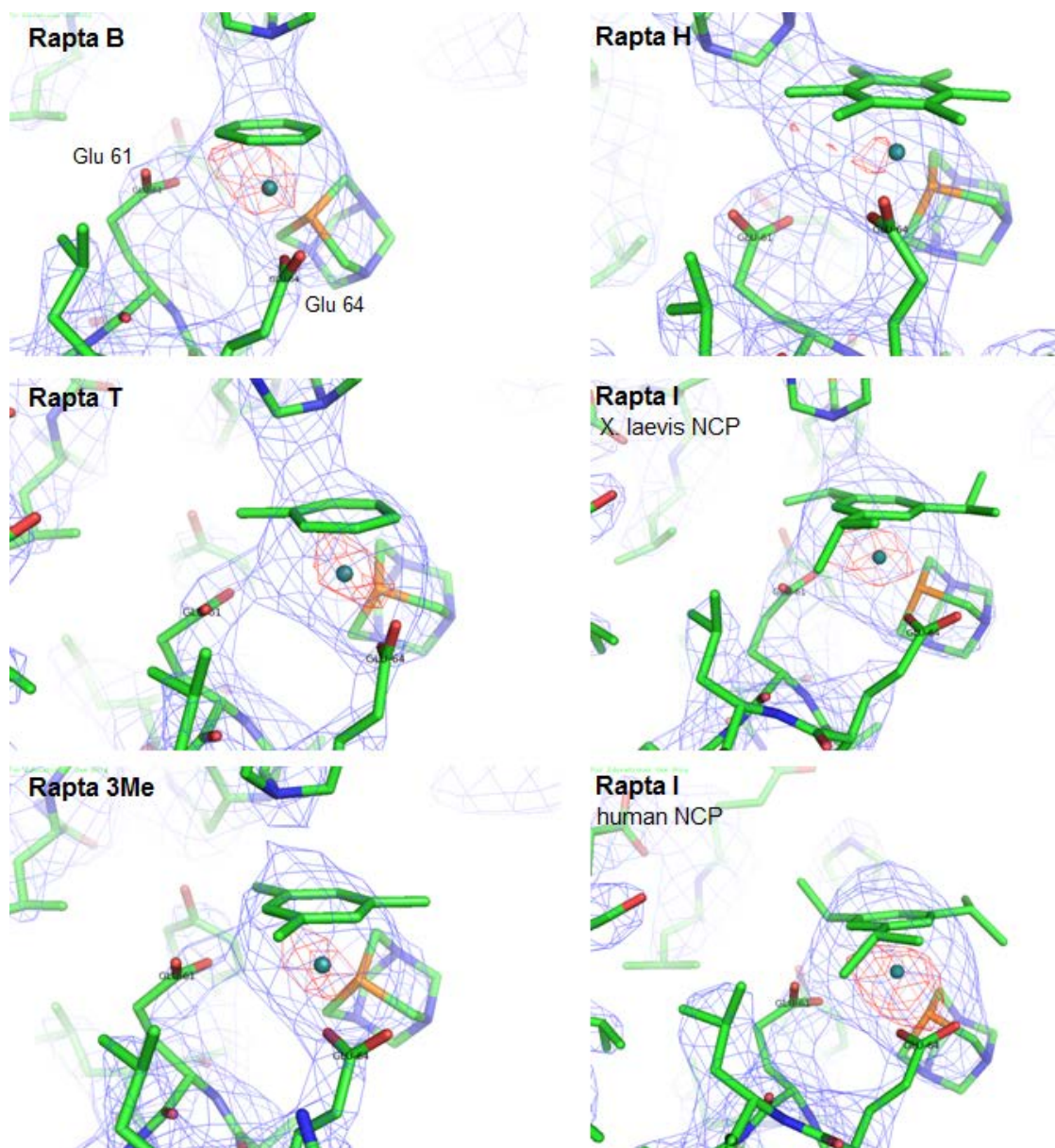


Figure 37. Interaction of RAPTA compounds with H2A Glu 61 and H2A Glu 64 (Site A). Both amino acids are positioned in the same orientation in all the images. 2Fo-Fc maps (blue mesh) are contoured at 1.5σ . Anomalous maps (red mesh) are contoured at 5σ .

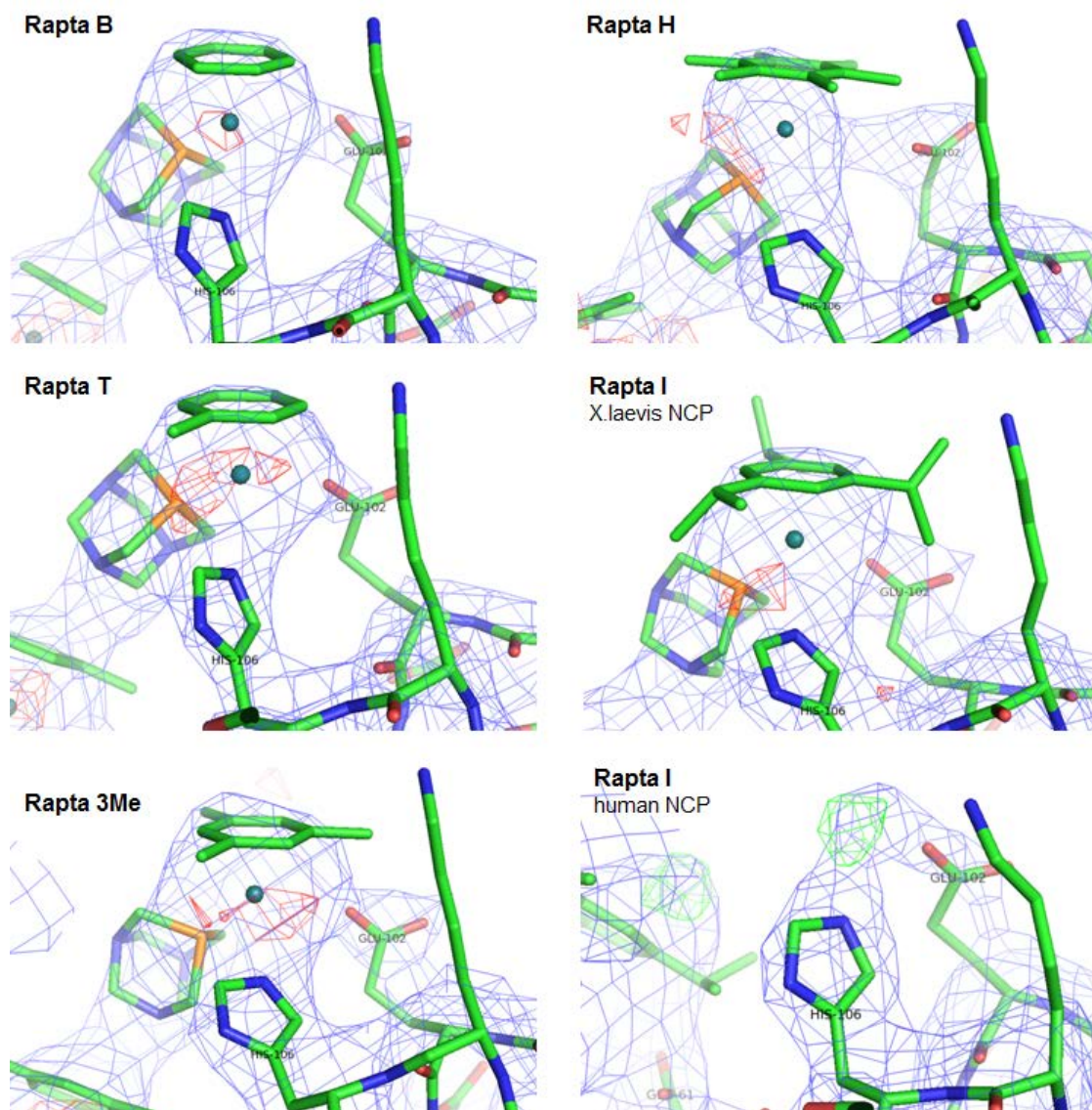


Figure 38. Interaction of RAPTA compounds with H2B His 106 and H2B Glu 102 (Site B). Both amino acids are positioned in the same orientation in all the images. 2Fo-Fc maps (blue mesh) are contoured at 1σ (RAPTA-3Me, RAPTA-H, RAPTA-I *X.laevis*) or 1.5σ (RAPTA-B, RAPTA-T). Anomalous maps (red mesh) are contoured at 5σ (RAPTA-B, RAPTA-T), 4σ (RAPTA-I *X.laevis*), or 3σ (RAPTA-H). RAPTA-I-treated human NCP does not show any anomalous signal at 3σ . Both 2fo-fc map (1σ) and fo-fc map (green mesh, 3σ) do not show any indication of strong RAPTA-I adduct formation in human NCP Site B.

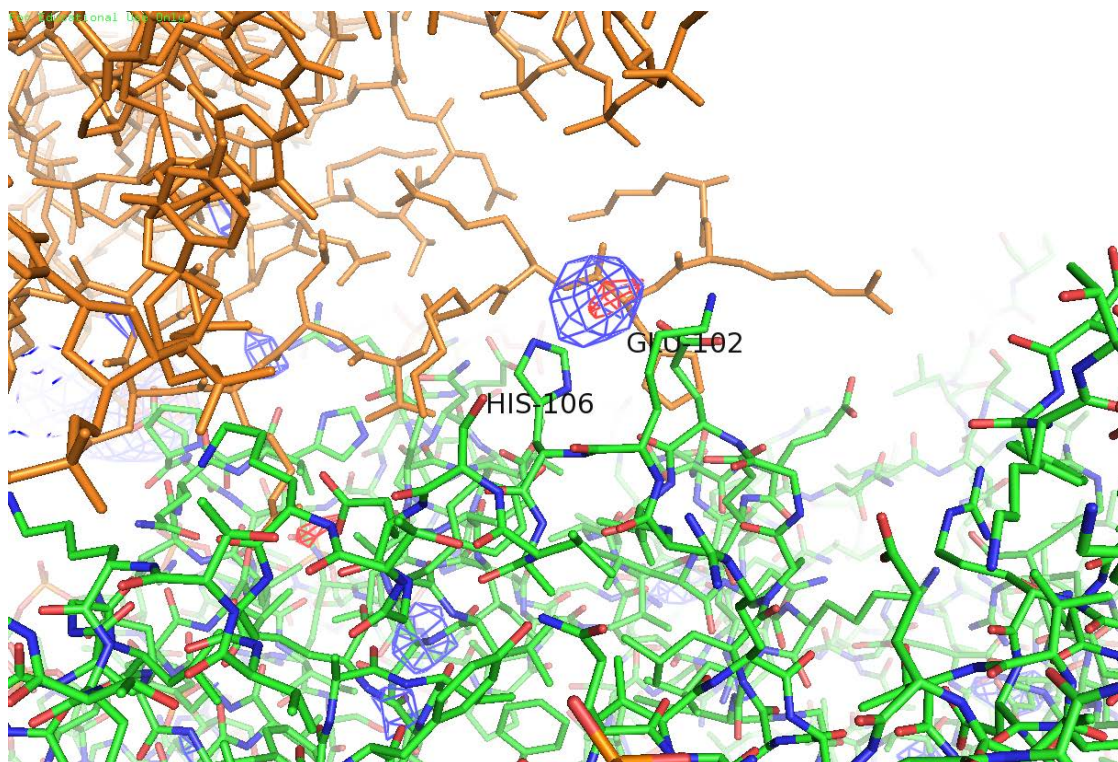


Figure 39. RAPTA-B treatment yields an anomalous signal on the symmetry-related location of Site B (Site B'). Both 2Fo-Fc (blue mesh) and anomalous (red mesh) maps are contoured at 4σ .

Chapter VI

HMGN Interaction with Nucleosome Core Particle

6.1. Result

6.1.1. HMGN Triggers Variant-specific Conformational Changes in Nucleosome Structure Without Apparent Loss of Histone Proteins

EMSA assessment of HMGNs in complex with NCP and nucleosomal arrays in the context of divalent ion-containing running buffers has revealed striking variant-specific differences in mobility. In particular, HMGN2 and truncated HMGN1 complexes showed profound retardation in mobility as compared to full length HMGN1, indicating a possibility of having a more open nucleosome conformation due to the action of either HMGN2 or truncated HMGN1. We termed the fast-migrating species corresponding to around 2:1 HMGN-to-NCP ratios as S1, and the much slower-migrating species as S2 (Figure 40).

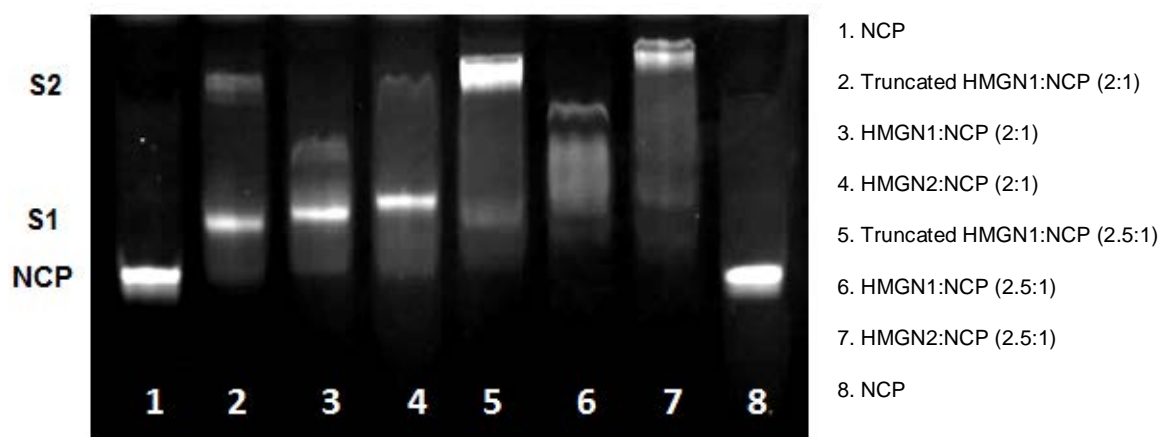


Figure 40. Formation of distinct slow-migrating species (S2) triggered by truncated HMGN1 (lane 5) and HMGN2 (lane 7), but not HMGN1 (lane 6) in the presence of divalent ion. Samples were run in native 6% PAGE gel, 1X TBE + 1 mM MgCl₂.

We came up with several theories on how this S2 species may form. The polymerization of nucleosomes, decompaction of the nucleosome due to histone loss or decompaction of the nucleosome due to the action of HMGN on the nucleosomal DNA termini were some of the theories. We conducted Western blot analysis on these assemblies to investigate qualitatively the content of histone for this S2 species. The result did not reveal any substantial changes in histone protein amount between the

two conformational states (S1 versus S2) for either NCP or nucleosome array (Figure 41). This ruled out the possibility of having one or some of the histones evicted from the nucleosome upon S2 formation. In a separate experiment, our group also found that S1 and S2 formation is reversible *in vitro*, thus supporting our previous conclusion about S2 formation (Ong *et al.*, 2010).

