Towards X-ray crystallographic characterization of nucleosome interactions with linker histones, FOXA proteins and high mobility group N factors

Ravi, Sailatha

2018

Ravi, S. (2018). Towards X-ray crystallographic characterization of nucleosome interactions with linker histones, FOXA proteins and high mobility group N factors. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/74380>

<https://doi.org/10.32657/10356/74380>

**Towards X-ray Crystallographic Characterization of
Nucleosome Interactions with Linker Histones, FOXA
proteins and High Mobility Group N Factors**

Sailatha Ravi

Supervisor: Associate Professor Curtis Alexander Davey

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

2018

Acknowledgements

At the outset, I sincerely appreciate the patient guidance provided by my advisor, Associate Professor Curt A. Davey throughout my PhD study. A great friend, philosopher and guide, he was always understanding, and was instrumental in ensuring my scientific focus and progress through the three very challenging projects. I would like to also thank my committee members, Dr. Sara Sandin and Dr. Newman Sze Siu Kwan, for their various inputs during my study. I was fortunate enough to have a great set of companions in my lab – Dr. Gabriela Davey, Dr. Deepti Sharma, Dr. Zenita Adhireksan, Dr. Sivaraman Padavattan, Shum Wayne, De Falco Louis, Rachael, Dr. Eugene Chua Yue Dao, Dr. Bao Quiye, Dr. Ma Zhujun, and Zhong Wee. Without their constant help and camaraderie, it would not have been possible to make significant progress in my research. Specific to my research work, I would like to thank Dr. Zenita for providing the basic linker histone purification protocol, Dr. Bao for providing the sticky end DNA constructs, and Dr. Deepti Sharma for optimizing the SPR platform to study nucleosome-protein interactions. I appreciate the timely valuable assistance of Wayne with the optimization of the sticky end DNA production, and the joint effort of Dr. Sivaraman, Louis and Rachael in conducting the linker histone project, and for taking over the project. A heartfelt gratitude to Dr. Gabriela Davey for being a great buffer during the most difficult times. I would like to thank Dr. Susana Geifman Shochat, and Chuah York Wiew for their collaboration with conducting the SPR experiments. I thank Dr. Anna Elisabet Jansson and her team of the protein production platform, at Biopolis, for their efforts. Most importantly, I appreciate the support that I received from my family members during the entire course of study. I thank my husband and daughter for being very patient and cooperative to support me; and extend my sincere regards to my parents and my parents-in-law for their numerous visits to Singapore to take care of my daughter, thereby giving me appreciably more time to concentrate on my research.

I extend my gratitude for NTU for the scholarship assistance and the other funding agencies: Singapore Ministry of Education Academic Research Fund Tier 2 Programme (MOE2015-T2-2-089), Ministry of Education Academic Research Fund Tier 3 Program (grant MOE2012-T3-1-001) and Ministry of Health National Medical Research Council (grant NMRC/1312/2011) for the financial support.

Table of Contents

Acknowledgements.....	2
Table of Contents.....	3
List of Figures.....	7
List of Tables and Flowchart	10
Abbreviations.....	11
Abstract.....	12
Chapter 1 Introduction	14
A brief account on chromatin history.....	16
Gene regulation in chromatin context.....	17
Implications in cancer	18
Overview of the major techniques in structural biology	20
Structural biology [X-ray crystallography] initiative in chromatin research.....	21
Nucleosome structure and dynamic interactions.....	21
Histone octamer core and the acidic patch	22
Histones as drug target.....	25
Strategy for crystallizing a complex of nucleosome with chromatin factors.....	25
Scope of this research.....	26
Chapter 2 Interactions of FOXA family of proteins with the nucleosome.....	27
Introduction	27
FOX family of proteins	27
FOXA subfamily	27
Role of FoxA proteins in cancer	28
Interactions of FoxA proteins with chromatin.....	28
FoxA proteins- classified as a pioneer factor	31
Interactions of FoxA proteins with DNA	32
DBD of FOXA3 resembles the globular domain of linker histones	32
Sequence specific and non-sequence specific interactions of FoxA with DNA	34
Kinetics- k_d measurements	34
Interactions of FoxA1 with nucleosomes.....	35
Scope of this project	38
Materials and methods.....	39
Core histones protein purification	39
Histone octamer refolding.....	41
FOXA protein purification	41
DNA production	43

Nucleosome reconstitution	45
EMSA binding studies	45
Radioactive gels.....	45
Surface Plasmon resonance	46
Results and discussion.....	47
Core Histone production	47
Histone Octamer purification.....	48
Purification of Full-length Human FoxA1 protein using bacterial expression.....	50
Screening for stable and soluble human FoxA1 truncation constructs	53
Creation of random truncation constructs.....	55
Purification of a C-terminal construct: 295-460.....	57
Purification of different length N- and C-terminal truncated constructs.....	57
Purification of constructs lacking N-terminal region.....	61
Purification of other FOXA family members	63
Tagged constructs for soluble FOXA expression	64
Characterization of binding to DNA.....	66
SPR measurements.....	75
Conclusions and Future Directions.....	77
Chapter 3 Interactions of High Mobility Group N [HMGN] family of proteins with nucleosome... ..	79
HMG [High Mobility Group] of proteins.....	79
HMGN family of proteins and their role in cellular events	79
Members of HMGN family	81
Characteristics of HMGN proteins.....	81
Interactions with chromatin.....	84
Interactions with nucleosomes	86
Existing models.....	87
Scope of this project.....	89
Materials and methods	90
Protein and DNA constructs	90
Protein Purification.....	90
DNA Production.....	91
EMSA binding studies	92
Crystallization setup	93
Data collection.....	93
Results and discussion.....	95
Purification of hHMGN1 and hHNGN2 constructs	95

Preparation of His tag cleaved hHMGN1t and hHMGN2.....	97
Other HMGN family members	97
DNA production	98
NCP reconstitution	99
Human HMGN proteins stably associate with the Nucleosome Core Particle	100
Crystallization trials with HT-hHMGN1	104
NCP binding of HMGN NBD [Nucleosome Binding Domain] Peptides	109
Conclusions and Future Directions	110
Chapter 4 Structural Insights into Chromatin Compaction by Linker Histone	
Proteins.....	112
Introduction	112
Dynamic interactions with chromatin	112
Linker histone confers protection to linker DNA ends in chromatin	112
Linker histone subtypes and variants	113
Role in gene regulation	114
Implications in cancer	116
Basic domain organization.....	116
Multiple sequence alignment and Phylogenetic tree analysis of the linker histone variants.....	117
Globular domain has a winged-helix structure.....	119
Role of N and C-terminal regions.....	121
Linker Histone purification.....	122
Maintenance of higher order chromatin structure	123
H1x-CMS169s structure:	127
Scope of this project	131
Materials and methods.....	132
Linker histone protein purification	132
DNA constructs	133
Histone octamer purification	133
Nucleosome reconstitution	134
EMSA binding studies.....	134
Crystallization.....	134
Data collection	134
Data analysis	135
Results and Discussion	136
Purification of Linker histone variants	136
Binding studies with 169s nucleosome.....	144

Crystallization setup	145
Cryogenic conditions	146
Data analysis.....	146
Conclusions and future directions.....	159
References.....	161
Appendix.....	176
DNA constructs.....	176
FOXA Protein sequences	177
Mass Spectrometry based molecular weight determination for recombinant human FOXA1 protein preparations.....	178
HMGN protein sequences	183
Mass Spectrometry based molecular weight determination for recombinant human HMGN protein preparations.	184
Linker histone protein sequences	188
Clustal omega multiple sequence alignment of human full-length linker histones....	190
Molecular weight estimation by Mass spectrometry for some of the recombinant human linker histone variants protein preparations	192
Secondary Structure Prediction of human H1T2.....	197
3D Structure prediction of human H1T2	198
Evaluating the appropriateness of molecular replacement using model 1	199
Evaluating the appropriateness of molecular replacement using model 3	200
Evaluating the appropriateness of molecular replacement using model 2	201
Bacterial Strains.....	202
Addgene plasmids for tagged constructs	202
Materials.....	202
Chemicals.....	203
Enzymes and Buffers	205
Media and Buffer Recipes	205

List of Figures

Figure 1.1 Chromatin state distributions during the different stages of the cell cycle. ...	15
Figure 1.2 Hierarchical packaging of DNA into chromatin.....	16
Figure 1.3 Structure of the nucleosome core particle (C. A. Davey et al., 2002).....	23
Figure 1.4 Interactions of different proteins with the acidic patch.....	24
Figure 2.1 Multiple Sequence Alignment of the DBD of FOXA family members	29
Figure 2.2 FoxA1 is associated with mitotic chromosomes in nocodazole-arrested cells.	30
Figure 2.3 Summary of FOXA binding sites on the Alb1 enhancer region.....	31
Figure 2.4 Winged helix structure of the DBD of FOXA3	33
Figure 2.5 Sequence comparison of globular domain of linker histone and DBD of FOXA1	37
Figure 2.6 Comparison of the DBD of FOXA3 with the globular domain of chicken H5 at specific residue locations.	38
Figure 2.7 Recombinant human core histone H2B and H4 purification	48
Figure 2.8 Purification of refolded human histone octamer	49
Figure 2.9 Optimizing bacterial expression conditions improved FOXA1 protein yield in the inclusion body.....	51
Figure 2.10 Representative SDS-PAGE analysis on attempts to purify full-length human FOXA1 protein.....	54
Figure 2.11 ExPASy tool GOR based secondary structure prediction.....	55
Figure 2.12 Depiction of the FOXA1 C-terminal and N-terminal truncated constructs....	56
Figure 2.13 Purification of recombinant HT-hFoxA1[295-460] harboring an N-terminal His tag with TEV cleavage site.	57
Figure 2.14 Total and Soluble expression of the various truncated FOXA1 constructs expressed in <i>E. coli</i> BL21DE3 Rosetta cells.	60
Figure 2.15 Summary of the recombinant expression and purification of FOXA1 truncation constructs.....	61
Figure 2.16 Purified shorter FOXA1 proteins.....	62
Figure 2.17 Attempts to check the recombinant expression of FOXA1 using Sf9 cells. ...	63
Figure 2.18 Expression screen for full length FOXA family members.....	64
Figure 2.19 Cloning, expression and purification of HT-MBP-N10-FOXA1.	66
Figure 2.20 Production of FOXA1-DNA containing a high affinity binding site	68
Figure 2.21 Binding of shorter FOXA1 proteins with FOXA1-DNA.....	70
Figure 2.22 EMSA based binding studies indicate sequence specific binding to cognate DNA	71
Figure 2.23 NCP reconstitution with FOXA1 DNA.....	72
Figure 2.24 Binding of Δ NT-FOXA1 with NCP reconstituted with FOXA1-DNA	73

Figure 2.25 SPR measurements of Δ NT-FOXA1-HT with cognate FOXA1 DNA	76
Figure 2.26 SPR measurements of Δ NT-FOXA1-HT with cognate FOXA1-NCP	76
Figure 3.1 Multiple Sequence Alignment of human HMGN proteins.	83
Figure 3.2 Proposed models for HMGN binding to NCP.	88
Figure 3.3 Purification of recombinant hHMGN1 expressed in bacterial cells	96
Figure 3.4 SDS-PAGE analysis of TEV treated HT-hHMGN1, HT-hHMGN1t and HT-hHMGN2 samples.....	97
Figure 3.5 Final purified product for different HMGN proteins	98
Figure 3.6 Purification of CMS169s – a sticky end DNA construct.	100
Figure 3.7 EMSA studies of xHMGN binding to NCP in the presence of 1 mM Mg^{2+}	101
Figure 3.8 Binding of HMGNs to NCP601L under 0.25x TBE conditions.	102
Figure 3.9 EMSA to study the effect of Mg^{2+} on human HMGN1t binding to NCP 601L.	103
Figure 3.10 hHMGN1 and hHMGN2 binding to different nucleosomes under 1x TBE conditions	104
Figure 3.11 Representative images of hHMGN1-601L crystals.....	106
Figure 3.12 Stoichiometry dependence of crystal formation.	106
Figure 3.13 Different morphologies of hHMGN-nucleosome crystals mounted and tested for diffraction at the synchrotron facility.....	107
Figure 3.14 Optimization of crystal stabilization buffer conditions and representative diffraction patterns.	108
Figure 3.15 Verification of the presence of HMGNs in the crystals	109
Figure 3.16 Binding of NBD peptides hNBD1 and hNBD2 with the NCP 601L construct.....	110
Figure 4.1 Multiple Sequence Alignment of the globular domain of the various linker histone variants in humans.	118
Figure 4.2 Phylogenetic tree analysis of linker histone family of proteins	119
Figure 4.3 Structures of the globular domain of linker histones.....	120
Figure 4.4 Crystal structure of a tetranucleosome.....	124
Figure 4.5 Cryo-EM structures of chromatin fiber formed from reconstituted nucleosome arrays.....	125
Figure 4.6 Lattice arrangement of nucleosomes in the H1x containing chromatosome crystal.	128
Figure 4.7 The asymmetric unit comprising two H1x molecules bound to a nucleosome pair.....	129
Figure 4.8 Linker histone interaction with multiple nucleosomes.....	130
Figure 4.9 Recombinant expression and purification of human H1.2 protein.....	137
Figure 4.10 Recombinant expression and purification of human H1.4.....	139
Figure 4.11 Recombinant expression and purification of human H1.5.....	140
Figure 4.12 Recombinant expression and purification of human H1t	141

Figure 4.13 Recombinant expression and purification of human H1T2	143
Figure 4.14 Recombinant expression and purification of human H1LS.....	144
Figure 4.15: 6% PAGE EMSA of linker histones binding to 169s nucleosome	145
Figure 4.16: Crystal images for the different linker histone-nucleosome complexes ...	147
Figure 4.17: Molecular replacement and preliminary refinement statistics of the H1T2 data set	150
Figure 4.18 Superimposition of the two models described in figure 4.17	151
Figure 4.19 Extra electron density consistent with bound linker histone in the H1T2-chromatosome crystal.	152
Figure 4.20 Extra electron density consistent with linker histone binding in the H1T2 chromatosome.....	153
Figure 4.21 Close view of H1T2 electron density in the chromatosome crystal	154
Figure 4.22 Model building trials for the H1T2 chromatosome	155
Figure 4.23 A Representative figure for the accommodation of globular domain of a linker histone [obtained from published pdb] into the extra density regions of H1T2-nucleosome data.....	156
Figure 4.24 Extra electron density consistent with linker histone binding in the H1LS chromatosome.....	158
Figure A 1 Clustal omega multiple sequence alignment of full-length human linker histone proteins.	191
Figure A 2 Secondary sequence element prediction of human H1T2 protein using GOR tool from Expasy	197
Figure A 3: A Pymol representation of a 3D structure prediction of human H1T2 protein using CASP (Moult et al., 2014).....	198
Figure A 4: Close view of the Electron density map to validate the appropriateness of using a paired-nucleosome model.....	199
Figure A 5: Close view of the Electron density map to validate the appropriateness of using one nucleosome model	200
Figure A 6: Close view of the Electron density map to validate the appropriateness of a modified paired-nucleosome model 2.....	201

List of Tables and Flowchart

Table 2.1 Affinity values of FOXA1 binding to different DNA probes	35
Table 2.2 List of possible combinations of the DNA sequences for the ‘RYMAAYA” FOXA1 cognate DNA binding site.	67
Table 3.1 Amino acid composition and biochemical properties within the HMGN family of proteins	82
Table 4.1 Data collection statistics for chromatosome crystals composed of different linker histone variants	148
Table 4.2 Summary of data processing results.....	148
Flowchart 2.1 Summary of the different approaches aimed at optimizing the bacterial expression and purification of human FOXA1 protein	52

Abbreviations

EM – Electron microscopy

NMR- Nuclear Magnetic Resonance

6x His – hexa histidine tag

FOXA – Forkhead box A protein

CTD – C-terminal domain

DBD – DNA binding domain

DNA – Deoxyribo nucleic acid

EMSA – Electromobility shift assay

H1.0, H1.1, H1.2, H1.3, H1.4, H1.5, H1x, H1oo, H1t, H1T2, HILS – Linker Histone variants

HO – Histone octamer

HMGN – High mobility group N

NCP – Nucleosome core particle

NTD – N-terminal domain

NUC – Nucleosome

PAGE – Polyacrylamide gel electrophoresis

SPR – Surface plasmon resonance

TD – Transactivation domain

TAD - Topologically Active Domain

Abstract

In the eukaryotic cell nucleus, genetic material is packaged into chromatin, consisting of arrays of nucleosomes that can be densely compacted. The three-dimensional organization of chromatin imparts a sophisticated temporally and spatially controlled regulation of accessibility to the genes. Globally, the chromatin landscape is compartmentalized into chromatin territories and further into topologically active domains. Locally, gene regulation events can be triggered by various factors including interactions of enhancer/silencer with promoter regions, role of chromatin architectural proteins that enhances gene accessibility. Histone and non-histone chromatin architectural proteins can recognize structural features of nucleosomes, which are the fundamental units of chromatin. In general, the binding of linker histone proteins to the linker DNA between nucleosomes facilitates the condensation of chromatin into higher order structures. Other chromatin architectural proteins, such as those in the FOXA and High Mobility Group N [HMGN] families, can interact with nucleosomes and help to maintain open and active chromatin states by competing with linker histone activities. Studying such interactions could provide a new perspective for understanding chromatin structure and activity, as well as shed light on the bases of developmental defects and cancer-related disorders. X-ray crystallography has provided a powerful platform to study biological macromolecules in great atomic detail. In this dissertation work, the major aim was to utilize X-ray crystallography to study the nucleosomal interactions of chromatin-interacting factors, to address the two opposing aspects of chromatin regulation.

The Fox family of proteins with a conserved winged-helix DNA-binding domain like the globular domain of the linker histones, recognizes a signature DNA sequence with an off-dyad position on the nucleosome. Owing to its highly unstable nature, recombinant bacterial expression and purification of full-length human FOXA proteins has proven especially challenging. We pursued random truncations and have purified a library of FOXA1 truncated protein constructs that could serve as controls for biochemical characterization. We have addressed the soluble purification of full-length FOXA1 using a maltose binding protein (MBP)-tag. Alternatively, we have purified a shorter, functionally relevant construct, lacking the N-terminus. Both constructs demonstrated sequence dependent binding to

cognate FOXA1-DNA. However, in contrast to previous work, stable binding to nucleosomes was not apparent, thereby warranting further optimization for *in vitro* binding studies.

Members of the HMGN subfamily of proteins, capable of dynamic interactions with chromatin, display a cooperative two-to-one stoichiometry of binding to the acidic patch of the nucleosome. Despite the highly conserved nucleosome binding domain between variants, under physiological conditions, the two HMGN molecules associated with a given nucleosome are always of the same variant type. We have optimized and purified all but one of the human HMGN variants in large quantity using recombinant bacterial expression systems. We have screened conditions to crystallize the complex of HMGN1 or HMGN2 with nucleosomes and obtained crystals that diffracted X-rays up to 7 Å resolution. Optimization of crystallization and crystal stabilization conditions could further improve the diffraction quality.

Different linker histone variants display distinct gene-modulating activities and localization on chromatin. Given our recent success in our lab solving the structure of an H1x-nucleosome assembly [Adhireksan et al., unpublished data], for this dissertation work we purified the human H1.5, H1t, H1T2 and HILS linker histone variants and crystallized the four nucleosome-linker histone assemblies. X-ray diffraction data sets for the different complexes range from 3.0 to 3.3 Å resolution. Although in some cases electron density maps indicate clearly the presence of bound linker histone, the degree of disorder in the linker histones is substantial. We propose that further experiments utilizing heavy atom labeling and phasing approaches should enable the solving of atomic models for these four different nucleosome-linker histone assemblies. This could reveal novel aspects of variant-specific linker histone structure and activity.

Here we have focused largely on developing structural biology platforms that could aid in the characterization of nucleosome-protein interactions with atomic detail. The above projects provide useful insights into purification of highly disordered human proteins, nucleosome assembly crystallization strategies, techniques to improve diffraction and strategies to address the challenges involved in solving large protein-DNA complex structures.

Chapter 1 Introduction

The mammalian cell nucleus, seemingly random, has a defined compartmentalization of nuclear bodies and displays a non-random genome organization (Medrano-Fernández & Barco, 2016) (Cattoni et al., 2015). The spatial and temporal regulation of the genome is an important aspect during development and for various cellular processes. Disruption of this architecture has major consequences in development and coincides with hallmarks of certain disorders. During a cell cycle, a cell spends most of its time in the interphase stage where the majority of transcription, replication and gene regulatory events takes place. In this phase the chromatin is relatively spread out and organized into chromosome territories, a complex domain with varied levels of condensation to dictate gene regulatory events (Cremer et al., 2006). Each chromosome territory is sub-compartmentalized into topologically active domains [TADs] that demonstrate frequent interactions within the domains (Dixon et al., 2012; Nora et al., 2012). The partitioning of the TADs influences gene expression patterns and their genomic distribution vary with cell-types and during development. The other minor phase [2 hours] of the cell cycle is mitosis, where all the genetic material is compacted and re-organized into highly condensed chromosomes [Figure 1.1]. As early as 1928, Heitz had categorized the chromatin segments broadly as euchromatin and dense heterochromatin based on their differential ability to stain (Heitz, 1928) (Fedorova & Zink, 2008). In addition to their characteristic density, they are defined molecularly based on post-translational modifications [such as acetylation, methylation patterns] on histones. In specific cell types, there seems to be a defined positioning of chromatin regions with respect to the nuclear center or periphery. Heterochromatin domains, both the constitutive and facultative types, are typically found near the nuclear or the nucleolar periphery while some in the chromocenters can be found in the interiors. The packaging of 2 m of DNA into the nucleus of about 6 μm in diameter and the spatial organization is a fascinating aspect of the eukaryotic cell. The occurrence of varying degrees of condensed genetic material inside the cell is locally and globally influenced by the cell type, functions, cell cycle, and stages of development.

Figure 1.1 Chromatin state distributions during the different stages of the cell cycle.

During the longest phase interphase, most of the chromatin is in the less condensed euchromatin state and the rest exists as dense heterochromatin. During mitosis, it takes the most condensed form of chromosomes.

The basic packaging involves wrapping of about 145 bp of DNA onto a histone octamer core to form the nucleosome core. The nucleosome core with flanking linker DNA regions on either side constitutes the nucleosome. The arrangement of several nucleosomes at a stretch displays a beads-on-a-string appearance, that has been observed under low salt conditions using an electron microscope (Olins & Olins, 1974). The linear array can fold and compact to varying degrees to form higher order chromatin structures [Figure 1.2]. This process is aided by linker histones and facilitated by inter-nucleosomal interactions.

Figure 1.2 Hierarchical packaging of DNA into chromatin

DNA in eukaryotes is wrapped into nucleosomes that is further condensed with the help of linker histones to fold into higher order dense structures. Figure adapted from Pierce Benjamin, *Genetics: A conceptual Approach*, Nature Education, 2013 (Pierce, 2013).

A brief account on chromatin history

The discovery of the basic components of chromatin dates to the late 1800s with the identification and the crude extraction of ‘nuclein’ and ‘protamin’ by Miescher, and subsequently of ‘histon’ by Kossel. By the year 1890, Flemming coined the term chromatin, an equivalent of nuclein. The book, ‘Chromatin’ by Van holde, 1989, provides a summary of ‘the first hundred years’ in the history of chromatin research. In the early 1900s, Huiskamp’s electro-mobility studies could isolate the acidic and basic components of chromatin based on their migration towards anode and cathode, respectively. Following skepticism owing to the changing nature of

chromatin during cell division [Figure 1.2], decades of long debate eventually lead to the final acceptance of nucleic acids as the heredity material. Further, with the structure of DNA being solved by Watson and Crick (Watson & Crick, 1953), the focus tilted towards elucidating the function of chromatin. How the gene is decoded and controlled in a time and tissue specific manner remains an intriguing question. In a pioneering study, (Stedman & Stedman, 1951) had separated the core histones and the lysine rich histone (linker histone) components of chromatin. A remarkable study by Steiner involved a low salt extraction procedure of chromatin with an intact 1:1 ratio of DNA to protein (Steiner, 1952). Further, he demonstrated that high salt conditions resulted in the loss of protein components and transformed it into an extended form. The advent of chromatographic and gel electrophoresis techniques came as a boon to the chromatin field. Allfrey & Mirsky showed that the addition of histones inhibited RNA synthesis from a chromatin template, thereby establishing their role as gene regulators (Allfrey & Mirsky, 1964). Histones were sequenced in the 1960s. A breathtaking conservation of histones across different species brought them to the spotlight. Several diffraction experiments and electron microscopy studies ensued to understand the structure of chromatin (Holde, 1989).

Gene regulation in chromatin context

There have been several ongoing studies to understand the structural intricacies of chromatin packaging and the 3D chromatin architecture. It raises an important biological question as to how the various gene regulatory factors specifically recognize the well buried DNA sequence in a timely manner (Luger et al., 2012). This aspect of gene regulation is important during development, and for regulation of various biochemical pathways. A complex interplay of spatial organization of the genome, the role of architectural proteins, epigenetic marks in different cell types facilitate gene regulation. A genome wide mapping of chromatin interactions in a differentiating embryonic stem cells revealed extensive chromosome reorganization resulting in altered active and inactive chromosomal regions in different lineages (Dixon et al., 2015). In the interphase cell nucleus, the chromosome is partitioned into topologically activating domains, referred to as TADs (L. Li et al., 2015) (Bonev & Cavalli, 2016). The TAD domains are conserved between different cell types and across species. These are further sub-

compartmentalized in up to six regions that can be re-arranged and demonstrate a cell-type dependent variation (Rao et al., 2014). Active genes and chromatin architectural factors are present between the borders of TAD regions. It has been demonstrated that during a temperature-induced stress in *Drosophila* resulting in repression of most genes, there is a significant rearrangement of the TAD borders to its interiors (L. Li et al., 2015). Occurrence of gene clusters in the human genome presents a role of chromatin architecture in facilitating gene events during development or in evolution and mis-regulation of these clusters have resulted in disorders (Lercher et al., 2002) (Gilbert et al., 2004). There are millions of *cis*-regulatory regions that are intergenic distal regulatory elements that can act over a distance from the promoter elements to activate a particular gene (Consortium, 2012) (Schmitt et al., 2016) (Ulianov et al., 2015). They act via chromatin looping and the gene regulatory process is further facilitated by the binding of pioneer factors and other architectural proteins (Fitzpatrick et al., 2015). Enhancers are one such *cis*-regulatory region, generally marked by H3K4Me1 and H3K27Ac, which show cell-type specific gene expression based on the methylation status (Plank & Dean, 2014). The spatial juxta positioning of enhancers to promoter regions favors physical interaction between the otherwise distal regions to trigger gene events (Y. Zhang et al., 2013) (Ulianov et al., 2015). Additionally, there are factors that can bring about local structural changes in chromatin and facilitate enhancer promoter interactions. Linker histone family of proteins mediate the condensation of chromatin to varying degrees. There are chromatin architectural proteins that can mediate gene regulatory events by recognizing the basic structure of chromatin, the nucleosome, and further triggering chromatin remodeling events. (Cubeñas-Potts & Corces, 2015).

Implications in cancer

Increased cell proliferation is a hallmark of cancer. Cancer research has focused on the expression and repression of specific genes. Current cancer perspective has diversified to address the three-dimensional landscape of chromatin that further influences the transcriptional profile of the genes (Wijchers & de Laat, 2011). Several genome wide mapping studies are directed towards understanding aberrant gene regulation at the chromatin architectural level. The advent of chromosome

capture techniques in recent times have enhanced the scope of our understanding of chromatin conformation in cancer (Jia et al., 2017). Mis-regulation of gene clusters has been implicated in different cancers and the global positioning of such chromosomal defects could help identify target oncogenes (Gilbert et al., 2004) (Caron et al., 2001). Gene translocation events occur in many different types of cancers (S.-H. Song & Kim, 2017). *c-myc* and *Igh*, that are prone for translocation in B-cell lymphomas, preferentially occur in the same RNA pol II mediated transcription factory within interphase nuclei (Osborne et al., 2007). Continuous DNA damage can alter the copy number of closely juxtaposed oncogenes leading to tumorigenesis (S.-H. Song & Kim, 2017). However, the exact mechanisms remain unknown. The inefficient DNA repair mechanism following a double stranded break could lead to mutations and chromosomal translocations that lead to cancer (Misteli & Soutoglou, 2009). DNA repair sites are marked by some histone modifications that are further recognized by chromatin remodelers. The spatial organization in the nucleus is important for the assembling of the necessary factors including DNA damage repair proteins for the recognition of histone variants with post translational modifications, recruitment of chromatin remodelers and downstream effector molecules for the DNA damage response (Downs et al., 2007). Several chromatin architectural proteins that mediate cellular events by recognizing the genome in the 3D chromatin landscape in a timely manner play a crucial role in development and cancer. There are factors that facilitate loop formation between enhancer or silencer regions with promoter regions involved in cancer-related genes. Such interactions and gene fusion events, resulting in increased cell growth and mis-regulated expression, have been implicated in prostate cancer, breast cancer, in leukemia and in colorectal cancers (Jia et al., 2017). Different cancers are associated with specific gene mutations, which in turn depend on local chromatin accessibility and histone modifications for each particular cell type (Polak et al., 2015). Understanding the mechanistic details of these interactions will provide insight into modulation and conformational changes of chromatin structure. This is supporting the emerging view of chromatin structure as a potential drug target (G. E. Davey & Davey, 2008) (Palermo et al., 2016).

Overview of the major techniques in structural biology

Three-dimensional structure information of biological macromolecules has paved way for understanding interactions, elucidating the mechanism and in drug development. In a broad sense X-ray crystallography, NMR and Electron microscopy have been the three major approaches with additional techniques including SAXS, neutron diffraction, FRET that have been useful in particular cases (Egli, 2010). With their own set limitations, the three main techniques can complement each other in elucidating molecular interactions of proteins, ligands and nucleic acids. Since the initial success stories of obtaining structural information on myoglobin (Kendrew et al., 1958), hemoglobin (Perutz et al., 1960) molecules, followed by the structure of DNA (Watson & Crick, 1953), X-ray crystallography has been emerging as one of the most successful techniques in solving structures of macromolecules. To date, in the protein data bank, there are more than 120,000 structures solved by X-ray crystallography [RCSB/PDB; <http://www.rcsb.org>; 122,922 structures as of February 4, 2018 (Berman et al., 2000)]. X-ray crystallography principle involves diffraction experiments performed on a regular arrangement of molecules, as in a crystal. The regular arrangement of atoms in a lattice intensifies the otherwise weak diffraction spots. Unlike light or electron beams, X-ray beam cannot be focused by any lens and the generation of three-dimensional image relies on mathematical principles. Although obtaining the amplitudes or the structure- factor information has been relatively easy, there is the lingering issue of losing the phase information that has to be addressed by additional means. Obtaining good quality diffracting crystals and the phase problem has been the two major limitations with X-ray crystallography (Egli, 2010). In most favorable cases, X-ray crystallography can help obtain a high-resolution structure and can uniquely provide extremely valuable structural information including solvent mediated interactions to great atomic detail. Examples of this achievement include, the 3D structures of F1 ATP synthase (Abrahams et al., 1994), photosynthetic reaction center (Ermler et al.), ribosome subunits (Ban et al., 2000) (Wimberly et al., 2000) (Schlunzen et al., 2000), proteasome (Löwe et al., 1995) and nucleosome (Luger et al., 1997).

In cases where X-ray crystallography approach proves difficult, electron microscopy has been a favored technique to solve complex structures, especially of large macromolecules >100 kDa. EM does not pose the phase problem and has seen a major breakthrough in development and has been successful in solving complex structures of large macromolecules, virions, ribosomes, membrane proteins (Egli, 2010) (Z. H. Zhou, 2011). However, the sample preparation, exposure to high intensity electron beam, data collection and processing has been challenging. At lower resolutions, this method is dependent on previously existing models of components derived from crystal structures or NMR models of different segments or domains. NMR spectroscopy is a powerful technique that can help characterize the conformational dynamics of proteins and nucleic acids [mostly of lower molecular weight typically <40 kDa] in solution to amino acid level in response to its environment (Feng et al., 2011). It does not require prior existing models and can directly determine the secondary structure elements. However, the resolution is often effectively less than that of X-ray crystallography. For a given data set, NMR techniques can elucidate multiple conformations and provide an ensemble of models and predictions of conformational freedom based on the degree of probability. The three powerful techniques can therefore complement each other to address the disadvantages of one technique over the other.

Structural biology [X-ray crystallography] initiative in chromatin research

Nucleosome structure and dynamic interactions

The crystal structure of the nucleosome core particle [Figure 1.3] has been solved and it reveals important specific contacts of DNA with the histone octamer. The first X-ray crystallography structure was solved by Luger et al., 1997. A 1.9Å higher resolution structure revealed additional solvent mediated contact information. (C. A. Davey et al., 2002; Luger et al., 1997). About 147 bp of DNA is wrapped $1\frac{2}{3}$ times, in a left-handed super-helical manner, around the histone octamer. Following the first crystal structure of the nucleosome core particle, there are about 150 PDB entries (Korolev et al., 2018) of NCPs containing histones from different organisms, histone variants, different factors or nucleosome arrays. A strongly positioning '601' Widom sequence has dominated the field for *in vitro* nucleosome studies

(Lowary & Widom, 1998). A 2.6 Å crystal structure of a nucleosome containing native MMTV promoter region has been recently reported (Frouws et al., 2016). In addition to the nucleosome structures with the highly conserved canonical histone, structures with histone variants has shed light on the variations in nucleosome structures. Crystal structure with H2A.Z (Suto et al., 2000) reveals an extended acidic patch to the nucleosome surface, and a crystal structure of CENP-A containing nucleosomes reveals the role of additional residues present in a loop region in stabilizing centromeric chromatin (Tachiwana et al., 2011). In addition to several crystal structures, recently there are cryo EM determined near atomic resolution structures of NCPs or nucleosome with linker histones (E. Y. D. Chua & Sandin, 2017) (Liu et al., 2017; Wilson et al., 2016).

Histone octamer core and the acidic patch

Two copies of each of the four core histones (H2A, H2B, H3 and H4) are assembled into an octamer. The core histones have a characteristic histone fold comprising of 3 alpha-helices and 2 loops. The folding into octamer involves a sequence of events with the formation of H3:H4 tetramer, followed by the sequential association of 2 H2A:H2B dimers (Dyer et al., 2004).

Around the interface of the H2A:H2B dimer, on either face of the nucleosome, is the acidic patch, which arises from the clustering of 6 H2A and 2 H2B aspartate/glutamate residues [Figure 1.3 and 1.4]. The acidic patch is considered to be important for inter-nucleosomal contacts (Luger et al., 1997) (J. Y. Fan et al., 2004), and it constitutes the most ubiquitous target for interactions with different chromatin factors (Kalashnikova et al., 2013). Structural studies of LANA [Kaposi's sarcoma herpes virus (KSHV) latency-associated nuclear antigen] peptide, RCC1 [Regulator of chromosome condensation 1], Sir3 [silent information regulator] and HMGN2 proteins have identified their mode of binding to the acidic patch. Mutations that interfere with this acidic patch binding abolishes the protein's potential to recognize the nucleosome.

Figure 1.3 Structure of the nucleosome core particle (C. A. Davey et al., 2002)

The 1.9 Å resolution NCP structure (C. A. Davey et al., 2002). The top panel shows a cartoon representation of the DNA wrapped around the histone octamer in two different views [H3 in blue, H4 in green, H2A in yellow and H2B in pink]. The bottom panel shows the same two different orientations of the NCP as above, but with the histone octamer rendered in space filling representation with electrostatic potential [red, electronegative; blue, electropositive]. Red arrows indicate the acidic patch regions on either side of the core.

Figure 1.4 Interactions of different proteins with the acidic patch.

The peptide regions of LANA (green), RCC1 (yellow), Sir3 (cyan) and HMGN2 (black) that interact with the acidic patch are overlaid together. The red shaded zone is the nucleosome acidic patch. NCP is shown in surface view. Figure adapted from (Kalashnikova et al., 2013).

Crystallography has been a platform to study the atomic details of some of the interactions and the minor changes brought about in the structures of nucleosomes (McGinty & Tan, 2016). LANA peptide is shown to form a hairpin structure, and fits into the hydrophobic elements of the acidic patch and makes specific interactions with the neighboring charged residues (Barbera et al., 2004). A 2.9 Å crystal structure of RCC1 (R. D. Makde et al., 2010) with the nucleosome reveals the great extent of charged interactions with the acidic patch together with an essential specific arginine-acidic patch contact. A 3 Å crystal structure of Sir3 BAH domain with the nucleosome (Armache et al., 2011) revealed the acidic patch mediated interactions that bring about some structural changes in the nucleosome as well as the BAH domain itself (McGinty & Tan, 2015). Although these proteins, with distinct biological roles and functions, target the same acidic patch of the nucleosome, there are distinctions in the mode of binding that could mediate variations in the binding response [Figure 1.4].

Histones as drug target

The structure of naked DNA is very different from that in the form of a nucleosome. Compared to drug binding to a free DNA that can adapt to varying configurations easily, nucleosomes are rigid and have a defined configuration. Therefore, by specifically targeting the histone component, alterations to the nucleosome could interfere with chromatin binding proteins, thereby affecting gene regulation process. Alternatively, drug targeting of specific histone modifications or the tails may interfere with internucleosomal interactions, which in turn is important for the maintenance of chromatin structure. Certain Ruthenium based compounds with anticancer properties associate with specific DNA targets on chromatin and can form histone adducts (Adhireksan et al., 2014) (B. Wu et al., 2011). A recent study from our lab (Adhireksan et al., 2017) shows that an allosteric mechanism is possible whereby the binding of one drug, causing histone adduct formation, could modulate the binding of another drug at a distant location on the nucleosome.