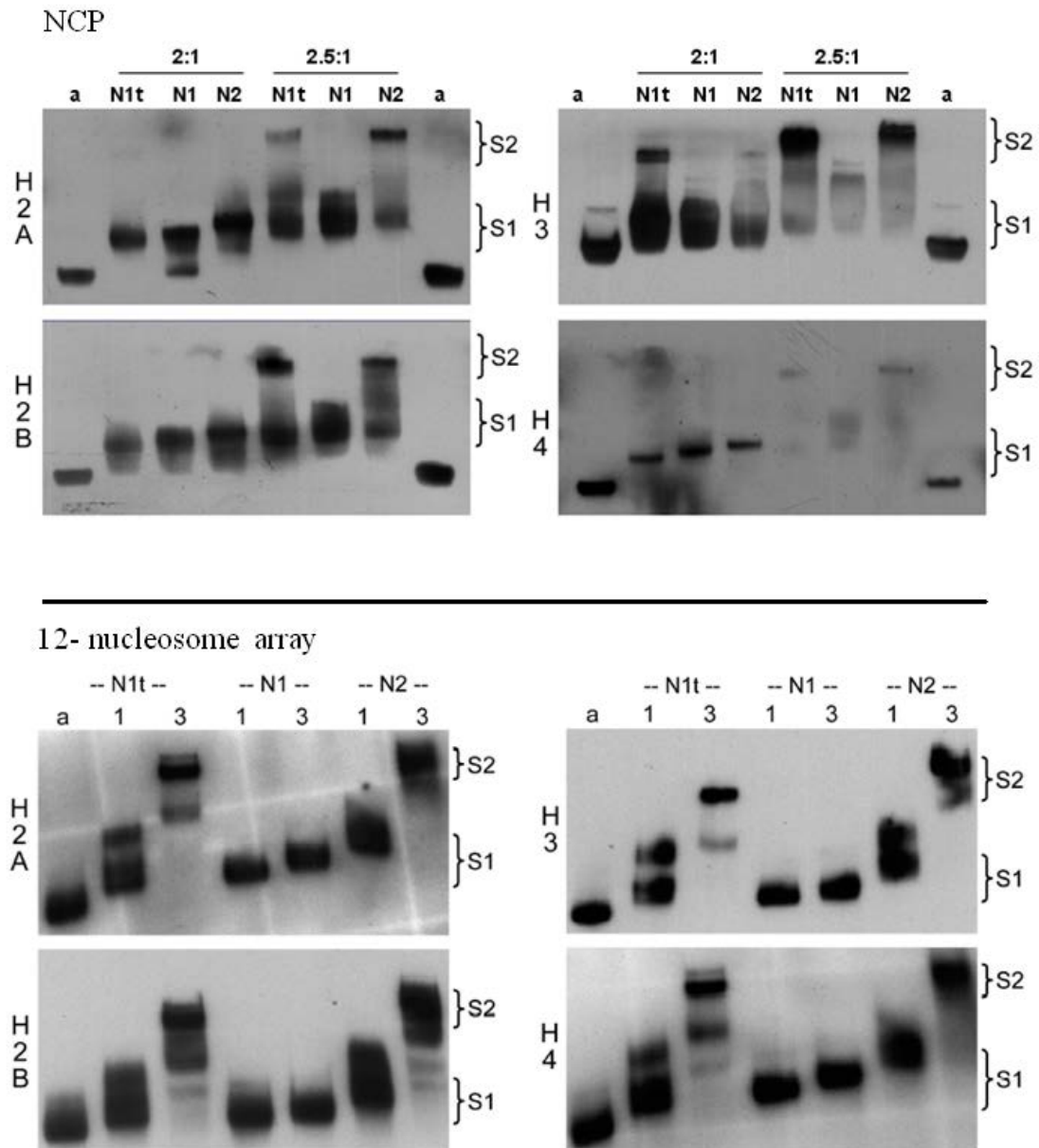


Figure 41. Western blot analysis of HMGN - NCP interactions. Histone antibodies are listed on the left side of the images. Similar to NCP and 2:1 HMGN:NCP species (S1), the slow-moving species (S2) does not show any loss of histone. HMGN-to-NCP ratios used on each lane are listed on top of the lanes. (NCP or 12-nucleosome array, a; truncated HMGN1, N1t; HMGN1, N1; HMGN2, N2)

6.1.2. Higher Stoichiometry of HMGN is Observed on S2 Species

We also conducted a fluorescence assay to further investigate the identity of the S2 species. We started our assay by optimizing the fluorescence intensity (red for NCP, green for HMGN) in a way that when HMGN-to-NCP ratio is 2:1 (S1) the band will turn yellow (Figure 42). Lesser than 2:1 HMGN-to-NCP stoichiometry, the band will be reddish, and inversely, higher than 2:1 stoichiometry, the band will be greenish. A 1:1 dye-to-HMGN2 ratio was used for subsequent fluorescence analysis to minimize the chance of the dye altering HMGN binding behaviour with the nucleosome.

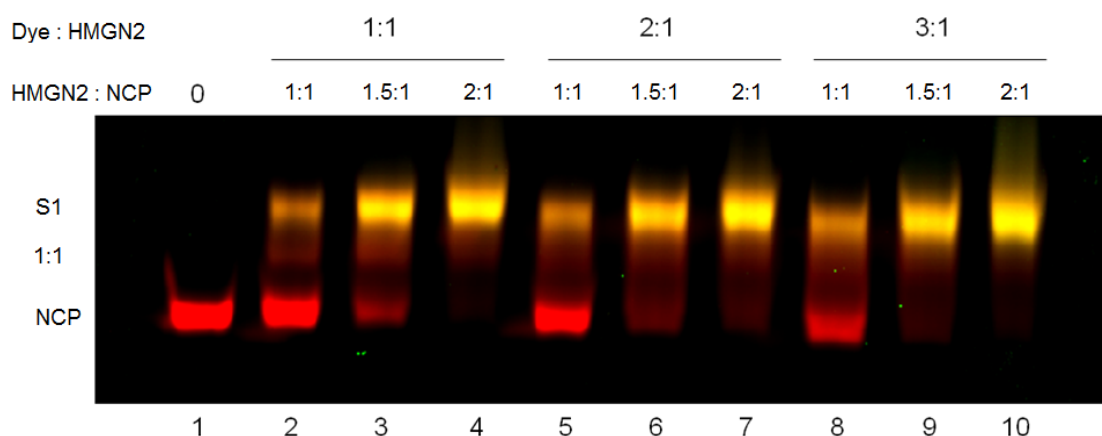


Figure 42. Optimization of fluorescence assay for HMGN2 – NCP studies. The free NCP band is red, 1:1 HMGN2:NCP band is red-orange, and S1 band is yellow. Samples were run under divalent ion-free condition.

In support of the color visualization, separate intensity measurements for both NCP and HMGN of each species (1:1 complex versus S1) were also done to confirm the integrity of our approach. We found that the HMGN-to-NCP intensity ratio for the S1 species is about double the HMGN-to-NCP intensity ratio for the 1:1 species (Table 7). This result is expected because the amount of HMGN involved in the S1 species is exactly twice the amount involved in the 1:1 species.

Intensity measurements of 1:1 complexes and S1 complexes

Lane	Band	Raw Intensity		Normalized Intensity		HMGN/NCP (Intensity ratio)	S1/1:1 complex (Intensity ratio)
		NCP	HMGN	NCP	HMGN		
2	1:1 complex	7485679	1548246	7036924	1228705.12	0.174608	1.840924
2	S1	14961466	4984513	14512711	4664972.55	0.32144	
3	1:1 complex	9518447	1972870	9069692	1653329.16	0.182292	1.941242
3	S1	31345933	11253196	30897178	10933655.6	0.353872	
6	1:1 complex	6593798	1514926	6145043	1195385.58	0.194528	1.890194
6	S1	33767720	12570806	33318965	12251265.5	0.367696	
4	S1	28198953	10230515	27750198	9910973.88	0.35715	
7	S1	33029490	12617912	32580735	12298371.3	0.377474	
9	S1	28070100	10939469	27621345	10619928.6	0.384483	
10	S1	28337299	11991188	27888544	11671647.3	0.41851	
Background Intensity		448755	319540.9				

Table 7. Intensity measurements of 1:1 complexes and S1 complexes. Raw intensity is normalized with respect to background intensity. HMGN/NCP column is the comparison of HMGN's and NCP's normalized intensities for each species (1:1 complex and S1). S1/1:1 complex column is the comparison of the HMGN/NCP values of S1 and 1:1 complexes present on the same lane.

Both colour visualization and intensity measurements confirm the capacity of our fluorescence assay to determine HMGN-to-NCP ratios on gels. Labeling NCP by attaching a fluorescence group to the DNA is also more effective than labeling histone proteins. It took more than three days to refold the HO and reconstitute the NCP, thus risking decay of the fluorescence dye. Moreover, HO refolding is more efficient when done in large scale. Fluorescence-labeled histone may become too little, and thus give weak fluorescence signal after the refolding and reconstitution process.

Our fluorescence analysis of the S2 species indicates the presence of extra HMGN on the S2 species. Fine screening of HMGN-to-NCP ratios was done in order to obtain the minimum stoichiometric ratio that still produces the S2 species. Our finding showed that the NCP is associated with more than two HMGN molecules even at the minimum HMGN-to-NCP ratio required for S2 generation (Figure 43, white arrows). Moreover, some of the NCP molecules are left without any HMGN bound to them (Figure 43, yellow arrows), while at the same time, some already transform to the S2 state (Figure 43, white arrows). We concluded that once two HMGNs bind to one NCP, that particular NCP becomes susceptible to transformation of S2, and this S2 species has elevated affinity for binding additional molecules of HMGN.

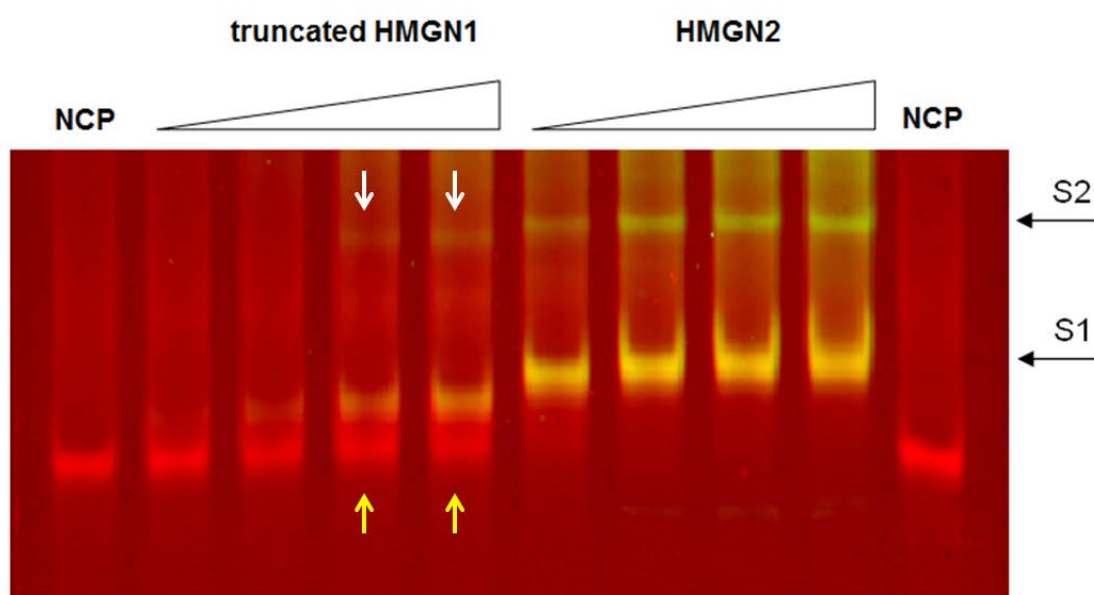


Figure 43. Fluorescence analysis of the S2 species.

Our fluorescence assay also ruled out the possibility of nucleosome polymerization in S2 formation. NCP bound to two HMGN molecules (S1) does not recruit free NCP for polymerization (S2 would be red in this scenario), nor recruit other S1 species (S2 would still be yellow). Formation of HMGN bridges consisting of three or more HMGN proteins that link multiple S1 species for S2 formation is also difficult to imagine as there is no evidence for HMGN-HMGN interaction in the absence of NCP. Taking both western blot and fluorescence assay results together, we hypothesize that HMGNs promote S2 formation by decompacting nucleosomes at the DNA termini. This nucleosome decompaction is propagated to the chromatin level, and the transition to the putative open conformation does not involve systematic histone eviction from the nucleosome.

6.1.3. Binding Assessment of HMGN NBD Truncation Series to Nucleosome Core Particle

In addition to full length HMGN proteins, we are also interested to study the interaction between the HMGN nucleosome binding domain (NBD) and NCP. We use full-length NBD peptides and three C-terminal truncated versions of the NBD (sequences shown in the Appendix) to dissect the role of NBD N-terminal and C-terminal elements separately. From EMSA results we found that all C-terminal truncated NBD peptides can still bind NCP, albeit increasing peptide-to-NCP ratios

are required as the peptides get shorter (Figure 44, left panel). More interestingly, in the presence of free DNA (Figure 44, right panel), excess amounts of C-terminal truncated NBD peptides (XN12, XN13, and XN14) do not cause non-specific DNA band smearing as extensively as the full length NBD, indicating their reduced affinity toward free DNA. From this result we hypothesized that the N-terminus is responsible for contacting and making specific interactions with the histone surface of the NCP, while the C-terminal part of the NBD may interact with nucleosomal DNA. The presence of highly conserved and critical amino acids for HMGN binding on the N-terminus of the NBD supports “N-terminus NBD” hypothesis (Ueda *et al.*, 2008).

EMSA result is also useful to determine appropriate peptide-to-NCP ratios for crystallization. We assumed the sharpest-looking shifted bands correspond to the native 2:1 peptide-to-NCP complex. Ratios that gave the best shifted bands were used for NBD-NCP co-crystallizations (Section 6.1.6).

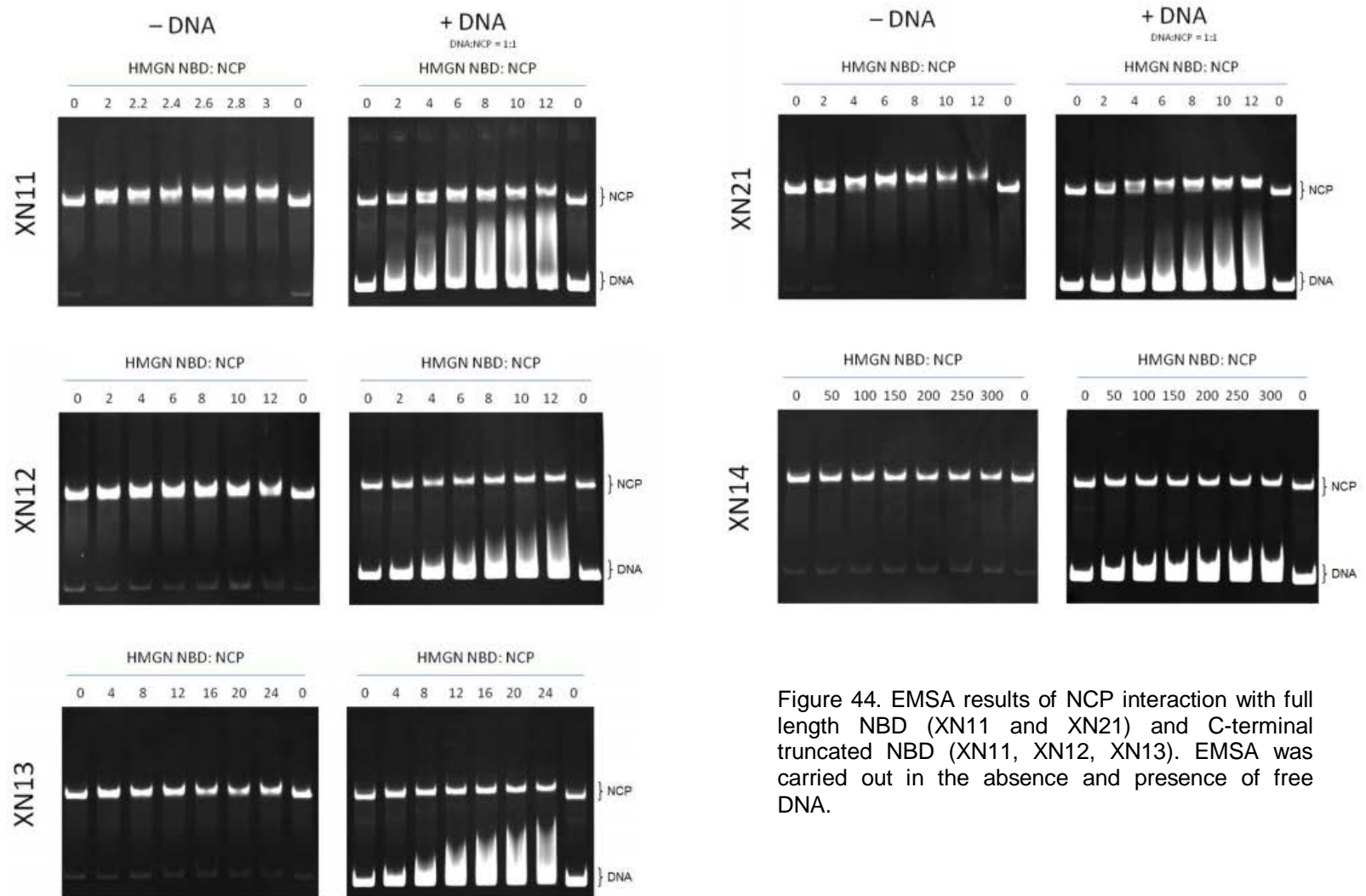


Figure 44. EMSA results of NCP interaction with full length NBD (XN11 and XN21) and C-terminal truncated NBD (XN11, XN12, XN13). EMSA was carried out in the absence and presence of free DNA.

6.1.4. Identification of the NBD's Histone Binding Partner

We aim to map to which histone proteins HMGN NBD binds. In order to capture these interactions, we utilized the commonly employed crosslinking agents, formaldehyde and glutaraldehyde. When we compared the two, we found that glutaraldehyde was more useful than formaldehyde. Being a one-carbon compound, formaldehyde can only covalently link very close protein contacts, while glutaraldehyde, which has a four-carbon linker, gives crosslinking flexibility to obtain coupling over longer distances.

We analyzed crosslinked NCP-NBD complexes on SDS PAGE gels to determine which histone proteins have retarded mobility due to HMGN NBD attachment. We observed two new bands above the uncrosslinked histone bands, but still in close proximity (Figure 45, red arrow). We reasoned that these two new bands correspond to covalent attachment of a single peptide molecule to two different individual histones. Bands with greater intensities above the crosslinked NBD-histone bands are thought to be multiple combinations of histone dimer, histone tetramer, HO, and unspecific higher molecular internucleosome crosslinking products. The resolution of the PAGE gel, however, does not allow us to identify any potential additional new bands which may overlap with uncrosslinked histone bands.

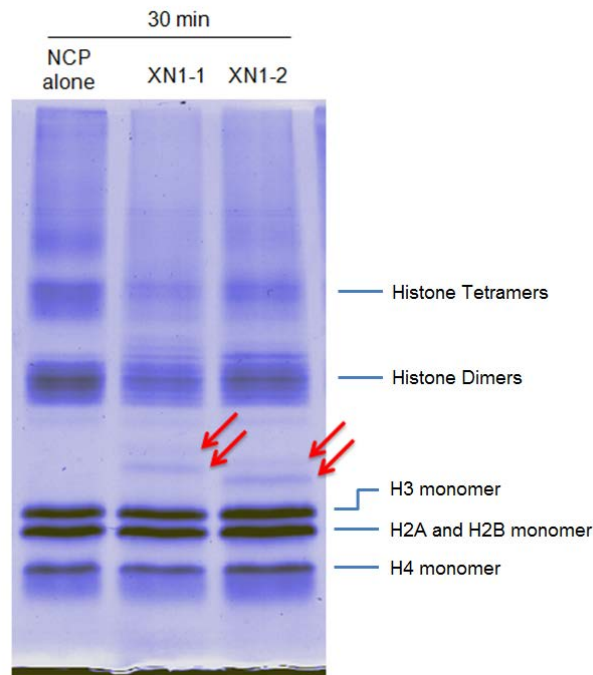
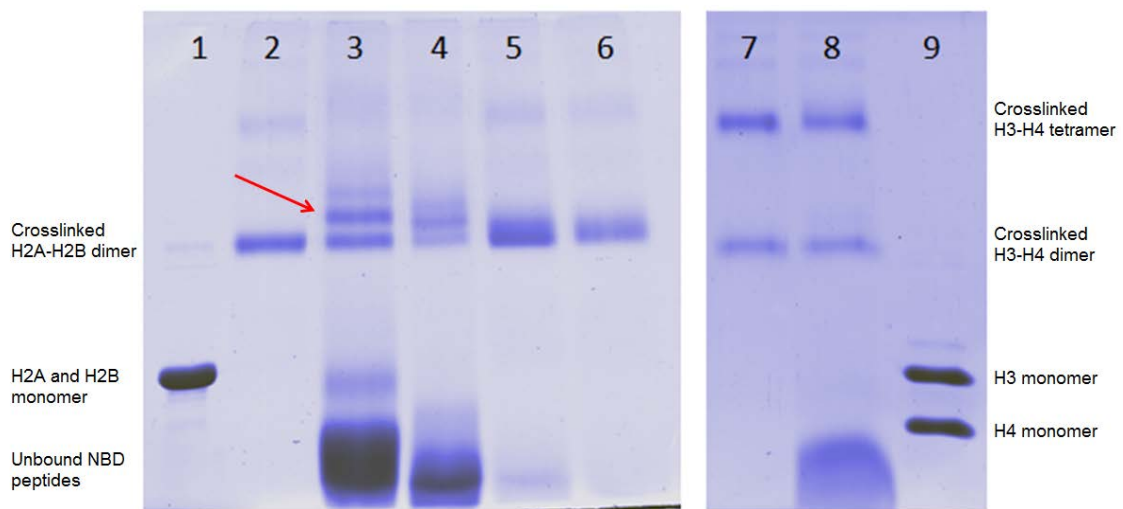


Figure 45. SDS PAGE gel of glutaraldehyde-crosslinked NCP – NBD samples

In addition to NCP crosslinking, we also tried to crosslink NBD peptides to both histone H2A-H2B dimer and H3-H4 tetramer separately. We found that crosslinking to H2A-H2B dimer generates higher molecular species that correspond to peptide-H2A-H2B species. On the other hand, NBD crosslinking to H3-H4 tetramer does not yield any peptide-H3-H4 species (Figure 46). This finding suggests the importance of the H2A-H2B dimer for HMGN binding to NCP. Nevertheless, we did not rule out the possibility of NBD or full HMGN making contact with H3-H4 tetramer because this crosslinking assay depends heavily on the presence of lysine side chains on both of the interaction sites. NBD may still interact with H3-H4 tetramer, but the interaction is probably less extensive than NBD contact with the H2A-H2B dimer.



- | | |
|---------------------------------|---------------------------------|
| 1. Uncrosslinked H2A-H2B dimer | 6. Crosslinked H2A-H2B and XN14 |
| 2. Crosslinked H2A-H2B | 7. Crosslinked H3-H4 tetramer |
| 3. Crosslinked H2A-H2B and XN11 | 8. Crosslinked H3-H4 and XN11 |
| 4. Crosslinked H2A-H2B and XN12 | 9. Uncrosslinked H3-H4 tetramer |
| 5. Crosslinked H2A-H2B and XN13 | |

Figure 46. SDS PAGE of glutaraldehyde crosslinking of H2A-H2B dimer (left panel) and H3-H4 tetramer (right panel) with NBD peptides. NBD peptides are shown to interact with H2A-H2B dimer (red arrow) but not with H3-H4 tetramer.

In order to identify the histone binding partner of NBD, we subjected crosslinked samples to three different approaches: mass spectrometry, individual histone labeling, and western blotting. The mass spectrometry approach was unsuccessful as we were

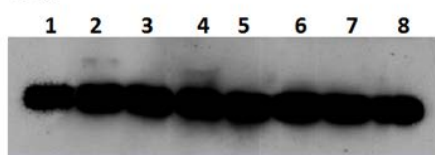
not able to observe any peaks having molecular weights consistent with histone-peptide species. Histone labeling was also unsuccessful because of the difficulties in producing NCP with labelled histone in sufficient quantity.

Using the western blot approach, we detected bands corresponding to crosslinking product between HMGN1 NBD truncation peptides with H2A, H2B, H3, but not with H4, while the HMGN2 NBD truncation peptides were found to bind predominantly to H2B, and to a much lesser extent to H3 (Figure 47, crosslinked NBD-histone bands appear above the uncrosslinked histone band, and have much weaker intensity than the uncrosslinked histone band). These findings support our previous conclusion from the histone dimer and histone tetramer crosslinking experiments. There may be some slight contact with H3, but HMGN NBDs predominantly interact with H2B.

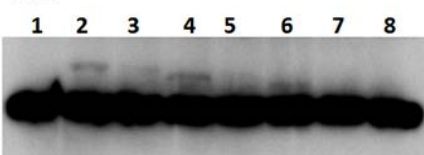
As a negative control, we also attempted to crosslink mutant NBD peptides (Δ) to the NCP. These mutants have one of the essential amino acids replaced, and have been shown to fail to bind specifically to NCP when analyzed by EMSA. Band intensities of the histone-peptide species were quantified and the relative intensity between mutant and native peptide was compared. Intensity measurement revealed that the native peptide (full-length NBD) bands were about 2-3 fold stronger than those of the mutant peptides (Table 8). This value is similar to binding affinity difference observed between wild type full length HMGN versus full length HMGN harbouring the same mutations we have for our mutant peptides (Ueda *et al.*, 2008). Unfortunately, HMGN2 NBD crosslinking could not be evaluated quantitatively due to insufficient data quality.

HMGN1 NBD Truncation Series

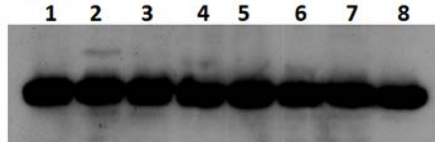
H2A



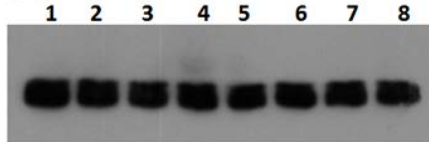
H2B



H3



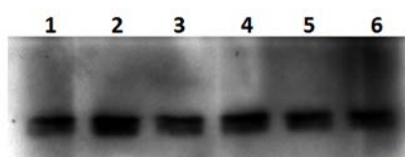
H4



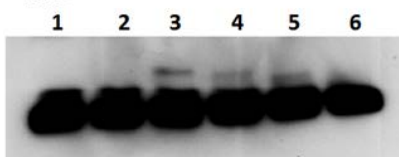
1. NCP alone
2. NCP-XN11
3. NCP-XN11Δ
4. NCP-XN12
5. NCP-XN12Δ
6. NCP-XN13
7. NCP-XN13Δ
8. NCP alone

HMGN2 NBD Truncation Series

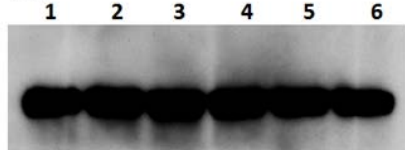
H2A



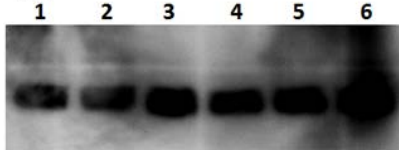
H2B



H3



H4



1. NCP alone
2. NCP-XN21Δ
3. NCP-XN21
4. NCP-XN22
5. NCP-XN23
6. NCP alone

Figure 47. Western Blot analysis of glutaraldehyde-treated NCP - NBD complexes. 'Δ' indicates a mutant version of the peptide, which does not bind specifically to the NCP. Very thick bands appearing on all western blots are bands corresponding to uncrosslinked histone. Crosslinked NBD-histone bands appear as a much weaker band above the thick uncrosslinked histone band. Note that prolonged crosslinking for H4 samples did not produce any additional band.

Western Blot Densitometry Measurements of Histone-HMGN NBD Crosslink

		Histone Intensity	Histone-Peptide Intensity	Blank	Normalized Histone Intensity	Normalized Hist-Pep Intensity	% Crosslinking	WT/ Δ
H2A	XN11	2.09	0.22	0.16	1.93	0.06	3.10880829	3.295336788
	XN11 Δ	2.3	0.2	0.18	2.12	0.02	0.943396226	
	XN12	2.24	0.36	0.22	2.02	0.14	6.930693069	2.108910891
	XN12 Δ	2.3	0.24	0.17	2.13	0.07	3.286384977	
	XN13	2.4	0.29	0.16	2.24	0.13	5.803571429	1.458147321
	XN13 Δ	2.18	0.25	0.17	2.01	0.08	3.980099502	
H2B	XN11	2.08	0.28	0.18	1.9	0.1	5.263157895	2.042105263
	XN11 Δ	2.12	0.23	0.18	1.94	0.05	2.577319588	
	XN12	2.22	0.35	0.19	2.03	0.16	7.881773399	1.950738916
	XN12 Δ	2.17	0.27	0.19	1.98	0.08	4.04040404	
	XN13	2.12	0.41	0.19	1.93	0.22	11.39896373	1.937823834
	XN13 Δ	2.04	0.28	0.17	1.87	0.11	5.882352941	
H3	XN11	1.77	0.24	0.17	1.6	0.07	4.375	3.45625
	XN11 Δ	1.76	0.2	0.18	1.58	0.02	1.265822785	
	XN12	1.69	0.27	0.19	1.5	0.08	5.333333333	1.36
	XN12 Δ	1.72	0.25	0.19	1.53	0.06	3.921568627	
	XN13	1.69	0.27	0.17	1.52	0.1	6.578947368	1.390977444
	XN13 Δ	1.66	0.25	0.18	1.48	0.07	4.72972973	
H4	XN11	1.37	0.13	0.13	1.24	0	0	
	XN11 Δ	1.26	0.13	0.13	1.13	0	0	
	XN12	1.68	0.22	0.14	1.54	0.08	5.194805195	
	XN12 Δ	1.52	0.13	0.13	1.39	0	0	
	XN13	1.62	0.13	0.13	1.49	0	0	
	XN13 Δ	1.34	0.13	0.13	1.21	0	0	

Table 8. Western blot densitometry measurements for quantification of degree of histone-NBD peptide crosslinking relative to native histone quantity. Normalized intensity equals the raw intensity with the background intensity (blank) subtracted (left). The % Crosslinking equals the fraction of normalized histone-peptide intensity relative to the normalized histone intensity. WT/ Δ corresponds to the % crosslinking of native peptide relative to the % crosslinking of its mutant counterpart.

6.1.5. Crystallization of HMGN-NCP Complex

Structures of the NCP alone have been solved previously up to 1.9 Å resolution. Our first attempt for crystallization, therefore, was to use the same standard crystallization buffer (Section 3.4) to co-crystallize HMGN with the NCP. Although crystals seemed to diffract to even higher resolution when grown in the presence of the HMGN, subsequent data analysis, however, indicated that there was no HMGN inside the crystal. Co-crystallizations with various crystallization buffers were also tried. Hexagonal crystals of HMGN1-NCP were produced in Hampton Research PEG Rx 1 #29 buffer (Figure 48) but the crystals were too fragile to be subjected to subsequent X-ray data collection.



Figure 48. HMGN1- NCP Crystals produced in 0.1 M Sodium citrate tribasic dihydrate (pH 5.5), 18% w/v Polyethylene glycol 3,350 (PEGRx[™] 1 No.29, Hampton Research)

The absence of HMGN in crystals from co-crystallization attempts prompted us to produce NCP with HMGN protein fused to one of the histones. Two-to-one HMGN to NCP ratio would be strictly maintained because there are two copies of each histone present in the NCP. We overexpressed two types of HMGN-fused histones; N1-H2A and N1-H3, both have HMGN1 protein fused to the N-terminus of H2A and H3, respectively. Histone production and NCP reconstitution were carried out in the same way as native histone and NCP production. EMSA analysis shows distinct bands corresponding to fusion HMGN NCP (Figure 49). Crystals were obtained after 4-5 hours under 65 mM MnCl₂, 40 mM KCl, 20 mM KCaco (pH 6.0) and 12.5% PEG 6000 condition. Unfortunately, crystals dissolved after one day (Figure 50). In order to slow down the equilibration, hence stabilizing the crystals, we tried to vary buffer concentration in the reservoir. Even when the ionic strength of the reservoir was almost the same as the ionic strength in the crystallization buffer, the crystals still disappeared after overnight incubation.

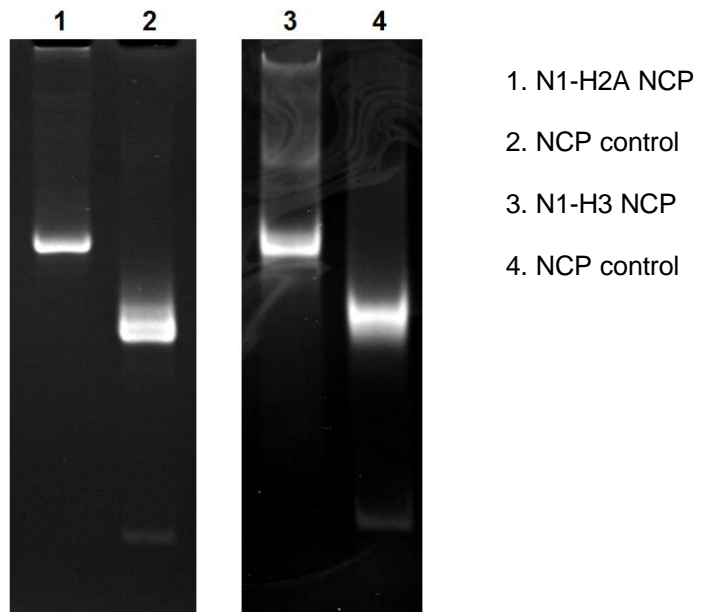


Figure 49. Fusion NCP after reconstitution.

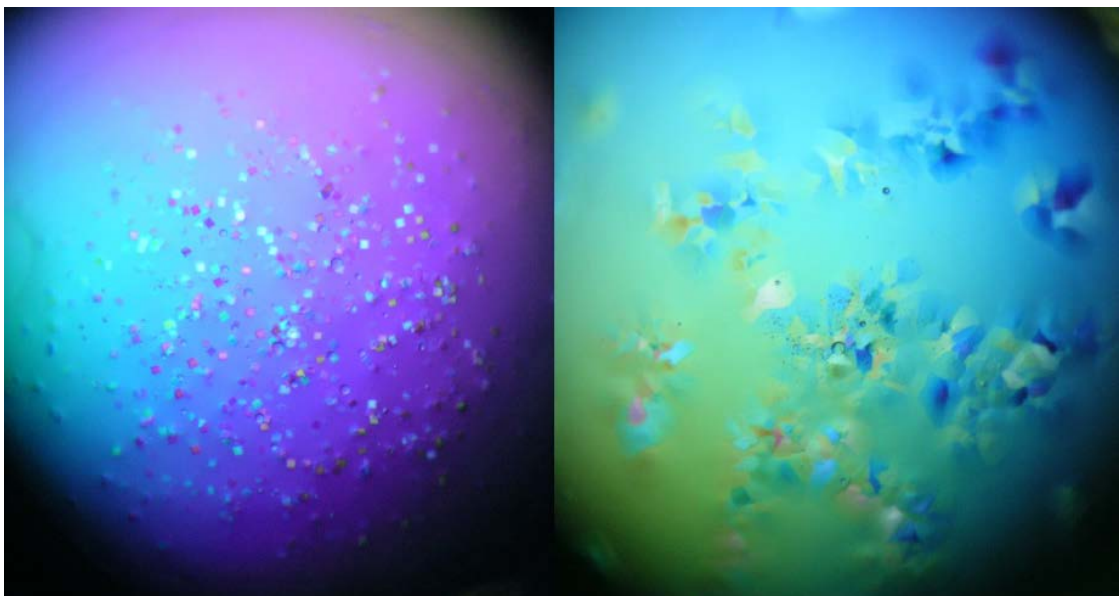


Figure 50. N1-H2A and N1-H3 NCP crystals grown under 65 mM MnCl₂, 40 mM KCl, 20 mM KCaco (pH 6.0) and 12.5% PEG 6000 condition (1X), equilibrated against reservoir filled with 0.5X buffer. Crystals formed after 4-5 hours of incubation at 18°C (left), but dissolved after one night (right).