Strategy for crystallizing a complex of nucleosome with chromatin factors

Crystallization of nucleosome with complexes involves several crucial steps (D. Makde & Tan, 2013). The chromatin factor under study must be prepared in large amounts (~5-10 mg) with high purity (>98%). For the sake of crystallization, high concentrations of these purified proteins must be achievable. There are often issues due to aggregation, precipitation and salt mediated effects. The protein factors of interest must be capable of forming stable complexes with the nucleosome. Depending on the purity of the protein preparation and feasibility of purifying and concentrating the complex, different strategies of making complex must be screened and considered. The choice of DNA and the length of the DNA ends in the nucleosome plays an important role in the formation of crystal contacts. (D. Makde & Tan, 2013), (McGinty & Tan, 2016). It is usually better to screen for varying ends for a factor to be able to form complex and to obtain good crystals. Further a broad screen of crystallization conditions for buffer pH and ionic strength should be pursued to maintain the complex and to obtain good three-dimensional crystals. Additives, especially divalent metals, are added considerations for obtaining well diffracting crystals. Further, an exhaustive screen for the choice of harvesting

buffers and post crystallization dehydration soaks using cryoprotectants is crucial for improving the diffraction quality of the crystals (Adhireksan et al., unpublished data; D. Makde & Tan, 2013).

Scope of this research

Chromatin's three-dimensional landscape provides a plethora of combinations for chromatin binding factors to recognize the varying degrees of condensed states. We are far from being able to generalize the governing principles of this seemingly complex process. Currently, genome wide studies and cancer research are oriented towards the role of epigenetic marks in the accessibility of chromatin, and chromatin as a drug target is a newly emerging theme. Structural biology initiatives could provide snapshots of the factor-chromatin interactions in an isolated manner from an otherwise crowded heterogeneous setting. Studying the unique and diversified mechanism of interaction of various chromatin factors, role of histone variants and the post translational modifications with the nucleosomes could shed light on a general mechanism of gene regulation in the chromatin context. These in the long run can provide insights into screening for drugs to harvest the therapeutic potential of targeting nucleosomes for various disorders.

The work presented here is aimed at characterizing the interactions of three factors, which have chromatin architectural functions, with nucleosomes by X-ray crystallography approaches. For this, we have considered 2 family of proteins that can bring about gene regulation events: 1) FOXA family of proteins that recognizes signature sequences on the nucleosome at specific positions and has an important role during development; 2) HMGN family of proteins that remain associated with the nucleosome in a cooperative fashion under physiological conditions. The other linker histone family of proteins bring about varying degrees of compaction of nucleosomes to pack them into chromatin. Accordingly, the remainder of the dissertation has been organized into three chapters that provides details of the literature and research work: 1) FOXA pioneer factors, 2) HMGN chromatin-unfolding proteins and 3) Linker histone chromatin-compacting proteins.

Chapter 2 Interactions of FOXA family of proteins with the nucleosome

Introduction

Early studies to understand gene regulation events had identified different factors associated with *cis* regulatory elements at various stages of liver development. The *ALB1* gene is the first to receive the earliest signals for the expression of liver specific genes, such as Transthyretin and alpha-1 antitrypsin. *In vivo* foot printing experiments identified several factors bound to the *ALB1* gene from the liver bud developmental stage onwards (Gualdi et al., 1996; McPherson et al., 1993). However, at the upstream endoderm stage, there were only 2 factors, FoxA1 and GATA1, found to be associated with the then silent *alb1* enhancer region (Bossard & Zaret, 1998). FoxA1 and A2, when recruited, can either displace or reposition nucleosomes to make them competent for transcription events to ensue (Lee et al., 2005). The *Drosophila* fly mutant of *FKH*, homolog of FoxA, has displayed a phenotype with fork head like structures (Weigel et al., 1989) and hence this family of proteins are referred to as Forkhead box or FOX proteins (Lai et al., 1991).

FOX family of proteins

Forkhead Box family of proteins are functionally and evolutionarily conserved from fly to mammals. In humans, FOX family of proteins constitutes 19 subtypes and comprises of 50 members. All the added family members have a highly conserved DNA binding domain [DBD], also referred to as a winged-helix motif structure (Lalmansingh et al., 2012). Based on phylogenetic analysis, these were classified into many subgroups. In 2000, the naming convention was standardized as Fox/FOX_subgroup_number, as in FOXA1 or FoxA1, for humans or mouse proteins, respectively (Kaestner et al., 2000). A list of all of the Fox family of proteins with their nomenclature is categorized (McPherson et al., 1993) at the website, <http://biology.pomona.edu/fox/foxbyspp.html>.

FOXA subfamily

The FoxA subfamily consists of 3 members, FoxA1, FoxA2 and FoxA3. In mammals, FoxA1 and A2 are expressed in the foregut endoderm and FoxA3 is majorly expressed in the midgut and hindgut endoderm. FoxA2 is present in the

primitive streak and is essential for node and notochord development that further forms neural tissues (Zaret, 2008) (Zaret et al., 2008). Mouse double knockdown models of FoxA1 and A2 results in embryonic lethality, and the one for FoxA3 has severe defects (Zaret, 2002). The Drosophila Fkh DBD shares over 90% identity with that of human FoxA proteins. Another FoxA homolog in *C. elegans*, *PHA4*, with over 75% sequence identity, is required for gut development. The DBD among the different FoxA family members has over 90% sequence identity [Figure 2.1]. These proteins recognize a DNA sequence pattern on the surface of the nucleosome (Z. Li et al., 2011). In addition to this, they all possess N- and a C-terminal transactivation domains. (Zaret, 2002). The great evolutionary conservation highlights the prominent role of these proteins during development. In addition to their role in liver development, the FoxA family of proteins can recruit factors such as steroid hormone receptors, GR, ER, AR, PR (Lam et al., 2013) to the target sites and have other biochemical roles.

Role of FoxA proteins in cancer

Mutations and improper regulation of FoxA proteins, mostly associated with the DBD, have been implicated in hormone dependent cancers and their response to therapy. These mutations exist in about 1.8% of breast cancers and 4% of prostate cancers (Lam et al., 2013). Further, these proteins can act as tumor suppressors or as oncoproteins. For example, overexpression of FoxA proteins has been observed in oesophageal, lung, prostate and breast cancers; loss of FoxA protein expression has a role in bladder cancer and with tumor proliferation (Lam et al., 2013). In addition to being transcriptional activators, FoxA1 and A2 can interact with Groucho related gene factors [Grg1 and Grg3] and recruit other co-repressors to the FoxA target genes (Santisteban et al., 2010).

Interactions of FoxA proteins with chromatin

Mitotic expression of FOXA1 and epigenetic signatures

During mitosis, transcriptional regulation is inhibited and most of the gene bound factors are moved out of the nucleus. However, there are certain factors which continue to remain associated with chromatin and are thought to be involved in epigenetic regulation. FoxA1 is one such factor, which associates with the highly

Figure 2.1 Multiple Sequence Alignment of the DBD of FOXA family members

The DBD of human FOXA proteins is highly conserved with over 90% sequence identity between members. **A.** Clustal Omega (Sievers & Higgins, 2014) multiple sequence alignment of the DBD of human FOXA1, FOXA2 and FOXA3 proteins. **B.** Tabulated percentage identity matrix comparing the DBD of two FOXA members is represented as the percentage score [last column]. The prefix ‘h’ of hFOXAx [x is 1 or 2 or 3] DBD indicates human protein. **C.** Weblogo3 (Crooks et al., 2004; Sharma et al., 2012) representation of the conserved DBD residues. The height of the letters represents the level of conservation.

mitotic cell, GFP tagged FoxA1 is visualized on mitotic chromosomes. This was further verified by its co-localization with an antibody directed to the mitotic marker, H3-phosphorylated on Serine 10 (Zaret et al., 2008). Additionally, FoxA1 can preferentially bind to DNA hypomethylated sites, triggering H3K4 methylation at these sites. Differential methylation patterns dictate the cell type specific recruitment of FoxA1 to the enhancer (Serandour et al., 2011). Thus, the epigenetic patterns can dictate the association of FoxA1 with the enhancer region.

Figure 2.2 FoxA1 is associated with mitotic chromosomes in nocodazole-arrested cells.

Compared to the controls of GFP with a Nuclear Localization Signal [NLS] or the C/EBP, GFP-tagged FoxA1 is associated with the mitotic chromosome in a nocodazole-arrested cell. **A** and **B** are negative controls showing no co-localization of GFP-NLS and GFP-C/EBP with mitotic chromosomes. **C**. Association of GFP-FoxA1 with mitotic chromosomes. **D**. Confirms the observation in panel C based on the co-localization [yellow arrow] of GFP-FoxA1 with the mitotic marker, phosphorylated Serine 10 on H3. Figure adapted from (Zaret et al., 2008).

Nuclear mobility

FRAP [Fluorescence Recovery After Photobleaching] experiments measure the time of recovery of GFP tagged nuclear factors following photo bleaching of the entire or partial regions of nucleus *in vivo* (Phair et al., 2004; Sekiya et al., 2009; Zaret et al., 2010). This duration of time spent for the recovery of GFP can help monitor the mobility of the factors within the nucleus. Hence it serves as an indirect measure to predict the nature of transient non-specific interactions of the factors with chromatin. The faster the movement, the lesser is the interaction with chromatin. For most transcription factors, this is in the range of seconds. Comparative studies employing FRAP of the various nuclear factors considered reveals that in sharp contrast to a rapidly moving GFP alone, FoxA1 was faster than

linker histones and HMGN1, but slower than most other factors such as c-myc. The intermediate slower mobility of FoxA1 suggested a high affinity for nucleosomes relative to the other factors considered.

FoxA1 interaction with nucleosome on alb1 enhancer region

The silent Alb1 enhancer region comprises 3 nucleosome-like particles: N1, N2 and N3 (McPherson et al., 1993). The N1 region contains high and low affinity sites for FoxA1 namely, eG and eH, respectively. Later, another binding site NS-A1 was identified (Cirillo et al., 1998) [Figure 2.3]. A hydroxyl radical foot printing assay, offering an increased sensitivity for tighter interactions, was used to map the interactions between FoxA1 DBD and the eG and eH sites. FoxA1 can interact with both DNA strands with a slightly better affinity [3-4 times] for the eG sequence. Replacing the possible binding sequence of eG with eH reduces the affinity by 30-50%. Binding to eG occurs at an earlier stage and it is possible that the protein uses affinity based interactions to mark the different stages of progression towards liver specific gene expression (Cirillo & Zaret, 2007).

Figure 2.3 Summary of FOXA binding sites on the Alb1 enhancer region. HNF3 is the alternative name for FOXA. 472 to 651 is the region of the Alb1 enhancer. eG and eH are the major binding sites of FOXA. NS-A1 is an additional FOXA binding site. Figure adapted from (Cirillo et al., 1998).

FoxA proteins- classified as a pioneer factor

Although FoxA1 can bring about the transcription of liver specific genes, it is referred to as a ‘pioneer factor’ and not as a transcription factor for the following reasons. Transcription factors [TFs] bind with a higher affinity to double-stranded DNA and do not have significant affinity for nucleosomes (Adams & Workman, 1995; McNally et al., 2000; Phair et al., 2004). They are usually not associated with the silent enhancer elements. TFs depend on prior chromatin remodeling activity by remodeling factors to recognize their exposed target region. On the other hand, FoxA1 can bind with a higher affinity to nucleosomes (Cirillo & Zaret, 1999).

FoxA1 has been shown to be able to relieve Afp [α -fetal protein]-mediated repression prior to the onset of liver development without the help of any chromatin remodelers (Cirillo et al., 2002). Association of FoxA1 with the nucleosome has been observed by both *in vitro* and *in vivo* studies (Chaya et al., 2001; Cirillo & Zaret, 1999; Sekiya et al., 2009).

Competitor binding assays with ^{32}P -labelled dinucleosomes containing albumin enhancer sequences were used to study FoxA1 binding interactions (Cirillo & Zaret, 1999). FoxA1 complexed with dinucleosomes or with free DNA were treated with excess unlabeled DNA and then assayed for DNase hypersensitivity at different time points. Their results revealed a stable association of the FoxA factor with nucleosomes, indicated by the presence of protected and hypersensitive regions even after longer duration (~9 min compared to only 20 s in the case of complex with free DNA). After its binding, FoxA1 can open compacted nucleosome arrays [either salt-induced or linker histone mediated (Cirillo et al., 2002; Zaret et al., 2008). This can facilitate the recruitment of other factors required to initiate transcription based events. MNase assays indicate that following the binding of the FoxA1 to the N1 region, there is re-positioning of nucleosomes in this region. Also, the region shows increased DNase and restriction enzyme sensitivity (Cirillo et al., 2002). Thus, the ability of the FoxA family of proteins to independently interact with and modify condensed chromatin and its ability to recruit other transcriptional factors, classifies them as a pioneer factor.

Interactions of FoxA proteins with DNA

DBD of FOXA3 resembles the globular domain of linker histones

Cross linking experiments involving bulk chromatin and nucleosomes with 5sRNA gene sequence DNA show FOXA has an asymmetrical binding to one side or edge of the nucleosome core (Hayes et al., 1994; Hayes & Wolffe, 1993) suggesting interactions with DNA and histones. Therefore, FoxA protein might demonstrate nucleosome binding properties like that of linker histone (Clark et al., 1993).

A co-crystal structure of the FoxA3 DBD with DNA (Clark et al., 1993) indicated that the DBD is a helix turn helix [HTH] structure flanked by 2 loops that gives an appearance of wings of a butterfly [Figure 2.3A]. Hence it is referred to as a

‘winged-helix’ structure. This structure resembles that of the linker histone (Ramakrishnan et al., 1993) [Figure 2.3B]. By interacting with major and minor groove, FoxA3 DBD associates on one side of the DNA. The HTH region makes base specific contacts with the DNA and the loop regions can interact with the phosphodiester backbone [Figure 2.4].

Figure 2.4 Winged helix structure of the DBD of FOXA3

A. FOXA3 DBD in complex with DNA [pdb id: 1VTN] shows a winged-helix structure. **B.** Structural comparison of the DBD of FOXA3 in complex with DNA [pdb id:1VTN] with that of the globular domain of linker histone H5 [pdb id:1HST].

Sequence specific and non-sequence specific interactions of FoxA with DNA

The DNA length used by (Clark et al., 1993) was insufficient to explain all the base specific interactions. (Cirillo & Zaret, 2007) studied specific interactions of the wing domains, that were missing in the crystal structure, with the eG and eH sites of varying affinities (Cirillo & Zaret, 2007). There was a reduction in affinity when the sequences were swapped. By SAAB analysis, (Overdier et al., 1994) identified a 7 bp core recognition site with 'RYMAAYA' [R = A/G; Y = C/T; M=A/C] sequence out of a 15 bp response element.

In vitro foot printing and FRAP analysis that measures nuclear mobility indicated nonspecific interactions of FoxA with nucleosomes (Sekiya et al., 2009). Mutation studies of the HTH and wing regions identified the residues responsible for the sequence and non-sequence specific interactions with the DNA and with the nucleosomes. With respect to nucleosome interactions, the rotational and translational position of the specific sequence could influence the affinity of the protein (Sekiya et al., 2009). The affinity for the D+2 position [D is the dyad axis position] was about 1.1 nM, which is slightly higher than the affinity of 1.8 nM with the consensus sequence close to the dyad. Though the native affinity of FoxA1 for its cognate sequences is higher, it can also have non-sequence specific interactions with the nucleosomes. This could rationalize the slower mobility observed *in vivo* by FRAP assays.

Kinetics- K_d measurements

Aung et al., 2014 employed EMSA, Fluorescent anisotropy method and gold nano particle based colorimetric assays for measuring MBP-tagged FoxA1 binding to DNA in a sequence dependent and independent manner (Aung et al., 2014). They considered 2 DNA probes containing FoxA1 cognate DNA sequences with different flanking regions. Another negative probe without the FoxA1 consensus sequence was used as a control [Table 2.1A]. Assuming that the preparation was 100% active, they had obtained K_d values. AuNP assay was the most sensitive one. EMSA and AuNPs could detect similar trends in differences between the two probes. However, different K_d values were obtained from different assays [Table 2.1B]. Therefore,

using SPR, a more sensitive method to obtain the k_{on} and k_{off} measurements would be a good strategy.

Table 2.1 Affinity values of FOXA1 binding to different DNA probes

A. The DNA probes used in the study. **B.** The corresponding K_d values of FoxA1 binding to DNA probes, obtained by 3 different methods. Table adapted from (Aung et al., 2014).

FoxA1 probe sequences.

A	Probe	Probe sequence ^a
	Probe 1	5'-GTACT <u>GTAATA</u> AAACT-3'
	Probe 2	5'-TGCCAAGTAAATAGTGCAG-3'
	Negative probe	5'-CACTT <u>TGTTG</u> CAAAGC-3'

^a FoxA1 consensus sequences are underlined.

B K_d (nM) values obtained by different methods.

Method	Probe 1	Probe 2
EMSA	47.82 ± 5.45	241.66 ± 5.24
FA	33.48 ± 10.76	84.67 ± 26.59
AuNPs	61.90 ± 17.40	341.10 ± 8.70

Interactions of FoxA1 with nucleosomes

FOXA enhances accessibility of nucleosomes at specific locations. A recent study (Iwafuchi-Doi et al., 2016) comparing MNase sensitivity of liver specific and ubiquitous enhancers establishes the retention of a DNase hypersensitive site of N1 nucleosomes at low MNase levels. Further, ChIP assays of cross linked chromatin confirms the presence of H2B and H3 at liver specific enhancer regions in contrast to active promoter sites. Thus, the liver specific enhancer harbors an accessible nucleosome that can be activated further for downstream events.

FoxA1 competes for linker histone binding site

The DBD of the FoxA3, as mentioned earlier, is analogous to the architecture of the globular domains of the linker histones. By MNase assay, (Cirillo et al., 1998) had identified another FoxA1 binding site, NS-A1, in addition to the eG and eH sites

[Figure 2.3]. This site also serves as a binding site for linker histones (Cirillo et al., 1998). It is believed that FoxA, which has a higher affinity for the sites, could compete with and displace the linker histones from the silent albumin enhancer during development (Cirillo et al., 2002; Cirillo et al., 1998).

Specific amino acid elements underlie the differential chromatin architectural properties of FoxA proteins and linker histones

Regions outside of the DBD define the chromatin interacting properties of FoxA proteins, and specific amino acid residues define the unique chromatin compaction property of linker histones. The C-terminal region is unique to FoxA1 and is different from linker histones. So, when the C-terminus of FoxA1 was attached to the linker histone's DNA recognizing region, it increased the otherwise slower nuclear mobility of GFP-tagged linker histone to approach that of FoxA1 (Sekiya et al., 2009). While linker histones cause the compaction of chromatin, FoxA1 causes opening of chromatin. The differences between the two have been attributed to the lack of the essential basic amino acid residues of linker histones (Goytisolo, Packman, et al., 1996) at fixed positions in the globular domain [Figure 2.5 and Figure 2.6]. *In vitro* studies of FoxA1 construct mutated to incorporate basic amino acids at these globular domain sites [at equivalent positions as in linker histones], caused compaction and triggered formation of chromatosomes (Cirillo et al., 1998).

The C-terminus of FoxA1 is important for interactions with nucleosomes

MNase and bead assays to study the binding interactions of FoxA1 truncation constructs with nucleosomes identified the C-terminal region along with the DBD to be important for chromatin binding effects. Notably, the C-terminus by itself or the rest of the construct without the C-terminus could not bring about this effect (Cirillo et al., 2002).

FRAP studies to measure nuclear mobility with N-terminal and C-terminal deletions of FoxA1 indicated that a C-terminal deletion mutant had increased mobility and an N-terminal deletion behaved similarly to the wildtype. Hence, the C-terminal could be more important for nucleosome binding. The C-terminal's potential interaction with core histones may interfere with the higher order chromatin packaging and create hypersensitive sites as observed by nuclease assays

Figure 2.6 Comparison of the DBD of FOXA3 with the globular domain of chicken H5 at specific residue locations.

Representation of the essential residues in the globular domain of H5 [pdb id: 1HST] and the corresponding residues in the DBD of FOXA3 [pdb id:1VTN]. Note that Ala 194 occurs in the FOXA3 structure (Clark et al., 1993), but FOXA1 has a Threonine at this position [also indicated in Figure 2.4].

Scope of this project

Studying FoxA interaction with nucleosomes and higher order chromatin structures could provide insight into the general mechanism associated with pioneer transcription factor activity and their mode of action in *cis*-acting regulatory events. Here, we aim to characterize the interaction of FoxA proteins with DNA and nucleosomes in order to understand the nature of DNA sequence dependent recognition in chromatin. FoxA proteins are highly disordered in solution and are unstable *in vitro*, and we have elucidated the contribution of different regions of the protein to stability and solubility using recombinant bacterial expression and purification.

Materials and methods

Core histones protein purification

Constructs

Human histones H2A, H2B, H3 and H4 with an N-terminal 6x histidine tag harboring a Thrombin cleavage site were courteously provided by Prof. H. Kurumizaka and by Dr. T.S. Kumaravel. The hH2A, hH2B and hH3 constructs were in pHCE vector, and were sub-cloned into pET28a, a kanamycin resistant plasmid. The hH4 construct was cloned into the pET15b vector.

Transformation

50 ng of plasmid containing a protein or DNA construct was transformed into 50 μ l of appropriate competent bacterial cells. The cells were incubated with the plasmid for about 10 min on ice, and were subjected to heat-shock treatment at 42°C for 45 seconds. After further incubation on ice for 5 min, the transformed cells were recovered in 1 ml of LB media at 37°C for 30 min. The successful transformants confer antibiotic resistance, and were positively selected for by plating the cell mixture onto LB agar plates containing appropriate antibiotic [100 μ g/ml of ampicillin; 50 μ g/ml of kanamycin].

Recombinant bacterial expression

Recombinant NT histidine tagged human core histone proteins were expressed in *E.coli*. H2A, H2B and H3 were expressed using competent BL21DE3 cells and H4 was expressed using JM109 bacterial strain. Large yield was possible by using an optimized auto induction protocol [Studier, 2005]. Positive transformants conferring antibiotic resistance were scraped off the agar plate and re-suspended in 10-20 ml of ZY media [1l containing 10 g of bacto tryptone, 5 g of yeast extract]. This was used to inoculate 500 ml of ZY media with 1mM MgSO₄, 1X NPS, 0.8% glucose and appropriate antibiotics. The volume was chosen depending on the number of colonies on the plate. Usually, the OD was maintained at 0.1-0.2. The starter culture was grown at 37 °C for several hours until it reached an OD of 0.6-0.8. This was then distributed at about 7-10 ml per 500 ml of auto induction media [ZY media containing appropriate antibiotic, 1 mM MgSO₄, 1X NPS, 10 mL of 50X 5052 to provide 2.5 g of glycerol, 250 mg of D. glucose and 1 g of α -Lactose].

The cells were further grown for 18-22 h at 37 °C. Upon the depletion of glucose for growth, auto induction occurred resulting in the expression of protein.

Core histone purification

The cells were harvested from the large-scale culture by centrifugation at 10,000 rpm for 15 min at 4°C. The cells were re-suspended in a sonication buffer [50 mM Tris-Cl pH 8.0, 500 mM NaCl, 5% glycerol], with added 5 mM BME, 1 mM EDTA, protease inhibitors, PMSF and benzamidine, and sonicated thrice at 40 amplitude, pulse on/off of 20 s/30 s for 5 min on ice. The mixture was clarified by centrifugation at 20,000 rpm for 15 min at 4 °C. The pellet was re-dissolved in a sonication buffer without EDTA, and the sonication and centrifugation steps were repeated as described above. The core histone proteins were expressed and were present in the insoluble inclusion body cell pellet. The pellet was then re-suspended in an unfolding buffer [sonication buffer containing 7 M Guanidinium chloride] and loaded onto an IMAC column for histidine tag-based affinity chromatography. A step-gradient with increasing imidazole concentrations [5 mM, 25 mM, 50 mM, 250 mM and 500 mM] in a 6 M urea containing buffer [6 M deionized urea, 50 mM Tris-Cl pH 8, 500 mM NaCl, 5% glycerol] eluted the histones at higher concentrations of imidazole. The fractions were almost pure owing to the increased expression of histones. The peak fractions were pooled and dialyzed against water overnight with 3 changes. On the following day, the histidine tag was removed by thrombin [1 unit of Thrombin/mg of protein, 10 mM Tris-Cl pH 7.5, 1 mM BME] digestion for 3 hours at room temperature. After 2 h, the extent of digestion was analyzed on 18% SDS-PAGE gel analysis run at 23 mA for 50 min in 1x TGS buffer. If it was over 80% complete, the additional 1 h during the gel run would take it to completion. PMSF and Benzamidine were added to inhibit the thrombin proteolytic activity. Throughout the procedure, the histones are maintained in either water or in unfolding buffer conditions. Any intermediate urea concentrations could enhance degradation issues. Therefore, just before loading it onto the column, the urea concentration of the sample is raised to 3 M. This way, with the column buffers at 6M urea concentration, the sample protein is maintained at over 3M urea during the purification process. The sample was loaded onto an ion exchange column [Resource S for H2A, H2B and H3 constructs and MonoS for H4]. A linear gradient 0 M salt to 1 M salt eluted the histone proteins at appropriate salt concentration.

The elution buffers comprised 6 M urea, 50 mM Tris-Cl pH 8.0 and salt [0 mM for buffer A and 1M for buffer B]. The initial peaks containing trace amounts of DNA contamination [checked based on the ratio of absorbance at A260/280 nm using nanodrop] were omitted, and the rest of the peak fractions corresponding to highly pure histone proteins were pooled and dialyzed against water overnight to remove urea. The samples were aliquoted and lyophilized.

Histone octamer refolding

Each aliquot of the core histones was dissolved in an unfolding buffer containing 7M Guanidinium, 10 mM Tris-Cl pH7.5 and 10 mM DTT, to a concentration of 1-2 mg/ml. Each of the core histones was then mixed in an equimolar ratio, and dialyzed [thrice] against a refolding buffer containing 50 mM Tris-Cl pH7.5, 1 mM EDTA, 2 M NaCl and 10 mM β -mercaptoethanol [BME]. The sample was concentrated to about 10 mg/ml using Amicon 10kDa molecular weight cut-off [MWCO] centrifuge filters. It was then loaded onto a gel filtration column [HiLoad 16/60 Superdex 200 preparation grade] and the octamer peak was separated from any dimer or tetramer contamination by size-exclusion chromatography. The fractions containing octamer was verified on an 18% SDS-PAGE gel, and the appropriate fractions were pooled. The sample was then concentrated using Amicon 10kDa spin concentrators to obtain a final concentration of about 8 mg/ml. It was then mixed with an equal volume of glycerol to obtain a final of 50% glycerol and stored at -20 ° C until further use (Dyer et al., 2004).

FOXA protein purification

Protein constructs

Human FOXA1, FOXA2 and FOXA3 constructs, codon optimized for bacterial expression, were synthesized and cloned into expression vector pET15b by EZ-Biolabs, USA. These constructs had either a cleavable N-terminal [NT] or a cleavable C-terminal [CT] hexa-histidine tag harboring a HRV3C protease site. Additionally, the attempts to purify hFoxA1 with the help of the Protein Production Platform [PPP], Biopolis involved a non-codon optimized construct of FoxA1 synthesized from human cDNA library [accession number BC033890]. The protein purification attempts by PPP employed a standard sub cloning procedure into pNIC-

Bse vector. Plasmids in pET background, compatible with ligation independent cloning protocol, for tagged constructs [MBP, GST, Trx, NusA and MBP with a spacer] were obtained from Addgene. Full length human FOXA1 codon optimized sequence was then cloned into these plasmids by ligation independent cloning protocol using the vendor recommended primers.

Expression and Purification of soluble proteins

The recombinant proteins were expressed in bacterial BL21-DE3 or RIPL cells, unless otherwise specified. The cells were grown at 37 °C, followed by induction with 0.4 mM IPTG at 18 °C for 2-3 hours. The bacterial cell pellet was harvested by centrifugation and lysed in a buffer containing 20 mM Tris pH 7.5, 5 mM BME, 5% glycerol, 500 mM NaCl, in the presence of PMSF, Benzamidine, protease inhibitor cocktail from Roche and pepstatin. The re-suspended cell pellet was homogenized and sonicated for 10 min on ice. The proteins expressed in the soluble fraction were separated by centrifugation at 20,000 g for 20 min at 4 °C. The first step employed histidine tag-based affinity chromatography using IMAC columns [elution by imidazole gradient from 0 to 500 mM] followed by ion exchange chromatography [elution by salt gradient from 150 mM to 1 M NaCl] using Resource S column. For the removal of the tag, HRV3 protease or TEV was used depending on the construct. Following overnight treatment with the proteases, the tag sample was passed through the IMAC column to separate the unbound pure tag cleaved protein from the column bound contaminants. The eluted fractions were tested on 15% SDS-PAGE gel to assess the purity at each stage of purification. The molecular weight of the final pure protein product was verified using MS analysis. All attempts made with the help of the PPP employed a common procedure mentioned on their website [www.proteins.sg].

Expression and Purification of insoluble proteins

For inclusion body preparation, the base buffer was composed of 20 mM Tris 7.5, 150 mM NaCl, 5 mM BME. The insoluble fraction of the lysed cell mass was re-suspended in an unfolding buffer [base buffer + 7 M Guanidinium chloride, 10% glycerol], clarified by centrifugation, and loaded onto an IMAC column for histidine tag-based affinity chromatography. A linear gradient of 0 to 100 % of buffer B equivalent to 0 to 500 mM imidazole was applied. Buffer A composition

included base buffer +6 M Urea and buffer B was composed of base buffer +6 M Urea, 500 mM Imidazole. The column bound sample was eluted at higher imidazole concentrations. Subsequently, the pure fractions were subjected to ion exchange chromatography using a salt gradient of 150 mM [Buffer A =base buffer +6 M urea] to 1 M salt containing buffer B [base buffer +6 M Urea, 1 M salt]. Various refolding strategies including stepwise reduction in urea concentration, drastic reduction in urea concentration, and refolding on column were tried [mentioned in the respective result section].

DNA production

DNA constructs

The FoxA1 cognate sequence containing 145 bp DNA construct [Appendix1] was designed by Assoc. Prof. Curt Davey, and was synthesized by EZbiolabs. The FOXA1-DNA has an EcoRV site at the two ends.

Cell lysis

Competent HB101 cells were transformed with about 20-50 ng of plasmid containing FOXA1-DNA using the standard transformation protocol. A small-scale starter culture with positive transformants was used to seed a culture for a large-scale preparation of about 2 – 6 l of commercially bought TB media. Following growth at 37 °C for 18 h, the cells were spun down at 7000 rpm for 7 min and subjected to lysis treatment. Alkaline lysis buffer I contained 25 mM Tris pH 8, 50 mM glucose, 10 mM EDTA; Alkaline lysis II buffer contained 0.2 M NaOH, 1% SDS; pre-chilled Alkaline lysis III buffer contained 4 M potassium acetate and 2M glacial acetic acid. For a 6 l culture, the cell mass was divided into 6 centrifuge bottles and the contents of each bottle was re-suspended with 50 ml of lysis I buffer to obtain a homogeneous preparation without lumps. 100 ml of alkaline lysis II was added and mixed vigorously by shaking and incubated on ice with intermittent shaking for 20 min. 125 ml of alkaline lysis III buffer was added to each of the bottles and mixed gently by swirling and inverting the bottles. The contents were incubated on ice for 20 minutes with intermittent swirling and mixing of the contents. The contents were then spun at 10,000 g for 20 min at 4 °C. After the removal of the cell debris, the supernatant was passed through layers of sterile gauge mesh filters to remove the particulates. The filtered supernatant was subjected to

isopropanol [0.52times] mediated precipitation, and allowed to stand for at least 30 min. Upon centrifugation at 16,000 g for 30 min at room temperature, the pellet was re-suspended in TE [10 mM Tris pH 7.5, 50 mM EDTA] transferred to polycarbonate tubes, and subjected to RNase treatment overnight.

Plasmid preparation

Visible debris was removed from the overnight RNase treated sample using a brief spin at 10,000 g for 10 min. The sample was then subjected to phenol chloroform extraction procedure to remove protein and other organic contaminants. Every 20 ml of the sample in Teflon tubes was first extracted with 10 ml of phenol [pre-equilibrated with Tris pH 8 buffer overnight to adjust the pH to 8]. The contents were mixed well by vortexing for a minute and spun at 27,000 g for 20 minutes at 20 °C. The top layer was carefully transferred to a fresh Teflon tube and the extraction process was repeated twice or until the interface becomes clear. The phenol contamination was removed by extraction with chloroform:isoamyl alcohol 24:1 mix [CIA] twice by centrifugation at 12000 g for 10 min at 20 °C. The top aqueous phase was then transferred to a fresh PPCO tubes for PEG precipitation. The plasmid material was precipitated with 10 % PEG6000 and 0.5 M NaCl and incubated on ice for 30 min. Spin at 15000 g for 15 min at 4 °C gave the final pure plasmid devoid of RNA contamination. The sample was then re-suspended in TE [10,0.1] buffer and the PEG was removed by extracting twice with CIA at 12000 g for 10 min at 20 °C. The aqueous top layer was then ethanol precipitated with 2.5x volume of cold 100 % ethanol and 1/10 volume of 3M sodium acetate pH 5.5. It was incubated on ice for 30 min and spun at 15000 g for 10 min at 4 °C. The pellet containing pure plasmid was air-dried and re-suspended in appropriate volume of TE [10,0.1] buffer. The plasmid concentration and quality was determined by measuring absorbance at 260 and 280 nm. An absorbance ratio of 260/280 close to 1.8 indicated a pure plasmid preparation.

DNA production

The FOXA1-DNA construct was designed to have flanking EcoRV sites at either end. The purified plasmid preparation was correspondingly digested with EcoRV to release the insert. The extent of digestion was verified on 10 % PAGE 1xTBE gel. The insert was then separated from the vector using a PEG fractionation using

9.5 % PEG 6000 and 0.5 M NaCl. It was kept on ice for 1 h, and spun at 27,000 g for 20 minutes at 4 °C. The aqueous phase was then subjected to ethanol precipitation as mentioned above and re-suspended in TE [10,0.1] buffer. The quality of the DNA was then verified on 10 % PAGE 1xTBE gel, run at 200 V for 30 min. All the gels were stained using ethidium bromide and visualized with UV trans illuminator of G-Box from Syngene.

Nucleosome reconstitution

A 4 µM reconstituted FOXA1-NCP was made by mixing different ratios of FOXA1-DNA to the purified histone octamer and performing a sequential salt dilution by dialysis (Dyer et al., 2004). The quality of the NCP was tested using 6 % native PAGE gels under 0.25x or 1xTBE conditions. Gel was run at 100V for 90 min at 4 °C in appropriate buffers [0.25x or 1x TBE] (Luger et al., 1999).

EMSA binding studies

Complexes of different FOXA constructs with FOXA1-DNA or FOXA1-NCP were made by incubating different ratios of protein with 1 µM DNA or NCP in 5ul of binding buffer containing 10 mM Tris-Cl pH 7.5, 1 mM MgCl₂, 40 mM KCl, 1 % glycerol and 1 mM DTT. The mixture was incubated for 15 min [for DNA] and for 1-2 h [for NCP] on ice. The complex formed by the binding interaction was tested using 6 % PAGE under 0.5 x TBE gel running conditions. All the gels were stained using Ethidium bromide and visualized with UV trans illuminator of G-Box from Syngene.

Radioactive gels

DNA was 5' end labeled with ³²P. The reaction was setup with about 1-3 µg of DNA, 1x T4 polynucleotide kinase (PNK) Buffer, 10 units of T4 PNK and 10 µCi of γ 32P-ATP in a total of 20 µl volume. The mixture was incubated for 1 hour at 37°C. The reaction mix was then passed through a DyeEx 2.0 spin column to remove the excess unbound γ 32P-ATP. The reaction was setup similar to the DNA binding studies mentioned above for EMSA using lower concentrations of DNA, about 50nM. Different molar ratios of protein to DNA was used to test for binding. The details are mentioned in the corresponding gel images.

Surface Plasmon resonance

SPR, a platform to study the binding interactions of chromatin factors with nucleosome, was optimized and established by Dr. Deepti Sharma from the lab of Assoc. Prof. Curt A. Davey. All SPR experiments and model fittings were carried out by York Wiew Chuah from the lab of Assoc. Prof. Susana Geifman from the School of Biological Sciences, NTU.

SPR experiments were run with a constant flow rate of $10 \mu\text{l min}^{-1}$ on a Biacore 3000 machine (BIAcore AB, GE Healthcare, USA) (Douzi, 2017) (Rich & Myszka, 2001) with a carboxymethylated dextran-coated sensor chip (CM5) at 25°C . The chip was activated and prepared using standard procedures. The binding reactions were studied in a buffer condition of 20 mM Tris-HCl pH 7.5, 200 mM NaCl at 25°C . The experiments were conducted in 2 different ways. The FOXA1-DNA was biotinylated and ethanol precipitated. The biotin-FOXA1-DNA was captured onto a neutravidin (Pierce, USA) coated CM5 chip. After capturing for a few 100 s of RUs, and once the baseline stabilizes, FOXA1 proteins are used as analyte in increasing concentrations. After each run, the surface is regenerated by injecting 2M NaCl followed by 75 mM EDTA for a minimum of 30 s. In another approach, histidine tagged FOXA1 protein was captured onto the anti-his antibody (Qiagen, Netherlands) coated chip. Increasing concentrations of DNA was used as the analyte and the response was monitored. The sensograms obtained for different samples were corrected for non-specific interactions by subtracting the values obtained for blank buffer. The data was processed and the affinity values were obtained by fitting the curve for a 1:1 binding.

Results and discussion

Core Histone production

Recombinant human core histones H2A, H2B, H3 and H4 expressed in the bacterial cells appears in the insoluble inclusion body. Figure 2.7 shows the purification process of core histones H2B and H4. The purifications of H2A and H3 are similar to that of H2B. The insoluble components of the inclusion body cell pellet are unfolded in 7 M Guanidinium buffer and passed through an IMAC column to which histidine tagged proteins remain bound. A subsequent 1-1.5 M high salt wash on the column and the 5 mM imidazole elution remove most of the contaminants. A step-wise gradient of imidazole elutes the bound proteins. Figure 2.7A shows the fractions that eluted between 25-250 mM imidazole were devoid of most contaminants. The pooled and dialyzed fractions subjected to thrombin proteolysis to remove the histidine tag. After a 2h thrombin treatment, the extent of digestion was tested on SDS-PAGE gel [Figure 2.7B] to be over 80% completed. Subsequent purification by Resource S or Mono Q column based ion exchange chromatography mainly removes any contaminating DNA and any additional protein bands [Figure 2.7C] that might be present. The first 1 or 2 fractions have DNA contamination that is observed as a higher A₂₆₀/A₂₈₀ [>0.8]. These fractions are discarded and only the rest are pooled. The final pure product is then dialyzed versus water, lyophilized and stored at -80°C.