6.1.6. Crystallization of an HMGN NBD-NCP Complex

We tried to co-crystallize an NBD-NCP complex in standard NCP crystallization buffer at optimal NBD-to-NCP ratios deduced from the EMSA experiments. Crystal

assessment from X-ray analysis, however, did not seem to indicate NBD peptide association to the NCP. The absence of peptide in the crystal directed us to try co-crystallizations at higher peptide to NCP ratios. Different stoichiometries of up to 50:1 peptide-to-NCP were attempted and the crystal morphologies were assessed under the microscope. At very high peptide ratios, the crystal form started to change, which we thought may be related to an alteration in crystal content. Nevertheless, diffraction data analysis of these crystals still did not indicate the presence of peptide in the crystals.

In addition to co-crystallization, we also tried soaking strategies to obtain a structure of the HMGN NBD-NCP complex. Native NCP crystals were transferred into and stored in buffers containing different concentrations of NBD peptide. Crystal morphology was monitored, and we found that at very high NBD peptide concentrations crystals started to crack and lose their diffraction quality. In spite of these morphological changes implying peptide binding within the crystals, diffraction analysis did not show any presence of peptide inside the NCP crystals. We hypothesized the peptides are not able to bind specifically to NCP inside the crystals under this high salt condition, thus no peptide electron density was observed after X-ray analysis.

HMGN is known to bind to chromatin with much higher affinity at low ionic strength conditions. We therefore reasoned that reducing the salt content of soaking buffers or searching for lower ionic strength conditions for co-crystallization trials may help to stabilize peptide binding to the NCP. We obtained rod-shaped crystals in buffer containing 1 mM MnCl_2 , 0.9 mM KCl, 20 mM KClO₄ (pH 6.0), 2.5% PEG MME 2000. We dissolved the crystals in EDTA buffer and ran the solution in a polyacrylamide gel. EMSA analysis of the crystals confirmed the presence of peptides inside the crystals. Crystallization in low ionic strength conditions, however, yielded poorly diffracting crystals (Figure 51). We did not get any meaningful electron density with regards to NBD peptides.

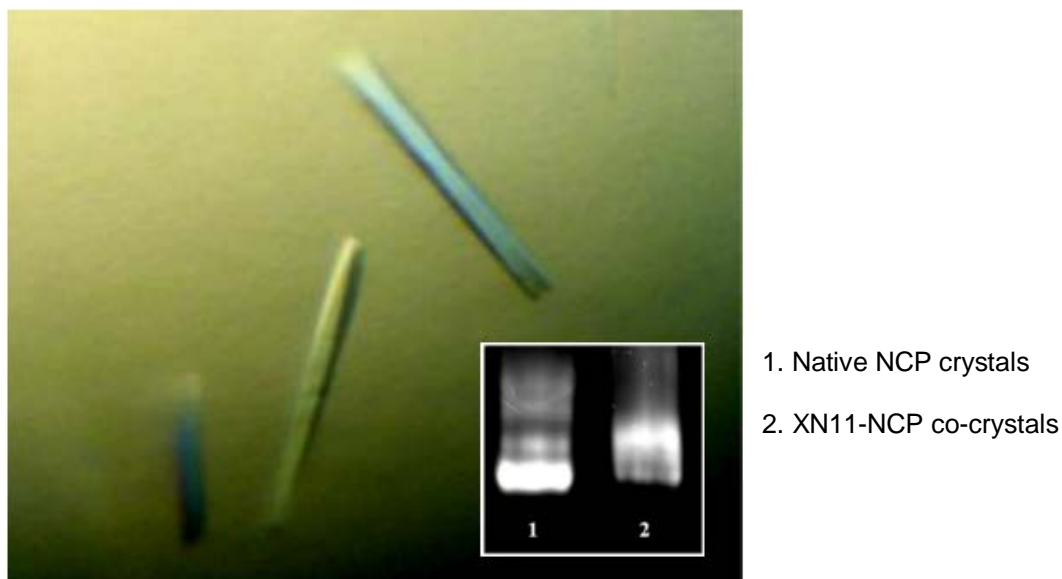


Figure 51. XN11-NCP co-crystals produced in low-ionic strength crystallization buffer. EMSA result confirmed the presence of NBD peptide inside the crystals (inset).

Nevertheless, our low-salt buffer yielded crystals with distinct unit cell parameters and crystal packing (Table 9). Low salt NCP crystals adopt the same $P2_12_12_1$ space group, but with one of the cell dimension tripled (unit length of c axis is tripled from 109.5 Å in native NCP crystals to 317.3 Å in low salt NCP crystals). Each asymmetric unit contains three NCPs, thus summing up to 12 NCPs in each unit cell. Due to the low resolution of the final data (~ 9 Å), we could not tell for sure the orientations of the NCPs within the asymmetric unit (Figure 52).

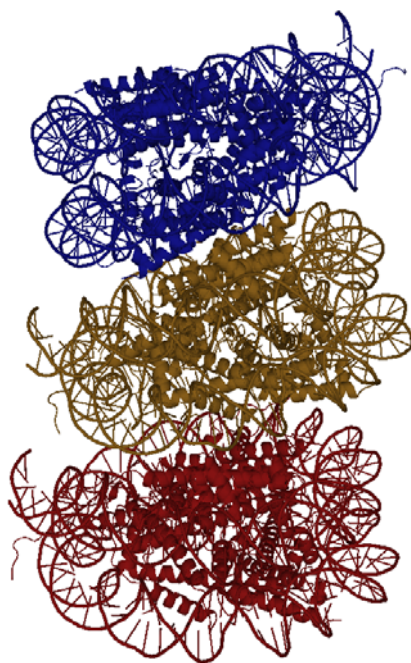


Figure 52. NCP packing in low-salt crystals. Three NCPs are present in each asymmetric unit.

Data Collection and Refinement Statistics for Low-salt NCP-XN11 Crystals

Data collection	
Resolution (Å)	60.00-8.50
Redundancy (last shell)	6.1 (6.3)
No. unique <i>hkl</i>	5,354
Completion % (last shell)	99.3 (100)
<i>R</i> _{meas} % (last shell)	19 (59)
Unit cell <i>P</i> 2 ₁ 2 ₁ 2 ₁	
<i>a</i> (Å)	100.8
<i>b</i> (Å)	177.0
<i>c</i> (Å)	317.3
Refinement	
Resolution (Å)	8.5
<i>R</i> -factor %	36
Free <i>R</i> -factor %	42
<i>r.m.s.d.</i> from ideality	
Bond lengths (Å)	0.007
Bond angles (deg.)	1.037
Chiral	0.053

Table 9. Data collection and refinement statistics for low-salt NCP-XN11 crystals grown in 1 mM MnCl₂, 0.9 mM KCl, 20 mM Kcac (pH 6.0), and 2.5% PEG MME 2000

6.1.7. Searching for NCP's Divalent ion-free Crystallization Buffers

Our standard NCP crystallization relies on the presence of manganese ion inside the crystallization buffer. From our previous EMSA results, we realized the presence of divalent ions promotes formation of non-homogenous decompacted NCP structures, thus making the NCP not suitable for crystallization. We therefore switched our strategy to search for novel NCP crystallization buffers that lack divalent ions. For initial screening, we tried to crystallize NCP only. Crystallization buffers that yield NCP crystals would be used for subsequent HMGN or NBD co-crystallization and crystal soaking experiments. Three different divalent ion-free crystallization buffers yielded crystals of the NCP (Figure 53). These three different crystals will be subjected to diffraction analysis in the future.

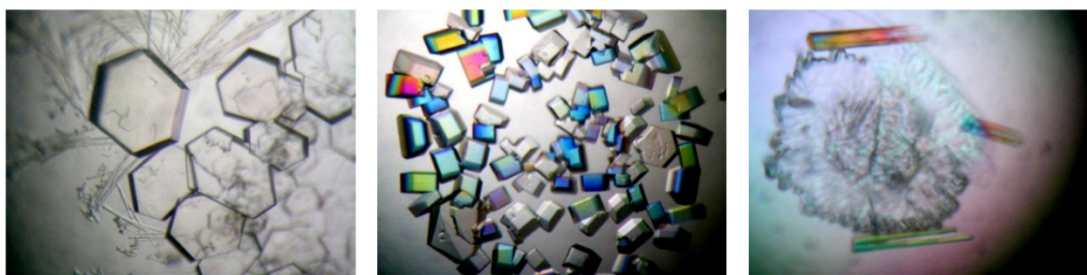


Figure 53. NCP crystals produced in divalent ion-free crystallization buffers. Crystallization buffer (2X) for the crystals are 0.1 M HEPES (pH 7.0), 12% w/v Polyethylene glycol 3,350 (PEGRx[™] 1 No.30) + 0.6X TBE (left), 0.1 M Sodium citrate tribasic dihydrate (pH 5.5), 38% v/v Polyethylene glycol 200 (PEGRx[™] 1 No.2) + 0.6X TBE (right), and 80 mM Bis Tris Propane (pH 9.0) + 14% Jeffamine ED2001 (pH 7.0) + 0.4X TBE (center).

Chapter VII

Discussion

7.1. Human and *Xenopus laevis* Nucleosomes Adopt Nearly Identical Structures

Crystallographic analysis of human NCP and *Xenopus laevis* NCP shows a high degree of similarity between them, indicating that, in addition to histone primary sequences, nucleosome structure is also highly conserved throughout evolution. We did observe differences in stretching location between human and *Xenopus laevis* NCP145. Yet, knowing the dynamics of nucleosome structure, we can argue that the energy difference between these two stretching configurations is likely very small (Wu *et al.*, 2011; Davey *et al.*, 2010). Crystallographic analysis, despite its robust power, could favor one particular structural conformation due to that particular structure's inherent crystallizability. Both stretching configurations are most likely present in equilibrium in both human and *Xenopus laevis* NCP145 solution. Once they form crystals, however, they are forced to adopt one particular stretching configuration in order to accommodate the crystal lattice.

Knowing that both NCP structures are quite similar, we hypothesize that *Xenopus laevis* and human NCP probably make use of their amino acid differences (most likely based on their post-translational modifications) in order to establish different details of genomic regulation (human and *Xenopus laevis* histone alignments can be found on the appendix of this thesis). Indeed, some of the substitutions are quite different, such as serine to glycine / alanine / histidine / cysteine. This amino acid substitution, in the case of serine, may lead to different serine phosphorylation pattern on human and *Xenopus laevis* NCP. Moreover, there is also a run of three amino acid substitutions on the H2B tail which is in close proximity to the DNA. Arginine to lysine, and threonine to lysine substitutions on this tail may take on distinct post-translational modifications, thus affecting downstream genomic activities. Lastly, amino acid differences between both species may affect nucleosome–nuclear protein interactions or chromatin condensation.

7.2. Steric Hindrance Governs RAPTA Compound - NCP Interactions

Our observations demonstrate RAPTA binding occurs mostly on one side of the symmetrical NCP. The reason for this most likely is due to accessibility of both Site A and Site B on one face of NCP, but not on the other symmetry-related half. Examination of NCP crystal packing shows the presence of a wide open space available only on one side of the NCP (Figure 54). The obstruction of equivalent spatial access on the other half of NCP by neighbouring NCP particles inhibits drug association at this face. Nevertheless, we still observed measurable anomalous signals on this obstructed side when we incubated NCP crystals with the smallest RAPTA compound, RAPTA-B. The absence of any modification on the arene ring perhaps allows RAPTA-B to diffuse into this area and bind specifically at both Site A' and Site B'.

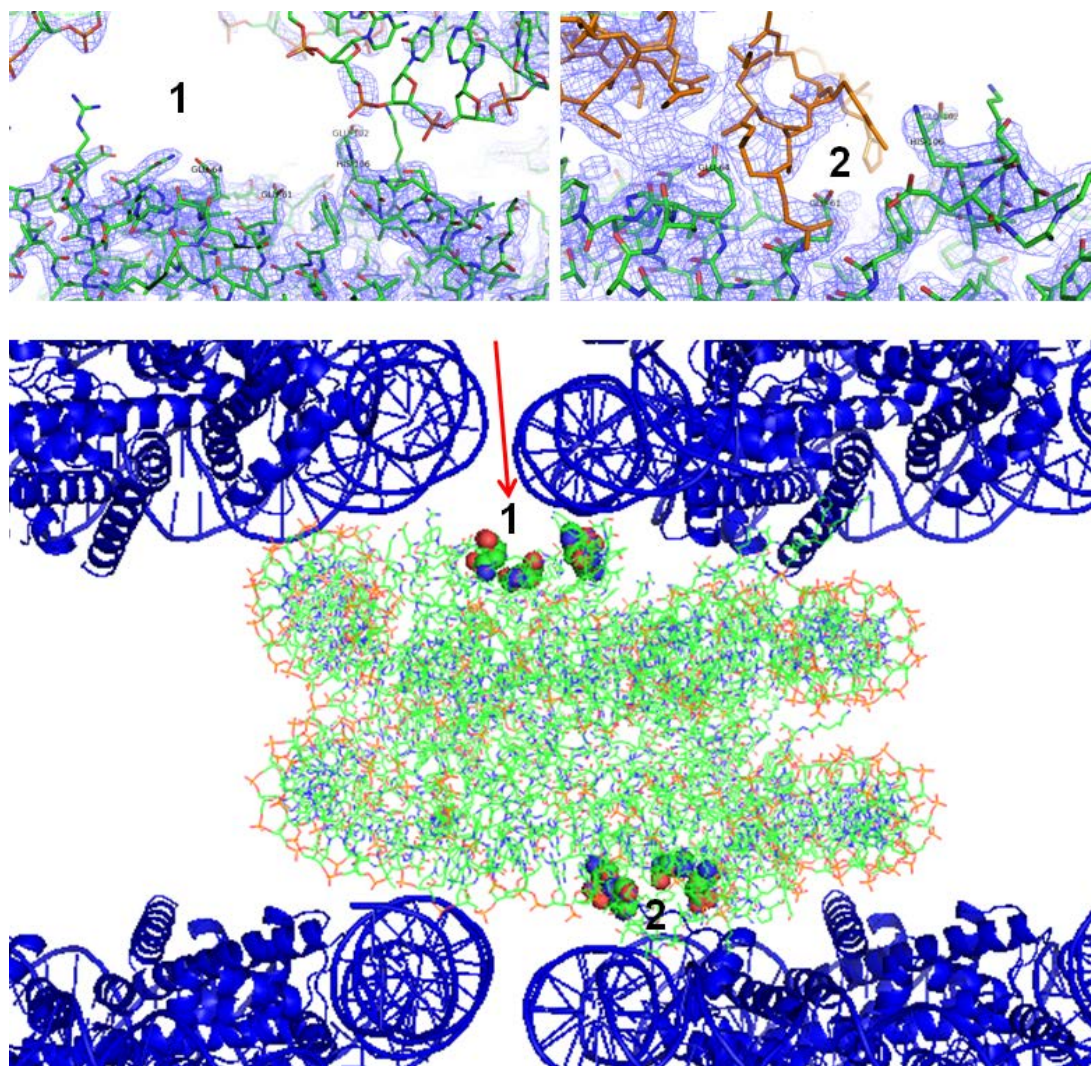


Figure 54. Different accessibility of RAPTA binding sites between the two NCP faces. H2A Glu 61, H2A Glu 64, H2B His 106, and H2B Glu 102 are depicted in space filling ball model (bottom panel). Position 1 and Position 2 are depicted in greater detail at top-left and top-right panels, respectively.

7.3. Arene Ring and PTA Ligand Confer Site Selectivity for RAPTA Binding

The presence of both an arene ring and a PTA group provides RAPTA compounds with greater preference towards histone proteins than to DNA. The arene ring is hydrophobic, while the PTA group is bulky and amphiphilic. Ruthenium ions alone should interact with DNA in a way similar to other heavy atoms (Mohideen *et al.*, 2010), however the presence of these two coordinated ligands impose steric restrictions to RAPTA binding selectivity. In support of that, other studies have suggested that the RAPTA mechanism of action is DNA-independent (Casini *et al.*,

2008) and they also have high reactivity and selectivity for different proteins (Casini *et al.*, 2009).

Crystal structures demonstrate that formation of RAPTA compound adducts on the surface of the histone core is mediated by glutamic acid and histidine side chains. This is consistent with previous results that RAPTA-C predominantly interacts with a histidine residue of cytochrome C (Casini *et al.*, 2007). In addition to those two coordinating amino acids that help stabilize each RAPTA compound at Site A and Site B, additional amino acids on the surrounding environment may also form interactions with RAPTA compounds (Figure 55, top panel). Side chains of H2A Leu 65 and H2B Leu 103 participate in Van der Waals hydrophobic interactions with the arene ring of RAPTA compounds at Site A. The attachment of hydrophobic groups on the arene ring could extend and facilitate further arene group interactions with these hydrophobic side chains. Amphiphilic PTA groups, on the other hand, form polar interactions with H2B Glu 110, and hydrophobic interactions with H2A Tyr 57, H2A Ala 60, and H2B Val 45. RAPTA adducts at Site B are involved in relatively fewer interactions with the surroundings. In addition to H2B His 106 and H2B 102 Glu that interact with the ruthenium (II) atom, the hydrocarbon portion of the H2B Lys 105 side chain may form hydrophobic interactions with the RAPTA arene ring (Figure 55, bottom panel). This unique environment that suits the RAPTA compound structure renders site selection for RAPTA binding. Development of subsequent RAPTA derivatives should be based on the optimization of these polar and hydrophobic interactions.

Interaction between the PTA ligand of RAPTA adducts at Site B and the arene ring of RAPTA adducts at Site A is thought to be one of the factors that promote the co-existence of neighbouring RAPTA molecules on a single NCP molecule. Both Van der Waals and hydrophobic interactions between carbon atoms on the arene group and PTA ligand help to stabilize both RAPTA compounds. Attachment of additional hydrophobic groups on the arene ring may extend the range of hydrophobic interactions, thus facilitating further interaction between these two compounds. Crystal inspection of RAPTA-I-treated NCP shows that one of the isopropyl groups attached to the RAPTA-I arene ring potentially forms additional interactions with the PTA ligand of the neighbouring RAPTA-I (Figure 55, bottom panel).

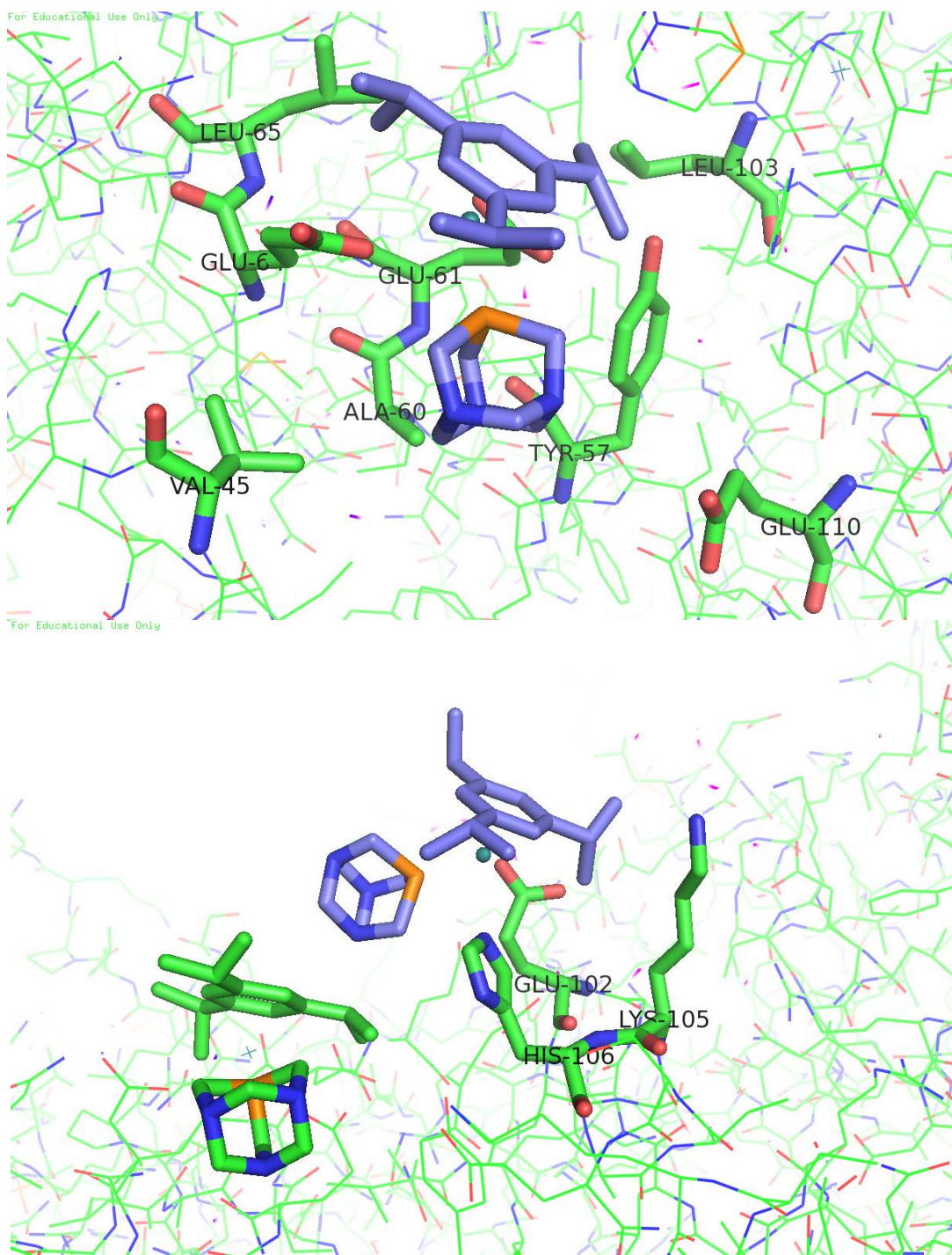


Figure 55. Important amino acids that shape the unique binding site for RAPTA compounds at Site A (top panel) and Site B (bottom panel). For clarity, carbon atoms of the RAPTA compound under inspection are coloured blue. Carbon atoms are coloured green on histone proteins and a neighbouring RAPTA adduct.

In contrast, six methyl groups attached to the RAPTA-H arene group seem to disturb interactions between neighbouring RAPTA-H adducts. Some of the methyl groups on the RAPTA-H arene ring are likely to cause steric clashes with PTA groups from the

other RAPTA-H, hence destabilizing the cooperativity between these two RAPTA-H sites. In addition to that, close proximity between hydrophobic methyl groups and the polar nitrogen atom from the PTA ligand may cause additional unfavourable interaction between these two compounds. Altogether these repulsive forces could push neighbouring RAPTA-H adducts away from each other. We tried to superimpose all RAPTA-treated NCP structures, and we indeed observed slight displacement of the arene ring of RAPTA-H from the consensus position (Figure 56).

The identity of the arene modification has been observed to play multiple roles in RAPTA compound efficacy. Cathepsin B, when treated with several RAPTA compounds *in vitro*, shows a wide range distribution of IC₅₀ values, indicating differential direct modulation of cathepsin B inhibition depending on the nature of arene ligand (Casini *et al.*, 2008). Moreover, aquation and anation reaction of RAPTA compounds, which are important for RAPTA compound activation inside the cells, are also found to be strongly affected by the identity of the arene ligand, suggesting that differences in the steric and electronic effects of the arene ligands also modulate ligand exchange reactions (Scolaro *et al.*, 2008; Wang *et al.*, 2003). Interestingly, another earlier study has shown that attachment of hydrophilic hydrogen-bonding groups to arene ligands of RAPTA compounds reduces their selectivity towards cancer cell lines as compared to hydrophobic benzene-containing RAPTA-B and toluene-containing RAPTA-T. RAPTA compounds bearing hydrophilic arene groups also show increased cytotoxicity towards normal cells, thus increasing potential side effects of the compounds (Scolaro *et al.*, 2005). Our findings show that modifications on the arene ring can help stabilization and cooperativity of RAPTA adduct formation on the surface of NCP, while the absence of any modification on the arene ring, as shown with RAPTA-B, enhances compound penetration into sterically occluded binding sites. Altogether, future RAPTA drug design should consider the necessity of adding new groups in order to strike the best balance between enhancing binding affinity and avoiding steric clashes.

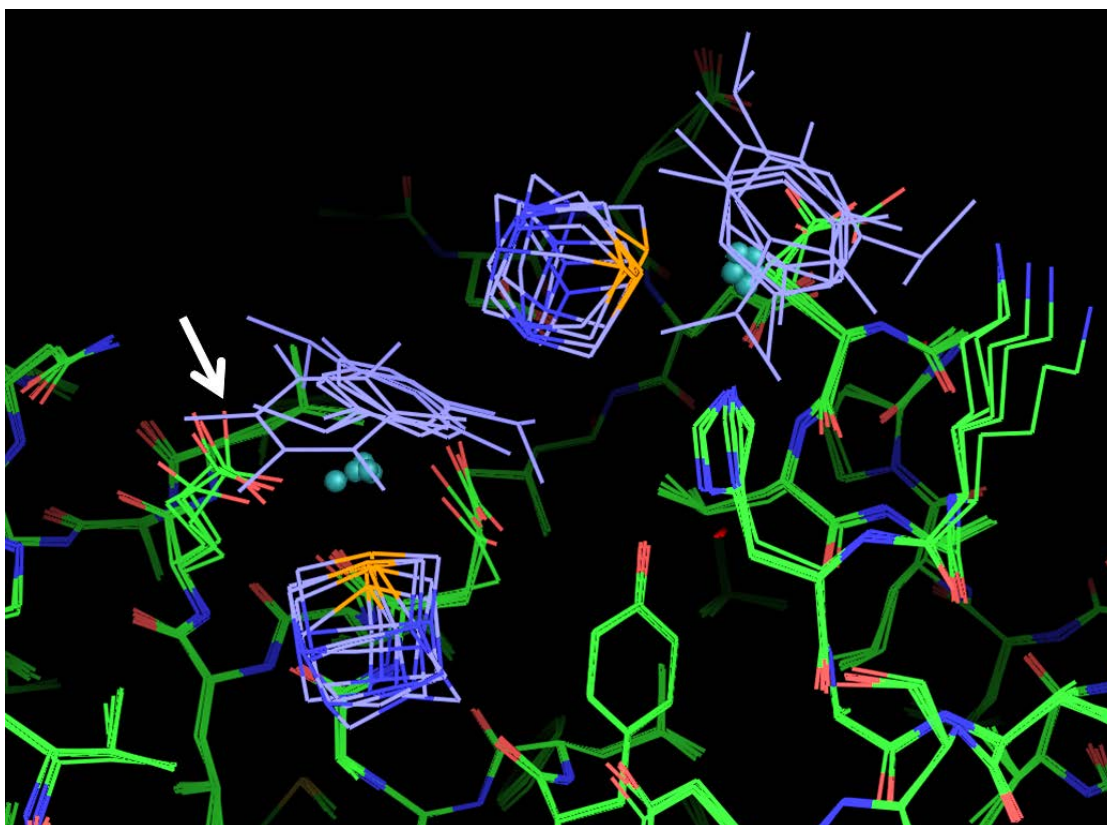


Figure 56. Displacement of the arene ring of RAPTA-H (white arrow) at Site A from the consensus position. All RAPTA compounds are superimposed together (blue structures) using NCP structure (green) as the reference.

7.4. RAPTA Compound Association to *Xenopus laevis* NCP and Human NCP

Our earlier RAPTA-C study has shown that RAPTA-C binds to three sites on the nucleosome surface (Wu *et al.*, 2011). Two of them are similar binding sites to what other RAPTA compounds bind to. On the other hand, the H4 Glu 63 - H4 Lys 59 site, which is occupied by RAPTA-C, was found unoccupied on other RAPTA-NCP crystals. No incidence of binding to this site was found with other Ruthenium and Osmium compounds investigated (unpublished data). This leads us to think perhaps RAPTA-C has the right kind of arene group to be recruited into this additional and highly specific binding site on the surface of the NCP. Another possibility, treatment of NCP crystals with drugs may hamper site identification as the drug slowly deteriorates the crystals, hence the quality of data. Therefore, we should always consider the relationship of data quality with crystal treatment.

It is interesting to see the differences found with regards to RAPTA-I presence in *Xenopus laevis* and human NCP. After 0.75 mM RAPTA-I two-day treatment, on *Xenopus* NCP, we found that both site A and B are occupied with moderately strong anomalous signals of 6.4σ and 4.7σ , respectively. On the other hand, only site A is occupied on human NCP, but with relatively higher anomalous signal, 9.6σ . One day RAPTA-I treatment for human NCP gave an even stronger anomalous signal, 11.1σ at site A, but none observed at site B. Moreover, RAPTA-I-treated human NCP crystals have better resolution (2.8 \AA) than RAPTA-I-treated *Xenopus laevis* NCP crystals (3.1 \AA), which suggest RAPTA-I binding difference between them could be real.

The observed RAPTA-I binding difference prompts us to further investigate distinctions between the *Xenopus laevis* and human NCP structures. Both amino acid conformation around Site B and crystal packing do not seem to be responsible for RAPTA-I exclusion from Site B on human NCP. The only difference we observed is in the DNA path from neighbouring NCP molecules. Human NCP has a slightly narrower major groove size (SHL -5.5) than that of *Xenopus* NCP near the site B position (Figure 57). *Xenopus* DNA has a negative roll value between cytosine -61 and adenine -62, while human DNA has a positive roll at this location (Figure 28 and Figure 57). These differences cause displacement of phosphodiester bond (Figure 57, red circle), leading to narrower major groove in human DNA at this location.

This small difference ($\sim 1 \text{ \AA}$) may obstruct the bulky RAPTA-I compound from accessing Site B on human NCP. However, it is also worth noting that DNA structure is quite flexible in this region. In some cases, drug binding has been shown to alter the surrounding DNA conformation (Davey *et al.*, 2010), thus favouring DNA rearrangements that promote drug binding. Therefore even though there is constriction imposed by the neighbouring nucleosomal DNA, the DNA could theoretically rearrange itself to allow bulky RAPTA-I to access Site B. Other than this theory, there is also a chance that the difference could be due to variations in the histone tail sequences or could be merely due to the sensitivity of data quality to drug treatment. More experiments are required to draw a decisive conclusion.

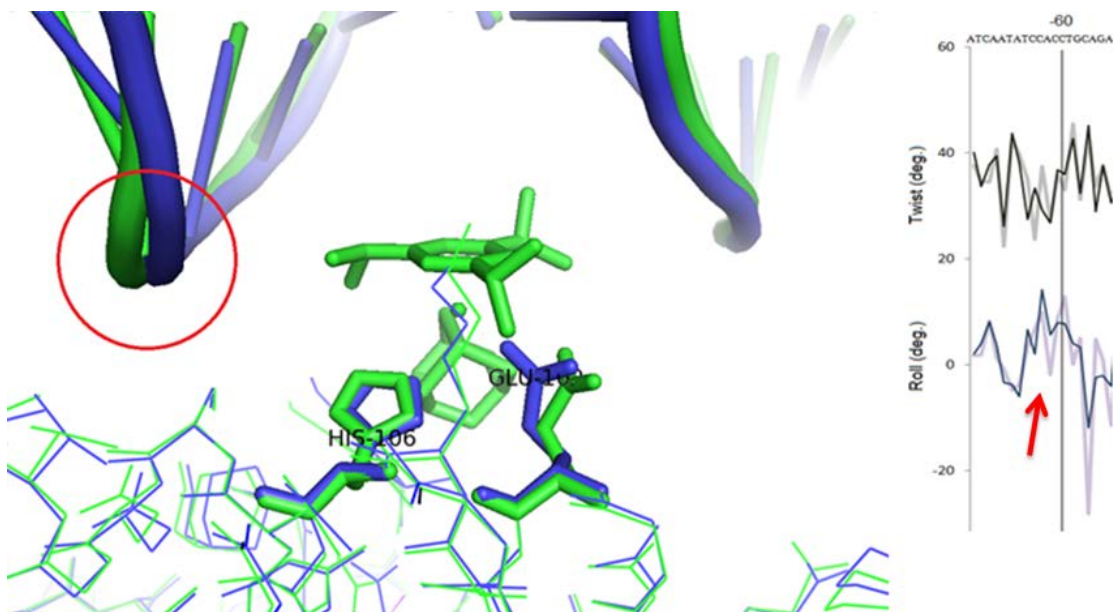


Figure 57. Superimposed structures of *Xenopus laevis* NCP (green) and human NCP (blue) at Site B. Both structures are from NCP crystals treated with 0.75 mM RAPTA-I for two days. RAPTA-I on site B is shown in the centre (green). The DNA major groove (SHL -5.5, depicted in tube form) from a neighboring NCP is shown. Human NCP has an ~1 Å narrower major groove than that of *Xenopus laevis* NCP. Displacement of phosphodiester backbone is highlighted (red circle) and the nucleosomal DNA bp parameters difference between human NCP145 (thin line) and *Xenopus laevis* NCP145 (thick line) is shown (red arrow).