Figure 2.7 Recombinant human core histone H2B and H4 purification
Purification of Human H2B [top panels A, B and C] and H4 [bottom panels A, B and C] under denaturing conditions. **A.** The peak fractions, following histidine tag based IMAC column purification. **B.** Extent of removal of the histidine tag following 2 hours of Thrombin digestion. HT-H2B or HT-H4 refers to histidine tagged and H2B or H4 is the tag cleaved protein. **C.** Final purity of the peak fractions following an ion exchange column purification. Resource S or Mono S column was used for the purification of hH2B or hH4, respectively. M is the protein molecular weight marker. The prefix 'h' refers to human proteins.

Histone Octamer purification

Each of the purified human core histones are re-suspended in an unfolding buffer containing guanidinium chloride and mixed in equimolar ratios. The components are refolded in a 2 M salt containing Tris pH 8.0 buffer. The refolded mixture was further purified by size-exclusion chromatography or a gel filtration column that elutes the sample components based on the size. Figure 2.8A is the UV trace of the eluting components (Rogge et al., 2013). The first peak, which is around 45ml, is due to some aggregates. The middle peak that elutes between 60-70 ml is the octamer peak. The trailing peak is due to the presence of dimer components. Figure 2.8B is the corresponding SDS PAGE analysis of the fractions. The octamer peak fractions are pooled [lanes 1-4]. The final pure product is shown in figure 2.8C. This is concentrated and stored in a buffer containing 50 % glycerol at -20 °C, until further use.

Figure 2.8 Purification of refolded human histone octamer

A. Gel filtration purification using Superdex HiLoad 16/60 column run showing the separation of octamer and dimer peaks. The blue trace is the UV absorbance of the elution peak. The peak at 45 ml is due to non-specific histone aggregates. The prominent peak at 65ml is due to octamer and the shorter one at 80 ml corresponds to H2A:H2B dimers. The in between shoulder at 70-75 ml is due to H3:H4 tetramers. **B.** 16 % SDS-PAGE analysis of the peak fractions as indicated. **C.** Final pool containing purified histone octamer. The bands corresponding to individual histone are shown.

Purification of Full-length Human FoxA1 protein using bacterial expression

Full-length human FOXA1 is a highly unstable protein prone to severe degradation and therefore, it has been challenging to obtain pure stable preparation of this protein for *in vitro* studies. This has hampered the accurate biochemical characterization of the protein. A methods paper for bacterial expression and purification of rat FoxA1 employed an N-terminally histidine tagged construct along with its 5' untranslated region which conferred better stability (Zaret & Stevens, 1995). Yet their yield was significantly affected at various stages. From a total starting material of 20 mg of FOXA1, they could obtain a final purified product of only 100-250 µg, of which only one-fifth remained active. Therefore for *in vitro* characterization, only low working concentrations [pM-nM] of the protein was achievable. Therefore, for obtaining structural insights and for accurate biochemical characterization, purification of active construct in considerable amount is a priority.

We tested the recombinant expression of human full length FOXA1 construct harboring a cleavable NT histidine [HT-hFoxA1] or C-terminal His tag [hFoxA1-HT] harboring HRV3C protease site in different bacterial expression strains. The expression was good in RIPL cells, when expressed at an optimum temperature of 37 °C and induced with 0.6 to 1 mM IPTG for 2-3 hours. Table 2.2 provides the summary of the results. Most of the FOXA1 protein remained insoluble and was present in the inclusion body cell pellet [Figure 2.9].

Following this, several attempts were tried to unfold and refold the protein, but the protein remained highly unstable and was prone to severe degradation. Flow chart 2.1 summarizes the various attempts made to improve the expression and purification of human full-length FOXA1.

Table 2.2 Summary of small scale bacterial expression studies of human FOXA1.

N-terminal his tagged construct (HT-hFoxA1) and C-terminal his tagged construct (hFoxA1-HT) were expressed in different bacterial strains. Good and low indicates that, in comparison to the background protein bands, the band corresponding to the molecular weight of FL-FoxA1 was intense and low, respectively. 'X' represents that there was no discernable band of appropriate mass.

Strain	Expression	
	HT-hFOXA1	hFOXA1-HT
BL21DE3	Good	X
BL21pLysS	X	X
Rosetta pLysS	Low	X
RIPL	Good	Low

Figure 2.9 Optimizing bacterial expression conditions improved FOXA1 protein yield in the inclusion body

Human FOXA1 [hFOXA1] protein expressed well at 37 °C with 0.4 mM IPTG in RIPL cells. The lanes represent the samples at different stages of lysis. FT is the flow through supernatant following different wash stages. IB pellet is the insoluble inclusion body cell pellet. The triangular arrow indicates the protein band corresponding to full-length FOXA1 protein.

Flowchart 2.1 Summary of the different approaches aimed at optimizing the bacterial expression and purification of human FOXA1 protein

FoxA1 is prone to severe aggregation during the refolding process

FoxA1 purification poses several challenges to maintain its stability and prevent aggregation through various stages of purification and during refolding attempts. The protein expressed in the insoluble fraction of inclusion body is extracted with 7M Guanidinium chloride containing buffer. His-tag based purification on IMAC columns was always the first step in the different purification attempts. However, there were several impurities and lower degradation bands. A considerable amount of protein was lost with each additional step of purification mostly due to degradation or aggregation or a combination of both.

For instance, figure 2.10A depicts a gel filtration effort, where only the lower bands eluted in the fractions [lanes 3-6], while the input had the protein of the correct size [lane 'Pre' of figure 2.10A]. Incorporating the insights from the different attempts,

we could obtain considerable amount of purified protein however with some shorter degraded products. The most recent effort involves IMAC purification followed by Resource S column based ion exchange chromatography, which removed most of the lower contaminating products [Figure 2.10B]. Throughout these steps, the protein remained denatured in 6 M urea with 20 mM Tris containing buffers. Following purification, another major hurdle was the refolding process. We tried refolding using sequential dialysis; refolding on column; refolding directly against 0 M urea. During sequentially dialyzing out urea from 6 M, 5 M, 4 M, 3 M, 2 M, 1 M, 0 M for 2.5 hours each, most of the material (>80%) precipitated out during the shift from 4M to 3M urea, suggesting misfolding. The main concern could be the cysteine residues. There are 5 cysteine residues in the full length with 3 cysteines present in the DBD. This refolding procedure needs to be optimized with a combination of reducing and oxidizing agents and their concentrations to guide proper folding. Alternatively, refolding on column with a buffer change to Hepes 6.5 [Buffer chosen based on the methods paper (Zaret & Stevens, 1995)] helped to obtain a full-length protein in soluble form [Figure 2.10B]. Yet, during concentration using spin concentrate devices, the entire material precipitated. The severe precipitation of refolded protein during concentration warrants further screening for stable conditions.

Screening for stable and soluble human FoxA1 truncation constructs

In parallel, we attempted to purify shorter truncation constructs that potentially retain the full-length FOXA1 nucleosome binding activity or can act as controls for future biochemical assays. Identifying a truncation construct is essential to establish a standard *in vitro* biochemical assay to measure FoxA1 activity. From the literature, we know that the DNA binding domain [DBD] and the C-terminal region [CT] is important for preserving the nucleosome binding activity. The C-terminus is important for binding in the nucleosome context but requires initial binding of the DBD to DNA (Cirillo et al., 2002).

Figure 2.10 Representative SDS-PAGE analysis on attempts to purify full-length human FOXA1 protein

A. Full-length FOXA1 expressed in the inclusion body was denatured and subjected to IMAC purification followed by refolding [dialyzing out urea] and gel filtration. The peak fractions indicate the presence of only lower minor bands remaining. **B.** The purified protein pool from an ion exchange chromatography purification following IMAC was refolded on the column to obtain a reasonable amount of protein. Black triangular arrows indicate the protein of the expected size. 'Pre' refers to the input material and 'FT' is the flow through. The lanes with numbers indicate the peak fractions.

Creation of random truncation constructs

Online ExPASy tool predicts an instability index of 55.56 for FOXA1 and the secondary structure prediction reveals very few helices with mostly unstructured regions. Figure 2.11 shows the secondary structure prediction using ExPASy tool GOR. Over 73% is random coil with the rest being alpha-helices [13%] or extended strands. We used the secondary structure analysis to guide us with designing constructs.

Figure 2.11 ExPASy tool GOR based secondary structure prediction.

A. The depicted sequence is that of the human FOXA1 protein and shows the corresponding secondary structure prediction. The codes for the different secondary structure elements are shown in panel B. **B.** The tabulated data shows the percentage occurrence of each of the different kinds of structural elements. **C.** The arrow representing the 425th residue marks the end of the last predicted helical structured region

Figure 2.12 Depiction of the FOXA1 C-terminal and N-terminal truncated constructs

The figure pictorially represents the varying length shorter human FOXA1 constructs. All of them harbor an N-terminal histidine tag, except for the ones marked as HT at the C-terminal side. The numbers indicate the starting and ending residue numbers. The pictorial representation above highlights the residues corresponding to DBD [central dark shaded region], trans activation domains [square boxes with crossed lines], and 425 as the last residue of the predicted structured region.

With the help of the Protein Production Platform [PPP], Biopolis, we created varied N-terminal [Δ NT] and C-terminal [Δ CT] truncations of the hFOXA1 protein. The target gene is from the human cDNA library clone and is not codon optimized for bacterial expression. Random truncations with spacings of 25 residues were made

from the N-terminal region. The C-terminal is essential for binding to histones, and the truncation constructs with CT deletions were trimmed in smaller fragments from the C-terminal end. A pictorial representation of the different length deletion constructs made is shown in figure 2.12. The residue numbers corresponding to the featured regions are indicated above the figure. This was used as a guide to design the constructs.

Purification of a C-terminal construct: 295-460

A CT construct [295-460], lacking the DBD, but with most of the C-terminus was successfully purified [Figure 2.13]. It can serve as a negative control for the binding studies with the DNA or with nucleosomes. Owing to its high negative charge, following IMAC affinity chromatography [Figure 2.13A], the construct was purified using a monoQ column [Figure 2.13B].

Figure 2.13 Purification of recombinant HT-hFoxA1[295-460] harboring an N-terminal His tag with TEV cleavage site.

A. Recombinant protein expressed in the soluble form was purified using His tag based IMAC column. 16 % SDS-PAGE analysis indicates the purity of the peak fractions [lanes 1-11 of IMAC]. **B.** Purification of protein pool from IMAC [pre] by MonoQ ion exchange chromatography. The white arrow indicates a highly pure fraction with about 2-3 mg of protein.

Purification of different length N- and C-terminal truncated constructs

Most of the longer truncation constructs still had the issue of insolubility and degradation. However, these purification attempts provided useful information regarding the contribution of different segments to the solubility and stability of

FOXA1 protein. Total expression of FOXA1 is feasible in BL21DE3 cells. But it did not appear in the soluble fraction, observed as an absence of the corresponding protein band in Figure 2.14A (i). Shorter constructs lacking the N-terminal region and starting from the DBD 158th residue could occur in the soluble fraction 2.14A (ii), lanes B8, A9 and B9. However, the longest construct of these, 158-472, lane A8, had troubled soluble expression. Our assumption was that the instability could be due to the random coil region on the C-terminal end. Based on the secondary structure prediction, the last residue of having a possible structured helical region terminates at position 425 [Figure 2.11] and the rest of the C-terminal portion of the protein remains as random coils. Truncating at this 425th position improved the stability of the protein. Since this protein lacks a major portion of the C-terminus, which may interact with histones (Cirillo et al., 2002), it may not be the most suitable candidate for studying the interactions with nucleosomes. However, this could be a potential control for affinity measurements. Therefore, additional trimming from the C-terminal end was considered [Figure 2.14A (iii)]. Using the insights from these constructs, residue length of up to 410 seemed to be promising for screening the N-terminal regions [Figure 2.14A(iv) and Figure 2.14B [i]]. Residue positions starting from 51 or 76 showed some soluble expression and truncation of first 25 residues had no soluble expression. Of the subsequent longer constructs tried, shown in Figure 2.14 B (ii) and (iii), the residue position terminating at 468 showed soluble expression. However, it had numerous lower molecular weight protein bands [figure 2.14B (iii) lane A8]. As a result, we considered using a C-terminal histidine tag to purify only the fully formed products from the bacteria [Figure 2.14C]. Figure 2.14C (i) starting from the DBD had some soluble expression, while Figure 2.14C(ii) of the full-length showed no soluble expression. Of the subsequent attempt to purify longer constructs, the construct 51-472 harboring a C-terminal histidine tag was quite promising [Figure 2.14C (iii) lane C3]. Based on these results, figure 2.15 provides a summary of these insights. N-terminal truncations of up to 25 remained insoluble while deleting the first 50

residues makes the protein express in a soluble form. However, the degradation issue persisted.

Figure 2.14 Total and Soluble expression of the various truncated FOXA1 constructs expressed in *E.coli* BL21DE3 Rosetta cells.

TE is Total expression; SE refers to soluble expression; Lane identifiers below each image are given in italics; residue numbers are indicated in bold with start position-end position; 6xHT indicates the tag is present at the C-terminal end. A i) Expression of full length FOXA1 protein. ii and iii) Expression of constructs starting from DBD with truncations at the C-terminal end. Iv) Expression of a construct missing the first 75 residues and ending at 410, [residue number chosen based on ii]. B. Comparison of N-terminal truncations of FOXA1 protein. i) Comparison of constructs missing the first 25 and 50 residues ending at 410 [similar to A. iv) gel image]. ii) and iii) are various C-terminal truncations of FOXA1 protein starting at 51 and 76, position respectively. C. Expression profile of different constructs harboring a C-terminal histidine tag. i) Starting from DBD and missing only the N-terminal regions of the full length-protein. ii) and iii) Expression of different constructs with trimmed N-terminal regions of FOXA1 protein harboring a C-terminal histidine tag.

Figure 2.16 shows the final pure product of the hFoxA1 truncation constructs starting with DBD [residue 158] and ending at varying lengths of CT. The different end residues were chosen based on the above studies on solubility and stability. Overall the purity of constructs in panel A-C is good. Panels D-F of longer DBD-CT constructs had lower degradation bands.

Figure 2.15 Summary of the recombinant expression and purification of FOXA1 truncation constructs

A pictorial representation on the summary of the FOXA1 truncation constructs and the corresponding inference on their recombinant bacterial expression and stability of the purified protein is mentioned on the right. The numbers indicate the corresponding amino acid residue number in the human FoxA1 sequence. DBD refers to the DNA binding domain. The green arrow representing residue 425 is the point where the stable helical structure ends, as per Expsy prediction in Figure 2.11.

Purification of constructs lacking N-terminal region

So far, the N-terminus has not been shown to be involved in any direct interactions with DNA or with histones and the deletion of NT region behaved like the wild type in these contexts (Cirillo et al., 2002). Therefore, we created constructs starting from the DBD and extensions of various lengths towards the end of the C-terminus [CT]. One construct that yielded promising results is the 158-472-HT [harboring CT His tag] construct [Figure 2.15], which starts from the DBD and ends at the last residue of the protein [panel H of Figure 2.16]. This construct will be designated as ΔNT-hFOXA1-HT, where HT-refers to

Figure 2.16 Purified shorter FOXA1 proteins

Purification of different FoxA1 truncation constructs starting from DBD extending towards C-terminus [PPP effort]. The numbers indicate the residues in the full length FoxA1 protein. **A-E** are N-terminally His tagged constructs with a TEV site and are not codon optimized. **A-F** and **H** are expressed in BL21DE3 bacterial cells. **G** is DBD to the end of C-terminus expressed in baculoviral based insect cell system. **H** is a C-terminal His tagged FoxA1 construct expressed in bacterial cells. All the above constructs were made by the Protein Production Platform, Biopolis.

a 6x histidine tag at the C-terminus. It remains soluble and relatively stable. It has been possible to obtain almost pure sample of this construct. It is important to note that the N-terminal His tagged construct 158-468, missing the last four residues [shown in panel F of Figure 2.16] still had severe degradation issues and these truncation constructs were from non-codon optimized template. In this regard, the low molecular weight protein bands seen from the purification of N-terminal His tagged construct is likely due to translation abort products. This was not seen previously because the C-terminal histidine tagged full-length protein failed to show good expression in a different kind of bacterial cells tested. With regards to stability, employing different tags, preferably C-terminal ones, could help obtain a soluble full-length form of the construct.

Baculovirus based insect cell expression in Sf9 cells failed to express any full-length protein [Figure 2.17A]. Based on the bacterial expression results [discussed in the following section]], shorter constructs have been tried as a control [Figure 2.17B and C]. In particular, the construct lacking the N-terminus [Figure 2.17 C] expressed well in bacteria. The expression levels for the full-length protein were too low and did not yield any protein upon trial purification. Since the aim was to

avoid refolding by *in vitro* approaches to obtain large amounts of the full-length protein, we did not pursue this further.

Figure 2.17 Attempts to check the recombinant expression of FOXA1 using Sf9 cells.

Below each image, Lane identifiers are in *italics*, followed by the residue numbers highlighted in **bold**. TE and SE refer to total expression and expression in the soluble fraction, respectively.

A. Expression of full-length FOXA1 in Sf9 cells. Red arrow indicates absence of a distinct protein band. B is the expression of a shorter FOXA1 missing few residues from the C-terminus end. Blue arrow with the bracket indicates the position where we expect the band corresponding to the protein to migrate. C is the expression of an N-terminal deletion construct, starting from the DBD to the end of the C-terminus. Blue arrow points to the expressed protein.

Purification of other FOXA family members

There are 3 different FOXA family members, FOXA1, FOXA2 and FOXA3. Their target DNA binding sequence is conserved. Despite a high degree of sequence conservation in the DBD domain, FOXA1 and FOXA2 are implicated in the early stages of liver development and FOXA3 in the later stages. FOXA2 is also involved in the neural and notochord development. Since the purification of FOXA1 DBD

was relatively simple and the issues with stability and solubility were caused due to the flanking regions, we set to purify the full-length of other FOXA family members. Codon optimized N-terminal and C-terminal His tagged constructs of full length FoxA2 and FoxA3 were screened for expression. Figure 2.18 is the SDS-PAGE gel image of bacterial total soluble expression of NT- histidine tagged FOXA1, FOXA2 and FOXA3 constructs. The issue of soluble expression prevails with these members as well.

Figure 2.18 Expression screen for full length FOXA family members

SDS PAGE analysis of the soluble recombinant expression of human FOXA proteins in bacterial cells. The respective lane ids are highlighted with a square [lane D1 for FOXA1, lane A12 for FOXA2 and lane C2 for FOXA3] in the respective panels. The corresponding lanes are also indicated above each gel image. As a control for comparison, lane C1 of FOXA1 and lane C1 of FOXA3 panels shows an unrelated protein that expresses well. Blue triangle indicates the position of a band corresponding to that of FOXA proteins expected to migrate.

Tagged constructs for soluble FOXA expression

Alternative to creating truncated FOXA1 proteins, we also considered using bulky tags to purify the full-length protein in the soluble form. This could eliminate the requirement to optimize refolding of this protein. For this purpose, we considered 4

different tags namely, MBP, GST, Trx and NusA (Costa et al., 2014). All of these had an additional histidine tag with a cleavable TEV or HRV3C protease site for easy purification on IMAC and for the removal of the bulky tag. The GST tag containing vector was readily available from our collaborator's lab and we prioritized that for a quick test for soluble expression without having to do any cloning. Once the ligation independent cloning worked, I considered MBP tag a priority because we had the tagged version and the option of introducing the spacer of additional amino acids between the bulky tag and the protein. This way we could eliminate the possibility of any steric hindrance due to a bulky N terminal tag. GST and MBP tags were considered as a priority.

With the GST tag, the full-length hFOXA1 protein remained in the insoluble fraction and the GST tag aggravated the problem due to degradation. Alternatively, a cleavable MBP- 6x histidine tagged construct with a HRV3C protease site to facilitate the tag removal was used. MBP is a 43kDa maltose-binding protein, expressed in the periplasmic space of *E.coli*. MBP presents a huge hydrophobic cleft with a conformational flexibility to bind and accommodate different polypeptides (Fox et al., 2001; Kapust & Waugh, 1999). The plasmid for this was purchased from Addgene and using ligation independent cloning procedure, full-length FOXA1 construct was cloned into the vector [Figure 2.19A]. Surprisingly, the recombinant expression in RIPL cells at 37°C for 3 hours with 0.5 mM IPTG, rendered the MBP-FOXA1 protein in the soluble fraction. However, the removal of the MBP-histidine tag rendered it unstable, which warrants optimization/search for stable conditions or binding partners that stabilize the full-length FOXA1 protein. Therefore, we decided to characterize FOXA1 protein in the presence of the tag. The MBP tag is a bulky one (~50 kDa) and could potentially interfere with the binding. Therefore, to reduce the potential for steric hindrance we used a construct with an additional spacer element of 10 amino acids between FOXA1 and the MBP tag. This construct is designated as the HT-MBP-N10- FOXA1 or as just MBP-N10-FOXA1. The expressed protein was purified on the Nickel-IMAC column, followed by a Heparin column purification. Several fractions of the peak had the protein of the correct size with minimal contaminants [Figure 2.19B]. This was successfully purified, although the yield is low [2-3 mg], but sufficient for biochemical characterization. NuSA, like MBP could have worked too. But since

MBP tag rendered the protein in the soluble fraction and the issue of stability upon removing the tag prevailed, we did not pursue other tags. We proceeded with testing the MBP tagged full-length proteins for its functionality.

Figure 2.19 Cloning, expression and purification of HT-MBP-N10-FOXA1.

A. Ligation independent cloning of HT-MBP-N10-FOXA1 construct. N is the negative control. 1 and 2 are the clones. Clone 1 had the insert of the correct size. The DNA sequence has been verified. **B.** Heparin column purification of HT-MBP-N10-FOXA1 protein following recombinant expression and purification using IMAC [not shown]. The left panel indicates the quality of every successive 4th fraction. Star at fraction 25 is the purest one. The panel on right indicates the quality of each fraction between 21 to 30.

Characterization of binding to DNA

FoxA1 binds with a higher affinity to nucleosomes than to free DNA (Cirillo & Zaret, 1999). The DNA binding domain can recognize a signature DNA site with ‘RYMAAYA’ sequence. Table 2.2 lists all the possible combination of this sequence. We use the binding as a measure to evaluate the activity of our different

protein preparations and to characterize its affinity for different sequences. These protein preparations will be subsequently used to obtain binding measurements with NCP or nucleosomes.

Table 2.2 List of possible combinations of the DNA sequences for the ‘RYMAAYA’ FOXA1 cognate DNA binding site.

Nucleotide code: R=A/G; Y=pyrimidine, M=A/C.

RYMAAYA	Reverse complement
ACAAACA	TGTTTGT
ACAAATA	TATTTGT
ACCAACA	TGTTGGT
ACCAATA	TATTGGT
ATAAACA	TGTTTAT
ATAAATA	TATTTAT
ATCAACA	TGTTGAT
ATCAATA	TATTGAT
GCAAACA	TGTTTGC
GCAAATA	TATTTGC
GCCAACA	TGTTGGC
GCCAATA	TATTGGC
GTAAACA	TGTTTAC
GTAAATA	TATTTAC
GTCAACA	TGTTGAC
GTCAATA	TATTGAC

Appendix 1 provides the FOXA1-DNA sequence containing eG high affinity binding site for FOXA protein [FOXA1-DNA]. The 145 bp FOXA1-DNA construct is made of 601L and α -satellite sequence elements along with an eG high affinity binding site (Cirillo & Zaret, 2007; Overdier et al., 1994) at a position just off the dyad. Dyad +2 position was determined to be the optimal location for FoxA1-NCP binding (Sekiya et al., 2009) . This construct was designed by Dr. Curt Davey and was synthesized by EZ Biolabs, USA. Figure 2.20A shows the EcoRV digestion of the plasmid containing FOXA1-DNA that cleaves the insert from the

vector. The mixture is then subjected to PEG fractionation to separate the insert from the vector [Figure 2.20B and C].

Figure 2.20 Production of FOXA1-DNA containing a high affinity binding site
10 % PAGE gel analysis of FOXA1-DNA production. **A.** The construct has EcoRV sites at either terminus. Upon enzyme digestion, the insert is cut from the vector. 'U' is the undigested plasmid sample. Lanes marked as EcoRV digested are three individual EcoRV digested samples of the same plasmid preparation. **B.** PEG fractionation of the sample. FOXA1-DNA is separated from the vector. Lane 1 is the PEG fractionation pellet containing vector and lane 2 is the supernatant containing the insert of the correct size. **C.** Final pure FOXA1-DNA. The red triangle indicates digested vector and the blue triangle indicates the 145bp FOXA1-DNA insert of the correct size. Lane 'L' is the 100 bp ladder.

The hFoxA1 DBD [residues 158-274] was tested for binding to different 145 bp DNA constructs [Figure 2.21A]. As a control for DNA binding, the strongly positioning 145 bp 601L DNA construct was used. DNA sequence information is provided in Appendix 1. The 601L construct does not harbor any of the combinations of 'RYMAAYA' sequence listed in the Table 2.2. DBD binds to the 145 bp FoxA1-DNA containing the eG high affinity site in a 1:1 molar ratio with some degree of non-specific binding. The nonspecific binding seems to be a bit more pronounced in the other 2 cases [α -satellite and 601L] with the appearance of a minor upper band and smeariness above the distinct band even at a limiting 1:1

molar ratio of protein to DNA. However, accurate biochemical measurements need to be done.

The binding of four different truncation constructs of DBD and DBD-CT of varying lengths was tested for binding to 145bp FOXA1-DNA using EMSA [Figure 2.21B]. For a given construct, with increasing protein: DNA molar ratios, i.e., comparing lanes 0.5 to 1.0 to 2.0, there is increased binding to the DNA, seen as one distinct band without any additional higher bands. On comparing with the different length protein samples, the band migration correlates with the increased molecular weight of the protein species used for binding. It is important to mention that there is always a proportion of free DNA even at a 2:1 of protein: DNA molar ratio. Previous works have considered this to be due to a partially active preparation of full-length protein (Zaret & Stevens, 1995). However, since this is observed with all the DBD and DBD-CT constructs, this could be attributed to the lower binding affinity to DNA (Cirillo & Zaret, 1999). This warrants accurate binding measurements. Some of the minor bands observed with longer constructs [158-410; 76-410] are due to the minor lower molecular weight degraded products present in the purified protein preparation.

The above characterization of DNA binding by DBD constructs was useful in the characterization of the longer constructs. This can be used as a validation tool for the purified preparation of the N-terminal deletion construct [Δ NT- FOXA1-HT] and for the MBP-tagged construct with the ten-amino acid spacer in between [MBP-N10-HT-FOXA1]. Figure 2.22 A and B shows the radioactive PAGE gel image to compare the binding of DBD and of Δ NT- FOXA1 with FOXA1-DNA and with the control 601L DNA. With increasing molar ratios of protein to DNA, we see a shifted band, indicating a 1:1 binding. At very high ratios, we see a super shifted band, that barely migrates into the gel, could be due to an aggregation of FOXA1-DNA complex. The figure reproduces the results of figure 2.20A in confirming the preferential sequence specific binding. Figure 2.22C compares the binding of the MBP-N10-His tagged-FOXA1 protein with FOXA1-DNA and with 601L DNA. The tagged protein displayed a pronounced sequence dependent binding to FOXA1-cognate DNA. However, there seems to be little binding to 601L NCP. Thus, the presence of the tag does not seem to interfere with the DNA binding.

Figure 2.21 Binding of shorter FOXA1 proteins with FOXA1-DNA

A. Analysis of DBD binding to different DNA constructs (harboring different sequences) by 6% PAGE, 0.5x TBE. FOXA1-DNA contains the cognate DNA sequence of the FOXA1 recognition element. The other 2 are the Human α -satellite and the super high affinity 601L construct. Different molar ratios of DBD: DNA are indicated at the top of the lanes. P is protein only control, which is not seen on this Ethidium bromide stained gel. '-' is the DNA only control. Blue and red triangles indicate free DNA and a complex of FOXA1-DNA, respectively. **B.** Analysis of binding of FoxA1 truncation constructs with DNA via 6% PAGE in 0.5x TBE running buffer. The numbers corresponding to FOXA1 constructs represent the residue numbers in full length FOXA1 protein sequence. Protein: FOXA1-DNA is the molar ratio. The arrowhead at the bottom indicates the FOXA1 DNA position. The purity of the FOXA1-DNA can be deduced from the gel image in A.

[i] with 601L DNA

[ii] with cognate FOXA1-DNA

A DBD of FOXA1

in nM 0 10 25 50 100 250 300 500

DBD of FOXA1

in nM 0 10 25 50 100 250 300 500

B ΔNT-FOXA1

in nM 0 10 25 50 100 250 300 500

ΔNT-FOXA1

in nM 0 10 25 50 100 250 300 500

C MBP- N10- HT- FOXA1

601L DNA

D 1:1 2:1 3:1

FOXA1- DNA

D 1:1 2:1 3:1

Figure 2.22 EMSA based binding studies indicate sequence specific binding to cognate DNA

EMSA binding experiments **A.** DBD with 50 nM of radiolabeled (i) 601L DNA or with (ii) FOXA1-DNA. **B.** ΔNT-hFOXA1-HT with 50 nM of radiolabeled (i) 601L DNA or, with (ii) FOXA1-DNA. **C.** Comparison of the binding of MBP-His tagged full-length FOXA1 [with a 10-amino acid spacer] with cognate DNA versus 601L DNA shows sequence specific binding.

Interestingly, from the results presented in the figure 2.22, it seems that the non-sequence specific binding to the 601L construct appears less pronounced in the case of N-terminal deletion construct [Figure 2.22B, left panel] and MBP-tagged full length construct [Figure 2.22C, left panel] than the DBD only [Figure 2.22A, left panel and Figure 2.21A]. This suggests that the regions outside of the DBD of FOXA1 might have a role in preventing non-sequence specific interactions with the DNA. However, this would have to be verified by further experiments with other controls.

Binding to NCP

Reconstituted nucleosome core particles containing 145 bp FOXA1-DNA and human histone octamers were obtained using standard established protocols [Figure 2.23].

Figure 2.23 NCP reconstitution with FOXA1 DNA

EMSA gel (6% PAGE under 0.25x TBE condition) indicating the quality of NCP reconstitutions with different DNA to hHO molar ratios [0.7, 0.8, 0.85, 0.9 and 1.0]. The ratio of 0.85 was chosen for large-scale reconstitution and used for subsequent binding studies. Blue and red triangle indicates band corresponding to NCP and free DNA, respectively.

The functionally relevant Δ NT-hFOXA1-HT construct was tested for binding with nucleosomes. The studies with full-length preparations, either the refolded FOXA1 or the MBP-tagged construct, failed to show a distinct band, had a smeared appearance [Data not shown] and had aggregation due to protein instability. With this Δ NT-hFOXA1-HT construct, we noticed that with increasing stoichiometry, a very slow migrating band appears at higher ratios.

The slowest migrating species, indicating a super-shift, migrates into the gel, thus eliminating the possibility of aggregation. This was observed on a 6 % native PAGE gel at 4 °C with 0.5x TBE [Figure 2.24]. The binding buffer had 40 or 80 mM salt. With increase in running time, the smearing effect is increased. The gel pore size and gel running conditions will need to be optimized. Also, there is the appearance of another minor intermediate band that might correspond to the actual complex.

Figure 2.24 Binding of Δ NT-FOXA1 with NCP reconstituted with FOXA1-DNA

Binding of Δ NT-hFOXA1 protein construct containing a C-terminal His tag with the nucleosomes bearing FoxA1 cognate binding sequence was performed under varying salt concentrations. The reaction was set up for 30 min to 1 h at 4 °C and then tested on 6 % PAGE gel with 0.5x TBE. Due to the poor staining and intense background of ethidium bromide stain, the gel was stained with Coomassie Instant blue protein stain. With increasing stoichiometry, the nucleosome band disappears completely, and a slower migrating species appears to dominate [Lanes 4 and 8 at 4:1 ratio].

Most studies and our DNA binding assay results have indicated the mode of binding of the FOXA1 protein to the nucleosome to be a monomer (Cirillo et al., 2002; Cirillo & Zaret, 1999; Sekiya et al., 2009) (Zaret et al., 2010). Consistent with this, we do not observe extra bands of varying sizes. Hence, the slowest migrating species might indicate an opening or a disruptive effect on the nucleosome, thereby enhancing the retardation on the gel. Alternatively, the slow migration could be due to the retardation imposed by the unstructured nature of the protein itself, as suggested by the slow migration of the protein-DNA complex. This might need optimization of gel pore size and running conditions. If this effect is due to disruption of the nucleosome core, then the feasibility of studying the structure of FOXA1-nucleosome complex by X-ray crystallography becomes questionable. Whether this binding means a specific alteration imposed on the nucleosome structure needs to be confirmed by alternate studies.

The appearance of a minor intermediate band could suggest a transition intermediate state where DBD binds the DNA first followed by the interactions of the CT with the rest of the nucleosome. Previous studies have indicated that the C-terminal's interaction with histones is contingent upon DBD recognizing the DNA element (Cirillo & Zaret, 1999). Hence there appears to be a sequential occurrence of events during the binding. This will require optimization of the duration of the binding under feasible conditions.

To address the unresolved issues with the EMSA studies, we considered SPR biochemical assay to measure the binding affinity of the protein to DNA and to the nucleosome. The major advantage being that it requires much less material compared to the ITC [Isothermal Titration Calorimetry]. SPR platform has been optimized for studying factor-nucleosome interactions in our lab [Sharma et al., in preparation]. We considered this since the monomeric binding mode of FOXA1 renders it less complicated to analyze in comparison to multimeric binding systems for obtaining affinity measurements. Additionally, the different length FOXA1 constructs will be extremely useful in providing mechanistic insights. Therefore, we set to validating the above hypothesis with binding affinity measurements using SPR and the other truncation constructs as controls.

SPR measurements

SPR is a sensitive and accurate technique to detect specific interactions (Jason-Moller et al., 2006). The platform has been optimized in our lab to study high affinity binding of factors with nucleosomes. The assay has been optimized for other high affinity [in the range of pM] binders to DNA and to nucleosomes [Sharma et al., in preparation]. The underlying principle involves capturing the substrate onto a chip and then the analyte is added in increasing concentrations. Specific interaction between the substrate and analyte causes a change in the refractive index of the surface, which can be monitored in real time. Based on the response curve and the concentrations of the substances used, the binding affinity and kinetic measurements can be obtained. To study FOXA1 interactions with DNA or with NCP-FoxA1, the experiment was performed in 2 ways. In one case, biotinylated DNA or NCP-FoxA1 was captured onto a Streptavidin coated CM5 chip and Δ NT-FOXA1-HT protein was used as the analyte. FOXA1 protein by itself was highly sticky to the streptavidin coated chip and there was a huge non-specific binding response to the chip. This potentially interfered with obtaining good response fits. With increasing concentrations of protein analyte, the response scales up exponentially in terms of the RU units, attributed to non-specific interactions of the protein to the chip. This adds much noise and the increase is more than that expected for binding to nucleosomes. Therefore, a confident affinity reading cannot be obtained in this case. Such stickiness was also observed with BLI [Biolayer interferometry], another biochemical technique that uses much less material.

In an alternative scenario, histidine tagged FOXA1 construct was captured using an anti-His antibody onto the chip and the DNA and the NCP-FOXA1 was used as the analyte. In this case, the response due to DNA or NCP-binding to the protein is small. However, with the Δ NT-FOXA1-HT construct, we could obtain a DNA binding affinity value of about 80 nM [Figure 2.25]. This number agrees well with the Aung et al., 2014 measurements. However, the response to NCP as analyte was too small and did not provide a good fit. The estimated affinity was approximately in the range of 200 nM [Figure 2.26]. This could be due to several alternative reasons, such as the lack of N-terminal residues, presence of a C-terminal Histidine tag, buffer conditions and the binding itself, making it difficult to draw conclusions.

Figure 2.25 SPR measurements of Δ NT-FOXA1-HT with cognate FOXA1 DNA

A. The Δ NT-FOXA1-HT was captured onto a CM5 chip coated with anti-histidine antibody. DNA in increasing concentrations was used as the analyte and the response was measured with time. **B.** A curve-fit of the different measurements provided an affinity value of 80 ± 10 nM, that matched the measurements made by Aung et al., 2014.

Figure 2.26 SPR measurements of Δ NT-FOXA1-HT with cognate FOXA1-NCP

A. The Δ NT-FOXA1-HT was captured onto a CM5 chip coated with anti-histidine antibody. FOXA1-NCP (with cognate binding site) in increasing concentrations was used as the analyte and the response was measured. **B.** A curve-fit of the different measurements provided an approximate affinity value of 230 ± 70 nM, lower than expected.

Conclusions and Future Directions

The FOXA family of proteins is classified as pioneer factors due to their ability to enhance accessibility to chromatin and having an essential role during development. With a highly conserved DBD that resembles the winged-helix structure of the globular domain of linker histone, it can compete with linker histones to open chromatin. It recognizes a signature of DNA recognition element positioned on the nucleosome surface. These features make this a potential factor to study by structural means.

FOXA1 being a very unstable protein posed several challenges during recombinant bacterial expression and purification. The full-length protein expressed in the inclusion body and posed issues with refolding. This hindered obtaining a biologically active protein. By creating a series of shorter truncated FOXA1 constructs, we have been able to successfully identify the characteristics of different regions in terms of rendering soluble expression. Based on literature, a functionally relevant N-terminal deletion construct could be purified in large quantities. In general, for understanding its chromatin architectural properties, it appears to be essential to study full-length FOXA proteins. Therefore, we are also considered purifying the FOXA2 and FOXA3 proteins. The purification remains equally challenging in these subtypes as well. In an alternative approach, the presence of an MBP-tag rendered full-length FOXA1 in the soluble fraction, eliminating the need to refold. However, the removal of the tag makes it prone to degradation. Screening for stable buffer conditions for a functional full-length protein needs to be pursued. In a scenario where the instability would be attributed to intrinsically disordered regions, considering orthologs of FOXA proteins in other species with diversified sequences could help in purifying a more stable protein.