7.5. RAPTA Binding Sites Overlap with Nuclear Protein Binding Sites on the NCP Surface

The RAPTA adducts reside in an area commonly referred as the acidic patch on the surface of NCP. This acidic patch serves as a binding platform for several nuclear and viral proteins (Figure 58), and also plays important roles in chromatin compaction (Armache *et al.*, 2011; Barbera *et al.*, 2006; Makde *et al.*, 2010; Zhou *et al.*, 2007). Among the proteins are the Sir3 complex which is important for gene silencing, RCC1 which is required during mitosis, and Kaposi Sarcoma antigen (LANA) which mediates viral genome attachment to the host chromosome. Disruption of RCC1 and Sir3 association to nucleosomes, and inhibition of chromatin condensation due to RAPTA binding could lead to disastrous cellular events such as apoptosis or senescence. In addition, association of RAPTA to the nucleosome core could inhibit viral genome attachment to the host genome. The specificity of RAPTA compounds towards cancer cells, perhaps due to selective ruthenium complexes uptake and

activation by cancer cells (Clarke *et al.*, 1988; Zorzet *et al.*, 2000), may help selectively target cancer cells for such inhibitory functions.

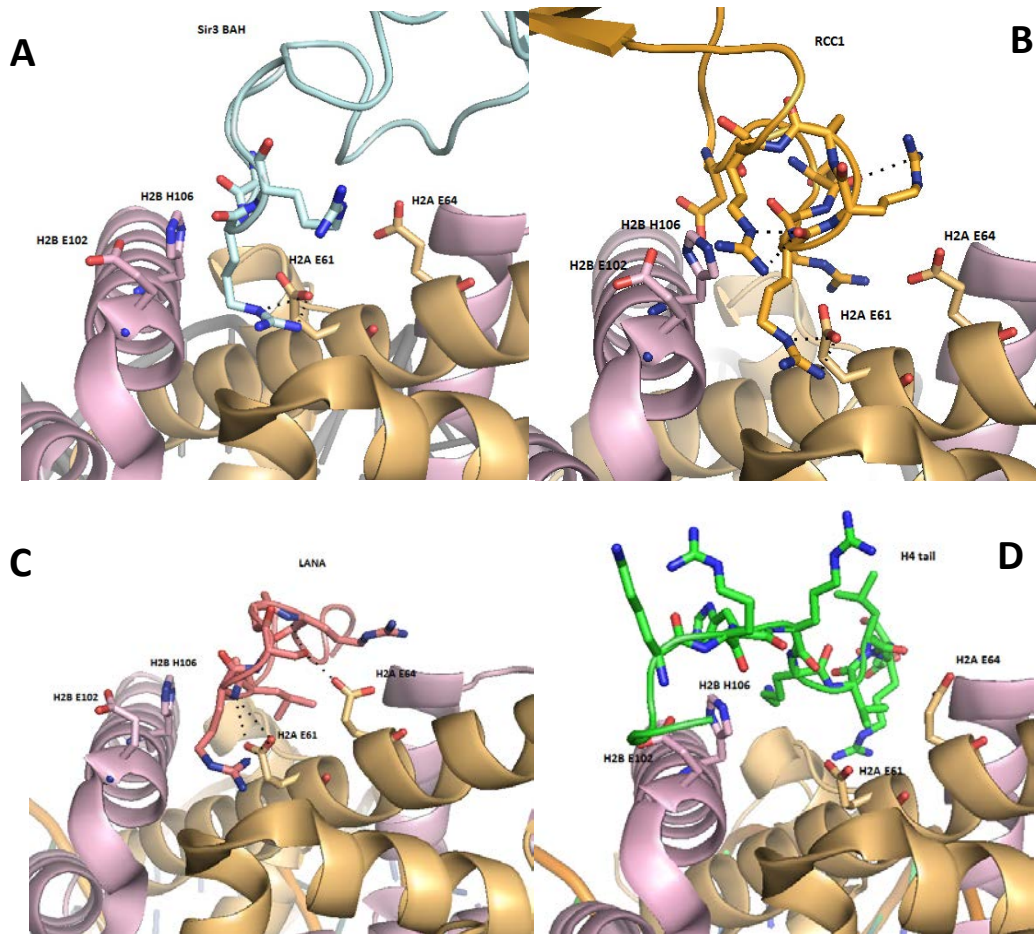


Figure 58. H2A glutamic acid 61 and 64, and H2B glutamic acid 102 involved in RAPTA binding also interact with other nuclear regulatory proteins such as Sir3 (A) and RCC1 (B). These amino acids are also recognized by Kaposi Sarcoma LANA antigen (C), and mediate chromatin condensation through interaction with the H4 tail (D). Pictures were adapted from Makde *et al.*, 2010. Structures are from the Protein Data Bank with accession codes 3TU4 (Sir3 BAH - NCP), 3MVD (RCC1 - NCP), 1ZLA (LANA - NCP), 1KX5(NCP-NCP).

7.6. HMGN Binding to Nucleosome

In vivo observation of HMGN interaction with chromatin has revealed the real time mobility of HMGNs. In contrast to EMSA results, wherein HMGN seems to associate stably with nucleosomes, *in vivo* studies showed that the interaction is just transient, only in the range of microseconds (Catez *et al.*, 2002). The dissociation constant for the binding of the first HMGN molecule is approximately 1.5 μ M, followed by the binding of the second HMGN molecule at approximately 0.17 μ M (Kato *et al.*, 2011).

Cooperative HMGN binding is apparent as the conformational changes induced upon the first HMGN binding to NCP promote the second variant-specific binding at a ~10-fold lower dissociation constant. The lifetime of the full ternary complex, however, is only about 20 ms. We speculate that this transient interaction probably stems from the reduced affinity of HMGN towards NCP as a consequence of the nucleosome having a new conformation induced by the two HMGN molecule associations.

Two closely related HMGN proteins, HMGN1 and HMGN2 share 45% sequence identity, with the NBD having most of the conserved elements. Both proteins have been associated with distinct biological processes *in vitro*, and the responsible domains have been located to the C-terminal regions of the HMGN. HMGN1 (but not HMGN2) inhibits the phosphorylation of both H3S10 and H3S28, whereas HMGN2 enhances H3K14 acetylation more robustly than HMGN1 (Ueda *et al.*, 2006). EMSA analysis of NCP–HMGN complexes at near physiological conditions demonstrates that the two HMGN types induce different NCP conformations upon binding. HMGN2, which is smaller in size than HMGN1, causes slower migration of the complex compared to that of HMGN1, implying a more open NCP conformation triggered by HMGN2 association. Furthermore, in the presence of divalent metal, formation of a distinctive NCP-HMGN complex species (S2) is observed only for HMGN2. This S2 species migrates much slower than the other HMGN-NCP complex at 2:1 stoichiometric ratio (S1 species). Taken together, these findings suggest that different genomic activities associated with different HMGN proteins may be regulated on the basis of distinct nucleosomal conformational changes induced by the binding of different types of HMGN proteins (Ong *et al.*, 2010).

The unique nucleosomal configuration associated with the S2 species, induced by HMGN2 association, perhaps has similarities with several other remodelled NCP configurations. SWI/SNF-triggered nucleosome remodelling, for example, creates nucleosomal DNA loops without really disturbing the intact nucleosome core configuration, thus allowing the nucleosome to easily go back to its original configuration once the remodelling complex is removed. Our western blot analysis of the S2 species and previous findings on the reversibility of S1-S2 interconversion *in vitro* may indicate the same line of approach utilized by HMGN2 for nucleosome remodelling. In this case, HMGN2 and divalent ion work hand in hand to open up nucleosome structure, allowing some part of the terminal nucleosomal DNA to freely interact with other regulatory factors, without causing any histone dissociation. Given the abundance of HMGN protein inside the nucleus, HMGN transient interaction with

nucleosome and their high mobility in moving from one nucleosome to another, this strategy is perhaps the most suitable way for HMGN to serve its function in nucleosome remodelling.

Several studies have shown that at one time only two HMGN proteins can associate specifically with the NCP. Nevertheless in our study, S2 formation is observed in EMSA only when HMGN2 or truncated HMGN1 is present in higher than 2:1 stoichiometry. This finding implies the presence of extra HMGN molecules in S2, which we demonstrated convincingly in our fluorescence studies. Extra HMGN proteins in S2, however, might be a consequence of having a more open NCP structure triggered by the first two HMGN molecules binding to the NCP. Using the same hypothesis as mentioned in the previous paragraph, we predict an increase in nucleosomal DNA breathing (the degree of spontaneous nucleosomal DNA wrapping and rewinding around histone core) upon the first two HMGN molecule associations. The extra HMGN proteins perhaps capture and bind non-specifically to nucleosomal DNA when the DNA termini partially dissociate from the NCP (Figure 59). The same phenomenon may also occur inside the living cell. HMGN proteins decompact the nucleosome, and subsequently facilitate other protein factors to access the loosened nucleosomal DNA.

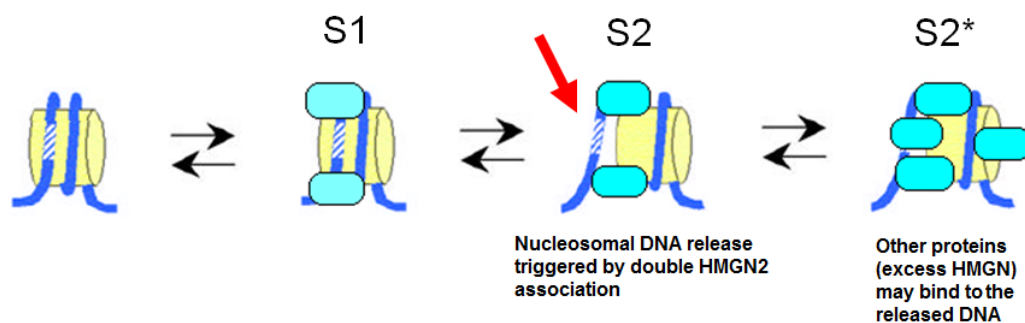


Figure 59. Hypothesis on S2 formation. Nucleosomal DNA is transiently unwrapped (red arrow), which subsequently provides binding site for other proteins.

7.7. NBD Binding Characterization

HMGN1 and HMGN2 have been found to interact with the NCP in a variant-specific manner. Only the same type of HMGN is found on a given NCP. Moreover, this

selectivity extends to the chromatin level, as both HMGN1 and HMGN2 localize to distinct clusters on chromatin *in vivo* (Postnikov *et al.*, 1997). Analysis of HMGN deletion mutants revealed the importance of the NBD in HMGN homodimer formation. It is intriguing as the mechanism for this homodimerization does not involve any direct contact between HMGN proteins. Crosstalk between two HMGN NBDs is perhaps mediated through an indirect allosteric mechanism propagated through the NCP structure (Postnikov *et al.*, 1995). Thus, unique conformational changes induced by specific HMGN variants again dictate the distinctive behaviour between HMGN proteins. It is important to note, however, that the homodimerization is abolished in low-ionic strength buffer conditions, indicating the sensitivity of HMGN behaviour towards environmental conditions (Crippa *et al.*, 1992).

Our EMSA and crosslinking-western blot experiments have shown that NBD truncation series of both HMGN1 and HMGN2 still maintain contact with the NCP. The shortest NBD truncation peptide (19 amino acids long) of HMGN1 protein has been shown to interact primarily with H2A, H2B, and H3 histones, whilst NBD truncation series from HMGN2, in contrast, are shown to crosslink only to the H2B protein. Sequence alignment of both NBDs shows that there is an extra Lysine residue present on the shortest HMGN1 NBD truncation series (Figure 60), which may be responsible for the difference observed in the HMGN1 NBD and HMGN2 NBD crosslinking profiles. Assuming that this was the case, we could infer the position of the second lysine from the N-terminus of the HMGN1 NBD to be somewhere around the H2A, H2B, and H3 interface. Differences in crosslinking profiles, however, could also arise from different structural changes induced by these two variant NBDs. In this case, the HMGN1 NBD and HMGN2 NBD may position their amino acids at slightly different locations, thus explaining the different crosslinking patterns observed. More advanced approaches, such as the newly developed technique of crosslinking/mass-spectrometry (Rappsilber, 2011), are required to test these two hypotheses, and pinpoint more precisely the HMGN binding sites.

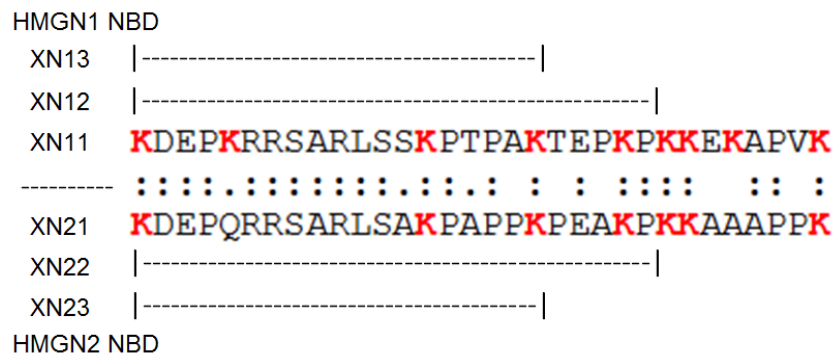


Figure 60. Sequence alignment of the HMGN1 NBD and HMGN2 NBD truncation series. Lysine residues present on the sequences are coloured red. These lysines participate in crosslinking reactions between HMGN and the NCP.

We show unambiguously that the N-terminal part of the NBD is responsible for histone core recognition and specific association to the NCP. In addition, although crosslinking experiments could be biased by lysine side chain distribution, the roughly uniform distribution and abundance of lysines in the NBD and histone proteins suggest that this strategy can be an informative approach to gain HMGN-NCP structural knowledge in the absence of a real atomic structure. It might be still difficult to infer the HMGN2 NBD binding site as the crosslinking was observed mainly to H2B. On the other hand, crosslinking of HMGN1 NBD to three of the histones, H2A, H2B, and H3, suggested that HMGN1 NBD binding is probably around the interface of these three histones. Subsequent inspection of the NCP crystal structure indicates the corresponding potential HMGN1 binding site to be nearby the H2A-H2B acidic patch (Figure 61). The presence of several positively-charged lysine and arginine residues, and polar serine side chains on the N-terminal part of NBD would appear to facilitate contact with the H2A-H2B acidic region. Moreover, it also helps to explain why phosphorylation of NBD's two critical serine residues on the highly conserved RRSARLSA motif leads to HMGN release from the NCP. The negative charge of the phosphate groups would cause electrostatic repulsion to the H2A-H2B acidic region. It is also very interesting to note that the same acidic patch has been known to be the association site for some other proteins factors, such as Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen (LANA) and Regulator of Chromosome Condensation (RCC1) protein (Barbera *et al.*, 2006; Makde *et al.*, 2010). More importantly, a recent NMR study of the HMGN2-NCP interaction also came to the same conclusion (Kato *et al.*, 2011).

From our previous EMSA analysis, we observed reduced smearing of free DNA bands as shorter C-terminal truncated NBD peptides were added as ligand. In addition to that, even at much higher peptide-to-DNA ratios, short NBD peptides (especially XN13 and XN14) still failed to produce the same level of DNA binding as compared to full-length NBD peptides (XN11). This finding indicates that NBD association to free DNA is not a simple random binding, as such association would allow short peptides to have a similar DNA binding level to the longer peptides when the short peptide concentration is increased. The deletion of the C-terminal part, therefore, seems to reduce NBD binding affinity to the DNA, indicating the possible interaction of the NBD C-terminal element with nucleosomal DNA in the HMGN-NCP complex. Multiple positively-charged lysine residues on the C-terminal part of the NBD perhaps facilitate the interaction with the negatively-charged nucleosomal DNA.

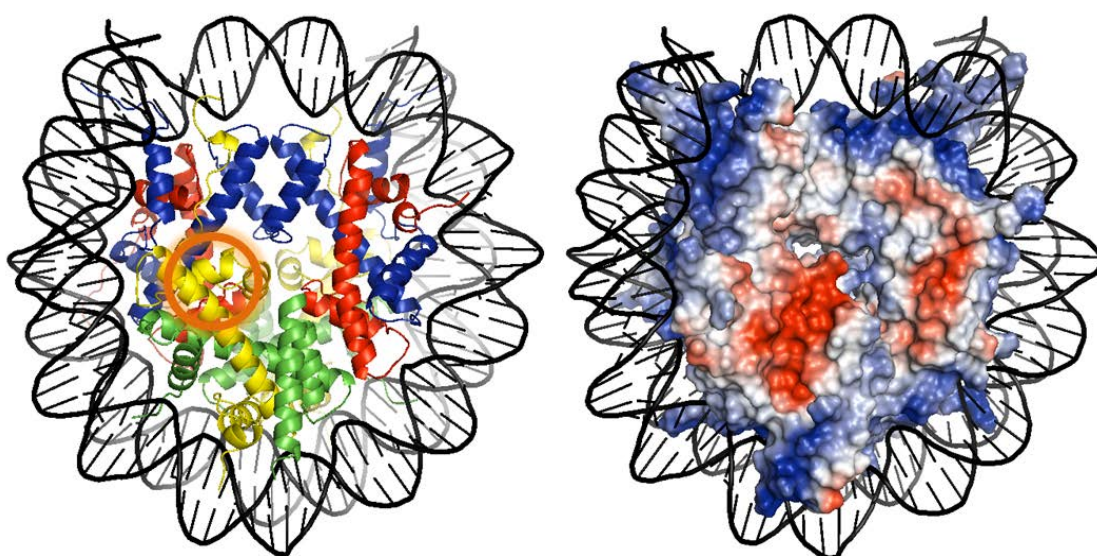


Figure 61. Prediction of HMGN1 NBD binding site on the NCP based on an HMGN1 NBD-NCP crosslinking - western blot profile [left panel; H2A (yellow), H2B (green), H3 (blue), H4 (red), DNA (black)]. The potential NBD binding site is indicated by a brown circle. This circle overlaps with the acidic patch present on the surface of H2A-H2B (right panel, red region).