For the functional characterization, the activity of the different FOXA1 protein preparations was characterized based on a DNA binding assay to selectively bind to cognate FOXA1-DNA versus 601L DNA. However, EMSA binding studies with the NCP failed to show any distinct complex species. Although it indicated an ~3-fold lower affinity for the NCP versus naked DNA, the SPR approach for obtaining accurate affinity measurements would need to be further optimized to establish this with certainty. The presence of varying linker DNA lengths could be an important

criterion for FOXA1 binding. Also, it might stabilize the binding to nucleosomes. There are several lingering questions regarding the work presented here. The extent of sequence specificity remains unclear. SPR based measurements could help in obtaining a quantitative assessment of the binding affinities. It might require the presence of linker DNA segments to stably associate with the nucleosome core. This can be tested by designing constructs of varying linker DNA lengths. Also, the FOXA protein could undergo some kind of transformation upon association with the nucleosome.

With the successes of the chromatosome project, presented in chapter 4, it would be worthwhile to try a similar purification protocol for the purification of full-length FOXA1. Since (Cirillo et al., 1998) had identified an additional NS-A1 binding site at the linker DNA regions [Figure 2.3], the binding studies can be conducted with nucleosome constructs harboring linker DNA regions as well, in case this is important for molecular recognition. The EMSA binding assays could be performed in a similar manner to the linker histones binding to the nucleosome. Alternatively, if full binding of FOXA1 actually has a disruptive effect on the nucleosomes, it would not be a good candidate for X-ray crystallography approaches. Depending on the extent of this effect, it could be studied by other means, including 2D EM using nucleosome arrays, optical tweezers or AFM. Overall, the work is inconclusive in determining FoxA1 interactions with nucleosome core particle. Given the nature of having a linker histone-like winged-helix DNA binding domain, FoxA1 could require the linker DNA regions flanking adjacent nucleosomes for binding. However, given its nature of interacting with enhancer regions in a timely manner and its role in development, it is a good candidate for understanding chromatin interactions at the very onset of gene regulation events. Also, because of the specific sequence dependent binding and its role in interacting with a few selected genes, it is a worthwhile drug candidate to address specific cancer targets.

Chapter 3 Interactions of High Mobility Group N [HMGN] family of proteins with nucleosome

HMG [High Mobility Group] of proteins

The HMG superfamily of proteins, present only in vertebrates, comprises the most abundant non-histone nuclear proteins. HMG proteins can independently interact with and modify the structure and function of chromatin. Therefore, along with linker histones, this group of proteins could be classified as chromatin architectural proteins. It consists of three distinct family members, HMGA, HMGB and HMGN.

HMGN family of proteins and their role in cellular events

Of the HMG superfamily of proteins, the HMGN family is capable of selectively binding to the nucleosome core. The members of the HMGN family consist of HMGN1, HMGN2, HMGN3 and its isoforms-HMGN3a, 3b, 3c and 3d, HMGN4 and HMGN5 [also referred to as NSBP1]. HMGN proteins bind independent of the underlying DNA sequence and can increase accessibility to the underlying gene in a densely packed chromatin (Crippa et al., 1992; Sandeen et al., 1980; Shirakawa et al., 2000; Trieschmann et al., 1995; Ueda et al., 2008). These events can drive various cellular processes such as replication, transcription, DNA repair, which are important for dictating cellular maintenance and the developmental state of a cell (Martinez de Paz & Ausio, 2016).

Role in Transcription

Binding of HMGNs to chromatin regulates and maintains an active chromatin structure. In a *Xenopus laevis* egg extract, HMGN1 and 2 increased the transcription of the 5srRNA gene, specifically, from a chromatin template during chromatin assembly (Trieschmann et al., 1995). In *HMGN*(-/-) mice, loss of HMGN affects the transcriptional profile in activated B-cell lymphocytes (S. Zhang et al., 2016). Incubation of permeabilized cells with competing NBD peptides decreased the polymerase II dependent transcription potential. HMGNs colocalized with a labeled nascent transcript that was introduced into permeabilized cells. The distribution of HMGNs within the nucleus correlates with transcriptionally active regions. HMGN proteins had a dispersed non-nucleolar pattern spread across the nucleus analogous

to the pattern of mono clonal antibodies directed to RNA pol II (Grande et al., 1997; Hock et al., 1998a; Hock et al., 1998b; Zeng et al., 1997). Monoclonal antibody confocal imaging and immunogold labeled electron microscopy studies in non-permeabilized cells revealed that on inhibition of transcription with actinomycin D or α -amanitin, the cellular localization of the HMGN proteins changed from a punctate dispersed distribution to a cluster in interchromatin granules (Hock et al., 1998b). Genome wide profiling of HMGN1 shows that the factor associates preferentially with DNase hypersensitive sites, promoters and enhancers (Cuddapah et al., 2011). These studies suggest that their cellular localization correlated with transcription events.

Role in replication

An *in vitro* study of the effect of HMGN2 on SV40 mini-chromosome containing the origin of replication showed a correlation between the HMGN's chromatin unfolding effect and an increase in replication efficiency. Particularly, this was evident in the chromatin template and not for free DNA (Ding et al., 1997).

Role in DNA repair

Mouse Embryonic Fibroblast knockout of HMGN1 showed increased UV hypersensitivity, which could be rescued by the ectopic expression of HMGN but not by the mutants that are deficient in nucleosome binding (Birger et al., 2003). The nuclear excision repair pathways seem to involve transcription coupled repair [TCR], which is associated with chromatin opening and acetylation of histones. HMGNs are capable of competing with linker histones, disrupting higher order chromatin structures and activating HATs [histone acetyl transferases], such as PCAF. These features drive the ability of HMGNs to direct the damage response pathways (Gerlitz, 2010). HMGNs can inhibit the chromatin remodeling activity mediated by ATP dependent remodelers such as ACF [ISWI ATPase subunit], BRG1, Mi-2, Rad54 +Rad51 (Rattner et al., 2009). Mouse fibroblasts study with *HMGN* (-/-) demonstrated the transcriptional repression activity of HMGN associated with *Sox9* and *N-cadherin* genes. (Furusawa et al., 2006; Rubinstein et al., 2005).

Members of HMGN family

HMGN1 and HMGN2 are ubiquitously present in all cell types while the expression of HMGNs 3, 4 and 5 are tissue and cell-type specific. HMGN1 and 2 are the most well studied proteins in this family. HMGN3 [Trip7 – Thyroid Hormone Receptor Interacting Protein 7] is expressed in the eye and brain of human and mouse and is involved in astrocyte function. HMGN3, also present in pancreatic cells in adult tissues, can affect insulin secretion (Gonzalez-Romero et al., 2015). HMGN4, a closely related member to HMGN2, also considered as a retro-pseudogene of HMGN2 lacking introns, is expressed more in the Thyroid gland, Thymus and lymph nodes. However, the expression is significantly lower than that of HMGN2 (Birger et al., 2001). HMGN5 is the most recent addition to the HMGN family and was formerly referred to as NSBP-1 [Nucleosome Binding Protein-1] or NBP-45 [Nucleosomal Binding Protein -45]. It has been rapidly evolving with a long C-terminus whose length and expression varies between species. The C-terminus of this protein is the major determinant of the chromatin binding properties and dictates the interactions with hetero versus euchromatin. The biological significance of this protein is still under study, although mutations are known to be associated with various cancers and tumorigenesis (Gonzalez-Romero et al., 2015).

Characteristics of HMGN proteins

Domain Organization

HMGNs are highly charged proteins with a strongly conserved nucleosome binding domain of about 30 amino acids and an acidic C-terminal domain that varies between the group members. With the exception of HMGN5, with a long acidic C-terminal consisting of 13 sequence repeats, the other members are less than 100 amino acids in length. (Rochman et al., 2010). Table 3.1 highlights the important amino acid composition and features of the different family members. Notably, they all comprise charged residues as indicated by the high pI values [except HMGN5] and the percent composition of acidic and basic residues. Also, they have a high composition of proline and glycine.

Table 3.1 Amino acid composition and biochemical properties within the HMGN family of proteins

	HMGN1	HMGN2	HMGN3		HMGN4	HMGN5
			3a	3b		
MW [Da]	10659	9393	10665	8377	9539	31525
pI	9.60	10	9.66	10.25	10.48	4.50
Number of amino acids	100	90	99	77	90	282
Negatively charged amino acids	19 %	14 %	18.2 %	14.3%	12.2 %	36.2 %
Positively charged amino acids	26%	26%	25.25%	29.9%	31.1%	21.27%
C	0	0	0	0	0	0
W, F, Y	0	0	0	0	0	0.4%
P	7%	12.2%	11.1%	13%	11.1%	2.5%
G	6%	11.1%	9.1%	9.1%	10%	10.3%
Instability index	55.13	35.05	58.46	63.93	44.28	58.67

Highly Conserved Nucleosome Binding Domain

Figure 3.1 shows the Multiple Sequence Alignment of the HMGN proteins and the charge distribution within each family member. Box 1 indicates the highly conserved basic nucleosome binding domain [NBD], and Box 2 indicates the identical ‘RRSARLSA’ region, which dictates the nucleosome binding property of these proteins. Any mutation in the arginine or serine residues, resulting in a change in conformation [mutating to proline] or a change in the net charge in this region, abolishes the nucleosome binding property (Prymakowska-Bosak et al., 2001)

(Ueda et al., 2008) (Shirakawa et al., 2000). The rest of the C-terminus consists of several acidic residues.

Figure 3.1 Multiple Sequence Alignment of human HMGN proteins.

ClustalO (Sievers & Higgins, 2014) sequence alignment [Top] of HMGN proteins. Box1 and Box2 indicate the conserved NBD domain and the “RRSARLSA” region, respectively. They all have an acidic C-terminus region with hHMGN5 having a long C-terminus. Bottom panel is the WebLogo3 (Crooks et al., 2004) representation of the conserved RRSARLSA domain. X-axis corresponds to residue numbers 14 -42 of HMGN1.

A recent study of the mapping of DNase hypersensitive regions at the enhancers has revealed that loss of either HMGN1 or HMGN2, the ubiquitous ones, does not affect the distribution (Deng et al., 2015) . However, loss of both HMGN1 and HMGN2 lead to a significant change in the regions marked with H3K4Me1 and H3K27Ac. Therefore, these members have overlapping functions and the loss of one can compensate for the other.

Interactions with chromatin

Nuclear mobility and cell cycle dependent cellular localization of HMGN

FRAP studies conducted *in vivo* indicated a faster mobility of the HMGNs and a transient interaction with chromatin (Catez et al., 2004; Phair & Misteli, 2000). As discussed above, *in vivo*, these proteins colocalize with regions of active transcription and are widely dispersed as punctate foci all over the nucleus and, on inhibition of transcription, they accumulate in the interchromatin granule clusters. These properties, along with their sequence independent mode of binding, demonstrate highly dynamic interactions and cell state dependent localization.

During interphase, the HMGNs colocalize with DNA (Cherukuri et al., 2008). But during mitosis, they are not associated with the chromatin. They re-enter the nucleus only on the formation of the nuclear membrane during the late telophase stage (Hock et al., 1998a; Prymakowska-Bosak et al., 2001). The re-entry into the nucleus is guided by a bipartite nuclear localization sequence and is directed by nuclear import signaling factors such as importin- α (Hock et al., 1998b). Additional evidence for their cell cycle dependent role is that mitotic phosphorylation of HMGN introduces a negative charge to the protein, rendering it refractory to nucleosome binding (Prymakowska-Bosak et al., 2001).

HMGNs disrupts higher order chromatin structure

Binding of HMGNs to the chromatin template in *Xenopus laevis* egg extract, which favors chromatin assembly, disrupts the higher order chromatin fiber structure, resulting in an extended conformation (Trieschmann et al., 1995). Neutron scattering studies on HMGN binding to chromatin revealed a decrease in the mass per unit length of the chromatin fiber (Graziano & Ramakrishnan, 1990). This means that the number of nucleosomes per repeat length is reduced resulting from

a change in the underlying chromatin structure. Study of the *in vivo* association of HMGN with cellular chromatin fiber revealed that the HMGN containing nucleosome regions were clustered into domains consisting of 6 contiguous HMGN₂-nucleosome complexes (Postnikov et al., 1997). These remain scattered throughout the non-nucleolar region. HMGN binding to nucleosomes and their ability to alter the level of histone modifications can regulate chromatin structure. HMGN1 can preferentially associate with chromatin regulatory regions such as enhancers, promoters. Additionally, HMGN1 and 2 can affect nucleosome positioning by intervening with ATP-dependent chromatin remodeling factors (Rattner et al., 2009). Overexpression of HMGN5 in mouse seems to enhance decompaction of heterochromatin regions in adult mice and causes disruption of nuclear lamina (Furusawa et al., 2015). Although at early stages, the nuclei of cardiomyocytes appears normal, the adult mice seem to develop a misshaped nuclei leading to arrhythmic heart and cardiac arrest.

Competing with linker histone

A quantitative study using crosslinking and immunoaffinity-based identification (Postnikov et al., 1991) compared the abundance of HMGN1/2 and linker histones in transcribed versus non-transcribed genes in chicken erythrocytes. HMGNs were more abundant in the transcribed regions in contrast to the linker histones that were less abundant compared to non-transcribed regions. In SV40 replicating minichromosomes, HMGN proteins can bring about the unfolding of chromatin structures (Ding et al., 1997). *In vivo* FRAP studies show that the presence of HMGNs decreases the residence time of linker histone. However, it cannot displace the linker histone already bound to chromatin. Mutations in the NBD or drugs that could displace HMGN from chromatin were not able to bring about this competing effect (Catez et al., 2002). A study using photo-crosslinking studies has shown that the binding of the globular domain of the linker histone remains unaffected upon prior binding of HMGN1 or HMGN2. However addition of HMGN1 or HMGN2 to H1 containing arrays in contrast to arrays lacking H1, significantly diminished the self-association potential of the arrays in the presence of 1.5-2.5M MgCl₂ (Murphy et al., 2017). Further, FRET experiment indicate that the HMGN binding can alter the C-terminal domain mediated H1 interactions with nucleosomes that have an important role in chromatin condensation.

Role of post translational modifications [PTMs]

PTMS, such as phosphorylation of specific Serine residues in the NBD of HMGNs, acetylation of lysine residues by histone acetylases and sumoylation by Sumo3-ligase, can decrease the binding affinity or completely abolish the nucleosome binding (Q. Zhang & Wang, 2008) (Bustin, 2001). Phosphorylation of the S10 residue of H3 disrupts the interactions with the DNA by the shielding of charges and could result in open chromatin structures for cellular events to occur. Loss of HMGN1 in mouse embryonic fibroblasts increases the phosphorylation of H3, and correspondingly binding of HMGN1 to nucleosomes inhibits the phosphorylation of H3 (Lim et al., 2002). During mitosis, phosphorylation of the S10 residue of H3 occurs in parallel with the condensation of chromatin. Phosphorylation of S6 of HMGN1 occurs before the S10 phosphorylation during mitosis. Also, HMGN2 can cause PCAF-mediated acetylation of H3 [H3K14 acetylation] (Lim et al., 2002; Ueda et al., 2006).

Interactions with nucleosomes

HMGNs bind to NCPs with a higher affinity than to DNA. Chromatin Immunoprecipitation and a modified RAPD [Random Amplification of Polymorphic DNA] method revealed that the HMGN interaction with chromatin does not have an underlying sequence specificity, although there seems to be a preference for low CG dinucleotide content (Shirakawa et al., 2000).

Cooperative binding under physiological ionic strength

HMGNs display a salt-dependent cooperative binding interaction with NCP (Postnikov et al., 1995; Sandeen et al., 1980). At low ionic strengths, < 0.1 M, HMGNs can transiently interact with the NCP, as observed by EMSAs and *in vivo* studies. At closer to physiological ionic strength, two molecules of HMGN interact with one molecule of NCP. The NBD, along with the acidic C-terminus, contributes to the high affinity binding to the NCP. Notably, there is no intermediate 1:1 species observed under physiological ionic strength conditions. This is also evidenced by the lack of HMGN heteropairing *in vitro* or *in vivo* i.e., only one variant type of HMGN is found associated with a given nucleosome (Postnikov et al., 1995; Sandeen et al., 1980)).

Stabilizing effect on NCP

Thermal denaturation assays and CD measurements indicate that binding of HMGN to NCP stabilizes the nucleosome core. In thermal denaturation experiments, upon binding to HMGN, the early melting peak of NCP disappears, which is also noted as the protection of DNA ends from hydroxyl radical treatment. CD measurements indicated an increase in winding angle of DNA on the NCP bound by HMGNs compared to NCP alone (Crippa et al., 1992; Gonzalez & Palacian, 1990; Paton et al., 1983; Sandeen et al., 1980; Yau et al., 1983). Another suggested model is that this stabilizing effect of HMGN makes the nucleosomes more resistant to the remodeling by ATP dependent chromatin remodelers (Rattner et al., 2009).

Existing models

HMGNs interact with nucleosomes near the dyad axis and protect the ends of the DNA from hydroxyl radical cleavage (Alfonso et al., 1994). Photo-crosslinking experiments show that the N-terminus interacts with a restricted region of H2B and the C-terminal end interacts with H3 (Trieschmann et al., 1998). An elegant NMR-Methyl TROSY study (Kato et al., 2011), monitoring changes in the side chains of Val, Leu, Ile residues, shows that one molecule of HMGN binds on each side of the nucleosome and can interact with the acidic patch formed by the H2A:H2B interface [Figure 3-2]. The C-terminus [CT] of the NBD contacts DNA near the entry and exit points on the NCP and the acidic C-terminus interacts with the N-terminal tail of H3. Based on this, it has been proposed that the HMGN2 binds by ‘stapling’ the DNA ends across the surface of the NCP. Also, the CT can compete with linker histones for chromatin binding (Kugler et al., 2012)

Figure 3.2 Proposed models for HMGN binding to NCP.

A. Nucleosome acidic patch [boxed region at left], with the corresponding residues indicated with black ovals at right. Adapted from (Kalashnikova et al., 2013). **B.** Left panel summarizes NBD interaction with the acidic patch and interactions that could interfere with linker histone interaction with the nucleosome. Right panel depicts that binding of HMGN could cause decompaction of higher order chromatin structures by interfering with linker histone association and internucleosomal interactions involving the NT tails of H3 and H4. Adapted from (Kugler et al., 2012).

Scope of this project

In summary, HMGNs are highly charged small molecular weight proteins capable of competing with linker histone activities and may interfere with core histone-mediated internucleosomal contacts. This could drive the decompaction of higher order chromatin structures. The general acidic patch mode of binding has been well established by a previous NMR study. Since HMGNs of only the same family member remain bound to a nucleosome in a cooperative manner, there could be an allosteric mechanism that results in subtle yet significant changes in the nucleosome that gets propagated to favor the binding of the second molecule of HMGN belonging to the same family. The mode of molecular recognition, in atomic detail, and the mechanism behind the cooperative binding behavior accompanying potential changes in nucleosome core structure remain unknown. This could well be studied by obtaining a high resolution structure of a complex of HMGNs bound to the nucleosome. Since HMGNs can exist in a cluster of nucleosomes as proposed by one study (Postnikov et al., 1997), within the context of crystals, we could also study any potential availability for interactions or cross-talk with the neighboring nucleosomes. In this project, by X-ray crystallographic means, we aim to address the mechanism of cooperativity and specificity exhibited by the HMGN family members in binding NCP and the basis for the increase in stability of the NCP upon binding HMGNs.

Materials and methods

Protein and DNA constructs

Human HMGN1, HMGN2, HMGN3b and HMGN4 constructs were made using the human cDNA library of the PPP, Biopolis. These constructs had a NT-6x histidine tag with a TEV protease site. Human HMGN3a and HMGN5 constructs, codon optimized for the bacterial expression studies, were synthesized and cloned into pUC57 vector by EZbiolabs. The protein purification attempts by PPP at Biopolis employed a standard sub-cloning procedure into pNIC-Bse vector. Peptides corresponding to the NBD regions were synthesized by the Peptide synthesis core facility at NTU.

The high affinity 601L DNA with 145 bp DNA, and the longer sticky-end DNA constructs, 169s, 173s (that were previously made and characterized in our lab), were used to study HMGN-nucleosome interactions.

Protein Purification

Transformation

50 ng of plasmid containing the histidine tagged HMGN proteins were transformed into 50 µl of bacterial BL21DE3 or RIPL competent cells. The cells were incubated with the plasmid for about 10 min on ice, and were subjected to heat-shock treatment at 42 °C for 45 seconds. After further incubation on ice for 5 min, the transformed cells were recovered in 1 ml of LB media at 37 °C for 30 min. The successful transformants were positively selected by plating all the cells onto LB agar plates containing kanamycin [50 µg/ml].

Cell culture

Recombinant hHMGN proteins were expressed in BL21 DE3 bacterial cells. For cell growth, IPTG induction protocol or auto induction protocol was used [details provided in the FOXA1 methods section]. HMGN proteins were expressed in the soluble fraction.

Purification of HMGN proteins

The cells were lysed in a sonication buffer containing 20 mM Tris pH 7.5, 5 mM BME and 500 mM salt and the soluble clarified portion was subjected to further purification. The supernatant containing the recombinant expressed proteins were bound to the IMAC column through their histidine tag. A brief [about 1 CV] 1-1.5 M NaCl wash step that helped eliminate a lot of contaminating proteins and DNA. The bound portion was then eluted by applying an imidazole gradient of 0 to 500 mM. The peak HMGN protein containing fractions were pooled and the final salt concentration was reduced to 150 mM by dilution or by dialysis. The staining and quantification of these proteins are discussed in the results section. The reduced salt content protein was then purified by ion exchange chromatography using Resource S or Heparin column. A linear salt gradient of 0 to 1 M eluted the protein at high salt. The pure peak fractions were then subjected to a TEV protease treatment overnight at 4 °C. The sample was passed through an IMAC and the flow through or fractions that eluted at low imidazole of [10 mM] contain the HMGNs of utmost purity. At this point, it was difficult to track the protein, owing to the low yield or a diluted sample. So, the pooled sample was concentrated on the Amicon centrifuge filters to detect the protein on SDS-PAGE by Coomassie staining or for the nanodrop measurements at 205 or 220 nm. The pure proteins were buffer exchanged with 20 mM K-cacodylate pH 6.0, aliquoted and stored at -80 °C.

DNA Production

Transformation, cell growth and plasmid preparation

The procedure used here is identical to that described in the Chapter 2 Methods section.

Palindromic, blunt-end DNA (601L) production & purification

Following plasmid isolation, we started with about 100 mg of plasmid for the palindromic DNA purification. EcoRV digestion released the half sites from the plasmid. A PEG fractionation with 9.5% PEG6000 and 0.5M salt separated the A and B half sites from the vector backbone. The half site inserts were then dephosphorylated by subjecting to alkaline phosphatase treatment. Phenol-CIA extraction was then followed to remove the enzyme and prevent any carryover of

proteins. The ethanol precipitated sample was then treated with Hinf1 enzyme that creates overhangs on side of the half-sites. The Hinf1 treated fragments were separated from the smaller digested contaminants by ion exchange chromatography using Resource Q column. The highly pure fractions were pooled together and concentrated by ethanol precipitation. Now the sample containing the two half-sites was ligated by T4 ligase treatment. The extent of ligation was tested on 10 % PAGE 1x TBE gel. Upon over 90 % completion of ligation, the sample was loaded onto a MonoQ column and eluted using a salt gradient. The final pure ligated DNA construct of the correct size was then ethanol precipitated and re-suspended in TE [10,0.1] buffer and stored at -20 °C until further use.

Palindromic, sticky-end DNA (165s, 169s, 173s) production & purification

The overall procedure for the palindromic sticky end DNA production is similar to the 601L purification. The designed sticky end constructs harbor a Pst1 digestion site and therefore, following plasmid purification, the repeated inserts [the half-sites] in the plasmid are released upon Pst1 enzyme digestion. Instead of PEG fractionation, the insert was separated from the vector by Resource Q column chromatography. The salt gradient had to be fine-tuned depending on the buffer compositions and the sample load. The separated inserts were then sequentially subjected to the alkaline phosphatase treatment, ethanol precipitation, Hinf1 digestion and purification on Resource Q column to obtain A and B sites, followed by ethanol precipitation, T4 ligase dependent ligation and MonoQ column purification. The fractions of utmost purity containing the correct size fragments and devoid of any half-site contamination were pooled, ethanol precipitated, and stored in TE [10,0.1] buffer until further use.

EMSA binding studies

The binding of HMGN proteins with NCP was tested in binding buffers containing 0.25xTBE or 1xTBE to verify the ionic strength dependent cooperative binding of HMGNs to nucleosomes. To test the effect of Mg^{2+} , buffer containing 1xTB with 1 mM Mg^{2+} was used. The mixture was incubated for 30 min on ice and tested on 6 % PAGE under 0.25x TBE or 1xTBE [or 1xTB with 1 mM Mg^{2+} in the case of studying the effect of Mg^{2+}] running buffer conditions, respectively. All the gels

were stained using Ethidium bromide and visualized with UV trans illuminator of G-Box from Syngene.

Crystallization setup

Initial screens for buffer conditions were done in a 96 well format and automated using the Art Robinsons Phoenix crystallization robot at Biopolis, Singapore. 0.2 μ l drops as 0.1 μ l of the complex and 0.1 μ l of 2x crystallization buffer was setup using the robots and allowed to form crystals against a reservoir containing 50 μ l of buffer at 2x concentration. The plates were sealed and imaged using a microscope.

Complex of hHMGN1 or peptides with 601L NCP were allowed to form in CCS buffer [20 mM Potassium Cacodylate pH 6.0 and 1 mM EDTA] by incubating the mixture on ice for 30 min. The concentration of the complex was between 3-5 mg/ml. Crystallization was setup using salting-out approach. 1 μ l of sample containing the complex of HMGN with NCP at 2.1:1 molar ratio was mixed with 1 μ l of 2x crystallization buffer [final concentration is 1x in the drop]. Crystals were allowed to form by hanging drop method against a reservoir containing 500 μ l of the corresponding 2x buffer.

Also, based on the existing constructs in the lab, crystallization was tried with the 169s nucleosome construct under similar conditions that yield chromatosome crystals. The set up involved 4 mg/ml of the complex containing 2.1:1 molar ratio of hHMGN1 to nucleosome and was allowed to crystallize using a salting-in approach. Sequentially, 2 μ l of the 4x crystallization buffer [final will be 2x in the drop] was added to 1 μ l of nucleosome and then 1 μ l of HMGN protein was added and mixed thoroughly. This drop was allowed to form crystals by hanging drop method against a reservoir containing crystallization buffer at 1x concentration.

Data collection

The crystals were harvested in the buffer conditions and exchanged with increasing concentrations of 2-methyl-1,3 propanediol (MPD) in 4 % increments up to 25 % v/v MPD. For controlled dehydration, Trehalose was also incorporated in the buffers increments up to 2 % w/v. This was achieved by making 2 stocks of base

buffers [10 mM sodium acetate 4.5, Potassium chloride, Calcium chloride]: 1) with 0 % MPD and 0 % Trehalose and 2) another buffer with 25 % MPD and 2 % trehalose. The 2 stocks were then mixed to achieve the intermediate increments of MPD and Trehalose.

Preliminary X-ray diffraction studies were conducted at the X-ray diffraction facility, A-Star, IMCB. This is a Bruker X8 PROTEUM machine with a PLATINUM135 CCD detector and a 4-circle KAPPA goniometer. Data sets for about 180 images were obtained with an oscillation range of 0.5° and an exposure time of 30 sec. X-ray diffraction data for analysis were collected at the beam line X06SA, with a PILATUS 2M-F detector, at the Swiss Light Source of the Paul Scherrer Institute, Villigen, Switzerland. Complete data sets were generated for 360 images collected with 0.5° oscillations and 1 sec exposures.

Results and discussion

Purification of hHMGN1 and hHNGN2 constructs

To study the complex of HMGN with NCP by X-ray crystallography, we require large quantities of highly pure protein. For the recombinant expression and purification of the full-length human HMGN proteins, we used constructs with a cleavable N-terminal His tag harboring a TEV protease site [HT-hHMGN]. In general, the IPTG induced or the auto-induced expression in the BL21 DE3 / RIPL cells concentrated the full-length HMGN proteins in the soluble fraction. Incorporation of a 1 M and a brief 2 M salt wash step during histidine affinity chromatography using IMAC column purification helped to eliminate a lot of DNA contamination [figure 3.3A].

Because of their high pI values, ion exchange chromatography performed using the Resource S column yielded highly pure fractions of the correct molecular weight that were pooled and concentrated using 3 kDa MWCO centrifuge filters [Figure 3.3B]. Overnight TEV protease treatment at 4 °C was highly efficient for histidine tag removal. Following this, the samples were passed through the IMAC column once and the proteins were in the flow through. The highly pure protein preparation was concentrated and stored as aliquots at -80 °C. The yield was about 10 mg from 2 liters of starting culture.

Owing to the lack of aromatic and cysteine residues, staining and quantifying these proteins posed difficulties and needed optimization. In general, standard protein Coomassie stain with a mild destaining process using only water worked. Comparing figures 3.4A with 3.4B shows the extent of staining difference between the commercial Instant blue versus Coomassie. In figure 3.4A, comparing lanes 1 with 2, 3 with 4 and 6 with 5, the instant blue stains TEV enzyme much better than the HMGNs irrespective of the loaded amount. However, in figure 3.4B, the Coomassie stains HMGNs readily. The histidine tag consists also of the TEV protease site, ENLYFQ, with 2 aromatic residues, F and Y. Upon removal of the tag, the issues with staining become more pronounced. In the case of instant blue staining, comparing lanes 1 with 2 [also 3 with 4 and 6 with 5] of figure 3.4A, the band corresponding to HMGN is essentially not visible. These are important considerations for these proteins, especially to interpret negative results.

Figure 3.3 Purification of recombinant hHMGN1 expressed in bacterial cells

Protein fractions analyzed on 15% SDS-PAGE gels. **A.** Cell lysis followed by purification on IMAC column. The numbered lanes are the peak eluted fractions. Rectangular box highlights the fractions that were pooled. **B.** Ion exchange chromatography using Resource S column. The peak fractions are indicated as numbers and the rectangular box highlights the pooled fractions. **C.** The purity of the pool from B is shown. **D.** TEV protease treatment. 'Pre' is the control sample and '+TEV' is the sample with added protease. **E.** The tag-cleaved protein was subjected to IMAC column purification. The step gradient and the corresponding imidazole concentrations are mentioned. **F.** Final human HMGN1 full length protein.

Due to the lack of Tryptophan, Tyrosine or Cysteines, HMGN proteins do not have any absorptivity at 280 nm. So, during the entire process of purification, it was difficult to track the proteins. For quantification using nanodrop, one approach was to measure the peptide bond absorbance at 205 nm in a buffer solution that has minimal absorption in this range (Scopes, 1974). By quantitative amino acid analysis, Shimahara et al., 2013, has determined the extinction coefficients at 220 nm for HMGN1 as $6550 \text{ M}^{-1} \text{ cm}^{-1}$ and for HMGN2 as $5900 \text{ M}^{-1} \text{ cm}^{-1}$ (Shimahara et al., 2013). We have used this in addition to the peptide bond quantitation to accurately quantify the proteins. As will be emphasized in the following sections,

the accurate quantification of these proteins is crucial for the stoichiometry dependent homogeneous complex formation.

Preparation of His tag cleaved hHMGN1t and hHMGN2

hHMGN1t and hHMGN2 were prepared using the same procedure as that of hHMGN1. The final pure products are shown in figures 3.4 and 3.5A.

Figure 3.4 SDS-PAGE analysis of TEV treated HT-hHMGN1, HT-hHMGN1t and HT-hHMGN2 samples

A. 15% SDS PAGE gel stained using Instant Blue stain. The stain could not stain the HMGNs properly. TEV enzyme [upper band] is stained very well. **B.** 15 % SDS-PAGE gel stained with Coomassie stain followed by a destaining process using water. The HMGNs stained very well in this case, much better than the TEV enzyme. The different species in the samples are indicated with the labelled arrows. 1- TEV enzyme; 2- Contaminants; 3- HT-HMGN or HMGN; 4- cleaved 6x histidine tag.

Other HMGN family members

In this study, we have also considered other HMGN family members, HMGN3a, HMGN3b, HMGN 4 and HMGN 5. Figure 3.5 shows the 15 % SDS PAGE analysis of the final pure protein product. HMGN5 with the long C-terminal tail failed to express in the bacterial cells. This requires additional optimization. We need to

determine the extinction coefficients for the other family members for accurate protein quantification, following tag removal. The molecular weight of the purified protein products has been verified by MALDI mass spectrometry.

Figure 3.5 Final purified product for different HMGN proteins

A. The final His tag cleaved products of hHMGN1t and hHMGN2. Purification of other hHMGN family members. Panels **B**, **C** and **D** show the final pure protein product of HT-hHMGN3a, HT-hHMGN3b and HT-hHMGN4, respectively.

DNA production

Some of the reasons that the X-ray crystallography study of HMGN complex with nucleosomes has not been successful so far was due to the failure to form a lattice that favors the presence of the protein in the crystal. Previous studies using *Xenopus* HMGN1 and HMGN2 conducted in our lab [unpublished] gave good diffracting crystals, but the factor was not lattice-incorporated. As a variation of this, one of the approaches is to use human HMGN subtypes to favor a protein mediated lattice arrangement. For this, we considered the Widom derivative, 601L DNA construct; the well characterized strongly positioning sequence of 145 bp length (Eugene Y. D. Chua et al., 2012). The 601L DNA was purified using standard protocols. In the meantime, sticky end DNA constructs of varying lengths were engineered in our lab for a favorable lattice packing. Since these were considered for the chromatosome project (Adhireksan et al., unpublished data), the library included longer DNA for reconstitution of nucleosomes with varying linker DNA lengths. One of the constructs, 169 bp sticky end [169s] DNA was successful for use in the

chromatosome project, and we considered this also to study HMGN complex with nucleosomes. Figure 3.6 shows the purification of the palindromic 169s DNA, containing 4-nucleotide overhangs from PstI digestion. This is representative of the palindromic DNA purifications with a slight modification step involving column purification to separate the insert from the vector, which in the case of blunt end preparation can be achieved by PEG precipitation. The rest of the purification is the same the between blunt end and the sticky end constructs.

NCP reconstitution

The histone octamer made using recombinant human histones was purified on a gel filtration column and the purified component was pooled and concentrated to about 4-5 mg/ml in buffer containing 50 % glycerol.

The reconstitution reaction was setup with different molar ratios of histone octamer to DNA and the ratio at which there was maximum NCP or nucleosome yield, with minimum DNA contamination, was utilized for subsequent tests or for large scale reconstitutions. Since the binding of HMGN to NCP is salt dependent, the reconstituted nucleosomes were tested for quality using EMSAs performed in 0.25x TBE and 1x TBE gel running conditions. A representative native PAGE gel image showing the quality of a 169s reconstituted nucleosome is shown in figure 3.6F.

Figure 3.6 Purification of CMS169s – a sticky end DNA construct.

Isolated plasmid is subjected to PstI digestion to create the sticky end overhangs.

A. Ion exchange chromatography based separation of the insert [half-sites of a palindromic DNA] from the vector. 'Pre' is the input material. **B.** Test ligation setup to test the extent of dephosphorylation of the insert following alkaline phosphatase treatment. The control lane is ligated and the lanes 1 and 2 are the dephosphorylated samples. **C.** Peak fractions comprising of the half-sites created following HinfI digestion. Pool represents the fractions considered based on A260/280 ratio. **D.** Ligation of the two half-sites. UL is the unligated control and 'L' is the ligated sample. **E.** Purification of the final ligated product [169s] using a monoQ column **F.** Representative 6 % PAGE analysis of the nucleosome reconstituted with the 169s DNA. 100 bp is the DNA ladder in increments of 100 bp.

Human HMGN proteins stably associate with the Nucleosome Core Particle

Previously published work from our lab (Ong et al., 2010) involved characterizing *Xenopus* HMGN1 and HMGN2 protein interactions with NCP. It also included a functionally relevant truncated HMGN1 construct, referred to as HMGN1t. The work studied the effect of divalent ions on the NCP binding (Ong et al., 2010). Analogous to previous observations from other studies, at physiological ionic

strength, the HMGN proteins associate with NCP as a homodimer (Sandeen et al., 1980) (Postnikov et al., 1995) . However, EMSA at physiological ionic strength, but with the additional presence of 1 mM Mg^{2+} or Ca^{2+} to test the binding of xHMGN1t and xHMGN2 with NCP, revealed a slower migrating band in addition to the 2:1 species. [Figure 3.7]. This was believed to be due to a destabilizing effect of HMGNs on NCP in the presence of divalent cations.

Figure 3.7 EMSA studies of xHMGN binding to NCP in the presence of 1 mM Mg^{2+} .

In the case of xHMGNs, under 1xTB with 1 mM Mg^{2+} conditions, there appears a slower migrating [S2] species in addition to the 2:1 HMGN-NCP species. Figure adapted from Ong et al., 2010.

In this current proposal, we aimed to characterize the human HMGNs. For this, we used hHMGN1, hHMGN2 and a truncated version of hHMGN1 which terminates either at D73 or L74. Since both of these truncation constructs demonstrated similar binding properties with NCP 601L [not shown], we used only the hHMGN1t D73 construct for subsequent studies. Please note that wherever it is denoted as hHMGN1t, it refers to hHMGN1t D73.

EMSAs at low ionic strength condition as that of 0.25x TBE indicate that hHMGN1, hHMGN1t and hHMGN2 proteins can bind to NCP in different ratios. Complexes corresponding to 1:1 and 2:1 HMGN: NCP were observed by the appearance of two

distinctly migrating bands on the gel. With increasing molar ratios to about 3, all the NCP is bound by protein and the 2:1 dimer species predominates [Figure 3.8A]. At near physiological ionic strength, as in 1xTBE gel running conditions [Figure 3.8B], the proteins bind to NCP only as a 2:1 complex. The intermediate species is not present. This has been observed by several previous studies, both *in vivo* and *in vitro* and has been well characterized [full references are provided in the literature section].

Figure 3.8 Binding of HMGNs to NCP601L under 0.25x TBE conditions.

A. A representative 6% PAGE native gel run under low salt, 0.25x TBE, conditions. Monomeric and dimeric binding of HMGN to NCP is indicated by small triangular arrows marked with 1 or 2, respectively. **B.** Native PAGE analysis under 1xTBE conditions. Consistent with the literature, there is no 1:1 band and only a 2:1 band appearance, indicated by a triangular arrow marked as 2.

To try to reproduce the results obtained from the *Xenopus* HMGNs with NCP [Figure 3.7], we did the binding studies including 1 mM Mg^{2+} in the buffers. The slower migrating band that was observed in the presence of Mg^{2+} with xHMGN1t and xHMGN2 was not been observed with the human HMGN1t protein [Figure 3.9]. We verified if this discrepancy is due to the presence of the histidine tag. Figure 3.9 is a representative 6 % PAGE gel image comparing the Mg^{2+} effect on binding of hHMGN1t [D73] to 601L-NCP with the histidine tag present or absent. The presence of the tag did not have any influence, and in both scenarios, there is no slower migrating component observed. Therefore, with the human proteins,

divalent conditions could be considered for the crystallization of the complex as well.

Figure 3.9 EMSA to study the effect of Mg²⁺ on human HMGN1t binding to NCP 601L.

Human HMGN1t with His tag [HT-hHMGN1t] and without tag [hHMGN1t] were compared for their binding to NCP. **A.** EMSA in the absence of Mg²⁺ **B.** EMSA in the presence of Mg²⁺. In both the cases, only one distinct slow migrating band species is observed.

Since human HMGNs did not show the Mg²⁺ mediated destabilizing effect observed in the case of *Xenopus* HMGNs, we proceeded to only pursue the study of the stable association of 2 HMGN proteins per NCP/nucleosome by crystallography, including with divalent metal conditions. Figure 3.10 shows the EMSA binding study conducted for hHMGN1 and hHMGN2 with NCPs/nucleosomes of varying DNA constructs and lengths. Under the 1x TBE conditions, there is only the 2:1 binding of HMGNs to nucleosomes observed.

Figure 3.10 hHMGN1 and hHMGN2 binding to different nucleosomes under 1x TBE conditions

Nucleosome core particle [601L] reconstituted with 145 bp and the nucleosomes reconstituted with sticky end DNA of varying lengths were tested for the 2:1 binding of HMGN1 [upper panels] or HMGN2 [lower panels]. The length of the DNA is noted above the gel images. Lane N corresponds to NCP or nucleosome only. The EMSAs were performed using 6% native PAGE gels run under 1xTBE conditions.

Crystallization trials with HT-hHMGN1

HMGN with NCP601L

The initial efforts were focused on HMGN1 protein, with a few additional setups using HMGN2. Concentrated hHMGN1 protein and NCP601L samples were buffer exchanged with CCS buffer. The two were premixed at a ratio of 2.2:1 to favor the formation of ternary species and to reduce heterogeneity. The mixture was incubated on ice for 1 hour and the samples were spun down at maximum speed to remove any precipitation. The final concentration of NCP in the complex was around 3.5-4 mg/ml.

Initially, a micro batch screen for the 27 conditions in the Hampton kit Helix Box 2 gave a few hits. Of the first 27 conditions listed in the kit, crystals started to appear within a day for the conditions: 0.05 M Lithium Sulphate 0.05 M Hepes (pH 6.5) with 18% PEG1000 or 10% PEG2000 MME or 15% PEG 350 MME. Comparing with the other similar conditions in the kit, pH and PEG concentration seemed important for the crystal appearance. Following this, I did a PEG screen with fixed 0.05 M Lithium Sulphate and 0.05M Hepes 6.5, which was setup as below.

8% - 16% PEG 550 MME

8% - 16% PEG 400

10% - 18% PEG 1000

6% to 2% PEG 2000 MME

Figure 3.11 shows the different morphology of the crystals obtained under some of the conditions mentioned above. Most of these were solid crystals but seemed to be in layers. These conditions gave reproducible crystals. As a control, NCP601L alone did not form any crystals under these conditions.

However, most of the crystals were rocky or very small for conducting diffraction experiments, warranting additional optimization. We tried using other DNA constructs in the lab derived from a 601 background but with linker DNA regions. In the meantime, since a nucleosome construct used for the chromatosome project, the 169s sticky-end DNA construct, gave crystals that diffracted, and the lattice arrangement seemed conducive for HMGN binding to the acidic patch, we decided to try this construct as well for the HMGN crystallizations. The crystallization conditions that gave diffracting chromatosome crystals also worked for HMGN-nucleosome complex. Figure 3.12 shows the stoichiometry dependence of crystal formation. Excess of HMGN or NCP hindered the process, and good crystal morphology was observed only when the setup was close to a 2:1 molar ratio of HMGN to nucleosome. Figure 3.13 shows some of the mounted, diffracting crystals at the synchrotron facility.

Figure 3.11 Representative images of hMGN1-601L crystals

The crystals were setup using a salting-out approach in 0.05M Lithium sulphate + 0.05M Hepes 6.5 buffer. The PEG concentrations are indicated above each of the images. These crystals appeared within a day under different conditions

Figure 3.12 Stoichiometry dependence of crystal formation.

HMG1 with 169s nucleosome was setup at increasing stoichiometry of 1.5 [left], 2 [middle] and 2.5 [right]. The most promising looking, large-single, crystals appeared only at close to 2:1 ratios.

Figure 3.13 Different morphologies of hHMGN-nucleosome crystals mounted and tested for diffraction at the synchrotron facility.

Figure 3.14 shows the diffraction patterns for some of the crystals tested. Figure 3.14A is a test diffraction experiment conducted along with screening for dehydration conditions. In general, at 25 % MPD, the diffraction quality of the crystals was poor. With increasing MPD concentrations, the diffraction quality improved. One of the crystals, in 45 % MPD, showed diffracted intensities up to nearly 4 Å resolution, but data could not be collected for this sample [due to limitations at the local source]. At the synchrotron facility, additional tests were conducted under strong dehydration conditions of up to 65% MPD. However, this did not yield better results, as the crystals diffracted only to about 6-8 Å resolution. One additional large difficulty is that crystal-to-crystal variations are substantial, making it difficult to obtain a single crystal with good diffraction potential. It appears likely that other NCP or nucleosome constructs may have better potential for yielding high resolution diffracting crystals in complex with HMGNs.

Figure 3.14 Optimization of crystal stabilization buffer conditions and representative diffraction patterns.

A., B. Diffraction experiments conducted at our in-house facility [A] and at the synchrotron facility [B]. The crystal dehydration conditions [MPD concentrations] are noted above each panel.

The presence of the target complex in the lattice was verified by harvesting some of the crystals and checking for the presence of HMGN proteins on an SDS-PAGE gel. Figure 3.15A depicts an Instant blue staining of crystals showing the presence of the histone octamer. Figure 3.15B shows a silver staining indicating the presence of HMGN1 in the crystals. Since the staining and destaining process are manually controlled, and owing to the differential staining abilities of different proteins, it has been difficult to interpret the stoichiometric ratios of these proteins based on the intensity of the bands. However, it indicates the presence of the HMGN protein in the crystals. In 3.15 panel C, the washed crystals were checked for the presence of the complex using a 6 % native PAGE. Compared to NCP control, the harvested

crystal samples [C1 and C2 lanes of figure 3.15C] have additional bands higher than the NCP, indicating the presence of the complex. Since, the crystals are harvested in a buffer containing half the concentration of the well buffer for harvesting and washing, there is a variation in the salt concentrations. Especially for the HMGN complex with the nucleosomes, observing a 2:1 ratio complex is contingent upon the sample preparation and gel running conditions. Hence, we observe additional bands in this case.

In conclusion, the stoichiometry dependent formation of good crystals, and the gel analysis is affirmative of the presence of HMGNs in the crystals.

Figure 3.15 Verification of the presence of HMGNs in the crystals

Representative SDS-PAGE [panels A and B] and native PAGE [panel C] gel analysis of crystals of hHMGN1-NCP complex, harvested and washed with a buffer containing half the salt concentrations of the well buffer to prevent crystal dissolution. **A.** Instant blue staining and **B.** Silver staining of an SDS-PAGE analysis of a harvested human HMGN-NCP complex crystal. Lanes hHO lane is the histone octamer control, hHMGN1 is the human HMGN1 control overloaded on the gel. Lane Crystal refers to the sample. The panel C is a representative 6% PAGE analysis of harvested crystals [lanes C1 and C2]. NCP lane is the NCP only control.

NCP binding of HMGN NBD [Nucleosome Binding Domain] Peptides

The approximately 30 amino acid NBD peptides of hHMGN1 and hHMGN2, designated as hNBD1 and hNBD2 respectively, were synthesized [NTU Peptide synthesis facility].

These peptides were tested for binding to the NCP 601L construct. Figure 3.16 shows the EMSA studies using 6 % PAGE with different concentrations of TBE. In

the binding with 601L [panels A and B], the peptides show a shifting of bands with increasing molar ratio of peptide to NCP. The nucleosome band is shifted completely by the 5:1 ratio.

Following this binding, crystallizations were setup with different ratios of hNBD1 or hNBD2 to NCP601L at similar conditions that gave crystals for the hHMGN1 complex with NCP 601L. We have not obtained crystals in these cases. Alternatively, we could try peptide-crystal soaking experiments, although these were unsuccessful in the past studies with the *Xenopus* system.

Figure 3.16 Binding of NBD peptides hNBD1 and hNBD2 with the NCP 601L construct.

A, B. Binding with NCP601L under 0.25xTBE [A] or 1xTBE [B] running buffer conditions show the shifting of bands with increasing amounts of peptide. The gel running conditions are indicated above the gel images. NCP control lane followed by increasing ratios of peptide to NCP is indicated above the lanes. hNBD1 and hNBD2 indicates NBDs of human HMGN1 and HMGN2, respectively.

Conclusions and Future Directions

HMGN proteins are small low molecular weight proteins that can stably associate with the nucleosome. An elegant NMR study revealed that the small nucleosome binding region of HMGN associates with the acidic patch of the nucleosome. Under physiological conditions, 2 HMGN proteins of only the same variant type bind to the nucleosome. This cooperative nature of binding could not be explained by the

NMR study. As a general theme HMGNs bind to the acidic patch like many other factors studied. Therefore, owing to the advantage of a stable association of the factor to the nucleosome, we pursued a crystallography approach to study the molecular details of the complex. In this work, we have successfully purified human HMGN proteins in large quantities. Contrary to our previous studies with *Xenopus* HMGNs (Ong et al., 2010), and in accordance with literature, human HMGN proteins could form a stable complex with the nucleosomes even in the presence of divalent cations. Therefore, we proceeded to use crystallographic approaches to study 2:1 HMGN: nucleosome complexes, including buffers containing divalent cations. We have screened through a variety of conditions and have been able to obtain some crystal hits. The stoichiometry dependency for the formation of the crystals and the PAGE gel analysis confirms the presence of the complex in the crystals. Currently, in one of the crystallization conditions HMGN1 or HMGN2 in complex with a nucleosome 169s, we obtained crystals that could diffract up to 6 or 7 Å resolution and possibly beyond. The diffraction could be further improved by optimizing cryogenic conditions. In parallel, we could search for other crystallization conditions to improve crystal quality. However, a major obstacle is crystal-to-crystal heterogeneity, and moreover it appears likely that other NCP or nucleosome constructs could work better with the HMGN system, since the 169s construct used here was developed initially for chromatosome work. Alternatively, though, with the success of the chromatosome crystals diffracting to 3 Å resolution and beyond, and with the apparent solvent accessibility of the acidic patches in this crystal form [For a more elaborate discussion, refer to chapter 4], it may be possible to soak NBD peptides into preformed chromatosome lattices. It has recently become apparent that linker histones and HMGNs can associate together with the same nucleosome. Since we have an existing chromatosome crystal system, studying HMGN and chromatosome as a complex could be advantageous. Understanding the interplay of HMGNs and linker histones in modulating nucleosome structure to control gene expression would provide novel insights into chromatin structure-mediated regulation. With having a chromatosome crystal platform [chapter 4] along with the success of purifying different linker variants and HMGN variants to large amounts provides a diverse opportunity to explore the different combinations of interactions and provide insights into the requirement for the variants function in modulating chromatin structure.

Chapter 4 Structural Insights into Chromatin Compaction by Linker Histone Proteins

Introduction

Epigenetic regulation of gene events is tightly controlled temporally and spatially, dictated by the accessibility of specific factors within the dense chromatin network. The structural organization of DNA into chromatin is facilitated by the binding of linker histone proteins to the adjoining linker DNA segments of the nucleosomes, and inter-nucleosomal contacts mediated by the core histone tails. However, the local degree of chromatin condensation is further determined by the complex interplay of several factors including the cell state, cell type, linker histone subtype, post translational modifications and the influence of pioneer factors. There have been ongoing strenuous efforts to understand the underlying principles of this packaging influenced by the linker histones. A brief account of the history of chromatin characterization has already been presented in Chapter 1.

Dynamic interactions with chromatin

The FRAP technique to measure the residence time of GFP-tagged linker histone showed real-time dynamic interactions with chromatin *in vivo*. At any given time, although most of the linker histone remained associated with the chromatin, it was dynamic and mobile. According to (Misteli et al., 2000), the GFP-tagged H1 remained associated for at least 220 s before exchange to a new binding site. Interestingly, modulating chromatin by increased acetylation of core histones [that causes opening of chromatin] or using drugs that affect phosphorylation of proteins reduced the linker histone residence time (Flanagan & Brown, 2016; Lever et al., 2000; Misteli et al., 2000). These studies suggested that linker histones are dynamic and are associated with both euchromatin and heterochromatin and that posttranslational modifications can alter their interactions with chromatin.

Linker histone confers protection to linker DNA ends in chromatin

The term chromatosome refers to the complex of linker histone bound to the nucleosome (~ 160 bp) (Simpson, 1978). Early MNase digestion assays, thermal denaturation assays and sedimentation analysis of chromatin indicates that at

physiological ionic strength, the presence of linker histone provides additional protection to the linker DNA regions flanking the nucleosome core (Carruthers et al., 1998; Finch & Klug, 1976; Harshman et al., 2013; Holde, 1989; Thoma et al., 1979). Furthermore, hydroxyl radical foot printing assays corroborates the linker histone mediated protection of an additional 20 bp; 10 bp on both ends near the dyad (Bednar et al., 1998). Increased salt concentrations result in the removal of linker histones. Chromatin, assayed under high salt conditions, renders the intermittent DNA regions adjoining two nucleosome core regions susceptible to nuclease digestion (Meyer et al., 2011; Simpson, 1978; Syed et al., 2010; Whitlock & Simpson, 1976).

Linker histone subtypes and variants

Linker histones are the lysine-rich histones that are dynamic regulators of chromatin. Unlike core histones, linker histones are less conserved among different organisms. They have undergone a great degree of diversification and their epigenetic roles are more pronounced in higher eukaryotic organisms. The presence of linker histones serves as an additional repressive barrier to gene regulation events. Genomes wide mapping studies of specific somatic linker histone variants have indicated that linker histone occupancy is low at transcription start sites, and linker histones are displaced from the cis-regulatory regions before the onset of gene activation (Cao et al., 2013; Izzo et al., 2013; Millan-Arino et al., 2016).

The linker histone family of proteins comprises eleven known variants in humans. H1.1, H1.2, H1.3 H1.4, H1.5, H1.0 and H1x are somatic variants, whereas gamete specific linker histones include H1t and H1T2 in spermatocytes, H1oo in oocytes, and a linker histone like protein HILS. The amino acid sequence varies among the different variants and some demonstrate cell-type specific roles. Their expression changes through the stages of development and cell cycle events. The genes for somatic linker histone variants H1.1, H1.2, H1.3, H1.4, H1.5 and for H1t are present as *HIST1* clusters on chromosome 6 along with the core histone genes. The somatic ones show DNA replication dependent expression. H1t has a tissue specific expression attributed to the differences in respective promoters. The genes of H1.0 and H1x are secluded and occur at chromosome 22 and 2, respectively. H1.0 is expressed in terminally differentiated cells. H1x is ubiquitously expressed

independent of the cell cycle stage (Happel & Doenecke, 2009; Harshman et al., 2013). The germ-line specific variants replace the somatic variants in specific cell types. During gametogenesis, chromatin undergoes severe condensation and changes during different stages. Transcription is shut down in these tissues during early stages of development. In different organisms, H1_{oo} is present during early embryogenesis until a stage where it gets replaced by somatic variants. The expression of testis-specific variants, H1_t, H1_{T2} and H1_{LS} vary during spermatogenesis (Pérez-Montero et al., 2016).

Role in gene regulation

Mutational studies have indicated that the H1 variants are functionally redundant and in the absence of one, other variants can compensate for the loss (Kowalski & Palyga, 2016). However, the binding affinities, differential distribution across active chromatin regions and cell type and state dependency suggest unique roles for each subtype (Mayor et al., 2015; Millan-Arino et al., 2016).

Mice null mutants generated for one or even 2 of the 6 somatic histone variants had maintained the normal linker histone to core histone levels and developed normally (Y. Fan et al., 2001; Sirotkin et al., 1995) indicating some compensatory overlapping functions among the variants. However, inactivation of 3 of the somatic variants H1_c, H1_d and H1_e demonstrated a 50% reduction in the linker histone content, which resulted in developmental defects leading to embryonic lethality (Y. Fan et al., 2003). Re-creating the triple mutant in mouse embryonic stem cells harboring reduced linker histone levels showed irregularities in chromatin organization with a significant change in the nucleosome repeat length. However, microarray analysis indicated that only a small subset of genes was affected by DNA methylation, suggesting a more complex gene regulatory role of linker histones (Y. Fan et al., 2005). This suggests that in addition to the general repressive function of the H1 family of proteins, they have specialized targets. This was further supported by the modulation of specific linker histone variants exerting gene specific differences (Alami et al., 2003; Brown et al., 1997). A subset of linker histone variants, specifically via the C-terminal regions, could directly interact with and recruit DNA methyl transferases to the target loci comprising imprinting genes (Yang et al., 2013). A study comparing the gene targets of H1_c and H1.0 identified

genes that are uniquely regulated by one of these variants and genes that could be regulated by both (Bhan et al., 2008).

The different linker histone variants show differences in their binding affinities to chromatin. Competitive binding assays to compare the binding affinities of different variants of somatic linker histones of rat brain to the rat liver chromatin sample indicated high binding affinities of H1e, H1d and H10; intermediate affinity for H1b and H1c with the least being that of H1a (Orrego et al., 2007). *In vivo* FRAP assays using GFP tagged proteins were particularly useful in comparing the binding affinities. $H1.4$ and $H1.5 > H1.3$ and $H1.0 > H1.2$ and $H1.1$ (Th'ng et al., 2005). (Clausell et al., 2009) also observed a similar trend of binding affinity using reconstituted mini chromosomes. Compared to the somatic variants, the degree of chromatin condensation achieved by H1t seems considerably less (Pérez-Montero et al., 2016).

The high lysine content of the C-terminal domain [CTD] of linker histones provides a net positive charge that facilitates higher order structure formation by DNA charge neutralization. It has been proposed that in the context of nucleosome binding, the CTD can fold into a secondary structure, which could further contribute to high affinity binding. Varying linker DNA length, charge, PTMs and the different CTD of linker histone variants suggest distinct roles in the regulation of higher order chromatin structure (Flanagan & Brown, 2016; Hendzel et al., 2004; Hutchinson et al., 2015; White et al., 2016).

Linker histone variants, in addition to their common targets, can also have a distinct set of protein targets. Histone chaperones, post translational modifications and other protein interactions could further regulate the stability and unique functions of the different linker histone variants. This remains largely unclear. For example, Tpr, a nuclear pore complex protein has been shown to specifically interact with and stabilize H1.1 and H1.2 proteins (P. Zhang et al., 2016). Moreover, specific linker histone variants were seen to be involved in the recruitment of epigenetic marks to target sites. This was observed as a direct interaction with DNA methyltransferases, DNMT1 and DNMT3B and interference with the activity of histone methyltransferases (Yang et al., 2013).

Implications in cancer

At a global level, chromatin organization into heterochromatin and euchromatin regions is shown to influence the mutation rates of the genome (Schuster-Bockler & Lehner, 2012) in different tissue types. A study on cancer heterogeneity and self-renewing ability establishes the role of H1.0 as a prognostic marker in determining the cancer cell fate in several types of cancer (Torres et al., 2016).. They demonstrate low H1.0 levels correlate with cancer progression and hierarchical status of cancer cells, and the effects could be reverted by re-expressing H1.0. There was also a tendency to negatively select for H1.0 expressing cells with cancer progression, suggesting that H1.0 levels could be indicative of survival status of patients. This is further supported by the methylation status of the HISTH1.0 and upregulation of genes essential for cancer cells progression

A shRNA mediated knockdown of H1.2 and H1.4 variants in breast cancer cell lines demonstrated that only specific genes were affected. H1.2 knockdown affected several cell-cycle specific genes, and in T4D4 cells, H1.4 depletion caused cell death (Sancho et al., 2008). A genome wide mapping study on the H1.2 variant indicated increased expression levels in cancer cells. H1.2 was associated with increased H3K27me3 marks on the nucleosomes, which in turn seemed to repress the genes that control growth (Kim et al., 2015).

Basic domain organization

Linker histone proteins have a tripartite structure comprising a structured globular domain flanked by intrinsically disordered N and C terminal regions. Although the globular fold is conserved among the variants, the N and C terminal regions are also important for binding and functionality. The intrinsically disordered nature provides conformational freedom and flexibility to bind varying targets to form specific stable complexes (Hansen et al., 2006). However, this domain organization and sequence identity is not strictly conserved, especially among lower organisms (Harshman et al., 2013; Izzo et al., 2008). *Saccharomyces cerevisiae* linker histone Hho1p has two folded globular domains with distinct functions (Ali & Thomas, 2004), whereas, in the amitotic dividing macronuclei of *Tetrahymena sp.*, the linker histone lacks the conserved globular domain (M. Wu et al., 1986).

Multiple sequence alignment and Phylogenetic tree analysis of the linker histone variants

Appendix 1 and Figure 4.1A shows the clustal omega (Sievers & Higgins, 2014) multiple sequence alignment [MSA] of the full-length protein and the globular domains of the linker histone family, respectively. Based on the globular domain alignment, a phylogenetic tree using a neighbor-joining method was generated using the clustal omega program and represented using Phylo.io (O. Robinson et al., 2016) [Figure 4.2]. For the phylogenetic tree analysis, considering less conserved regions introduces noise, therefore, only the globular domain sequence analysis has been considered. The evolutionary divergence is measured as genetic distance, based on the number of mutations that have occurred in each sequence during their divergence from a common ancestor. In general, the analysis indicates that the linker histone somatic variants are grouped together with the H1t testis-specific subtype, which is thus closer to the somatic variants. This is in accordance with their gene location on the *HIST1* cluster. The globular domains of H1x and H1.0 are similar. H1LS and the H1oo are the most divergent ones. The WebLogo3 (Crooks et al., 2004; Sharma et al., 2012) representation highlights some of the most conserved residues in the globular domain of the linker histones variants [Figure 4.1B]. Thus, the linker histone proteins have been evolving to perform specialized diverse functions in higher eukaryotes.

A

hGH100, 50-129	RRHPPVLRMLVLEALQAGEQRRGTSVAAILK-YILHKYPTVDVLRFKYLLKQALATGMRRG
hGH1t2, 55-124	SSVLRVSQVLVQAI---STHKGLTLAALKK-ELRNAG--YEVRRKSGRHEA--PRGQAKA
hGH1x, 43-118	NQPGKYSQLVVEITIRRLGERNGSSLAKIYTEAKKVPW--FDQQNGRTYLYKYSIKALVQND
hGHILS, 112-177	CQKPSTSKVILRAVADKGTCKYVSLATLKK-AVSTTG--YDMARNAYHFKRVLKGLVDKG
hGH10, 23-97	TDHPKYSDMIVAATQAEKNRAGSSRSIQK-YIKSHY--KVGENADSQIKLSIKRLVTTG
hGH1t, 39-113	VPNLSVSKLITEALSVSQERVGMSLVALKK-ALAAAG--YDVEKNNSRIKLSKSLVSKG
hGH1.1, 38-112	PAGPSVSELIVQAASSSKERGGVSLAALKK-ALAAAG--YDVEKNNSRIKLGKSLVSKG
hGH1.2, 35-109	ASGPPVSELITKAVAASKERSGVSLAALKK-ALAAAG--YDVEKNNSRIKLGKSLVSKG
hGH1.3, 36-110	ASGPPVSELITKAVAASKERSGVSLAALKK-ALAAAG--YDVEKNNSRIKLGKSLVSKG
hGH1.4, 35-109	ASGPPVSELITKAVAASKERSGVSLAALKK-ALAAAG--YDVEKNNSRIKLGKSLVSKG
hGH1.5, 38-112	ATGPPVSELITKAVAASKERNGLSLAALKK-ALAAAG--YDVEKNNSRIKLGKSLVSKG
	: : : : :
hGH100, 50-129	LLARPLNSKARGATGSFKLVP
hGH1t2, 55-124	TL---LRVSGSDAAGYFRVWK
hGH1x, 43-118	TL---LQVKGTTGANGSFKLNR
hGHILS, 112-177	-----SAGSFTLGK
hGH10, 23-97	VL---KQTKGVGASGSFRLAK
hGH1t, 39-113	IL---VQTRGTGASGSFKLSK
hGH1.1, 38-112	TL---VQTKGTGASGSFKLNK
hGH1.2, 35-109	TL---VQTKGTGASGSFKLNK
hGH1.3, 36-110	TL---VQTKGTGASGSFKLNK
hGH1.4, 35-109	TL---VQTKGTGASGSFKLNK
hGH1.5, 38-112	TL---VQTKGTGASGSFKLNK
	: * * :

B

Figure 4.1 Multiple Sequence Alignment of the globular domain of the various linker histone variants in humans.

A. Clustal omega based multiple sequence alignment generated for the globular domains of the human linker histone variants. The prefix hG represents the human globular domain and the numbers indicate the corresponding residue numbers of the linker histone variants. **B.** WebLogo3 representation of the alignment presented in panel A. The height of the letters indicates the level of conservation of the residues among the variants.

Figure 4.2 Phylogenetic tree analysis of linker histone family of proteins

Phylo.io representation of the phylogenetic tree generated using Neighbor-Joining method of Clustal omega program. The tree has been generated based on the sequence alignment for only the globular domains of the linker histone variants. The distance number represents the genetic distance, an indicator for the number of mutations that has occurred to diversify from the common ancestor [node].

Globular domain has a winged-helix structure

Two structures of the globular domain of chicken H5 [pdb: 1HST, (Ramakrishnan et al., 1993) and of human H1x [pdb:2LSO] were solved by X-ray crystallography and by NMR studies, respectively. Figure 4.3 shows the winged-helix motif consisting of 3 α -helices with a β -hairpin. The conserved globular domain of the linker histone family is essential for binding to the linker DNA of the nucleosome.

Figure 4.3 Structures of the globular domain of linker histones

A. X-ray crystal structure of the globular domain of H5 [pdb id: 1HST] from chicken erythrocytes. **B.** NMR structure of the globular domain of human H1x [pdb id: 2LSO]. Both structures indicate a winged-helix motif.

Dynamic interactions with DNA

Structural and conformational studies indicate that the globular domain of the linker histone H5 binds asymmetrically to 2-DNA duplexes (Goytisolo, Gerchman, et al., 1996; Ramakrishnan et al., 1993). A MNase study has shown that the globular domain can protect 2 full turns of a superhelical DNA.

(Y. B. Zhou et al., 1998) suggested that the linker histone position favors interactions with the linker DNA and the DNA near the dyad axis. Recent experiments with H1.5 show binding at the minor groove of the nucleosome dyad axis (Syed et al., 2010). (Hamiche et al., 1996) has proposed that upon binding of the linker histones, the geometry of the terminal linker DNA is altered.

In vitro models and the proposed mode of binding of linker histones to nucleosomes

In vitro studies of chromatosome have suggested on-dyad and off dyad modes of binding of the globular domain of linker histones to the nucleosome. A Cryo-EM and crystallography study that investigated the mode of binding of *X. laevis* H1b and *human* H1.5 vertebrate linker histones with a 197 bp nucleosome harboring 25 bp of linker DNA on either side, suggested an on-dyad binding mode. This gave

insights regarding a significant conformational change in the linker DNA region upon linker histone binding and a partial imposition of asymmetry to the nucleosome due to the CTD binding to one of the linker DNA sections (Bednar et al., 2017). Hydroxyl radical foot printing assays, giving single base pair nucleosome binding mapping, suggests an on-dyad binding mode for linker histones (Syed et al., 2010).

On the contrary, photo crosslinking experiments of the globular domain of chicken H5 suggest an off-dyad binding (Y. B. Zhou et al., 1998). An *in vivo* analysis of linker histone binding using photobleaching experiments with modeling suggest binding near the dyad (Brown et al., 2006). Also, a 11 Å Cryo-EM structure of nucleosomal array harboring H1.4 reveals an asymmetric off-dyad mode of binding of linker histone (F. Song et al., 2014).

The differences in the mode of binding to the nucleosome have been attributed to linker histone variants and species-specific differences. NMR studies indicate that the globular domain of chicken H5 binds on-dyad to the nucleosome (B. R. Zhou et al., 2013), while that of the *Drosophila* H1 binds off-dyad (B. R. Zhou et al., 2015). In a follow-up NMR spin-labeling study (B. R. Zhou et al., 2016), mutation of 5 DNA interacting residues in the globular domain of H5, replaced with that of *Drosophila* H1, indicated an off-dyad binding mode. There was a change in the sedimentation coefficient of this mutated construct to mimic that of *Drosophila* H1. Human linker histone H1.0, the mammalian homologue to H5, also indicates an on-dyad binding mode with NMR analysis.

Role of N and C-terminal regions

Studies on the N and C-terminal regions flanking the globular domains have led to proposals for their roles in the formation of higher order structures and other subtype specific functions (Allan et al., 1986). Deletion of the N-terminal region of a bovine linker histone showed marked reduction in the affinity for chromatin (Hendzel et al., 2004; Oberg & Belikov, 2012). Nucleosome repeat length [NRL] is a measure of the characteristic linker histone interactions *in vivo*. A heterologous expression of human H1.4 in *X. laevis* oocyte extract that lacks somatic linker histone variants showed altered NRL that had a value corresponding to that of H1.4

binding. The FRAP approach with domain swapping experiments conducted to study the role of the N terminal region of mouse H1o and H1c revealed an exchange of their chromatin binding affinities (Vyas & Brown, 2012). These studies suggested that the N-terminal domain [NTD] plays a role in proper binding to nucleosomes, however this warrants additional studies.

The C-terminal domain [CTD] that accounts for more than half of the linker histone sequence varies substantially among the different variants. With an average amino acid composition of about 40% lys, 20-35% ala, 15% pro, this region remains intrinsically disordered in solution (Hansen et al., 2006). The CTD has been shown to be essential for the formation of higher order structures (Allan et al., 1986; Lu et al., 2009), (Hendzel et al., 2004). In the context of nucleosomes, the CTD can interact with an additional 21 bp of linker DNA. Furthermore, deletion and truncations of CTD showed altered affinity to nucleosomes, with some shorter truncations having increased affinity. This has been attributed to the intrinsically disordered nature of the CTD and the binding events being influenced by the enthalpy and entropy related modulations caused by the cellular environment. (Caterino et al., 2011). CTD swapping experiments with sedimentation analysis have suggested that the amino acid composition of the domain is important for its chromatin based function (Lu et al., 2009). Other studies support the theory of CTD taking a folded conformation upon binding to DNA at physiological salt concentrations (Fang et al., 2012).

Linker Histone purification

Purification of full-length linker histones in high amounts using bacterial (Hayashihara et al., 2010; Pyo et al., 2001) or yeast expression systems (Albig et al., 1998) has proven difficult (Gerchman et al., 1994). The major issues have been low expression, association with chromatin DNA affecting yield and lots of lower molecular weight contaminants due to either partly degraded or incompletely made proteins. Therefore, one of the challenging aspects for studying chromatosome by X-ray crystallography approach has been to obtain large amounts of highly pure linker histone proteins.

Optimization of first five codons seemed to improve the expression and purification of chicken H5 (Gerchman et al., 1994). (Pyo et al., 2001) expressed and purified human H1.5 using bacterial *E.coli* system. They had employed a C-terminal Histidine tag and purified using nickel affinity chromatography and Heparin agarose. A summary of the methods for purifying non-tagged linker histones under non-denaturing conditions is presented in (Hayashihara et al., 2010). A modified protocol for purifying recombinant linker histones expressed in bacterial BL21 DE3 plysS cells exploited the high affinity binding of linker histones to DNA. The nucleoprotein component was separated from the rest of the bacterial cell lysate components by centrifugation, followed by sonication, hydroxyapatite chromatography and ion exchange chromatography. Small molecular weight contaminants were present at earlier stage like in other studies. In general, this approach was helpful in improving the yield.

Maintenance of higher order chromatin structure

The basic level of packaging of DNA is into nucleosomes, resulting in a beads-on-a-string appearance, which can further fold into 10 nm, 30 nm chromatin fibers and beyond that into higher level structures, ultimately up to highly condensed chromosomes [Figure 1.2]. Early electron microscopy studies and light scattering measurements have established the formation of salt-dependent higher order structures of chromatin. Linker histone proteins contributed to the stabilization and regulated arrangement of nucleosomes in the presence of Mg^{2+} or high concentrations of monovalent ions (Thoma et al., 1979). A single molecule imaging technique to study reconstituted nucleosome array showed the salt-dependent transition from a beads-on-a-string appearance at low salt [50mM] to partially aggregated 20-30nm structures at about 100 mM salt (Hizume et al., 2005).

Existing Chromatin models of 30nm structure

An early model of digested native chromatin based on cryo-EM study is the solenoidal model (Finch & Klug, 1976). In the presence of Mg^{2+} ions and linker histones, the nucleosome filaments folded into supercoiled structures of a solenoidal type. Other models include the two-start helical ribbon and the twisted cross-linker model. This proposes a zig-zag arrangement of adjacent nucleosomes with a straight linker DNA. The orientation of the DNA is either parallel to the axis, as in a helical

ribbon model or perpendicular to the axis, as in a crossed linker model. In a two-start model, the fiber dimensions could vary linearly with the length of the linker DNA, unlike a solenoid model. A crystal structure of a tetranucleosomal array has two stacks of nucleosome arrangement, but the array is too short to form a solenoidal arrangement (Schalch et al., 2005).

A Cryo-EM study employed a homogeneous *in vitro* system with reconstituted nucleosomes of varying linker DNA lengths (F. Song et al., 2014). The 11 Å resolution structure showed an asymmetric binding of linker histone [human H1.4] that favors a zig-zag stacking of nucleosomes in a left-handed helical manner. The tetranucleosomal arrangement was formed by 2 stacks of two nucleosomes [Figure 4.4], consistent with the appearance of dinucleosomes resulting after DNaseI digestion of native chromatin (Staynov, 2008). The linker DNA was in a straight extended form and an increase [10bp] in the linker length increased the fiber dimensions without affecting the overall structure, supporting the proposed two-start helix model. The presence of the linker histone seemed to introduce a twist to accommodate the varying linker DNA lengths [Figure 4.5].

Figure 4.4 Crystal structure of a tetranucleosome.

Three tetranucleosome units are shown in shades of red, yellow and blue, indicated by 1, 2 and 3, respectively [Staynov et al., 2008].

Figure 4.5 Cryo-EM structures of chromatin fiber formed from reconstituted nucleosome arrays.

Chromatin fiber models to accommodate varying linker DNA lengths are depicted.

A-C. Summary of cryo-EM results for different types of arrays [indicated above images]. Scale bar represents 11 Å. Adapted from Song et al., 2014.

Other cryo-EM studies on the dimensions of compacted chromatin structure revealed interdigitation that occurs due to inter-nucleosomal interactions (P. J. Robinson & Rhodes, 2006). This was based on an *in vitro* uniformly distributed 72 nucleosome array containing varying linker DNA lengths, in the range of 10–70 bp, to compact into a uniform chromatin fiber. An interesting observation was that increasing lengths did not have yield a linear rise in fiber diameter, but modulated nucleosome density within the fiber. A comparative study with 197 bp and 167 bp containing arrays suggested NRL dependent linker histone mediated compaction in determining the fiber dimensions (Routh et al., 2008). Modeling and EM-assisted nucleosome interaction capture [EMANIC] study (Grigoryev et al., 2009), indicated a heteromorph model of chromatin fiber consisting of both two-start and solenoid

type arrangements. The model indicates that linker histones mediate interactions with alternate ($i\pm 2$) nucleosomes, and the presence of divalent cations induce DNA bending at linker DNA that favor interactions between adjacent nucleosomes ($i\pm 1$, also $i\pm 3$). The nucleosome chain folds into an energetically favorable chromatin fiber with a dimension of about 32 nm. This could pave way for tighter interactions towards higher order compacted structures. A novel Micro-C technique, a modified Hi-C method using micrococcal nuclease assay, was developed to map the genome in yeast at single nucleosome resolution. This revealed the presence of short-range interactions *in vivo*, consistent with the tri-or tetra-nucleosomal zig-zag model (Chen & Li, 2017; Hsieh et al., 2015). Another *in vivo* study conducted in human intact interphase nuclei, using RICC-seq [Ionizing Radiation-induced Spatially Correlated cleavage of DNA accompanied with a deep-sequencing] technique demonstrated variable longitudinal chromatin compaction. It supported the two-start helical model comprising tri-or tetranucleosomal units in heterochromatin regions (Risca et al., 2016). There are several Monte Carlo simulation studies aimed at understanding the role of linker histones, core histone tails (Arya & Schlick, 2006), nucleosome repeat length and the salt-dependency of chromatin compaction (Beard & Schlick, 2001). A simulation study (Stehr et al., 2008) identified the strength of internucleosomal interactions as a crucial factor in the compaction of chromatin to higher order structures.