7.8. Crystallization Strategy

Several factors are thought to hamper our HMGN-NCP crystallization attempts. Examination of HMGN binding affinity under different buffers demonstrated that affinity is greatly affected by the ionic strength of the buffer solution. With high concentrations of both divalent and monovalent ions in our native crystallization

buffers, HMGN might be readily excluded from the crystal due to weakened binding affinity. In this case, the overall free energy associated with HMGN-NCP crystals is perhaps less favourable than that of NCP-only crystals, making NCP crystals the favoured outcome for crystallizations in high ionic strength buffer solutions. Subsequent co-crystallization attempts should consider low-ionic strength buffer conditions, and most preferably in the absence of divalent ions that promotes nucleosome unfolding in the presence of HMGN2 and truncated HMGN1. Possibility of having different mode of HMGN binding to nucleosome; cooperative versus non-cooperative bindings, however, should be addressed when the crystallization is conducted in low-ionic strength buffers. Should there be any necessity to crystallize the complex in high-ionic strength buffer, HMGN1-NCP complex perhaps is more promising because HMGN1 does not promote formation of the S2 species.

The HMGN dissociation constant towards NCP is known to be relatively modest, only in the range of micromolar, and the affinity difference between HMGN specific and nonspecific association with the nucleosome is only about 3-fold (Ueda *et al.*, 2008). This moderate affinity makes purification of HMGN-bound NCP difficult as the HMGN may readily dissociate from the NCP during the purification process. Moreover, the small affinity difference between specific and non-specific binding makes the system very susceptible to the occurrence of non-specific interactions, especially when high concentrations of NCP and HMGN are used, such as in the crystallization process. In addition, the situation is exacerbated when stoichiometry of 2:1 HMGN-NCP is not achieved, as excess HMGN would just randomly bind to the NCP, and non-saturating HMGN levels would yield a mixture of HMGN-NCP species. Crystallization of HMGN-NCP fusions is perhaps a valuable approach for HMGN-NCP crystallization due to its precise 2:1 HMGN-to-NCP stoichiometry.

In addition to its role in providing a perfect 2:1 HMGN-to-NCP ratio, HMGN-NCP fusion also increases the effective concentration of the complex as the linkage between HMGN protein and NCP keeps both of them in close proximity. This higher effective concentration is likely to promote stable association between HMGN and NCP. Moreover, in the case of crystallization, protein fusion will ensure that HMGN is present inside the crystal, eliminating the possibility of having HMGN excluded from the crystals. Caution regarding linker length, however, needs to be observed. Amino acid linkers should be long enough to allow HMGN to bind the NCP in the appropriate fashion, but not too long that it will start causing problems for crystallization. Amino

acid identity of the linker may be an additional factor that affects the crystallization process.

The presence of divalent ion, even as low as one milimolar concentration, produces significant alterations in HMGN-induced NCP conformations. Closer inspection of EMSA results shows that this new conformation does not adopt a rigid and solid structure, as indicated by band smearing in the HMGN-NCP complexes. This very flexible structure may impede crystal formation since a well-defined structure is required to build proper crystal contacts between molecules. Our HMGN-NCP fusion approach yielded very small crystals after 4-5 hours of incubation, yet these went back into solution after ~12 hours. We hypothesized this is perhaps due to the intrinsic dynamics of the HMGN-NCP system. It is therefore very important to try crystallizing the NCP with truncated HMGNs or with certain HMGN domains that do not induce too much flexibility on the overall NCP structure.

Earlier studies have shown that the NBD domain works independently, and its binding to NCP still upholds HMGN variant-specific pairing *in vitro* even though there appears to be no direct contact between the NBDs within an individual nucleosome (Postnikov *et al.*, 1995). As we stated previously, this may come about from variant-specific conformational changes induced within the nucleosome by a particular NBD. Knowing the structure of an HMGN NBD-NCP complex could give insight into the nature of the conformational change required for formation of the variant-specific complex. Our studies have shown that, in addition to full-length NBD, series of NBD truncation peptides are also able to maintain association with the NCP. Therefore, crystallization of NCP-NBD or NCP-truncated NBD complexes might be fruitful as the NBD domain alone in theory does not cause NCP structural rearrangements as extensive as full length HMGN, but at the same time would still give good structural knowledge about NBD variant-specific pairing. Crystallizations of NCP-NBD and NCP-truncated NBD in standard NCP crystallization buffers, however, seemed to exclude NBD peptide from the crystals. Closer examination of NCP crystal contacts shows that at least one of the potential binding sites should be accessible to NBD or truncated NBD peptides (Figure 62). Crystallization of NBD-NCP fusions could also be helpful for future crystallization trials.

EMSA analysis of low-ionic strength NBD-NCP crystals indicated the presence of NBD peptide inside the crystals. Data resolution however prevented us from finding electron density associated with the NBD peptides. In this case, derivatized NBD

peptides containing one or more heavy atoms (incorporated as modified amino acid) perhaps could help identification of NBD peptide inside the crystals. Heavy atoms generate much stronger and specific signals (anomalous signals) than light atoms when diffraction data are collected at heavy atom's absorption edge x-ray wavelength. This strategy has a good chance of working even with low resolution data.

HMGN proteins are intrinsically disordered, and a defined structure is established only after they bind to nucleosomes (Crippa *et al.*, 1992). This nature of HMGN proteins prevents crystallization of HMGN protein alone. HMGN or NBD crystallization in the presence of their minimum binding site, probably the H2A-H2B dimer, may be a good approach, as this minimal histone motif may stabilize HMGN in its defined structure. This strategy has been employed to elucidate interactions of several protein factors with histone motifs (Taverna *et al.*, 2007).

Lastly, the search for novel NCP crystallization buffers should be continued. Interactions of several molecules with the NCP have been investigated by means of NCP crystal soaking (Barbera *et al.*, 2006; Davey *et al.*, 2010; Wu *et al.*, 2008; Wu *et al.*, 2011). The nature of HMGN-NCP interaction however appears to prevent a crystal soaking approach being useful in our current high ionic strength crystallization buffers. We have found several leads for novel NCP crystallization buffers. In any case, further optimization of crystallization buffer conditions is still required. These new NCP crystallization buffers may be very promising for crystal soaking experiments in the future.

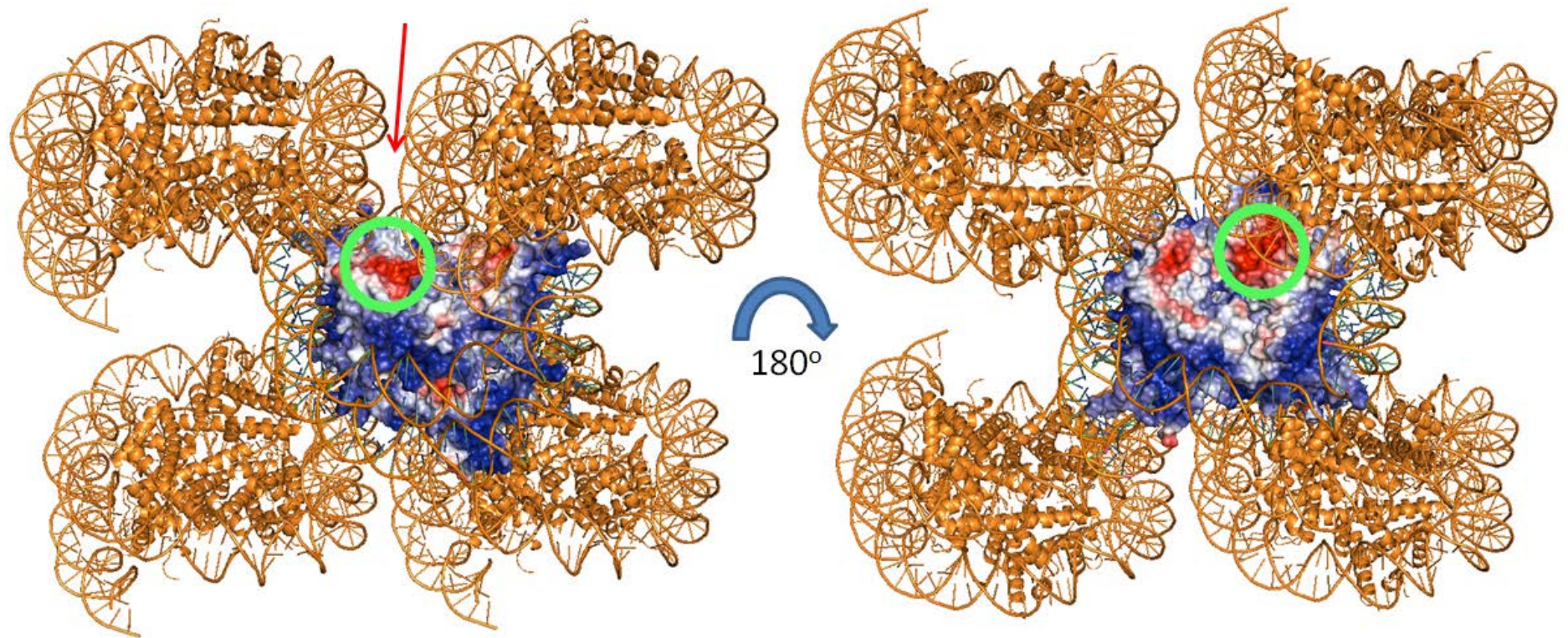


Figure 62. Location of potential NBD binding sites inside NCP crystals (green circle). One of the binding sites (as shown by NCP structure on the left panel) is accessible to peptide binding. NBD peptides may diffuse in through the available channel (red arrow) during soaking of NCP crystals with NBD peptide solutions. On the other hand, the opposite symmetry-related potential binding site is blocked by a crystal contact with neighboring NCP (right panel). Neighboring NCPs in crystal lattice and DNA of NCP in the centre are colored gold. Histone octamer of NCP in the centre is depicted in space-filling mode and colored based on its negative (red) and positive (blue) charge distributions.

Appendix

DNA

146b: atctccaaatatcccttgcggatcgtagaaaaagtgtgtcaaactgcgctatcaaagggaacttcaactgaattcagttgaagtttccctttgatagcgcagttgacacacttttctacgatccgcaagggatattggagat

147 : atcaatatccacctgcagatactacaaaaagtgtatttggaaactgctccatcaaaaggcatgttcagctggaatcagctgaacatgccttttgatggagcagttccaaatcaccttttgtagtatctgcaggtggatattgat

145 : atcaatatccacctgcagatactacaaaaagtgtatttggaaactgctccatcaaaaggcatgttcagctgaatcagctgaacatgccttttgatggagcagttccaaatcaccttttgtagtatctgcaggtggatattgat

601L145:

atcagaatcccggtgccgaggccgctcaattggtcgtagacagctctagcaccgcttaaacgcacgtacggaatccgtacgtgcgttaagcgggtgctagagctgtctacgaccaaattgagcggcctcggcaccgggatttgtat

DNA bases with major groove facing the histone octamer are colored black, while ones with minor groove facing the histone octamer are colored red.

H2A

H2A

X.laevis 1 SGRGKQGGKTRAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVY LAAVLEYLTAETLELAGNAARDNKKTRIIPRHLQLAVRNDEELNKLGRV TIAQGGVLPNIQSVLLPKKTES-SKSKSK 128
|||||.|||||||:|||||||:|||||||:|||||||:|||||||.|||.||
H.sapiens 1 SGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVY LAAVLEYLTAETLELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLGRV TIAQGGVLPNIQAVLLPKKTESHHKAKGK 129

H2B

[illegible]

H3

X.laevis 1 ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALRE IRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYL VALFEDTNLCAIHAKRVTIMPKDIQLARRIGERA 135
|||||
H.sapiens 1 ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALRE IRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEACEAYL VGLFEDTNLCAIHAKRVTIMPKDIQLARRIGERA 135
|||||

H4

X.laevis	1	SGRGKGGKGLGKGGAKRHRKVL	RDNIOGITKPAIRRLARRGGVKRISGLIYEETR	GVLVFLENVIRDAVITYTEHAKRKT	VTAMDVVYALKRQGR	ILYGF	GG	102
H.sapiens	1	SGRGKGGKGLGKGGAKRHRKVL	RDNIOGITKPAIRRLARRGGVKRISGLIYEETR	GVLVFLENVIRDAVITYTEHAKRKT	VTAMDVVYALKRQGR	ILYGF	GG	102

Xenopus laevis HMGN Proteins and Peptides

HMGN1

PKRKQVNADV ADAKDEP KRRSARLSSK PTPAKTEPKP KKEKAPVKEKPEE KEKKVPAKGK KGAKGKQTEE ANKEEANEDQ PSENGETKSD EAPASDGGDK
ESKSE

HMGN1-truncated protein

PKRKQVNADV ADAKDEP KRRSARLSSK PTPAKTEPKP KKEKAPVKEKPEE KEKKVPAKGK KGAKGKQTEE ANKEEANEDQ

HMGN1 NBD (XN11) KDEP KRRSARLSSK PTPAKTEPKP KKEKAPVK

HMGN1 NBD (XN12) KDEP KRRSARLSSK PTPAKTEPKP

HMGN1 NBD (XN13) KDEP KRRSARLSSK PTPAK

HMGN1 NBD (XN14) KDEP KRRSARLSSK P

HMGN1 mutated NBD (XN11Δ) KDEP KRR**E**ARL**E**SK PTPAKTEPKP KKEKAPVK

HMGN1 mutated NBD (XN12Δ) KDEP KRR**E**ARL**E**SK PTPAKTEPKP

HMGN1 mutated NBD (XN13Δ) KDEP KRR**E**ARL**E**SK PTPAK

HMGN2

PKRKADGDSK AEKAKAKDEP QRRSARLSAK PAPPKPEAKP KAAAAPPKA DKAPKGKKGK ADSGKDSSNA AENGEAKSDQ AQKAETGDTK

HMGN2 NBD (XN21) KDEP QRRSARLSAK PAPPKPEAKP KAAAAPPK

HMGN2 NBD (XN22) KDEP QRRSARLSAK PAPPKPEAKP

HMGN2 NBD (XN23) KDEP QRRSARLSAK PAPPK

HMGN2 mutated NBD (XN21Δ) KDEP KRR**E**ARL**E**AK PAPPKPEAKP KAAAAPPK

HMGN1-H2A fusion protein

PKRKQVNADVADAKDEPKRRSARLSSKPTPAKTEPKPKKEKAPVKEKPEEKEKKVPAKGKKGAKGKQTEEANKEEANEDQPSENGETKSDEAPASDGGDKESKSEGGSGGGSSGSGRGKQGGKTRAKAKTRSSRAGLQFPVGRVHLLRKGNYAERVGAGAPVYLAAVLEYLTAEILELAGNAARDNKKTRIIPRHLQLAVRNDDEELNKLGRVTIAQGGVLPNIQSVLLPKKTESSKSKSK

HMGN1-H3 fusion protein

PKRKQVNADVADAKDEPKRRSARLSSKPTPAKTEPKPKKEKAPVKEKPEEKEKKVPAKGKKGAKGKQTEEANKEEANEDQPSENGETKSDEAPASDGGDKESKSEGGSGGGSSGGARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLCAIHAKRVTIMPKDIQLARRIGERA

References

- Abbott, D.W., Chadwick, B.P., Thambirajah, A.A., Ausio, J., 2005. Beyond the Xi: macroH2A chromatin distribution and post-translational modification in an avian system. *J Biol Chem* 280, 16437-16445.
- Adam, M., Robert, F., Larochelle, M., Gaudreau, L., 2001. H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. *Mol Cell Biol* 21, 6270-6279.
- Agresti, A., Scaffidi, P., Riva, A., Caiolfa, V.R., Bianchi, M.E., 2005. GR and HMGB1 interact only within chromatin and influence each other's residence time. *Mol Cell* 18, 109-121.
- Altaf, M., Auger, A., Covic, M., Cote, J., 2009. Connection between histone H2A variants and chromatin remodeling complexes. *Biochem Cell Biol* 87, 35-50.
- Anderson, J.D., Widom, J., 2001. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol Cell Biol* 21, 3830-3839.
- Armache, K.J., Garlick, J.D., Canzio, D., Narlikar, G.J., Kingston, R.E., 2011. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* 334, 977-982.
- Avery, O.T., Macleod, C.M., McCarty, M., 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type Iii. *J Exp Med* 79, 137-158.
- Bannister, A.J., Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell Res* 21, 381-395.
- Bao, Y., Konesky, K., Park, Y.J., Rosu, S., Dyer, P.N., Rangasamy, D., Tremethick, D.J., Laybourn, P.J., Luger, K., 2004. Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *EMBO J* 23, 3314-3324.
- Barbera, A.J., Chodaparambil, J.V., Kelley-Clarke, B., Joukov, V., Walter, J.C., Luger, K., Kaye, K.M., 2006. The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* 311, 856-861.
- Black, B.E., Cleveland, D.W., 2011. Epigenetic centromere propagation and the nature of CENP-a nucleosomes. *Cell* 144, 471-479.
- Casini, A., Gabbiani, C., Michelucci, E., Pieraccini, G., Moneti, G., Dyson, P.J., Messori, L., 2009. Exploring metallodrug-protein interactions by mass spectrometry: comparisons between platinum coordination complexes and an organometallic ruthenium compound. *J Biol Inorg Chem* 14, 761-770.