Lack of evidence for 30 nm fiber in vivo

Ultra SAXS and cryo-EM studies of mitotic chromosomes of human HeLa cells failed to detect any regular arrangement of chromatin fibers over 11 nm (Nishino et al., 2012). However, these studies argue against a requirement for this regular arrangement, although 30 nm fiber could be transiently present in a low population depending on the cell state and cell type. Nucleosome organization may be fractal in nature, and heterochromatin appears as heterogeneous ‘clutches’ of nucleosomes in interphase. Specifically, this *in vivo* (Ricci et al., 2015) study of chromatin fiber appearance in human fibroblasts interphase nuclei used the STORM technique. They analyzed the nucleosome count and distribution in fixed and living cells. There was a noticeable difference in the nucleosome patterns in human fibroblasts and in Trichostatin [TSA] [an inhibitor of histone deacetylase that leads to increased

acetylation and more decondensation of chromatin] treated cells. Nucleosome localization was closer to the periphery and was significantly reduced in the TSA treated cells. Moreover, there was no noticeable 30 nm fiber type of nucleosomal arrangement, but rather the distribution involves small clusters [clutches] of nucleosomes, which were more prevalent in the heterochromatin areas. This was also confirmed by staining for centromere regions. Larger nucleosome clutches had significantly higher amounts of linker histone and lower RNA polymerase II activity compared to the less dense regions.

H1x-CMS169s structure:

Our most recent success of obtaining chromatosome crystals and solving its structure (Adhireksan et al., unpublished data) provided the basic framework for this thesis project. The 2.7 Å resolution structure of a chromatosome comprising human H1x with a nucleosome of 169 bp DNA reveals significant information on the role of linker histones, core histones and divalent cations in chromatin compaction. This system is applicable to the study of other linker histone variants.

The chromatosome crystals, although homogeneous, provide a good system for achieving a densely packed environment. The use of sticky end DNA constructs favored the lattice contacts between two nucleosome ends. This formed an ordered lattice arrangement of a pair of adjacent nucleosomes in close contact, resulting in an asymmetric unit comprising a dinucleosome.

Figure 4.6A shows the lattice arrangement of dinucleosomes in one direction. The nucleosomes have a face-to face interaction along the lattice with the adjacent dinucleosomes from different layers positioned in opposite orientation. This results in a zig-zag arrangement of paired nucleosomes. Furthermore, the paired nucleosomes in the orthogonal direction are offset to fit in the available space between the paired nucleosomes [Figure 4.6B]. The system is robust in terms of providing a conducive compact environment to study linker histone-mediated interactions.

Figure 4.6 Lattice arrangement of nucleosomes in the H1x containing chromatosome crystal.

The asymmetric unit is formed by paired-nucleosomes, indicated in green color.

A. Face-to-face stacking of the nucleosomes in one direction. **B.** Interdigitation and tight packing interactions in the orthogonal direction.

The binding mode of the globular domain of the H1x linker histone on the paired nucleosome indicated a 1:1 stoichiometry of binding. Figure 4.7 shows the binding of two linker histones to the paired nucleosomes. Interestingly, each linker histone can contact nucleosomes across several layers and is able to interact with up to 7 different nucleosomes. These interactions can be distinguished as host, lateral, proximal and distal. In turn, each nucleosome in the lattice acts as a host, lateral,

proximal and distal target of different linker histones [Figure 4.8]. Thus, a complex mapping of linker histone in a dense chromatin context was possible. Binding of linker histone to one side of the nucleosome imparts asymmetry to an otherwise symmetric nucleosome. There is a huge surface area associated with linker histone protein interactions. The details of these interactions are presented in the (Adhireksan et al., unpublished data). Given the apparently favorable lattice arrangement in the H1x crystals, we pursued further study of the properties of nucleosome binding and chromatin condensation imparted by other linker histone variants.

Figure 4.7 The asymmetric unit comprising two H1x molecules bound to a nucleosome pair.

A pymol generated view of H1x-169s chromatosome structure (Adhireksan et al., unpublished data) displaying the asymmetric unit. The 1:1 stoichiometry of H1x to nucleosome is still maintained. The linker histones present a very large surface area of interactions. The two nucleosomes are represented as cartoon and the linker histones are represented as a transparent surface view. The two Linker histone-169s complexes are shown in either brown or blue color.

Figure 4.8 Linker histone interaction with multiple nucleosomes

Depiction of complex interactions throughout the lattice. **A.** Each linker histone interacts with several nucleosomes across layers. **B.** Multiple linker histones can differentially interact with different regions of a nucleosome [bottom panel]. DNA [cartoon representation] of nucleosomes in the same layer is shown with similar colors. Different linker histones are represented using surface view.

Scope of this project

Linker histones remain a major component of chromatin organization. However, their mode of interaction(s) and mechanism(s) of compaction have long been speculated over and debated. In the following work, we present a system for studying the complex of linker histones with nucleosome by X-ray crystallography. This lays the foundation to study chromatin in a compacted setting.

The previous work from our lab was focused on improving lattice order with engineered DNA constructs of varying lengths and optimized for linker histone binding to nucleosomes in the presence of divalent ions. We successfully obtained a 2.7 Å resolution dataset for the H1x-nucleosome complex crystallized under near physiological concentrations of calcium ions (Adhireksan et al. (unpublished data). This provides important insights which we pursue further to understand variant-specific similarities and differences and explore the basic mechanisms underlying the formation of condensed chromatin.

The interactions being highly dynamic, studying by structural means will help capture some of the important dominating interactions between the nucleosome and the linker histones. Studying different variants can help discern the general principles in the mode of recognition of nucleosome by linker histones and highlight some of the variant specific differences.

Materials and methods

Linker histone protein purification

Constructs

Constructs of human linker histone variants, codon optimized for bacterial expression, were cloned into expression vector pET15b by Genescript. These constructs had either a cleavable N-terminal [NT] 6x histidine tag harboring a HRV3C or a TEV protease site.

Recombinant bacterial expression

Linker histone constructs containing plasmids were transformed into competent BL21DE3 or BL21DE3 pLysS cells using standard transformation protocol [methods section of FOXA1 chapter]. A starter culture of about 100 ml was started and distributed to 6 l of 2XTY media. Following growth until an OD of 0.6 to 0.8, the cells were induced with 1uM IPTG for 3 hours at 37 °C.

Protein purification

The cells were spun down at 6000 rpm for 6 minutes at 4°C and re-suspended in H1 lysis buffer [50mM Tris-Cl pH 8.0, 2 % (v/v) TritonX-100] with 0.4 g of lysozyme supplemented with PMSF and Benzamidine, and allowed to stir for 1 hour at 4 °C.

Following the lysozyme treatment, the contents were spun at 20,000 rpm for 20 min at 4°C. The pellet was re-dissolved in H1 lysis buffer with 1M NaCl [+PMSF +Benzamidine]. A brief sonication for 3 min at amplitude of 35, 30 s on/30 s off pulse, helped with the consistency. The contents were allowed to stir overnight at 4 °C. This helped in partitioning the DNA bound linker histones to appear in the soluble fraction.

Following the solubilization in high salt, the contents were spun at 20,000 rpm for 15 min at 4°C. The supernatant contained the linker histone proteins although there may be some protein in the insoluble fraction as well.

The contents were loaded onto an IMAC column for histidine tag based affinity chromatography. Once bound to the column, a pre-wash with high salt buffer of

about 2 M NaCl helped in removing a lot of contaminating proteins and DNA. The bound protein was then eluted with an imidazole gradient of 0 mM to 500 mM in buffers containing 50 mM Tris-Cl pH 8 and 500 mM NaCl.

The peak fractions were pooled and the salt concentration was diluted from 500 mM to about 150-250 mM. The sample was loaded onto a Heparin column and eluted with a salt gradient of 150 mM to 1M in 50 mM Tris-Cl pH 8 buffer. Linker histones eluted at very high concentrations of salt, close to 1 M. Different variants eluted at different concentrations, over 70 % of the gradient.

The peak was broad and the fractions were tested on SDS-PAGE gels. There were multiple bands. The protein tended to run a bit higher than its molecular weight. The upper most band that eluted at the tailing end of the peak [usually the last few fractions] contains the protein of the correct size. Only highly pure fractions were pooled and subjected to overnight TEV/ HRV3C protease treatment to cleave off the Histidine tag. On the following day, the sample was passed through the IMAC column and the linker histone proteins eluted in the flow through or at low imidazole concentrations [10 mM]. The protein sample was concentrated and buffer exchanged to 20 mM K-Cacodylate pH 6.5 using the Amicon spin concentrators. The final pure protein was stored as small aliquots directly suitable for crystallization.

DNA constructs

The palindromic CMS 169s construct was used. The sticky end DNA production protocol has already been described in the Chapter 3 Methods.

Histone octamer purification

Equimolar amounts of each of the core histones in unfolding buffer were mixed and allowed to dialyze versus 2 M salt containing refolding buffer [details presented in the Chapter 2, Methods section]. The refolded octamer was purified by size exclusion chromatography, concentrated using spin concentrators, and stored in 50 % glycerol containing buffer at -20 °C, until further use.

Nucleosome reconstitution

A 4uM reconstituted NCP was made by mixing different ratios of DNA to purified histone octamer and performing a sequential salt dilution by dialysis (Dyer et al., 2004). The quality of the NCP was tested using 6 % native PAGE gels under 0.25x TBE conditions.

EMSA binding studies

Binding of linker histone constructs with nucleosomes were tested in binding buffer containing 20 mM K-cacodylate pH 6.0. Protein preparations were added in increasing molar ratios to a nucleosome sample containing buffer. The mixture was incubated for 10 min on ice and tested on 6 % PAGE under 0.25x TBE running buffer conditions. All the gels were stained using Ethidium bromide and visualized with UV trans illuminator of G-Box from Syngene.

Crystallization

Crystallizations were attempted with the CMS169s nucleosome construct. The set up involved 4 mg/ml of the complex containing 1.2:1 molar ratio of recombinant human linker histone proteins to nucleosome, which was crystallized using a salting-in approach. In this case, 1 μ l of nucleosome was mixed with 2 μ l of the 4x crystallization buffer [final will be 2x in the drop] and 1ul of linker histone protein with thorough mixing. The 1x crystallization buffer, which was added to the well, contained 10mM Na-acetate pH 4.5, 25 mM KCl, 40-45 mM CaCl₂. Crystal optimization centered around the primary crystallization conditions, especially by varying the concentration of calcium ions, size of the drop, and concentration of the complex in the initial setup. The crystals were formed by hanging drop method against a reservoir containing crystallization buffer at 1x concentration.

Data collection

The crystals were harvested in 1x buffer conditions and exchanged with increasing concentrations of 2-methyl-1,3 propanediol (MPD) in 4 % increments to up to 25 % MPD. To prevent rapid crystal drying during mounting, Trehalose was also incorporated in the final buffer to 2 %. This was achieved by making 2 stocks of base buffers [10mM sodium acetate pH 4.5, Potassium chloride, Calcium chloride]

with 0 % MPD and 0 % Trehalose and another buffer with 25 % MPD and 2 % trehalose. The 2 stocks were then mixed to achieve the intermediate increments of MPD and Trehalose.

Preliminary X-ray diffraction studies were conducted at the X-ray diffraction facility, A-star, IMCB. It is a Bruker X8 PROTEUM machine with a PLATINUM135 CCD detector and a 4-circle KAPPA goniometer. Data sets for about 180 images were obtained for an oscillation of 0.5° with an exposure of 30 sec.

X-ray diffraction data for analysis were collected at the beam line X06SA with a PILATUS 2M-F detector, at the Swiss Light Source of the Paul Scherrer Institute, Villigen, Switzerland. Complete data sets were generated for 360 images collected at 0.5° oscillations, with 1 sec exposures.

Data analysis

The data collected at the SLS of PSI was automatically processed using the “go.com” and the output files were obtained. Alternatively, the data was integrated and scaled using iMOSFLM (Powell et al., 2017) and Scala software of the CCP4 suite (Collaborative Computational Project) (CCP4, 1994)). Molecular replacement was conducted with different existing nucleosome and linker histone globular domain structures. A rigid body and restrained refinement using Refmac5, CCP4 was used to evaluate the fit of the models used. Model building was carried out using Coot software (Emsley et al., 2010).

Results and Discussion

Purification of Linker histone variants

Although the basic protocols were set, recombinant expression and purification of specific linker histone variants using bacterial expression system required additional optimization. Owing to the differences in the sequence, each linker histone subtype behaved differently in terms of their expression, elution profile in heparin, sensitivity to protein stain and absorbance to UV. We have used a cleavable N-terminal 6x histidine tagged codon optimized constructs for recombinant bacterial expression. Purification involved majorly nickel affinity chromatography and heparin column purification.

Human H1.2

The overall purification of human H1.2 was compatible with the basic linker histone protocol. A hexahistidine tagged human H1.2 construct was expressed using RIPL cells and was subjected to lysis with lysozyme treatment followed by an overnight high salt extraction in protease containing buffers rendered the linker histones in the soluble fraction. Following IMAC purification, the linker histone proteins were concentrated in the peak. The peak fractions contained a lot of lower bands [Figure 4.9 A]. The pooled fractions were then subjected to the heparin column purification. This helped get rid of DNA contamination and the protein peak was broad. The lower molecular weight bands got separated in the initial fractions and the trailing end of the peak contained the protein of the correct size [Figure 4.9 B]. This pool was subjected to an overnight TEV protease treatment [Figure 4.9C]. Upon the removal of the histidine tag, the sample was concentrated on Heparin column using a steep gradient. We could obtain highly pure protein preparation [Figure 4.7D]. However, the final yield was affected due to the membrane concentrators and issues with degradation. This needs to be verified with a different batch or type of membrane concentrators.

Figure 4.9 Recombinant expression and purification of human H1.2 protein
 16 % SDS PAGE gel analysis. **A.** Clarified, salt extracted supernatant following lysis procedure purified on IMAC column. P is the input. Lanes 1-8, the peak fractions have additional lower molecular weight bands. **B.** Heparin column purification of the pooled fractions from IMAC, lane 'P'. FT is the flow through. Lanes 1-15 are the peak fractions. Trailing end of the peak marked as pool has the correct molecular weight protein, indicated by the blue triangle. **C.** Overnight TEV protease treatment to remove the histidine tag. HT-H1.2 and H1.2 refers to the protein with and without tag, respectively. **D.** Final Heparin column run to concentrate and remove the contaminating bands yields highly pure protein [lanes 3-9]. M- Marker.

Human H1.4

Following the basic linker histone protocol for recombinant expression and purification we could concentrate the recombinant human H1.4 protein after IMAC elution [Figure 4.10A]. In general, for all linker histones, during the heparin column purification, the highest MW species of the correct size elutes at the far end of the gradient in high salt [1 M NaCl]. In the case of H1.4 the relative proportion of this is low compared to other lower molecular weight species [Figure 4.10B and Inset 1]. Following the removal of tag [Figure 4.10C] and another heparin column purification to separate the two protein bands, we pooled the fractions of utmost purity of one of the species [Figure 4.10D and Inset 2]. Since this higher molecular weight was present before the TEV protease treatment, inset 1, this is not due to incomplete tag removal. The molecular weight of this pool containing pure protein was compared against H1t as a control and it runs a bit lower. We suspect that the upper band that co-elutes with the lower high yield protein band could be the full-length H1.4 protein. Therefore, the expression and the purification need to be further optimized to skew towards making more of the full-length H1.4 protein. Alternatively, using mass spectrometry approach we could map the correct size of the purified substance.

Figure 4.10 Recombinant expression and purification of human H1.4

16% SDS PAGE gel analysis of **A**. Clarified, salt extracted supernatant following lysis procedure purified on IMAC column. The peak fractions [lanes 1-10] have additional lower molecular weight bands. **B**. Heparin column purification of the peak fractions from IMAC [indicated as Pre and FT is the late Flow through. Lanes 1-17 are the peak fractions. Trailing end of the peak marked as 'pool' has the correct molecular weight protein, indicated by a blue triangle in the inset 1, zoomed in view of lanes 16 and 17. **C**. Overnight TEV protease treatment of his tagged H1.4, represented as HT hH1.4 and the tag cleaved protein, represented as hH1.4. **D**. Final Heparin column run to concentrate and remove the contaminating bands. 'Pool' indicates the pooled fractions. Inset 2 is a zoomed in view of lanes 8 and 9. **E**. Final shorter product of human H1.4 compared with human H1t.

hH1.5

This linker histone subtype could be easily purified using the standard linker histone purification protocol. IMAC pure, Heparin pure fractions are indicated in the Figure 4.11 A and B, respectively. Figure 4.11C shows the TEV treated samples. The gel needed to be run longer to see complete separation [not shown]. The protein band indicates the purity and the molecular weight have been verified by MALDI-TOF Mass spectrometry.

Figure 4.11 Recombinant expression and purification of human H1.5

16% SDS PAGE gel analysis. **A.** Affinity tag-based IMAC column purification. The peak fractions, lanes 1-5, have additional lower molecular weight bands. **B.** Heparin column purification of the peak fractions from IMAC [lane P]. LFT is the late flow through. Lanes 4-14 are the peak fractions. Trailing end of the peak marked as pool has the correct molecular weight protein. **C.** Overnight TEV protease treatment of his tagged H1.5, represented as HT hH1.5 and the tag cleaved protein, represented as hH1.5.

Purification of gamete specific linker histone variants

hH1t

For human H1t, recombinant expression and purification required a few Heparin column runs in addition to the basic linker histone purification protocol. This requirement is based on the expression, yield and the presence of contaminating bands. Figure 4.12 is a representative of one such purification. In addition to the major band there had been a small fraction of a closer degradation band. However, the molecular weight is in the correct range, verified using MALDI/TOF.

Figure 4.12 Recombinant expression and purification of human H1t

16% SDS PAGE gel analysis. **A.** Affinity tag based IMAC column purification. The peak fractions, lanes 1-9, have additional lower molecular weight bands. **B.** Heparin column purification of the peak fractions from IMAC. Trailing end of the peak marked as pool has the correct molecular weight protein. **C.** Overnight TEV protease treatment of histidine tagged H1t, indicated as HT hH1t and the tag cleaved protein, indicated as hH1t. **D.** Final pure H1t product of this preparation.

hH1oo

Human H1oo recombinant expression and purification proved difficult. The basic linker histone protocol failed at different levels. After optimization of expression, it has been difficult to render the protein in the soluble fraction, instead it ends up in the inclusion body cell pellet. A modification of core histone and linker histone protocol might work for this variant. It is a work still in progress.

hH1T2 and hHILS

The two gamete-specific linker histone variants, H1T2 and HILS have a high level of sequence diversity from the other linker histones. Their sequence contains more cysteine and aromatic residues compared to the other variants. The basic linker histone protocol worked. The issues encountered during purification are the same for these 2 linker histone proteins. The expression and the yield have been low. This requires further optimization of expression to improve the yield. Due to the presence of Cysteine residues, the purification involved addition of TCEP in all the buffers. Following TEV protease treatment, it has been difficult to estimate the extent of tag removal based on SDS analysis, especially because these proteins tend to run much higher than their estimated molecular weight on the SDS PAGE gels. The quality of the purified proteins seems to vary between preparations. It has been difficult to obtain an accurate molecular weight estimation by mass spectrometry.

For H1T2 purification, IMAC and the heparin column runs [Figure 4.13 A and B] yielded highly pure protein. TEV protease treatment did not seem to cleave off the tag. The DNA sequence of the construct has been verified to confirm the in-frame presence of the TEV protease site following histidine tag. Since the tag removal proved difficult, we did another heparin column purification to remove the added TEV protease [Figure 4.13 C].

For HILS purification, IMAC and the heparin column runs [Figure 4.14 A and B] yielded highly pure protein. Just like the H1T2 construct, TEV protease treatment did not seem to cleave off the tag. The DNA sequence of this construct has also been verified to confirm the in-frame presence of the TEV protease site following

histidine tag. However, based on the H1x chromatosome and our EMSA studies, the presence of the tag does not seem to interfere with the linker histone behavior.

Figure 4.13 Recombinant expression and purification of human H1T2

16 % SDS PAGE gel analysis. **A.** Affinity tag based IMAC column purification. Lanes 1-10 indicate the peak fractions. **B.** Heparin column purification of the peak fractions from IMAC [pre]. Lanes 1-7 are the eluted fractions. **C.** Overnight TEV protease treatment of histidine tagged H1T2, represented as HT hH1T2 and +TEV indicates addition of TEV protease enzyme. **D.** Final heparin purification to remove TEV enzyme. Lanes 1-6 are the eluted fractions M- Marker, Pre- input, LFT- later flow through.

Figure 4.14 Recombinant expression and purification of human HILS

16% SDS PAGE gel analysis of **A.** Affinity tag based IMAC column purification. Pre-input material. The peak fractions are from lanes 1-9. **B.** Heparin column purification of the peak fractions from IMAC [pre]. Blue triangle indicates the protein of the correct molecular weight. The subsequent purification steps are mentioned in the flowchart to the right.

Binding studies with 169s nucleosome

By EMSA, the purified protein preparations were characterized for the binding to 169s containing nucleosomes using 6 % PAGE under 0.25x TBE gel running conditions. The binding reaction was setup in 20 mM K.Cacodylate 6.5 buffer. Figure 4.15 shows a representative EMSA study of linker histones H1.5, H1T2 and H1t binding to the reconstituted 169s nucleosome. There is a single shifted nucleosome band with increasing ratios of protein to nucleosome [blue triangle pointer in Figure 4.15]. With increasing amounts of protein, there is an accumulation of aggregates in the well that does not migrate into the gel [red triangle pointer in Figure 4.15]. The different affinity and dynamic interactions of different variants, presence of cysteine residues in H1T2 and HILS, and differences in the protein preparations could contribute to the different behavior observed on the EMSA gel retardation analysis. In general, the protein: nucleosome molar ratio of 1.2 was chosen for crystal setups.

Figure 4.15: 6% PAGE EMSA of linker histones binding to 169s nucleosome
A, B, C and D panels represent the 169s nucleosome binding studies of linker histones H1.5, H1T2, H1t and HILS, respectively. N is the 169s nucleosome only control. The numbers above the lane represent the molar ratios of protein to nucleosome. Blue triangle points to linker histone-169s nucleosome complex. The red triangle indicates the aggregates that accumulate at higher molar ratios.

Crystallization setup

The details of the crystallization setup are provided in the methods section. The crystal morphology was heterogeneous and remained crowded in a single well. It was difficult to isolate one type to represent. The crystal images presented below are representative of the different morphologies mounted and tested for diffraction [Figure 4.16].

Cryogenic conditions

Our previous work on H1x-chromatosome crystals indicated that the quality of diffraction depends on the cryo-processing techniques. Formed under the same crystallization conditions, compared to the control 169s nucleosome crystals [can withstand up to 25 % MPD], the chromatosome crystals were able to withstand extensive dehydration [up to 65 % MPD]. The presence of the linker histone H1x contributed to this ability to withstand pronounced dehydration and correspondingly allows for a fantastic gain in diffraction quality. The same procedure was applied to the other linker histone-specific chromatosome crystals.

Due to the highly heterogeneous population of the crystals, not all the crystals survive extensive dehydration conditions. The susceptibility of the crystals to withstand longer dehydration seems to depend on the specific linker histone variant present. Although most of the smooth surface crystals could withstand the extensive dehydration, good diffraction showed a time dependence of exposure to extensive dehydration. For example, H1.5-chromatosome crystals could diffract up to 3-3.5 Å even after prolonged exposure to dehydration conditions (for even a day), compared to the other variants that were more sensitive. With H1t containing crystals, although having good appearance, the chances of displaying good diffraction ability was low. H1T2 or H1LS could not survive longer cryogenic treatments. They seem to demonstrate a bell-curve time dependency behavior with the maximum diffraction observed after 2-3 h of extensive dehydration. Much longer incubations with high MPD concentrations meant crystals did not diffract very well. These general observations need to be further verified in our follow up diffraction studies.

Data analysis

The data was processed in 3 different ways: 1) Using the automated go.com program of the synchrotron facility, 2) With XDS or 3) Using iMosflm and scala and the statistics were compared. Representative data collection statistics for the different chromatosome crystals is presented in Table 4.1 and a summary of the processing results of individual crystals for the various complexes is provided in Table 4.2.

H15 chromosome crystals

H1t chromosome crystals

H1T2 chromosome crystals

H1LS chromosome crystals

Figure 4.16: Crystal images for the different linker histone-nucleosome complexes
 Different chromosome crystals that were tested for diffraction are shown. The labels above each lane indicate the linker histone variant present in the chromosome crystals. The yellow rectangle coincides with dimensions of 50 X 90 μM .

Table 4.1 Data collection statistics for chromosome crystals composed of different linker histone variants

	H1T2	H1LS	H1t	H1.5
Space group	P2 ₁	P2 ₁	P1	P1
Cell dimensions				
<i>a</i> (Å)	104.08	104.43	127.86	103.67
<i>b</i> (Å)	102.65	102.63	144.07	174.70
<i>c</i> (Å)	216.08	216.35	148.68	200.77
$\alpha/\beta/\gamma$ (deg.)	90/98.21/90	90/98.03/90	74.64/73.1/65	109.17/94.04/95.83
Resolution (Å)	3.22–49.23 (3.22–3.39)	3.25–49.34 (3.25–3.42)	3.26–49.31 (3.26–3.44)	3.02–49.16 (3.02–3.18)
Unique reflections	72642	70316	137747	250261
<i>R</i> _{work} (%)	5.5 (94.5)	9.0 (85.6)	8.3 (90.3)	7.4(62.7)
<i>I</i> / σ	12.1 (1.0)	9.2(1.1)	8.5(1.1)	8.3 (1.0)
CC½ (%)	99.9 (37.0)	99.2 (46.1)	99.9 (41.4)	99.7 (54.5)
Completeness (%)	98.7 (93.7)	98 (94.8)	97.5 (93.0)	96.9 (91.2)
Redundancy	3.2 (2.7)	3.3 (3.2)	1.8 (1.8)	1.8 (1.7)

Single crystal data. Values in parentheses are for the highest-resolution shell.

Table 4.2 Summary of data processing results

Human Linker Histone constructs	Resolution	Unit cell dimensions	Number of nucleosomes in the asymmetric unit
H15	3.26Å	P1 [103.66,174.70,200.77];[109.17,94.04,95.83]	5-6
H1t	3.25Å	P1 [127.85,144.07,148.67]; [74.63,73.11,65.01]	4
H1T2	3.57Å 3.25Å 3.4Å 3.24Å 3.32Å	P1 [102.42,102.65,111.92];[81.50,72.98,88.50] P1 P1 P2 ₁ [104.08,102.66,216.08];[90.0,98.21,90.0] P2 ₁	2
H1LS	3.3Å 3.25 Å	P1 [103.31, 103.74, 113.22];[81.22,73.06,88.49] P2 ₁ [104.43, 102.63, 216.35];[90.0,98.03,90.0]	2

Based on the number of molecules in an asymmetric unit, of the gamete-specific linker histone variants, I focused on the H1T2 and H1LS chromosome diffraction data. The initial structural solution was achieved with the automated Phaser molecular replacement routine in CCP4. For this, we used the nucleosome model [lacking any disordered core histone tail regions and the linker histones] of the asymmetric unit of the previously solved H1x chromosome structure (Adhireksan et al., unpublished data). The lattice arrangement of the nucleosomes is like that of the H1x chromosome crystals, described in Figure 4.6A and B.

H1T2 chromosome structural analysis

For the H1T2 chromosome data, rigid body refinement with restrained refinement of this initial model indicated differences in the orientation of the linker DNA regions relative to the H1x chromosome (linker histone-free) model used for molecular replacement. If we use the asymmetric unit comprising two nucleosomes, depicted in Figure 4.6 A as the model, one of the nucleosomes did not fit in the refined electron density map [Appendix figure A4]. This is also evident by the high R values shown in the refinement statistics of molecular replacement using model 1 in figure 4.17. If we use just one nucleosome as the starting model and search for 2 copies, we could obtain better R value statistics [Appendix figure A5]. However, phaser does not model the asymmetric unit as paired nucleosomes. Instead the two nucleosomes are placed in an 'X' orientation, where the linker DNA ends are facing in the opposite directions [Appendix figure A5]. The linker DNA regions were moved into the density of refined electron density maps. Using the generated symmetry mates, a paired-nucleosome model was generated in Pymol. This H1x-chromosome-like structure with altered linker DNA ends was subsequently used as a model for an additional round of Phaser-based molecular replacement. The result of this comparison is summarized in figure 4.18. Model 2, with the modified linker DNA gives a much better fit, as indicated by the R values [Appendix figure A6]. Hence at a preliminary level all this suggests a fundamental change in the linker DNA conformations imparted by the different linker histone variants. The two models are superimposed, as shown in Figure 4.18. The overall nucleosome core structure is very similar, and only the linker DNA ends appear to be very different.

Molecular replacement Model 1			
Nucleosome of solved H1x-chromatosome structure			
	Initial	Final	
R factor	0.4533	0.4104	
R free	0.4604	0.4516	
Rms BondLength	0.0126	0.0073	
Rms BondAngle	1.5911	1.2209	
Rms ChirVolume	0.1004	0.0725	

Molecular replacement Model 2			
Nucleosome of solved H1x-chromatosome structure with modified linker DNA region			
	Initial	Final	
R factor	0.2547	0.2086	
R free	0.2535	0.2418	
Rms BondLength	0.0111	0.0077	
Rms BondAngle	1.7722	1.4391	
Rms ChirVolume	0.1186	0.0914	

Figure 4.17: Molecular replacement and preliminary refinement statistics of the H1T2 data set

Model 1 is the paired nucleosome [asymmetric unit] of the H1x-chromatosome structure without any linker histones [this was the original model used for molecular replacement with the H1T2 data]. Model 2 is a variation of Model 1 with linker DNA regions moved into the density corresponding to the H1T2 data. Model 2 was used for an additional molecular replacement trial. The refinement statistics following rigid body and restrained refinement performed using Refmac are shown to indicate the appropriateness of the models for use in the final structural solution.

Figure 4.18 Superimposition of the two models described in figure 4.17

Model 1 [magenta] is the nucleosome pair of the H1x-chromatosome structure, and Model 2 [cyan] is that of the H1T2-chromatosome [i.e., linker DNA modified to fit H1T2 electron density]. **A.** Superimposition of the nucleosome pairs from the asymmetric units of the H1x and H1T2 chromatosome structures. **B.** Close view of the linker DNA regions.

Following this, we manually searched to find regions of extra density near the linker DNA of different nucleosomes [like in the case of H1x chromatosome] that could coincide with a bound linker histone. Close to one of the linker DNA sticky end base pairing regions and the dyad, there is extra density on either side [Figure 4.19].

The extra electron density, consistent with bound linker histone proteins, is clearly visible in Figures 4.19-4.21.

Figure 4.19 Extra electron density consistent with bound linker histone in the H1T2-chromatosome crystal.

The model stems from molecular replacement with Model 2 of figure 4.17 and 4.18, followed by rigid body and restrained refinement. A 2Fo-Fc electron density map in green, contoured at 0.58σ , is shown superimposed. The DNA at the center represents the base pairing region of the single-stranded termini from two nucleosomes.

Figure 4.20 Extra electron density consistent with linker histone binding in the H1T2 chromatosome

The model stems from molecular replacement with Model 2 of figure 4.17 and 4.18, followed by rigid body and restrained refinement. The top panel shows the region of interest with the host nucleosome shown in brown and the neighboring nucleosomes shown in green. In the bottom panel, an $2F_o - F_c$ electron density map in blue, contoured at 0.41σ , is shown superimposed. The oval shapes indicate regions of extra density corresponding to H1T2 binding.

Figure 4.21 Close view of H1T2 electron density in the chromosome crystal

A close view of a region of extra density depicted in figure 4.20. A 2Fo-Fc electron density map in green, contoured at 0.41σ , is shown superimposed on the partially refined model. The black oval indicates a region of extra electron density that appears to coincide with an alpha-helical element of H1T2.

At present, in the initial phases of model building, we could identify density for 2 alpha-helical regions, as shown in figure 4.22. We also notice extra density for a helix situated further away at a symmetric position [not shown]. At present we are not able to locate density that is consistent with a ‘classic’ globular domain. This could be due to the H1T2 sequence divergence from that of other linker histones. From the multiple sequence alignment shown in figure 4.1 and by comparing the secondary structure prediction [Appendix figure A2] and 3D model predicted using CASP [Appendix Figure A3], it indicates that the globular domain of linker histone comprising the three helices may not be conserved. It seems possible to have an additional helical region in the C-terminal region, marked by residues 197-207 in the appendix figure A3. For instance, H1T2 has a proline residue [P101] at a position that could disrupt the third helix [Custal omega sequence alignment shown figure A1]. A manual building of fitting helices into the density regions looking for clues was attempted, but there appeared to be modeling bias and the result could not be validated. Since we do not have a solved structure for the H1T2 globular domain, we are in parallel pursuing experimental phasing and heavy atom labeling

experiments to target specific sites within the nucleosome or linker histones to obtain additional data for solving the structure.

Figure 4.22 Model building trials for the H1T2 chromosome

A close view of regions of extra electron density coinciding with bound H1T2, depicted also in figure 4.20. A 2Fo-Fc map is shown in red, contoured at 0.41σ , and superimposed on the model. A modeled alpha-helix is shown in green, highlighted with a green oval. The black oval indicates another region of extra electron density that could also be helical.

Further, by manually placing the existing published globular domain of the linker histone structure, we could infer that there is enough space to accommodate the globular domain of a linker histone [Figure 4.23A]. Although there is extra density consisted with helices and other protein elements [Figure 4.23B], it has been difficult to directly fit the globular domain into the density. This may indicate that the globular domain structure of H1T2 is different from existing models of other variants.

Figure 4.23 A Representative figure for the accommodation of globular domain of a linker histone [obtained from published pdb] into the extra density regions of H1T2-nucleosome data.

A. A cartoon representation of the space accommodation of a globular domain of a linker histone. gLH is the globular domain of linker histone from the published H1x structure. The host nucleosome in cartoon representation and the immediate neighboring nucleosome in stick representation are indicated.

B. 2Fo-Fc map of the representation in A indicating regions of extra density surrounding the implanted gLH. Blue ovals indicate additional extra density regions surrounding gLH. The figure is only to indicate the space available to accommodate a linker histone.

HILS chromosome structural analysis

Following initial data processing for the HILS chromosome dataset, we used the H1T2 asymmetric unit [paired nucleosome] model for molecular replacement and searched for extra electron density near the linker DNA regions. Figure 4.24 indicates a region of extra density, that could accommodate a globular domain of the Hils linker histone. Similar to the case of H1T2, the sequence divergence and lack of existing protein structure for hHILS, posed severe model bias and hampered the progress with structure solving.

Hence will still require more work to build atomic models for H1T2, HILS and the other variants. Added complications are that the sequence space of these germ-line variants is so diverse and indeed the secondary structure predictions indicate the helices to be of different lengths. Additional difficulties arise from the high B-factors and highly dynamic unstructured regions of the linker histone proteins. Therefore, solving the structure by molecular replacement alone can pose a substantial model bias issue. Hence, it may help to obtain experimental phase information by heavy atom approaches (Pike et al., 2016). For this, we intend to obtain experimental phase data and site labeling information by using metal compounds that have specific targets on the nucleosome. Currently we are screening different metal compounds in terms of their solubility, concentration dependent sensitivity of the crystals and X-ray absorption wavelength accessibility at the synchrotron beamline.

Figure 4.24 Extra electron density consistent with linker histone binding in the H1LS chromatosome

A 2Fo-Fc map in blue, contoured at 0.41σ , is shown superimposed on the H1T2-chromatosome derived molecular replacement model. The black ovals indicate several regions of extra density consistent with H1LS binding.

Conclusions and future directions

Potential insights from chromosome structures with different linker histone variants may shed light on their differential activities in nucleosome recognition, gene regulation and chromatin structure... Histone variants, post translational modifications and varying linker DNA lengths provide an additional level of complexity. Linker histone with nucleosome complex forms the fundamental unit of a chromosome structure. We have a solved H1x-nucleosome unpublished structure from previous studies in our lab. As an immediate next level, I pursued to study chromosome structures harboring different linker histone variants. This would also serve as controls for the validation of existing structures. There are lots of lingering questions regarding the mode of interactions of linker histone binding, modulation of their interactions with varying linker DNA lengths and existing species-specific differences. X-ray crystallography approach of crystallizing the various linker histones with nucleosomes could provide snapshots of the dominating interactions pertaining to that specific variant. This could shed light on some of the apparent contradictions in the existing literature. As a result, we have crystallized H1.5, H1.t, H1T2 and HILS containing chromosomes. H1.2 and H1.4 have been suspended due to time consideration and purification failures to obtain large amounts of full length protein. Of the successful crystals, we collected good diffraction data set for H1.5, H1t, H1T2 and HILS. Preliminary analysis of the data suggested H1.5 containing crystals had multiple molecules in the asymmetric unit. The H1t containing crystals do not seem to exhibit a defined linker histone density. Hence, I focused on the H1T2 and HILS chromosome structures. We could spot extra density for linker histones near the linker DNA in these datasets. Additionally, for the chromosome with the testis-specific variants, H1T2 and HILS, we observe a change in the linker DNA region compared to the H1x-chromosome, which suggests an at least somewhat different mode of binding and DNA bending. The sequence divergence of these two variants from H1x and the lack of previously solved structures for at least their globular domains means that the initial electron density maps must be quite clear to allow building the linker histone regions de novo in atomic detail. To assist with this, we will be pursuing different phasing and heavy atom labeling experiments to help obtain more accurate electron density maps. Also, solving the globular domains of these linker histones by itself would

provide us with models for molecular replacement. With these being the immediate goals, *in vivo* chromatin interactions are much more complex and degree of chromatin condensed state is influenced by post translational modifications, histone tail mediated interactions, presence of chromatin modelers and proximity to preferred DNA sites. With the current success with chromatosome project, it is feasible to obtain crystal structures of all the linker histone variants, by addressing the issues of linker histone disorder. The overall process of obtaining diffracting crystals has been optimized and could provide a framework for studying additional factor interactions with chromatin in the presence of specific linker histones and for the role of different post translational modifications.

References

- Abrahams, J.P., Leslie, A.G., Lutter, R. & Walker, J.E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature*, 370(6491), 621-628.
- Adams, C.C. & Workman, J.L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol*, 15(3), 1405-1421.
- Adhireksan, Z., Bao, Q., Padavattan, S. & Davey, C.A. (unpublished data). Linker Histone Links Nucleosomes in the Crystal Structure of a Chromatosome.
- Adhireksan, Z., Davey, G.E., Campomanes, P., Groessl, M., Clavel, C.M., Yu, H., Nazarov, A.A., Yeo, C.H., Ang, W.H., Droge, P., Rothlisberger, U., Dyson, P.J. & Davey, C.A. (2014). Ligand substitutions between ruthenium-cymene compounds can control protein versus DNA targeting and anticancer activity. *Nat Commun*, 5, 3462.
- Adhireksan, Z., Palermo, G., Riedel, T., Ma, Z., Muhammad, R., Rothlisberger, U., Dyson, P.J. & Davey, C.A. (2017). Allosteric cross-talk in chromatin can mediate drug-drug synergy. *Nat Commun*, 8, 14860.
- Alami, R., Fan, Y., Pack, S., Sonbuchner, T.M., Besse, A., Lin, Q., Greally, J.M., Skoultschi, A.I. & Bouhassira, E.E. (2003). Mammalian linker-histone subtypes differentially affect gene expression in vivo. *Proc Natl Acad Sci U S A*, 100(10), 5920-5925.
- Albig, W., Runge, D.M., Kratzmeier, M. & Doenecke, D. (1998). Heterologous expression of human H1 histones in yeast. *FEBS Lett*, 435(2-3), 245-250.
- Alfonso, P.J., Crippa, M.P., Hayes, J.J. & Bustin, M. (1994). The footprint of chromosomal proteins HMG-14 and HMG-17 on chromatin subunits. *J Mol Biol*, 236(1), 189-198.
- Ali, T. & Thomas, J.O. (2004). Distinct properties of the two putative "globular domains" of the yeast linker histone, Hho1p. *J Mol Biol*, 337(5), 1123-1135.
- Allan, J., Mitchell, T., Harborne, N., Bohm, L. & Crane-Robinson, C. (1986). Roles of H1 domains in determining higher order chromatin structure and H1 location. *J Mol Biol*, 187(4), 591-601.
- Allfrey, V.G. & Mirsky, A.E. (1964). Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. *Science*, 144(3618), 559.
- 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(6058), 977-982.
- Arya, G. & Schlick, T. (2006). Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proceedings of the National Academy of Sciences*, 103(44), 16236.
- Aung, K.M., New, S.Y., Hong, S., Sutarlie, L., Lim, M.G., Tan, S.K., Cheung, E. & Su, X. (2014). Studying forkhead box protein A1-DNA interaction and ligand inhibition using gold nanoparticles, electrophoretic mobility shift assay, and fluorescence anisotropy. *Anal Biochem*, 448, 95-104.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. (2000). The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*, 289(5481), 905-920.
- Barbera, A.J., Ballestas, M.E. & Kaye, K.M. (2004). The Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen 1 N terminus is essential for chromosome association, DNA replication, and episome persistence. *Journal of virology*, 78(1), 294-301.

- Beard, D.A. & Schlick, T. (2001). Computational modeling predicts the structure and dynamics of chromatin fiber. *Structure*, 9(2), 105-114.
- Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., Menoni, H., Papin, C., Skoufias, D.A., Kurumizaka, H., Lavery, R., Hamiche, A., Hayes, J.J., Schultz, P., Angelov, D., Petosa, C. & Dimitrov, S. (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell*, 66(3), 384-397.e388.
- Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J. & Woodcock, C.L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A*, 95(24), 14173-14178.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res*, 28(1), 235-242.
- Bhan, S., May, W., Warren, S.L. & Sittman, D.B. (2008). Global gene expression analysis reveals specific and redundant roles for H1 variants, H1c and H1(0), in gene expression regulation. *Gene*, 414(1-2), 10-18.
- Birger, Y., Ito, Y., West, K.L., Landsman, D. & Bustin, M. (2001). HMGN4, a newly discovered nucleosome-binding protein encoded by an intronless gene. *DNA and cell biology*, 20(5), 257-264.
- Birger, Y., West, K.L., Postnikov, Y.V., Lim, J.H., Furusawa, T., Wagner, J.P., Laufer, C.S., Kraemer, K.H. & Bustin, M. (2003). Chromosomal protein HMGN1 enhances the rate of DNA repair in chromatin. *The EMBO journal*, 22(7), 1665-1675.
- Bonev, B. & Cavalli, G. (2016). Organization and function of the 3D genome. *Nature Reviews Genetics*, 17, 661.
- Bossard, P. & Zaret, K.S. (1998). GATA transcription factors as potentiators of gut endoderm differentiation. *Development*, 125(24), 4909-4917.
- Brown, D.T., Gunjan, A., Alexander, B.T. & Sittman, D.B. (1997). Differential effect of H1 variant overproduction on gene expression is due to differences in the central globular domain. *Nucleic Acids Res*, 25(24), 5003-5009.
- Brown, D.T., Izard, T. & Misteli, T. (2006). Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nat Struct Mol Biol*, 13(3), 250-255.
- Bustin, M. (2001). Chromatin unfolding and activation by HMGN(*) chromosomal proteins. *Trends in biochemical sciences*, 26(7), 431-437.
- Cao, K., Lallier, N., Zhang, Y., Kumar, A., Uppal, K., Liu, Z., Lee, E.K., Wu, H., Medrzycki, M., Pan, C., Ho, P.-Y., Cooper, G.P., Jr., Dong, X., Bock, C., Bouhassira, E.E. & Fan, Y. (2013). High-Resolution Mapping of H1 Linker Histone Variants in Embryonic Stem Cells. *PLOS Genetics*, 9(4), e1003417.
- Caron, H., Schaik, B.v., Mee, M.v.d., Baas, F., Riggins, G., Sluis, P.v., Hermus, M.-C., Asperen, R.v., Boon, K., Voûte, P.A., Heisterkamp, S., Kampen, A.v. & Versteeg, R. (2001). The Human Transcriptome Map: Clustering of Highly Expressed Genes in Chromosomal Domains. *Science*, 291(5507), 1289.
- Carruthers, L.M., Bednar, J., Woodcock, C.L. & Hansen, J.C. (1998). Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays: mechanistic ramifications for higher-order chromatin folding. *Biochemistry*, 37(42), 14776-14787.
- Caterino, T.L., Fang, H. & Hayes, J.J. (2011). Nucleosome linker DNA contacts and induces specific folding of the intrinsically disordered H1 carboxyl-terminal domain. *Mol Cell Biol*, 31(11), 2341-2348.

- Catez, F., Brown, D.T., Misteli, T. & Bustin, M. (2002). Competition between histone H1 and HMGN proteins for chromatin binding sites. *EMBO Rep*, 3(8), 760-766.
- Catez, F., Yang, H., Tracey, K.J., Reeves, R., Misteli, T. & Bustin, M. (2004). Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin. *Mol Cell Biol*, 24(10), 4321-4328.
- Cattoni, D.I., Valeri, A., Le Gall, A. & Nollmann, M. (2015). A matter of scale: how emerging technologies are redefining our view of chromosome architecture. *Trends in Genetics*, 31(8), 454-464.
- CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, 50(Pt 5), 760-763.
- Chaya, D., Hayamizu, T., Bustin, M. & Zaret, K.S. (2001). Transcription factor FoxA (HNF3) on a nucleosome at an enhancer complex in liver chromatin. *J Biol Chem*, 276(48), 44385-44389.
- Chen, P. & Li, G. (2017). Structure and Epigenetic Regulation of Chromatin Fibers. *Cold Spring Harb Symp Quant Biol*.
- Cherukuri, S., Hock, R., Ueda, T., Catez, F., Rochman, M. & Bustin, M. (2008). Cell cycle-dependent binding of HMGN proteins to chromatin. *Molecular biology of the cell*, 19(5), 1816-1824.
- Chua, E.Y.D. & Sandin, S. (2017). Advances in phase plate cryo-EM imaging of DNA and nucleosomes. *Nucleus*, 8(3), 275-278.
- Chua, E.Y.D., Vasudevan, D., Davey, G.E., Wu, B. & Davey, C.A. (2012). The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res*, 40(13), 6338-6352.
- Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M. & Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*, 9(2), 279-289.
- Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K. & Zaret, K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *The EMBO journal*, 17(1), 244-254.
- Cirillo, L.A. & Zaret, K.S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell*, 4(6), 961-969.
- Cirillo, L.A. & Zaret, K.S. (2007). Specific interactions of the wing domains of FOXA1 transcription factor with DNA. *Journal of Molecular Biology*, 366(3), 720-724.
- Clark, K.L., Halay, E.D., Lai, E. & Burley, S.K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, 364(6436), 412-420.
- Clausell, J., Happel, N., Hale, T.K., Doenecke, D. & Beato, M. (2009). Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF. *PLoS One*, 4(10), e0007243.
- Consortium, T.E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57.
- Costa, S., Almeida, A., Castro, A. & Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Frontiers in Microbiology*, 5, 63.
- Cremer, T., Cremer, M., Dietzel, S., Muller, S., Solovei, I. & Fakan, S. (2006). Chromosome territories--a functional nuclear landscape. *Curr Opin Cell Biol*, 18(3), 307-316.
- Crippa, M.P., Alfonso, P.J. & Bustin, M. (1992). Nucleosome core binding region of chromosomal protein HMG-17 acts as an independent functional domain. *Journal of Molecular Biology*, 228(2), 442-449.

- Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res*, 14(6), 1188-1190.
- Cubeñas-Potts, C. & Corces, V.G. (2015). Architectural proteins, transcription, and the three-dimensional organization of the genome. *FEBS Lett*, 589(20 Pt A), 2923-2930.
- Cuddapah, S., Schones, D.E., Cui, K., Roh, T.Y., Barski, A., Wei, G., Rochman, M., Bustin, M. & Zhao, K. (2011). Genomic profiling of HMGN1 reveals an association with chromatin at regulatory regions. *Mol Cell Biol*, 31(4), 700-709.
- 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 1.9Å Resolution††We dedicate this paper to the memory of Max Perutz who was particularly inspirational and supportive to T.J.R. in the early stages of this study. *Journal of Molecular Biology*, 319(5), 1097-1113.
- Davey, G.E. & Davey, C.A. (2008). Chromatin - a new, old drug target? *Chem Biol Drug Des*, 72(3), 165-170.
- Deng, T., Zhu, Z.I., Zhang, S., Postnikov, Y., Huang, D., Horsch, M., Furusawa, T., Beckers, J., Rozman, J., Klingenspor, M., Amarie, O., Graw, J., Rathkolb, B., Wolf, E., Adler, T., Busch, D.H., Gailus-Durner, V., Fuchs, H., Hrabe de Angelis, M., van der Velde, A., Tessarollo, L., Ovcherenko, I., Landsman, D. & Bustin, M. (2015). Functional compensation among HMGN variants modulates the DNase I hypersensitive sites at enhancers. *Genome Res*, 25(9), 1295-1308.
- Ding, H.F., Bustin, M. & Hansen, U. (1997). Alleviation of histone H1-mediated transcriptional repression and chromatin compaction by the acidic activation region in chromosomal protein HMG-14. *Mol Cell Biol*, 17(10), 5843-5855.
- Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenko, V.V., Ecker, J.R., Thomson, J.A. & Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539), 331-336.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), 376-380.
- Douzi, B. (2017). Protein-Protein Interactions: Surface Plasmon Resonance. *Methods Mol Biol*, 1615, 257-275.
- Downs, J.A., Nussenzweig, M.C. & Nussenzweig, A. (2007). Chromatin dynamics and the preservation of genetic information. *Nature*, 447(7147), 951-958.
- 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.
- Egli, M. (2010). Diffraction Techniques in Structural Biology: Overview for unit 7 “Biophysical Analysis of Nucleic Acids” in: *Current Protocols in Nucleic Acid Chemistry*. Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.], CHAPTER 7, Unit-7.13.
- Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D*, 66(4), 486-501.
- Ermler, U., Fritsch, G., Buchanan, S.K. & Michel, H. Structure of the photosynthetic reaction centre from *Rhodobacter sphaeroides* at 2.65 Å resolution: cofactors and protein-cofactor interactions. *Structure*, 2(10), 925-936.
- Fan, J.Y., Rangasamy, D., Luger, K. & Tremethick, D.J. (2004). H2A.Z alters the nucleosome surface to promote HP1α-mediated chromatin fiber folding. *Mol Cell*, 16(4), 655-661.

- Fan, Y., Nikitina, T., Morin-Kensicki, E.M., Zhao, J., Magnuson, T.R., Woodcock, C.L. & Skoultchi, A.I. (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Mol Cell Biol*, 23(13), 4559-4572.
- Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L. & Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, 123(7), 1199-1212.
- Fan, Y., Sirotkin, A., Russell, R.G., Ayala, J. & Skoultchi, A.I. (2001). Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. *Mol Cell Biol*, 21(23), 7933-7943.
- Fang, H., Clark, D.J. & Hayes, J.J. (2012). DNA and nucleosomes direct distinct folding of a linker histone H1 C-terminal domain. *Nucleic Acids Res*, 40(4), 1475-1484.
- Fedorova, E. & Zink, D. (2008). Nuclear architecture and gene regulation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1783(11), 2174-2184.
- Feng, W., Pan, L. & Zhang, M. (2011). Combination of NMR spectroscopy and X-ray crystallography offers unique advantages for elucidation of the structural basis of protein complex assembly. *Science China Life Sciences*, 54(2), 101-111.
- Finch, J.T. & Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A*, 73(6), 1897-1901.
- Fitzpatrick, D.J., Ryan, C.J., Shah, N., Greene, D., Molony, C. & Shields, D.C. (2015). Genome-wide epistatic expression quantitative trait loci discovery in four human tissues reveals the importance of local chromosomal interactions governing gene expression. *BMC Genomics*, 16, 109.
- Flanagan, T.W. & Brown, D.T. (2016). Molecular dynamics of histone H1. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1859(3), 468-475.
- Fox, J.D., Kapust, R.B. & Waugh, D.S. (2001). Single amino acid substitutions on the surface of Escherichia coli maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein Sci*, 10(3), 622-630.
- Frouws, T.D., Duda, S.C. & Richmond, T.J. (2016). X-ray structure of the MMTV-A nucleosome core. *Proceedings of the National Academy of Sciences*, 113(5), 1214.
- Furusawa, T., Lim, J.H., Catez, F., Birger, Y., Mackem, S. & Bustin, M. (2006). Down-regulation of nucleosomal binding protein HMGN1 expression during embryogenesis modulates Sox9 expression in chondrocytes. *Mol Cell Biol*, 26(2), 592-604.
- Furusawa, T., Rochman, M., Taher, L., Dimitriadis, E.K., Nagashima, K., Anderson, S. & Bustin, M. (2015). Chromatin decompaction by the nucleosomal binding protein HMGN5 impairs nuclear sturdiness. *Nat Commun*, 6, 6138.
- Gerchman, S.E., Graziano, V. & Ramakrishnan, V. (1994). Expression of chicken linker histones in E. coli: sources of problems and methods for overcoming some of the difficulties. *Protein Expr Purif*, 5(3), 242-251.
- Gerlitz, G. (2010). HMGNs, DNA repair and cancer. *Biochim Biophys Acta*, 1799(1-2), 80-85.
- Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P. & Bickmore, W.A. (2004). Chromatin Architecture of the Human Genome: Gene-Rich Domains Are Enriched in Open Chromatin Fibers. *Cell*, 118(5), 555-566.
- Gonzalez-Romero, R., Eirin-Lopez, J.M. & Ausio, J. (2015). Evolution of high mobility group nucleosome-binding proteins and its implications for vertebrate chromatin specialization. *Mol Biol Evol*, 32(1), 121-131.

- Gonzalez, P.J. & Palacian, E. (1990). Structural and transcriptional properties of different nucleosomal particles containing high mobility group proteins 14 and 17 (HMG 14/17). *J Biol Chem*, 265(14), 8225-8229.
- Goytisolo, F.A., Gerchman, S.E., Yu, X., Rees, C., Graziano, V., Ramakrishnan, V. & Thomas, J.O. (1996). Identification of two DNA-binding sites on the globular domain of histone H5. *The EMBO journal*, 15(13), 3421-3429.
- Goytisolo, F.A., Packman, L.C. & Thomas, J.O. (1996). Photoaffinity labelling of a DNA-binding site on the globular domain of histone H5. *European journal of biochemistry / FEBS*, 242(3), 619-626.
- Grande, M.A., van der Kraan, I., de Jong, L. & van Driel, R. (1997). Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *Journal of cell science*, 110 (Pt 15), 1781-1791.
- Graziano, V. & Ramakrishnan, V. (1990). Interaction of HMG14 with chromatin. *Journal of Molecular Biology*, 214(4), 897-910.
- Grigoryev, S.A., Arya, G., Correll, S., Woodcock, C.L. & Schlick, T. (2009). Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc Natl Acad Sci U S A*, 106(32), 13317-13322.
- Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J.R. & Zaret, K.S. (1996). Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev*, 10(13), 1670-1682.
- Hamiche, A., Schultz, P., Ramakrishnan, V., Oudet, P. & Prunell, A. (1996). Linker histone-dependent DNA structure in linear mononucleosomes. *J Mol Biol*, 257(1), 30-42.
- Hansen, J.C., Lu, X., Ross, E.D. & Woody, R.W. (2006). Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J Biol Chem*, 281(4), 1853-1856.
- Happel, N. & Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene*, 431(1-2), 1-12.
- Harshman, S.W., Young, N.L., Parthun, M.R. & Freitas, M.A. (2013). H1 histones: current perspectives and challenges. *Nucleic Acids Res*, 41(21), 9593-9609.
- Hayashihara, K., Zlatanova, J. & Tomschik, M. (2010). Simplified method for recombinant linker histone H1 purification. *Mol Biotechnol*, 44(2), 148-151.
- Hayes, J.J., Pruss, D. & Wolffe, A.P. (1994). Contacts of the globular domain of histone H5 and core histones with DNA in a "chromatosome". *Proc Natl Acad Sci U S A*, 91(16), 7817-7821.
- Hayes, J.J. & Wolffe, A.P. (1993). Preferential and asymmetric interaction of linker histones with 5S DNA in the nucleosome. *Proc Natl Acad Sci U S A*, 90(14), 6415-6419.
- Heitz, E. (1928). Das Heterochromatin der Moose. *Jahrb Wiss Botanik*, 69, 762-818.
- Hendzel, M.J., Lever, M.A., Crawford, E. & Th'ng, J.P. (2004). The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *J Biol Chem*, 279(19), 20028-20034.
- Hizume, K., Yoshimura, S.H. & Takeyasu, K. (2005). Linker histone H1 per se can induce three-dimensional folding of chromatin fiber. *Biochemistry*, 44(39), 12978-12989.
- Hock, R., Scheer, U. & Bustin, M. (1998a). Chromosomal proteins HMG-14 and HMG-17 are released from mitotic chromosomes and imported into the nucleus by active transport. *The Journal of cell biology*, 143(6), 1427-1436.
- Hock, R., Wilde, F., Scheer, U. & Bustin, M. (1998b). Dynamic relocation of chromosomal protein HMG-17 in the nucleus is dependent on transcriptional activity. *The EMBO journal*, 17(23), 6992-7001.
- Holde, K.E.v. (1989). *Chromatin*. New York: Springer

- Hsieh, T.-Han S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N. & Rando, Oliver J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*, 162(1), 108-119.
- Hutchinson, Jordana B., Cheema, Manjinder S., Wang, J., Missiaen, K., Finn, R., Gonzalez Romero, R., Th'ng, John P.H., Hendzel, M. & Ausió, J. (2015). Interaction of chromatin with a histone H1 containing swapped N- and C-terminal domains. *Bioscience Reports*, 35(3), e00209.
- Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H. & Zaret, K.S. (2016). The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol Cell*, 62(1), 79-91.
- Izzo, A., Kamieniarz-Gdula, K., Ramirez, F., Noureen, N., Kind, J., Manke, T., van Steensel, B. & Schneider, R. (2013). The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. *Cell Rep*, 3(6), 2142-2154.
- Izzo, A., Kamieniarz, K. & Schneider, R. (2008). The histone H1 family: specific members, specific functions? *Biol Chem*, 389(4), 333-343.
- Jason-Moller, L., Murphy, M. & Bruno, J. (2006). Overview of Biacore systems and their applications. *Curr Protoc Protein Sci*, Chapter 19, Unit 19.13.
- Jia, R., Chai, P., Zhang, H. & Fan, X. (2017). Novel insights into chromosomal conformations in cancer. *Molecular Cancer*, 16(1), 173.
- Kaestner, K.H., Knochel, W. & Martinez, D.E. (2000). Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev*, 14(2), 142-146.
- Kalashnikova, A.A., Porter-Goff, M.E., Muthurajan, U.M., Luger, K. & Hansen, J.C. (2013). The role of the nucleosome acidic patch in modulating higher order chromatin structure. *Journal of the Royal Society, Interface / the Royal Society*, 10(82), 20121022.
- Kapust, R.B. & Waugh, D.S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*, 8(8), 1668-1674.
- 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(30), 12283-12288.
- Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. & Phillips, D.C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*, 181, 662.
- Kim, J.-M., Kim, K., Punj, V., Liang, G., Ulmer, T.S., Lu, W. & An, W. (2015). Linker histone H1.2 establishes chromatin compaction and gene silencing through recognition of H3K27me3. *5*, 16714.
- Korolev, N., Lyubartsev, A.P. & Nordenskiöld, L. (2018). A systematic analysis of nucleosome core particle and nucleosome-nucleosome stacking structure. *Scientific Reports*, 8(1), 1543.
- Kowalski, A. & Palyga, J. (2016). Modulation of chromatin function through linker histone H1 variants. *Biol Cell*, 108(12), 339-356.
- Kugler, J.E., Deng, T. & Bustin, M. (2012). The HMGN family of chromatin-binding proteins: dynamic modulators of epigenetic processes. *Biochim Biophys Acta*, 1819(7), 652-656.
- Lai, E., Prezioso, V.R., Tao, W.F., Chen, W.S. & Darnell, J.E., Jr. (1991). Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the Drosophila homeotic gene fork head. *Genes Dev*, 5(3), 416-427.

- Lalmansingh, A.S., Karmakar, S., Jin, Y. & Nagaich, A.K. (2012). Multiple modes of chromatin remodeling by Forkhead box proteins. *Biochim Biophys Acta*, 1819(7), 707-715.
- Lam, E.W., Brosens, J.J., Gomes, A.R. & Koo, C.Y. (2013). Forkhead box proteins: tuning forks for transcriptional harmony. *Nature reviews. Cancer*, 13(7), 482-495.
- Lee, C.S., Sund, N.J., Behr, R., Herrera, P.L. & Kaestner, K.H. (2005). Foxa2 is required for the differentiation of pancreatic alpha-cells. *Developmental biology*, 278(2), 484-495.
- Lercher, M.J., Urrutia, A.O. & Hurst, L.D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet*, 31(2), 180-183.
- Lever, M.A., Th'ng, J.P., Sun, X. & Hendzel, M.J. (2000). Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, 408(6814), 873-876.
- Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H.Q., Ong, C.T., Cubenas-Potts, C., Hu, M., Lei, E.P., Bosco, G., Qin, Z.S. & Corces, V.G. (2015). Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol Cell*, 58(2), 216-231.
- Li, Z., Schug, J., Tuteja, G., White, P. & Kaestner, K.H. (2011). The nucleosome map of the mammalian liver. *Nature structural & molecular biology*, 18(6), 742-746.
- 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(23), 20774-20782.
- Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. (2017). Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature*, 544(7651), 440-445.
- 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(1), 19-42.
- Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995). Crystal Structure of the 20S Proteasome from the Archaeon *T. acidophilum* at 3.4 Å Resolution. *Science*, 268(5210), 533-539.
- Lu, X., Hamkalo, B., Parseghian, M.H. & Hansen, J.C. (2009). Chromatin condensing functions of the linker histone C-terminal domain are mediated by specific amino acid composition and intrinsic protein disorder. *Biochemistry*, 48(1), 164-172.
- Luger, K., Dechassa, M.L. & Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol*, 13(7), 436-447.
- 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(6648), 251-260.
- Luger, K., Rechsteiner, T.J. & Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol*, 304, 3-19.
- Makde, D. & Tan, S. (2013). Strategies for crystallizing a chromatin protein in complex with the nucleosome core particle. *Anal Biochem*, 442(2), 138-145.
- 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(7315), 562-566.
- Martinez de Paz, A. & Ausio, J. (2016). HMGNs: The enhancer charmers. *Bioessays*, 38(3), 226-231.
- Mayor, R., Izquierdo-Bouldstridge, A., Millan-Arino, L., Bustillos, A., Sampaio, C., Luque, N. & Jordan, A. (2015). Genome distribution of replication-independent histone

- H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions. *J Biol Chem*, 290(12), 7474-7491.
- McGinty, R.K. & Tan, S. (2015). Nucleosome structure and function. *Chem Rev*, 115(6), 2255-2273.
- McGinty, R.K. & Tan, S. (2016). Recognition of the nucleosome by chromatin factors and enzymes. *Curr Opin Struct Biol*, 37, 54-61.
- McNally, J.G., Muller, W.G., Walker, D., Wolford, R. & Hager, G.L. (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science*, 287(5456), 1262-1265.
- McPherson, C.E., Shim, E.Y., Friedman, D.S. & Zaret, K.S. (1993). An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. *Cell*, 75(2), 387-398.
- Medrano-Fernández, A. & Barco, A. (2016). Nuclear organization and 3D chromatin architecture in cognition and neuropsychiatric disorders. *Molecular Brain*, 9, 83.
- Meyer, S., Becker, N.B., Syed, S.H., Goutte-Gattat, D., Shukla, M.S., Hayes, J.J., Angelov, D., Bednar, J., Dimitrov, S. & Everaers, R. (2011). From crystal and NMR structures, footprints and cryo-electron-micrographs to large and soft structures: nanoscale modeling of the nucleosomal stem. *Nucleic Acids Res*, 39(21), 9139-9154.
- Millan-Arino, L., Izquierdo-Bouldstridge, A. & Jordan, A. (2016). Specificities and genomic distribution of somatic mammalian histone H1 subtypes. *Biochim Biophys Acta*, 1859(3), 510-519.
- Misteli, T., Gunjan, A., Hock, R., Bustin, M. & Brown, D.T. (2000). Dynamic binding of histone H1 to chromatin in living cells. *Nature*, 408(6814), 877-881.
- Misteli, T. & Soutoglou, E. (2009). The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nature Reviews Molecular Cell Biology*, 10, 243.
- Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T. & Tramontano, A. (2014). Critical assessment of methods of protein structure prediction (CASP) — round x. *Proteins*, 82(0 2), 1-6.
- Murphy, K.J., Cutter, A.R., Fang, H., Postnikov, Y.V., Bustin, M. & Hayes, J.J. (2017). HMGN1 and 2 remodel core and linker histone tail domains within chromatin. *Nucleic Acids Res*, 45(17), 9917-9930.
- Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A.S., Imamoto, N., Ishikawa, T. & Maeshima, K. (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO journal*, 31(7), 1644-1653.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Bluthgen, N., Dekker, J. & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), 381-385.
- Oberg, C. & Belikov, S. (2012). The N-terminal domain determines the affinity and specificity of H1 binding to chromatin. *Biochem Biophys Res Commun*, 420(2), 321-324.
- Olins, A.L. & Olins, D.E. (1974). Spheroid chromatin units (v bodies). *Science*, 183(4122), 330-332.
- 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.

- Orrego, M., Ponte, I., Roque, A., Buschati, N., Mora, X. & Suau, P. (2007). Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biol*, 5, 22.
- Osborne, C.S., Chakalova, L., Mitchell, J.A., Horton, A., Wood, A.L., Bolland, D.J., Corcoran, A.E. & Fraser, P. (2007). Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol*, 5(8), e192.
- Overdier, D.G., Porcella, A. & Costa, R.H. (1994). The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. *Mol Cell Biol*, 14(4), 2755-2766.
- Palermo, G., Magistrato, A., Riedel, T., von Erlach, T., Davey, C.A., Dyson, P.J. & Rothlisberger, U. (2016). Fighting Cancer with Transition Metal Complexes: From Naked DNA to Protein and Chromatin Targeting Strategies. *ChemMedChem*, 11(12), 1199-1210.
- Paton, A.E., Wilkinson-Singley, E. & Olins, D.E. (1983). Nonhistone nuclear high mobility group proteins 14 and 17 stabilize nucleosome core particles. *J Biol Chem*, 258(21), 13221-13229.
- Pérez-Montero, S., Carbonell, A. & Azorín, F. (2016). Germline-specific H1 variants: the “sexy” linker histones. *Chromosoma*, 125(1), 1-13.
- Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G. & North, A.C.T. (1960). Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å Resolution, Obtained by X-Ray Analysis. *Nature*, 185, 416.
- Phair, R.D., Gorski, S.A. & Misteli, T. (2004). Measurement of dynamic protein binding to chromatin in vivo, using photobleaching microscopy. *Methods Enzymol*, 375, 393-414.
- Phair, R.D. & Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature*, 404(6778), 604-609.
- Pierce, B.A. (2013). *Genetics: A Conceptual Approach* (5th ed.). New York: W. H. Freeman 2013-12-27.
- Pike, A.C., Garman, E.F., Krojer, T., von Delft, F. & Carpenter, E.P. (2016). An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr D Struct Biol*, 72(Pt 3), 303-318.
- Plank, J.L. & Dean, A. (2014). Enhancer function: mechanistic and genome-wide insights come together. *Mol Cell*, 55(1), 5-14.
- Polak, P., Karlic, R., Koren, A., Thurman, R., Sandstrom, R., Lawrence, M.S., Reynolds, A., Rynes, E., Vlahovicek, K., Stamatoyannopoulos, J.A. & Sunyaev, S.R. (2015). Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature*, 518(7539), 360-364.
- Postnikov, Y.V., Herrera, J.E., Hock, R., Scheer, U. & Bustin, M. (1997). Clusters of nucleosomes containing chromosomal protein HMG-17 in chromatin. *Journal of Molecular Biology*, 274(4), 454-465.
- Postnikov, Y.V., Shick, V.V., Belyavsky, A.V., Khrapko, K.R., Brodolin, K.L., Nikolskaya, T.A. & Mirzabekov, A.D. (1991). Distribution of high mobility group proteins 1/2, E and 14/17 and linker histones H1 and H5 on transcribed and non-transcribed regions of chicken erythrocyte chromatin. *Nucleic Acids Res*, 19(4), 717-725.
- Postnikov, Y.V., Trieschmann, L., Rickers, A. & Bustin, M. (1995). Homodimers of chromosomal proteins HMG-14 and HMG-17 in nucleosome cores. *Journal of Molecular Biology*, 252(4), 423-432.
- Powell, H.R., Battye, T.G.G., Kontogiannis, L., Johnson, O. & Leslie, A.G.W. (2017). Integrating macromolecular X-ray diffraction data with the graphical user interface iMosflm. *Nat Protoc*, 12(7), 1310-1325.

- Prymakowska-Bosak, M., Misteli, T., Herrera, J.E., Shirakawa, H., Birger, Y., Garfield, S. & Bustin, M. (2001). Mitotic phosphorylation prevents the binding of HMGN proteins to chromatin. *Mol Cell Biol*, 21(15), 5169-5178.
- Pyo, S.H., Lee, J.H., Park, H.B., Hong, S.S. & Kim, J.H. (2001). A large-scale purification of recombinant histone H1.5 from *Escherichia coli*. *Protein Expr Purif*, 23(1), 38-44.
- Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. & Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, 362(6417), 219-223.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A., Machol, I., Omer, A.D., Lander, E.S. & Aiden, E.L. (2014). A three-dimensional map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7), 1665-1680.
- 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(5), 620-626.
- Ricci, M.A., Manzo, C., Garcia-Parajo, M.F., Lakadamyali, M. & Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, 160(6), 1145-1158.
- Rich, R.L. & Myszka, D.G. (2001). BIACORE J: a new platform for routine biomolecular interaction analysis. *J Mol Recognit*, 14(4), 223-228.
- Risca, V.I., Denny, S.K., Straight, A.F. & Greenleaf, W.J. (2016). Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. *Nature*, 541, 237.
- Robinson, O., Dylus, D. & Dessimoz, C. (2016). Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. *Mol Biol Evol*, 33(8), 2163-2166.
- 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(3), 336-343.
- Rochman, M., Malicet, C. & Bustin, M. (2010). HMGN5/NSBP1: a new member of the HMGN protein family that affects chromatin structure and function. *Biochim Biophys Acta*, 1799(1-2), 86-92.
- Rogge, R.A., Kalashnikova, A.A., Muthurajan, U.M., Porter-Goff, M.E., Luger, K. & Hansen, J.C. (2013). Assembly of Nucleosomal Arrays from Recombinant Core Histones and Nucleosome Positioning DNA. *Journal of Visualized Experiments : JoVE*(79), 50354.
- 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(26), 8872-8877.
- Rubinstein, Y.R., Furusawa, T., Lim, J.H., Postnikov, Y.V., West, K.L., Birger, Y., Lee, S., Nguyen, P., Trepel, J.B. & Bustin, M. (2005). Chromosomal protein HMGN1 modulates the expression of N-cadherin. *The FEBS journal*, 272(22), 5853-5863.
- Sancho, M., Diani, E., Beato, M. & Jordan, A. (2008). Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet*, 4(10), e1000227.
- Sandeén, G., Wood, W.I. & Felsenfeld, G. (1980). The interaction of high mobility proteins HMG14 and 17 with nucleosomes. *Nucleic Acids Res*, 8(17), 3757-3778.
- Santisteban, P., Recacha, P., Metzger, D.E. & Zaret, K.S. (2010). Dynamic expression of Groucho-related genes *Grg1* and *Grg3* in foregut endoderm and antagonism of differentiation. *Developmental dynamics : an official publication of the American Association of Anatomists*, 239(3), 980-986.

- 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(7047), 138-141.
- Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. & Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, 102(5), 615-623.
- Schmitt, A.D., Hu, M. & Ren, B. (2016). Genome-wide mapping and analysis of chromosome architecture. *Nature Reviews Molecular Cell Biology*, 17, 743.
- Schuster-Bockler, B. & Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*, 488(7412), 504-507.
- Scopes, R.K. (1974). Measurement of protein by spectrophotometry at 205 nm. *Anal Biochem*, 59(1), 277-282.
- Sekiya, T., Muthurajan, U.M., Luger, K., Tulin, A.V. & Zaret, K.S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev*, 23(7), 804-809.
- Serandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R., Salbert, G. & Eeckhoutte, J. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res*, 21(4), 555-565.
- Sharma, V., Murphy, D.P., Provan, G. & Baranov, P.V. (2012). CodonLogo: a sequence logo-based viewer for codon patterns. *Bioinformatics*, 28(14), 1935-1936.
- Shimahara, H., Hirano, T., Ohya, K., Matsuta, S., Seeram, S.S. & Tate, S. (2013). Nucleosome structural changes induced by binding of non-histone chromosomal proteins HMGN1 and HMGN2. *FEBS Open Bio*, 3, 184-191.
- Shirakawa, H., Herrera, J.E., Bustin, M. & Postnikov, Y. (2000). Targeting of high mobility group-14/-17 proteins in chromatin is independent of DNA sequence. *J Biol Chem*, 275(48), 37937-37944.
- Sievers, F. & Higgins, D.G. (2014). Clustal omega. *Curr Protoc Bioinformatics*, 48, 3.13.11-16.
- Simpson, R.T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry*, 17(25), 5524-5531.
- Sirotkin, A.M., Edelmann, W., Cheng, G., Klein-Szanto, A., Kucherlapati, R. & Skoultschi, A.I. (1995). Mice develop normally without the H1(0) linker histone. *Proc Natl Acad Sci U S A*, 92(14), 6434-6438.
- Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P. & Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science*, 344(6182), 376-380.
- Song, S.-H. & Kim, T.-Y. (2017). CTCF, Cohesin, and Chromatin in Human Cancer. *Genomics Inform*, 15(4), 114-122.
- Staynov, D.Z. (2008). The controversial 30 nm chromatin fibre. *Bioessays*, 30(10), 1003-1009.
- Stedman, E. & Stedman, E. (1951). The Basic Proteins of Cell Nuclei. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 235(630), 565-595.
- Stehr, R., Kepper, N., Rippe, K. & Wedemann, G. (2008). The Effect of Internucleosomal Interaction on Folding of the Chromatin Fiber. *Biophysical Journal*, 95(8), 3677-3691.
- Steiner, R.F. (1952). Physico-chemical studies on the components of thymus cell nuclei. *Transactions of the Faraday Society*, 48(0), 1185-1196.

- Suto, R.K., Clarkson, M.J., Tremethick, D.J. & Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat Struct Biol*, 7(12), 1121-1124.
- Syed, S.H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M.S., Hayes, J.J., Everaers, R., Angelov, D., Bednar, J. & Dimitrov, S. (2010). Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. *Proc Natl Acad Sci U S A*, 107(21), 9620-9625.
- 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(7359), 232-235.
- Th'ng, J.P., Sung, R., Ye, M. & Hendzel, M.J. (2005). H1 family histones in the nucleus. Control of binding and localization by the C-terminal domain. *J Biol Chem*, 280(30), 27809-27814.
- Thoma, F., Koller, T. & Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of cell biology*, 83(2 Pt 1), 403-427.
- Torres, C.M., Biran, A., Burney, M.J., Patel, H., Henser-Brownhill, T., Cohen, A.S., Li, Y., Ben-Hamo, R., Nye, E., Spencer-Dene, B., Chakravarty, P., Efroni, S., Matthews, N., Misteli, T., Meshorer, E. & Scaffidi, P. (2016). The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science*, 353(6307).
- Trieschmann, L., Alfonso, P.J., Crippa, M.P., Wolffe, A.P. & Bustin, M. (1995). Incorporation of chromosomal proteins HMG-14/HMG-17 into nascent nucleosomes induces an extended chromatin conformation and enhances the utilization of active transcription complexes. *The EMBO journal*, 14(7), 1478-1489.
- Trieschmann, L., Martin, B. & Bustin, M. (1998). The chromatin unfolding domain of chromosomal protein HMG-14 targets the N-terminal tail of histone H3 in nucleosomes. *Proc Natl Acad Sci U S A*, 95(10), 5468-5473.
- 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(9), 2872-2883.
- Ueda, T., Postnikov, Y.V. & Bustin, M. (2006). Distinct domains in high mobility group N variants modulate specific chromatin modifications. *J Biol Chem*, 281(15), 10182-10187.
- Ulianov, S.V., Gavrilov, A.A. & Razin, S.V. (2015). Nuclear compartments, genome folding, and enhancer-promoter communication. *Int Rev Cell Mol Biol*, 315, 183-244.
- Vyas, P. & Brown, D.T. (2012). N- and C-terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isoforms H1(0) and H1c. *J Biol Chem*, 287(15), 11778-11787.
- Watson, J.D. & Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738.
- Weigel, D., Jurgens, G., Kuttner, F., Seifert, E. & Jackle, H. (1989). The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell*, 57(4), 645-658.
- White, A.E., Hieb, A.R. & Luger, K. (2016). A quantitative investigation of linker histone interactions with nucleosomes and chromatin. 6, 19122.
- Whitlock, J.P., Jr. & Simpson, R.T. (1976). Removal of histone H1 exposes a fifty base pair DNA segment between nucleosomes. *Biochemistry*, 15(15), 3307-3314.