- Casini, A., Mastrobuoni, G., Ang, W.H., Gabbiani, C., Pieraccini, G., Moneti, G., Dyson, P.J., Messori, L., 2007. ESI-MS Characterisation of Protein Adducts of Anticancer Ruthenium(II)-Arene PTA (RAPTA) Complexes. *ChemMedChem* 2, 631-635.
- Casini, A., Gabbiani, C., Sorrentino, F., Rigobello, M.P., Bindoli, A., Geldbach, T.J., Marrone, A., Re, N., Hartinger, C.G., Dyson, P.J., Messori, L., 2008. Emerging protein targets for anticancer metallodrugs: inhibition of thioredoxin reductase and cathepsin B by antitumor ruthenium(II)-arene compounds. *J Med Chem* 51, 6773-6781.
- Catez, F., Brown, D.T., Misteli, T., Bustin, M., 2002. Competition between histone H1 and HMGN proteins for chromatin binding sites. *EMBO Rep* 3, 760-766.
- Catez, F., Lim, J.H., Hock, R., Postnikov, Y.V., Bustin, M., 2003. HMGN dynamics and chromatin function. *Biochem Cell Biol* 81, 113-122.
- CCP4, 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.
- Chodaparambil, J.V., Barbera, A.J., Lu, X., Kaye, K.M., Hansen, J.C., Luger, K., 2007. A charged and contoured surface on the nucleosome regulates chromatin compaction. *Nat Struct Mol Biol* 14, 1105-1107.
- Clarke, M.J., Galang, R.D., Rodriguez, V.M., Kumar, R., Pell, S., Bryan, D.M., 1988. Chemical Considerations in the Design of Ruthenium Anticancer Agents. In: Nicolini, M. (Ed.), *Platinum and Other Metal Coordination Compounds in Cancer Chemotherapy*. Springer US, pp. 582-600.
- Costanzi, C., Pehrson, J.R., 2001. MACROH2A2, a new member of the MARCOH2A core histone family. *J Biol Chem* 276, 21776-21784.
- Crippa, M.P., Alfonso, P.J., Bustin, M., 1992. Nucleosome core binding region of chromosomal protein HMG-17 acts as an independent functional domain. *J Mol Biol* 228, 442-449.
- Danford, A.J., Wang, D., Wang, Q., Tullius, T.D., Lippard, S.J., 2005. Platinum anticancer drug damage enforces a particular rotational setting of DNA in nucleosomes. *Proc Natl Acad Sci U S A* 102, 12311-12316.
- Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., 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.
- Davey, G.E., Davey, C.A., 2008. Chromatin - a new, old drug target? *Chem Biol Drug Des* 72, 165-170.
- Davey, G.E., Wu, B., Dong, Y., Surana, U., Davey, C.A., 2010. DNA stretching in the nucleosome facilitates alkylation by an intercalating antitumour agent. *Nucleic Acids Res* 38, 2081-2088.
- Durrenberger, M., Bjornsti, M.A., Uetz, T., Hobot, J.A., Kellenberger, E., 1988. Intracellular location of the histonelike protein HU in *Escherichia coli*. *J Bacteriol* 170, 4757-4768.

- Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., Luger, K., 2004. Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* 375, 23-44.
- Eberharter, A., Becker, P.B., 2002. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep* 3, 224-229.
- Eirin-Lopez, J.M., Ishibashi, T., Ausio, J., 2008. H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. *Faseb J* 22, 316-326.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Felsenfeld, G., Groudine, M., 2003. Controlling the double helix. *Nature* 421, 448-453.
- Furuyama, T., Henikoff, S., 2009. Centromeric nucleosomes induce positive DNA supercoils. *Cell* 138, 104-113.
- Gerlitz, G., Hock, R., Ueda, T., Bustin, M., 2009. The dynamics of HMG protein-chromatin interactions in living cells. *Biochem Cell Biol* 87, 127-137.
- Gillis, A.J., Schuller, A.P., Skordalakes, E., 2008. Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* 455, 633-637.
- Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., de Lange, T., 1999. Mammalian telomeres end in a large duplex loop. *Cell* 97, 503-514.
- Hargreaves, D.C., Crabtree, G.R., 2011. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* 21, 396-420.
- Harp, J.M., Hanson, B.L., Timm, D.E., Bunick, G.J., 2000. Asymmetries in the nucleosome core particle at 2.5 Å resolution. *Acta Crystallogr D Biol Crystallogr* 56, 1513-1534.
- Hock, R., Furusawa, T., Ueda, T., Bustin, M., 2007. HMG chromosomal proteins in development and disease. *Trends Cell Biol* 17, 72-79.
- Huth, J.R., Bewley, C.A., Nissen, M.S., Evans, J.N., Reeves, R., Gronenborn, A.M., Clore, G.M., 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat Struct Biol* 4, 657-665.
- Jiang, C., Pugh, B.F., 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10, 161-172.
- Jin, C., Felsenfeld, G., 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* 21, 1519-1529.
- Jones, P.A., Baylin, S.B., 2007. The epigenomics of cancer. *Cell* 128, 683-692.

- Jung, Y., Lippard, S.J., 2007. Direct cellular responses to platinum-induced DNA damage. *Chem Rev* 107, 1387-1407.
- Kato, H., van Ingen, H., Zhou, B.R., Feng, H., Bustin, M., Kay, L.E., Bai, Y., 2011. Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A* 108, 12283-12288.
- Landry, J., Sutton, A., Hesman, T., Min, J., Xu, R.M., Johnston, M., Sternglanz, R., 2003. Set2-catalyzed methylation of histone H3 represses basal expression of GAL4 in *Saccharomyces cerevisiae*. *Mol Cell Biol* 23, 5972-5978.
- Lavelle, C., Recouvreux, P., Wong, H., Bancaud, A., Viovy, J.L., Prunell, A., Victor, J.M., 2009. Right-handed nucleosome: myth or reality? *Cell* 139, 1216-1217; author reply 1217-1218.
- Lei, M., Podell, E.R., Cech, T.R., 2004. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat Struct Mol Biol* 11, 1223-1229.
- Leslie, A.G., 2006. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* 62, 48-57.
- Lim, J.H., Bustin, M., Ogryzko, V.V., Postnikov, Y.V., 2002. Metastable macromolecular complexes containing high mobility group nucleosome-binding chromosomal proteins in HeLa nuclei. *J Biol Chem* 277, 20774-20782.
- Lin, J.C., Jeong, S., Liang, G., Takai, D., Fatemi, M., Tsai, Y.C., Egger, G., Gal-Yam, E.N., Jones, P.A., 2007. Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. *Cancer Cell* 12, 432-444.
- Lipps, H.J., Rhodes, D., 2009. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol* 19, 414-422.
- Lotze, M.T., Tracey, K.J., 2005. High-mobility group box 1 protein (HMGB1): nuclear weapon in the immune arsenal. *Nat Rev Immunol* 5, 331-342.
- Lowary, P.T., 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.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251-260.
- Makde, R.D., England, J.R., Yennawar, H.P., Tan, S., 2010. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* 467, 562-566.
- Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., Pugh, B.F., 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* 18, 1073-1083.

Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M.M., Wu, C., 2007. Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell* 129, 1153-1164.

Mohideen, K., Muhammad, R., Davey, C.A., 2010. Perturbations in nucleosome structure from heavy metal association. *Nucleic Acids Res* 38, 6301-6311.

Neumann, H., Hancock, S.M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., Chin, J.W., 2009. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol Cell* 36, 153-163.

Nightingale, K., Dimitrov, S., Reeves, R., Wolffe, A.P., 1996. Evidence for a shared structural role for HMG1 and linker histones B4 and H1 in organizing chromatin. *EMBO J* 15, 548-561.

Ong, M.S., Richmond, T.J., Davey, C.A., 2007. DNA stretching and extreme kinking in the nucleosome core. *J Mol Biol* 368, 1067-1074.

Ong, M.S., Vasudevan, D., Davey, C.A., 2010. Divalent metal- and high mobility group N protein-dependent nucleosome stability and conformation. *J Nucleic Acids* 2010, 143890.

Oudet, P., Gross-Bellard, M., Chambon, P., 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* 4, 281-300.

Pisano, S., Leoni, D., Galati, A., Rhodes, D., Savino, M., Cacchione, S., 2010. The human telomeric protein hTRF1 induces telomere-specific nucleosome mobility. *Nucleic Acids Res* 38, 2247-2255.

Poirier, M.G., Oh, E., Tims, H.S., Widom, J., 2009. Dynamics and function of compact nucleosome arrays. *Nat Struct Mol Biol* 16, 938-944.

Postnikov, Y.V., Trieschmann, L., Rickers, A., Bustin, M., 1995. Homodimers of chromosomal proteins HMG-14 and HMG-17 in nucleosome cores. *J Mol Biol* 252, 423-432.

Postnikov, Y.V., Herrera, J.E., Hock, R., Scheer, U., Bustin, M., 1997. Clusters of nucleosomes containing chromosomal protein HMG-17 in chromatin. *J Mol Biol* 274, 454-465.

Rappsilber, J., 2011. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* 173, 530-540.

Rattner, B.P., Yusufzai, T., Kadonaga, J.T., 2009. HMGN proteins act in opposition to ATP-dependent chromatin remodeling factors to restrict nucleosome mobility. *Mol Cell* 34, 620-626.

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., Bonner, W., 2002. Histone H2A variants H2AX and H2AZ. *Curr Opin Genet Dev* 12, 162-169.

Reeves, R., Beckerbauer, L., 2001. HMG1/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim Biophys Acta* 1519, 13-29.

- Reeves, R., Edberg, D.D., Li, Y., 2001. Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells. *Mol Cell Biol* 21, 575-594.
- Richmond, T.J., Davey, C.A., 2003. The structure of DNA in the nucleosome core. *Nature* 423, 145-150.
- Robinson, P.J., Rhodes, D., 2006. Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* 16, 336-343.
- Robinson, P.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., Rhodes, D., 2008. 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J Mol Biol* 381, 816-825.
- Routh, A., Sandin, S., Rhodes, D., 2008. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci U S A* 105, 8872-8877.
- Santos-Rosa, H., Caldas, C., 2005. Chromatin modifier enzymes, the histone code and cancer. *Eur J Cancer* 41, 2381-2402.
- Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., Kouzarides, T., 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407-411.
- Satchwell, S.C., Drew, H.R., Travers, A.A., 1986. Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol* 191, 659-675.
- Schalch, T., Duda, S., Sargent, D.F., Richmond, T.J., 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436, 138-141.
- Schwabish, M.A., Struhl, K., 2007. The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. *Mol Cell Biol* 27, 6987-6995.
- Scolaro, C., Hartinger, C.G., Allardyce, C.S., Keppler, B.K., Dyson, P.J., 2008. Hydrolysis study of the bifunctional antitumour compound RAPTA-C, [Ru(eta⁶-p-cymene)Cl₂(pta)]. *J Inorg Biochem* 102, 1743-1748.
- Scolaro, C., Geldbach, T.J., Rochat, S., Dorcier, A., Gossens, C., Bergamo, A., Cocchietto, M., Tavernelli, I., Sava, G., Rothlisberger, U., Dyson, P.J., 2005. Influence of Hydrogen-Bonding Substituents on the Cytotoxicity of RAPTA Compounds. *Organometallics* 25, 756-765.
- Seet, B.T., Dikic, I., Zhou, M.M., Pawson, T., 2006. Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 7, 473-483.
- Segal, E., Widom, J., 2009. What controls nucleosome positions? *Trends Genet* 25, 335-343.
- Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., Widom, J., 2006. A genomic code for nucleosome positioning. *Nature* 442, 772-778.

- Sekulic, N., Bassett, E.A., Rogers, D.J., Black, B.E., 2010. The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. *Nature* 467, 347-351.
- Shay, J.W., Bacchetti, S., 1997. A survey of telomerase activity in human cancer. *Eur J Cancer* 33, 787-791.
- Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., Peterson, C.L., 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844-847.
- Sinden, R.R., Pettijohn, D.E., 1981. Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. *Proc Natl Acad Sci U S A* 78, 224-228.
- Subramanian, M., Gonzalez, R.W., Patil, H., Ueda, T., Lim, J.H., Kraemer, K.H., Bustin, M., Bergel, M., 2009. The nucleosome-binding protein HMGN2 modulates global genome repair. *FEBS J.*
- Sun, Z.W., Allis, C.D., 2002. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* 418, 104-108.
- Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.Y., Kimura, H., Kurumizaka, H., 2011. Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* 476, 232-235.
- Tagami, H., Ray-Gallet, D., Almouzni, G., Nakatani, Y., 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116, 51-61.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., Zhao, Y., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1016-1028.
- Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., Patel, D.J., 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* 14, 1025-1040.
- Thomas, J.O., 1999. Histone H1: location and role. *Curr Opin Cell Biol* 11, 312-317.
- Thomas, J.O., Travers, A.A., 2001. HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci* 26, 167-174.
- Tims, H.S., Gurunathan, K., Levitus, M., Widom, J., 2011. Dynamics of nucleosome invasion by DNA binding proteins. *J Mol Biol* 411, 430-448.
- Ueda, T., Postnikov, Y.V., Bustin, M., 2006. Distinct domains in high mobility group N variants modulate specific chromatin modifications. *J Biol Chem* 281, 10182-10187.
- Ueda, T., Catez, F., Gerlitz, G., Bustin, M., 2008. Delineation of the protein module that anchors HMGN proteins to nucleosomes in the chromatin of living cells. *Mol Cell Biol* 28, 2872-2883.

- Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., Sidow, A., 2011. Determinants of nucleosome organization in primary human cells. *Nature* 474, 516-520.
- Verdaasdonk, J.S., Bloom, K., 2011. Centromeres: unique chromatin structures that drive chromosome segregation. *Nat Rev Mol Cell Biol* 12, 320-332.
- Wang, F., Chen, H., Parsons, S., Oswald, I.D.H., Davidson, J.E., Sadler, P.J., 2003. Kinetics of Aquation and Anation of Ruthenium(II) Arene Anticancer Complexes, Acidity and X-ray Structures of Aqua Adducts. *Chemistry – A European Journal* 9, 5810-5820.
- Watson, J.D., Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737-738.
- West, K.L., Bustin, M., Zlatanova, J., Leuba, S.H., 2004. The role of HMGN proteins in chromatin function, *New Comprehensive Biochemistry*. Elsevier, pp. 135-154.
- White, C.L., Suto, R.K., Luger, K., 2001. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *Embo J* 20, 5207-5218.
- Williams, J.S., Hayashi, T., Yanagida, M., Russell, P., 2009. Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. *Mol Cell* 33, 287-298.
- Wu, B., Davey, C.A., 2008. Platinum drug adduct formation in the nucleosome core alters nucleosome mobility but not positioning. *Chem Biol* 15, 1023-1028.
- Wu, B., Droge, P., Davey, C.A., 2008. Site selectivity of platinum anticancer therapeutics. *Nat Chem Biol* 4, 110-112.
- Wu, B., Davey, G.E., Nazarov, A.A., Dyson, P.J., Davey, C.A., 2011. Specific DNA structural attributes modulate platinum anticancer drug site selection and cross-link generation. *Nucleic Acids Res* 39, 8200-8212.
- Wu, B., Ong, M.S., Groessl, M., Adhireksan, Z., Hartinger, C.G., Dyson, P.J., Davey, C.A., 2011. A ruthenium antimetastasis agent forms specific histone protein adducts in the nucleosome core. *Chemistry* 17, 3562-3566.
- Yau, P., Imai, B.S., Thorne, A.W., Goodwin, G.H., Bradbury, E.M., 1983. Effect of HMG protein 17 on the thermal stability of control and acetylated HeLa oligonucleosomes. *Nucleic Acids Res* 11, 2651-2664.
- Zakian, V.A., 1995. Telomeres: beginning to understand the end. *Science* 270, 1601-1607.
- Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., Struhl, K., 2009. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol* 16, 847-852.
- Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., Pugh, B.F., 2011. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* 332, 977-980.

Zhou, J., Fan, J.Y., Rangasamy, D., Tremethick, D.J., 2007. The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat Struct Mol Biol* 14, 1070-1076.

Zorzet, S., Bergamo, A., Cocchietto, M., Sorc, A., Gava, B., Alessio, E., Iengo, E., Sava, G., 2000. Lack of In vitro cytotoxicity, associated to increased G(2)-M cell fraction and inhibition of matrigel invasion, may predict In vivo-selective antimetastasis activity of ruthenium complexes. *J Pharmacol Exp Ther* 295, 927-933.