- Wijchers, P.J. & de Laat, W. (2011). Genome organization influences partner selection for chromosomal rearrangements. *Trends Genet*, 27(2), 63-71.
- Wilson, M.D., Benlekhir, S., Fradet-Turcotte, A., Sherker, A., Julien, J.P., McEwan, A., Noordermeer, S.M., Sicheri, F., Rubinstein, J.L. & Durocher, D. (2016). The structural basis of modified nucleosome recognition by 53BP1. *Nature*, 536(7614), 100-103.
- Wimberly, B.T., Brodersen, D.E., Clemons Jr, W.M., Morgan-Warren, R.J., Carter, A.P., Vonnrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, 407, 327.
- 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(13), 3562-3566.
- Wu, M., Allis, C.D., Richman, R., Cook, R.G. & Gorovsky, M.A. (1986). An intervening sequence in an unusual histone H1 gene of *Tetrahymena thermophila*. *Proc Natl Acad Sci U S A*, 83(22), 8674-8678.
- Yang, S.M., Kim, B.J., Norwood Toro, L. & Skoultchi, A.I. (2013). H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc Natl Acad Sci U S A*, 110(5), 1708-1713.
- 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(9), 2651-2664.
- Zaret, K.S. (2002). Regulatory phases of early liver development: paradigms of organogenesis. *Nature reviews. Genetics*, 3(7), 499-512.
- Zaret, K.S. (2008). Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. *Nature reviews. Genetics*, 9(5), 329-340.
- Zaret, K.S., Caravaca, J.M., Tulin, A. & Sekiya, T. (2010). Nuclear mobility and mitotic chromosome binding: similarities between pioneer transcription factor FoxA and linker histone H1. *Cold Spring Harb Symp Quant Biol*, 75, 219-226.
- Zaret, K.S. & Stevens, K. (1995). Expression of a highly unstable and insoluble transcription factor in *Escherichia coli*: purification and characterization of the fork head homolog HNF3 alpha. *Protein expression and purification*, 6(6), 821-825.
- Zaret, K.S., Watts, J., Xu, J., Wandzioch, E., Smale, S.T. & Sekiya, T. (2008). Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb Symp Quant Biol*, 73, 119-126.
- Zeng, C., Kim, E., Warren, S.L. & Berget, S.M. (1997). Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *The EMBO journal*, 16(6), 1401-1412.
- Zhang, P., Branson, O.E., Freitas, M.A. & Parthun, M.R. (2016). Identification of replication-dependent and replication-independent linker histone complexes: Tpr specifically promotes replication-dependent linker histone stability. *BMC Biochemistry*, 17(1), 18.
- Zhang, Q. & Wang, Y. (2008). High mobility group proteins and their post-translational modifications. *Biochim Biophys Acta*, 1784(9), 1159-1166.
- Zhang, S., Zhu, I., Deng, T., Furusawa, T., Rochman, M., Vacchio, M.S., Bosselut, R., Yamane, A., Casellas, R., Landsman, D. & Bustin, M. (2016). HMGN proteins modulate chromatin regulatory sites and gene expression during activation of naive B cells. *Nucleic Acids Res*, 44(15), 7144-7158.
- Zhang, Y., Wong, C.H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., Mulawadi, F.H., Sung, W.K., Nicolis, S., Ahituv, N., Ruan, Y. &

- Wei, C.L. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature*, 504(7479), 306-310.
- Zhou, B.R., Feng, H., Ghirlando, R., Li, S., Schwieters, C.D. & Bai, Y. (2016). A Small Number of Residues Can Determine if Linker Histones Are Bound On or Off Dyad in the Chromatosome. *J Mol Biol*, 428(20), 3948-3959.
- Zhou, B.R., Feng, H., Kato, H., Dai, L., Yang, Y., Zhou, Y. & Bai, Y. (2013). Structural insights into the histone H1-nucleosome complex. *Proc Natl Acad Sci U S A*, 110(48), 19390-19395.
- Zhou, B.R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T.S. & Bai, Y. (2015). Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell*, 59(4), 628-638.
- Zhou, Y.B., Gerchman, S.E., Ramakrishnan, V., Travers, A. & Muyldermans, S. (1998). Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature*, 395(6700), 402-405.
- Zhou, Z.H. (2011). Atomic resolution cryo electron microscopy of macromolecular complexes. *Adv Protein Chem Struct Biol*, 82, 1-35.

Appendix

DNA constructs

DNA sequences

FOXA1-145

ATCA**CAAT**CCCGGT**GCCGA**AGGCCG**CTCAA**TTGGT**CGTAG**ACAGCT**TCT**
AGCACCGCT**TAAAC**CGCAC**GTAGG**GAATG**TTTGT**TCCTA**TTTAAGCGC**
AC**CTAGAGCTCG****CTACT**TCGCATTCTACGATCCGCAAGGGATA**TTTGG**
AGAT

601L

ATC**ACAAT**CCCGGT**TGCCG**AGGCCG**CTCAA**TTGGT**CGTAG**ACAGCT**TCT**
AGCACCGCT**TAAAC**CGCAC**GTACG**GA**A**TCAGTTCCGATATC**ACAATCC**
CGG**TGCCG**AGGCCG**CTCAA**TTGGT**CGTAG**ACAGCT**CTAG**CACCGCT**T**
AAACGCACGTACGGAT**T**TCAGTTCCGAT

FOXA Protein sequences

hFoxA1

MLGTVKMEGHETSDWNSYIYADTQEAYSSVPVSNMNSGLGSMNSMNTYMTMNTMTTSGNMTPASFNMSYANPGL
GAGLSPGAVAGMPGGSAGAMNSMTAAGVTAMGTALSPSGMGAMGAQQAAASMNGLGPYAAAMNPCMSPMAYAPS
NLGRSRAGGGGDAKTFKRSYPHAKPPYSYISLITMAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIR
HSLSFNDCFVKVARSPDKPGKGSYWTLHPDSGNMNFENG CYLRRQKRFKCEKQPGAGGGGSGSGSGGSAKGGPE
SRKDPGASNP SADSPLHRGVHGTGQLEGAPAPGPAASPQTLDHSGATATGGASELKTPASSTAPPISSGPG
ALASVPASHPAHGLAPHESQLHLKGDPHYSFNHPFSINNLMSSSEQQHKLDFKAYEQALQYSPYGSTLPASLP
LGSASVTTRSPIEPSALEFAYYQGVYSRPVLNTS

hFoxA2

MLGAVKMEGHEPSDWSSYYAEPEGYSVSNMAGLGMNGMNTYMSMSAAAMGSGSGNMSAGSMNMSSYVGAGM
SPSLAGMSPGAGAMAGMGGSAGAAGVAGMGPHLSPLSLPLGGQAAGAMGGLAPYANMNSMSPMYGQAGLSRAR
DPKTYRRSYTHAKPPYSYISLITMAIQQSPNKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFND CFLK
VPRSPDKPGKGSFWTLHPDSGNMNFENG CYLRRQKRFKCEKQLALKEAAGAAGSGKAAAGAQASQAQLGEAAG
PAETPAGTESPHSSASPCQEHKRGGLGELKGTAAAALSPPEPAPSPGQQQAAAHL LGPPHHPLPPEAHLK
PEHHYAFNHPFSINNLMSSSEQQHHHSHHHHQPHKMDLKAYEQVMHYPGYGSMPGSLAMGPVTNKTGLDASPL
AADTSYYQGVYSRPIMNSS

hFoxA3

MLGSKVMEAHDLAEWSYYPEAGEVYSPVTPVPTMAPLNSYMTLNPLSSPYPPGGPLPASPLPSGGLAPPAPAAP
LGPTFPGLGVSGGSSSSGYGAPGPGLVHGKEMPKGYRRPLAHAKPPYSYISLITMAIQQAPGKMLTLSEIYQW
IMDLFPYYRENQQRWQNSIRHSLSFNDCFVKVARSPDKPGKGSYWALHPSSGNMFENG CYLRRQKRFKLEEKV
KKGSGAATTRNGTGSAASTTTAATVTSPPQPPPPAPEPEAQGGEDVGALDCGSPASSTPYFTGLELPGEL
KLDAPYNFNHPFSINNLMSEQTPAPPKLDVGFGGYGAEGGEPGVYYQGLYSRSLNAS

FoxA1_DBD 169-263

AKPPYSYISLITMAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIRHSLSFNDCFVKVARSPDKPGKG
SYWTLHPDSGNMNFENG CYLRRQ

FoxA2_DBD 159-252

KPPYSYISLITMAIQQSPNKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFND CFLKVPRSPDKPGKGS
FWTLHPDSGNMNFENG CYLRRQ

FoxA3_DBD 116-210

AKPPYSYISLITMAIQQAPGKMLTLSEIYQWIMDLFPYYRENQQRWQNSIRHSLSFNDCFVKVARSPDKPGKG
SYWALHPSSGNMFENG CYLRRQ

Mass Spectrometry based molecular weight determination for recombinant human FOXA1 protein preparations.

1. HT-DBD [D158-A274] of human FOXA1 protein

Expected molecular mass is 16530 Da. Observed MW is 16531 Da.

Qualitative Analysis Report

Data Filename	FoxA1-HUMAN-vc009.d	Sample Name	FoxA1-HUMAN-vc009
Sample Type	Sample	Position	P1-A9
Instrument Name	LCMS TOF6224	User Name	
Acq Method	Protein18min4degrC.m	Acquired Time	11/26/2012 9:39:45 PM
IRM Calibration Status	All Ions Missed	DA Method	autoreportMSstandard.m
Comment	FoxA1-HUMAN-b004		

User Chromatograms

Integration Peak List

Start	RT	End	Height	Area	Area %
6.577	6.838	7.164	53674435	1121590569	17.14
7.164	7.419	8.417	206496017	6544372464	100

User Spectra

Spectrum Source	Fragmentor Voltage	Collision Energy	Ionization Mode
Peak (1) in "+ TIC Scan Smo - FoxA1-HUMAN-vc009.d"	180	0	ESI

Qualitative Analysis Report

Peak List

m/z	z	Abund
130.1617	1	25963.8
239.1091	1	78228.9
634.5986		18512.3
676.8307		19950.3
725.1368	1	25179.9
729.1499		19530.7
780.8263		32047.7
783.6109	2	19434.1
845.8391		26578.9
848.8135	2	17470.5

Peak List

m/z	Abund	Score (DB)
10138.4	1305382	10
10174.53	763138.9	7
10210.8	450186.9	0
16019.75	625047.7	9
16055.24	208941.8	9
16092.91	39479.3	9
20275.91	27688.7	10
32038.36	41872.3	8
48057.26	22921.9	10
48164.88	18200.1	10

Spectrum Source: Fragmentor Voltage: Collision Energy: Ionization Mode:

2. ΔNT-FOXA1-HT of human FOXA1 protein

Expected molecular mass is 44641 Da. Observed MW is 44642 Da.

Qualitative Analysis Report

Data Filename	FoxA1-HUMAN-vc065.d	Sample Name	FoxA1-HUMAN-vc065
Sample Type	Sample	Position	P1-C6
Instrument Name	Genomax QTOF	User Name	
Acq Method	Intact protein_0-5.m	Acquired Time	4/17/2015 9:46:02 AM
IRM Calibration Status	Success	DA Method	Report format.m
Comment			
Sample Group	Info.		
Acquisition SW	6200 series TOF/6500 series		
Version	Q-TOF B.05.01 (B5125.1)		

User Chromatograms

User Spectra

Peak List

m/z	z	Abund
721.4657	1	31505.51
769.4964	1	43169.87
769.7244	1	24265.42
824.3897	1	49231.36
824.6319	1	25448.14
887.7034	1	43415.03
887.9122	1	30873.51
961.4886	1	30310.53

Qualitative Analysis Report

961.5722	1	34317.68
961.6546	1	33535.09
961.7378	1	30663.86
961.8208	1	24055.83
1048.8043	1	27928.26
1048.8953	1	30198.37
1048.9867	1	28836.21
1049.0767	1	26295.97
1153.5845	1	26342.57
1153.6847	1	29546.76
1153.7845	1	29922.49
1153.8841	1	25873.51

Peak List

m/z	Abund	Score (DB)
46111.93	205007.68	9
46172.41	5484.71	10
46196.92	12174.07	8
46822.16	4101.23	8
57640.77	34365.91	8
57747.84	9880.52	8
58530.55	5023.51	7
69170	16263.67	5
69296.53	4145.15	8
69391.35	5197.67	8
70232.11	4345.81	8
74932.35	4592.85	6
79898.58	5387.2	9
80680.24	18542	7
80700.98	16391.41	7
81384.61	4951.96	9
81404.07	4405.58	9
81933.99	5765.55	9
81946.06	7019.31	9
89750.14	4267.26	9

Spectrum Source	Fragmentor Voltage	Collision Energy	Ionization Mode
Peak (5) in "+ TIC Scan"	175	0	ESI

A representative MALDI-Protein id determination for a recombinant FOXA1 purified protein construct.

Reported By: 11/14/2013 10:31:19 - admin

Analysis Information

Report Type	Protein-Peptide Summary by Spot	Analysis Type	Combined (MS+MS/MS)
Sample Set Name	130814	Database	Swissprot
Analysis Name	Lim_Rajan_Ravi_Ma_human	Creation Date	11/14/2013 10:20:12
Reported By	11/14/2013 10:31:20 - admin	Last Modified	11/14/2013 10:23:24
MS Acq. : Proc. Methods	(Unspecified) : (Unspecified)		
Interpretation Method	(Unspecified)		

Gel Idx/Pos		382/P9	Instr./Gel Origin		ak072/130814		Process Status		Analysis Succeeded	
Plate [#] Name		[1] mw1		Instrument Sample Name		Spectra		11		
Rank	Protein Name		Species	Accession No.	Protein MW	Protein Score	Protein Score C. I. %	Total Ion Score	Total Ion C. I. %	Confirmed
1	Hepatocyte nuclear factor 3-alpha OS=Homo sapiens GN=FOXA1 PE=1 SV=2			sp P55317 FOXA1_HUMAN	49344.4	81	99.984	48	99.806	.T.
Peptide Information										
	Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Sequence Seq.	Ion Score	C. I. %	Modification	Rank Result Type
	803.4158	803.3543	-0.0615	-77	214	219 WQNSIR	18	0		Mascot
	803.4158	803.3543	-0.0615	-77	214	219 WQNSIR				Mascot
	875.4329	875.3755	-0.0574	-66	151	160 SRAGGGDAK				Mascot
	1353.6256	1353.5801	-0.0655	-48	220	230 HSLSFNDCFVK	11	0	Carbamidomethyl (C)[8]	Mascot
	1353.6256	1353.5801	-0.0655	-48	220	230 HSLSFNDCFVK			Carbamidomethyl (C)[8]	Mascot
	1607.7408	1607.6531	-0.0877	-55	296	311 DPSSGASNPADSPHHR				Mascot
	1679.8323	1679.7534	-0.0789	-47	220	233 HSLSFNDCFVKVAR			Carbamidomethyl (C)[8]	Mascot
	1735.8358	1735.7194	-0.1164	-67	295	311 KDPSGASNPADSPHHR	19	0		Mascot
	1735.8358	1735.7194	-0.1164	-67	295	311 KDPSGASNPADSPHHR				Mascot
	2482.2185	2482.0642	-0.1543	-62	190	208 MLTLSEIQWIMDLFPYYR				Mascot
	2482.2185	2482.0642	-0.1543	-62	190	208 MLTLSEIQWIMDLFPYYR				Mascot
	2498.2134	2498.0349	-0.1785	-71	190	208 MLTLSEIQWIMDLFPYYR			Oxidation (M)[1]	Mascot
	2504.0757	2503.9072	-0.1685	-67	241	261 GSYWTLHPDPSGNMFENGCYLRL			Carbamidomethyl (C)[18]	Mascot

MengWei130814Lim_Rajan_Ravi_Ma_human

1 of 7

HMGN protein sequences

hHMGN1

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVANQETKED
LPAEN
GETKTEESPASDEAGEKEAKSD

hHMGN2

MPKRKAEGDAKGDKAKVKDEPQRRSARLSAKPAPPKPEPKPKKAPAKKGEKVPKGKKGKADAGKEGNNPAENG
DAKTD
QAQKAEGAGDAK

hHMGN3a

MPKRKSPENTEGKDGSKVTKQEPTRRSARLSAKPAPPKPEPKPRKTSAKKEPGAKISRGAKGKKEEKQEAGKE
GTAP
SENGETKAEEAQKTESVDNEGE

hHMGN3b

MPKRKSPENTEGKDGSKVTKQEPTRRSARLSAKPAPPKPEPKPRKTSAKKEPGAKISRGAKGKKEEKQEAGKE
GTEN

hHMGN4

MPKRKAKGDAKGDKAKVKDEPQRRSARLSAKPAPPKPEPRPKKASAKKGEKLPKGRKGKADAGKDGNNPAKNR
DASTL
QSQKAEGTGDAK

hHMGN5

MPKRKAAGQGDMRQEPKRRSARLSAMLVPTPEVKPKRTSSSRKMKTksDMMEENIDTSAQAVAETKQEAVVE
EDYN
ENAKNGEAKITEAPASEKEIVEVKEENIEDATEKGGEKKEAVAAEVKNEEEDQKEDEEDQNEEKGEAGKEDKD
EKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKNGEDGKEKGEDEKEEEDRKETGDGKENEDGKEKGDKKEG
KDVVKVEDEKEREDGKEDEGGNEEEAGKEKEDLKEEEEGKEEDEIKEDDGKKEEPQSIV

HMGN1_NBD

KEEPKRRSARLSAKPPAKVEAKPKKAAAK

HMGN2_NBD

KDEPQRRSARLSAKPAPPKPEPKPKKAPAK

Mass Spectrometry based molecular weight determination for recombinant human HMGN protein preparations.

A representative report of Mass spectrometry based molecular weight determination for human HMGN1

Expected molecular weight is 13212 Da. The observed molecular weight is 13212 Da.

Qualitative Analysis Report

Data Filename	HMGN1-HUMAN-vc001-r002.d	Sample Name	HMGN1-HUMAN-vc001
Sample Type	Sample	Position	P1-C8
Instrument Name	Genomax QTOF	User Name	
Acq Method	Intact protein_all ms.m	Acquired Time	8/7/2014 4:21:52 PM
IRM Calibration Status	Success	DA Method	MPP_DA_MFE_CEF_Metabolites.m
Comment			
Sample Group	Info.		
Acquisition SW	6200 series TOF/6500 series		
Version	Q-TOF B.05.01 (B5125.1)		

User Chromatograms

User Spectra

Peak List

m/z	z	Abund
61.0401		12158.58
118.0858		16435.42
121.0509	1	212018.56
122.0537	1	13402.25
123.0801		17212.42
154.9894		9082.88
171.0622		11975.98
173.0781		12112.95
189.0732		12033.97

Qualitative Analysis Report

203.0887		23600
217.1044		9870.42
239.106	1	56474.3
251.0676	1	33187.91
261.0878	1	131932.35
262.0906	1	14833.29
281.0782		14721.76
283.0694	1	57191.56
393.2964	1	17996.53
922.0099		134621.64
923.0128		26279.99

Peak List

<i>m/z</i>	Abund	Score (DB)
10409.7	14657.03	6
10620.8	19649.14	4
10691.35	13241.17	5
11120.79	12192.59	3
11350.26	17011.45	4
11708.49	17354.49	4
11797.38	13085.68	4
11893.15	30981.3	6
11964.87	12655.75	4
12424.44	35137.45	6
12553.18	44235.09	6
12681.27	13543.98	5
12882.65	16493.3	4
13211.77	119484.56	7
13231.58	11773.22	1
13250.48	23184.33	3
13253	21822.32	1
13338.16	10434.94	0
13390.22	22481.12	5
25757.46	11891.42	8

Fragmentor Voltage 250 **Collision Energy** 0 **Ionization Mode** ESI

A representative report of Mass spectrometry based molecular weight determination for human HMGN1

Expected molecular weight is 11945 Da. The observed molecular weight is 11946Da.

Qualitative Analysis Report

Data Filename	HMGN2-HUMAN-vc001-r001.d	Sample Name	HMGN2-HUMAN-vc001
Sample Type	Sample	Position	P2-E7
Instrument Name	Genomax QTOF	User Name	
Acq Method	Intact protein_all ms.m	Acquired Time	8/11/2014 9:27:54 AM
IRM Calibration Status	Success	DA Method	MPP_DA_MFE_CEF_Metabolites.m
Comment			

Sample Group	Info.
Acquisition SW	6200 series TOF/6500 series
Version	Q-TOF B.05.01 (B5125.1)

User Chromatograms

User Spectra

Peak List

m/z	z	Abund
114.0916	1	22240.77
203.0887	1	22771.83
239.1057	1	116075.27
251.0676	1	392437.32
252.0708	1	36791.77
261.0878	1	128794.43
273.0491	1	26214.29
281.0783	1	26663.47
283.0693	1	53773.46

Qualitative Analysis Report

922.011	1	160102.59
923.0133	1	34022.43

Peak List

<i>m/z</i>	Abund	Score (DB)
11929.02	13000.67	8
11945.81	297491.43	10
11966.71	36569.42	7
11982.03	25867.51	7
11986.4	21566.83	8
12006.21	12457.97	4
12043.75	8849.89	6
12123.74	33637.79	8
12203.49	13308.1	8
12224.15	9444.53	7
15345.61	9653.04	4
15506.94	10300.53	5
16646.13	8122.41	3
17167.12	9686.82	3
18206.78	8573.2	3
19507.5	8462.78	2
22107.41	12252.95	5
23148.79	13046.1	7
25488.52	9012.13	7
28567.51	8102.15	7

Spectrum Source	Fragmentor Voltage	Collision Energy	Ionization Mode
Peak (4) in "+ TIC Scan"	175	0	ESI

Linker histone protein sequences

hH1.0

GPTENSTSAPAAKPKRAKASKKSTDHPKYSDMIVAAIQAEKNRAGSSRSIQKYIKSHYKVGENADSQIKLSI
KRLVTTGVLKQTKGVGASGSFRLAKSDEPKKSVAFKKTKEIKKVATPKKASKPKKAASKAPTCKPKATPVKK
AKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVKPKAKSSAKRAGKKK

hH1x gi|5174449|ref|NP_006017.1| histone H1x [Homo sapiens]

MSVELEEALPVTTAEGMAKKVTKAGGSAALSPSKKRKNSKKKNQPGKYSQLVVETIRRLGERNGSSLAKIYTE
AKKVFWFDQQNGRTYLYKYSIKALVQNDTLLQVKGTGANGSFKLNRRKKLEGGGERRGAPAAATAPAPTAHKAKK
AAPGAAGSRADKKPARQGKPEQRSHKKGAGAKKDKGGKAKKTAAGGKKVKKAAPSVKVPKGRK

hH1.1 sp|Q02539|H11_HUMAN Histone H1.1 OS=Homo sapiens GN=HIST1H1A PE=1 SV=3

MSETVPPAPAAASAAPEKPLAGKKAKKPAKAAAASKKKPAGPSVSELIVQAASSSKERGGV
SLAALKKALAAAGYDVEKNNSRIKLGKISLVSKGTLVQTKGTGASGSFKLNKKASSVETK
PGASKVATKTKATGASKKLKATGASKKSVKTPKKAKKPAATRKSSENPKPKTVKPKKV
AKSPAKAKAVKPKAAKARVTKPKTAKPKKAAPKKK

hH1.2 gi|4885375|ref|NP_005310.1| histone H1.2 [Homo sapiens]

MSETAPAAPAAAPPAEKAPVKKKAAKKAGGTTPRKASGPPVSELITKAVAASKERSGVSLAALKKALAAAGYDV
EKNNSRIKLGKLSLVSKGTLVQTKGTGASGSFKLNKKAASGEAKPKVKKAGGTPKKPVGAACKPKKAAGGAT
PKKSAKKTTPKKAKKPAAATVTKKVAKSPKKAHVAKPKKAASAAKAVKPKAAKPKVVKPKKAAPKKK

hH1.3 sp|P16402|H13_HUMAN Histone H1.3 OS=Homo sapiens GN=HIST1H1D PE=1 SV=2

MSETAPLAPTIPAPAEKTPVKKKAKKAGATAGKRKASGPPVSELITKAVAASKERSGVSL
AALKKALAAAGYDVEKNNSRIKLGKLSLVSKGTLVQTKGTGASGSFKLNKKAASGEKPK
AKKAGAAKPRKPAGAAKPKKVAGAATPKKSIKKTTPKKVKKPATAAGTKKVAKSARKVKT
PQPKKAAKSPAKAKAPKPKAAKPKSGKPKVTKAKKAAPKKK

hH1.4 gi|4885379|ref|NP_005312.1| histone H1.4 [Homo sapiens]

MSETAPAAPAAPAPAEKTPVKKKARKSAGAAKRKASGPPVSELITKAVAASKERSGVSLAALKKALAAAGYDV
EKNNSRIKLGKLSLVSKGTLVQTKGTGASGSFKLNKKAASGEAKPKAKKAGAAKAKKPAGATPKKAKKAA
PKKSAKKTTPKKAKKPAAAAGAKKAKSPKKAKAAKPKKAPKSPAKAKAVKPKAAKPKTAKPKAAKPKKAAAKK

hH1.5 gi|4885381|ref|NP_005313.1| histone H1.5 [Homo sapiens]

MSETAPAEATATPAPVEKSPAKKKATKKAAGAGAAKRKATGPPVSELITKAVAASKERNGLSLAALKKALAAGG
YDVEKNNSRIKLGKLSLVSKGTLVQTKGTGASGSFKLNKKAASGEAKPKAKKAGAAKAKKPAGATPKKAKKAA
GAKKAVKKTTPKKAKKPAAAGVKKVAKSPKKAKAAKPKKATKSPAKPKAVKPKAAKPKAAKPKAAKPKAAKAK
KAAAKK

hH1oo gi|24475863|ref|NP_722575.1| histone H1oo isoform 1 [Homo sapiens]

MAPGSVTSDISPSTSTAGSSRSPSEKPGPSHGGVPPGGPSHSSLPVGRRHPPVLRMVLEALQAGEQRRGTS
VAAIKLYILHKYPTVDVLRFKYLLKQALATGMRRGLLARPLNSKARGATGSFKLVPKHKKKIQPRKMAPATAP
RRAGEAKGKGPKKPSEAKEDPPNVGKVKKAAKRPKVQKPPPKPGAATEKARKQGGAADTRAQSGEARKVPP
KPDKAMRAPSSAGLSRKAKAKGSRSSQDAEAYRKTKAESKSKPTASKVKNGAASPTKKKVAKAKAPKAG
QGPNTKAAAPAKGSGSKVVPVPAHLRSKTEAPKGP RKAGLPKASSSKVSSQRAEA

Shorter isoform

gi|815891235|ref|NP_001295191.1| histone H1oo isoform 2 [Homo sapiens]

MAPATAPRRAGEAKGKGPKKPSEAKEDPPNVGKVKKAAKRPAAKVPKPPKPGAATEKARKQGGAAKDTRAQSG
EARKVPPKPDKAMRAPSSAGGLSRKAKAKGSRSSQGDAAEAYRKTKAESKSSKPTASKVKNGAASPTKKKVAK
AKAPKAGQGPNTKAAAPAKGSGSKVPAHLRSKTEAPKGPRKAGLPIKASSSKVSSQRAEA

hH1t gi|20544168|ref|NP_005314.2| histone H1t [Homo sapiens]

MSETVPAASASAGVAAMEKLPTKKRGRKPAGLISASRKVPNLSVSKLITEALSVSQERVGMSLVALKKALAAA
GYDVEKNNSRIKLSLKSILVNKGILVQTRGTGASGSFKLSKKVIPKSTRSKAKKSVSAKTKKLVLSDSKSPKT
AKTNKRAKKPRATTPKTVRSRGRKAKGAKGKQQQKSPVKARASKSKLTQHHEVNVNRKATSKK

hH1T2 gi|32401437|ref|NP_861453.1| testis-specific H1 histone [Homo sapiens]

MEQALTGEAQRWPRRGSGGAMAEAPGPSGESRGHSATQLPAEKTVGGPSRGCSVLRVSQLVLQAISTHKGLTLAALK
KELRNAGYEVRRKSGRHEAPRGQAKATLLRVSGSDAAGYFRVWVKPVRKPRRKPGRARQEEGTRAPWRTPAAPRSSRRRRQP
LRKAARKAREVWRRNARAKANARARRRRARPRAKEPPCARAKEEAGATAADEGRGQAVKEDTTPRSGKDKRRSSKPR
EEKQEPKKPAQRTIQ

hHILS gi|38372285|sp|P60008.1|HILS1_HUMAN RecName: Full=Spermatid-specific linker histone H1-like protein

MLHASTIWHLRSTPPRRKQWGHCDPHRILVASEVTTEITSPTPAPRAQVCGGQPWVTVLDPLSGHTGREAEERH
FATVSIISAVELKYCHGWRPAGQRVPSKTATGQRTCAKPCQKPKSTSKVILRAVADKGTCKYVSLATLKKAVSTT
GYDMARNAYHFKRVLKGLVDKGSAGSFTLGKKQASKSKLVKVRQRQRWRSGQRPFQHRSLLGSKQGKHRLI
KGVRRVAKCHCN

Clustal omega multiple sequence alignment of human full-length linker histones

```

hHILS      MLHASTIWLRLSTPPRRKQWGHCDPHRILVASEVTTEITSPTAPRAQVCGGQPWVTVLD 60
hH1T2      -----MEQ-ALTGEAQSRLWPR-----RGGG-----G 20
hH1oo      -----MAPGSVTSDISPSSTS-----TAGSSR-----S 23
hH1x       -----MSVELEEALPVTTA-----E-GM-----A 18
hH1.0      -----G 1
hH1t       -----MSETVPAASA-----SAGV-----A 15
hH1.1      -----MSETVPPAPA-----AS-----A 13
hH1.5      -----MSETAPAETA-----TP-----A 13
hH1.3      -----MSETAPLAPT-----IP-----A 13
hH1.2      -----MSETAPAAPA-----AA-----P 13
hH1.4      -----MSETAPAAPA-----AP-----A 13

hHILS      PLSGHTGREAEERHFATVSIISAVELKYCHGWRPAGQRPVPSKTATGQRTCAKPCQKPKSTSKV 120
hH1T2      AMAEAPGPSGES-----RGHSATQLPAE--KTVG-GPSRG-----CSSSVLRV 60
hH1oo      PESEKPGPS-----H---GG--VPPG-GPSHSSLPVGRHPPVLRM 58
hH1x       KKVTKAGGS-----A---AL--SPSK---KRKNSKKKNQPGKYSQL 51
hH1.0      PTENSTSA-----P---AA--KP-K---RAKASKKSTDHPKYSDM 32
hH1t       AMEKLPTKK-----R---GR--KPA---GLISASRKVPNLSVSKL 47
hH1.1      APEKPLAGK-----K---AK--KPAK---AAAASKKKPAGPSVSEL 46
hH1.5      PVEKSPAKK-----K---AT--KKA---GAGAAKRAKATGPPVSEL 46
hH1.3      PAEKTVPKK-----K---AK--KAG---ATAGKRKASGPPVSEL 44
hH1.2      PAEKAPVKK-----K---AA--KKA---GGTPRKASGPPVSEL 43
hH1.4      PAEKTVPKK-----K---AR--KSA---GAAKRKASGPPVSEL 43

      .                               ;

hHILS      ILRAVADKGTCKYVSLATLKK-AVSTTGMDMARNAYHFKRVLKGLVDK----- 167
hH1T2      SQLVLQAISTHKLTLAALKK-ELRNAGYEVRRKSGRHEAP-----RGQAKATL---LR 110
hH1oo      VLEALQAGEQRRGTSVAAILK-YILHKYPTVD--VLRFKYLLKQALATGMRRGLLARPLN 115
hH1x       VVETIRRLGERNGSSLAKIYTEAKKVPWFQQNGRTYLYKYSIKALVQN---DTL---LQ 104
hH1.0      IVAAIQAEKNRAGSSRQSIQK-YIKSHYKVENADSQIKLSIKRLVTT---GVL---KQ 84
hH1t       ITEALSVSQERVGMSLVALKK-ALAAAGYDVEKNNSRIKLSLKLNVN---GIL---VQ 99
hH1.1      IVQAASSSKERGGVSLAALKK-ALAAAGYDVEKNNSRIKLGKLSLVSK---GTL---VQ 98
hH1.5      ITKAVAASKERNGLSLAALKK-ALAAAGYDVEKNNSRIKLGKLSLVSK---GTL---VQ 98
hH1.3      ITKAVAASKERSGVSLAALKK-ALAAAGYDVEKNNSRIKLGKLSLVSK---GTL---VQ 96
hH1.2      ITKAVAASKERSGVSLAALKK-ALAAAGYDVEKNNSRIKLGKLSLVSK---GTL---VQ 95
hH1.4      ITKAVAASKERSGVSLAALKK-ALAAAGYDVEKNNSRIKLGKLSLVSK---GTL---VQ 95

      .           :           :           ;

hHILS      ---GSAGSFTLGKKQASKSLKVKRQRQ---RWRS--GQ---RPFQGHRSLLGSKQ- 213
hH1T2      VSGSDAAGYFRVWKVPKPRRKPGRRARQEEGTRAPWRTPAAPRSSRRRRQPLRKAA-RKAR 169
hH1oo      SKARGATGSFKLVPKHKKKIQPRKMAPATA---PRR--AGEAKGKGPKKPS---EA---K 164
hH1x       VKGTGANGSFKLNRKKLEGGGERRGAPAAA-TAPAP-T-----AHKAKKAAPGAAGSR-- 155
hH1.0      TKGVGASGSFRLAKSDEPKKSVAFKKT--K-KEIKK-VATPKKASKPKKAA-SKAPTCKP 139
hH1t       TRGTGASGSFKLSKKVIPKSTRSKAKKSVS-AKTKK-LVLSRDSKSPKT----- 146
hH1.1      TKGTGASGSFKLNKKASSVETKPGASKVAT-K--TK-ATGA--SKKLKKAT----- 143
hH1.5      TKGTGASGSFKLNKKAASGEAKPKAKKAGA-AKAKK-PAGA--TPK--KAK-KAA----- 146
hH1.3      TKGTGASGSFKLNKKAASGEAKPKAKKAGA-AKPRK-PAGA--AKPKKVA-GAA----- 146
hH1.2      TKGTGASGSFKLNKKAASGEAKPKVKKAGG-TKPKK-PVGA--AKPKKAA-GGA----- 145
hH1.4      TKGTGASGSFKLNKKAASGEAKPKAKKAGA-AKAKK-PAGA--AKPKKAT-GAA----- 145

      .: * * :

```

hH1LS	-----GH-KRL-----IKGVRRVAKCHCN-----	231
hH1T2	EWRRRNARAKAKANARARRTRRARPRAKEPPCARAKEEAGATAADEGRGQAVKEDTTPRS	229
hH1oo	ED-PPNVGKVKA-----AKRPAKVQPPPKPGAATEKAR--KQGG--AA-KDTRAQS	211
hH1x	-----RA-----DKKPARGQKPEQRSHKKGAGAKKDKGGK--AK-KTA----	190
hH1.0	KA-TPVKKAKKKL-----AATPKKAKKPKTVKAKP---VKASK-----	173
hH1t	-----AKTNKRAKKPRATTP-----KTVRSGR--KA-K-----	171
hH1.1	-----GASKKS-----VKTPKKAKKPAATRKS-----KNPK--KP-K-----	173
hH1.5	-----GAKKAV-----KKTTPKKAKKPAAA-GVK---KVAKSPK--KA-KA-----	179
hH1.3	-----TPKKS-----KKTTPKKVKKPATAAGTK---KVAKSAK--KV-K-----	179
hH1.2	-----TPKKS-----KKTTPKKAKKPAAATVTK---KVAKSPK--KA-K-----	178
hH1.4	-----TPKKS-----KKTTPKKAKKPAAAAGA---KKAKSPK--KA-K-----	177
: :		
hH1LS	-----	231
hH1T2	GKDK-----RRSSKPREEKQE	245
hH1oo	GEARKVPPKPDKAMRAPSSAGGLSRKAKAKGSRSSQGDAAEYRKTKAESKSSKPTASKVK	271
hH1x	-----AAGGKKVKK---A-----AKPSVPKVPKGR	212
hH1.0	-----PKK-----AKPVKPKAK---	185
hH1t	-----GAKGKQQQKSPVK-----ARASKSLTQHH	196
hH1.1	-----TVKPKKVAKSPAK-----AKAVKPKAAKAR	198
hH1.5	-----AAKPKKATKSPAK-----PKAVKPKAAKPK	204
hH1.3	-----TPQPKKAAKSPAK-----AKAPKPKAAKPK	204
hH1.2	-----VAKPKKAAKSA-----AKAVKPKAAKPK	201
hH1.4	-----AAKPKKAPKSPAK-----AKAVKPKAAKPK	202
hH1LS	-----	231
hH1T2	PKK--PAQRTIQ-----	255
hH1oo	NGAASPTKKKVAKAKAPKAGQGPNTKAAAPAKGSGSKVVPAPHLSRKTEAPKGPRAKGLP	331
hH1x	K-----	213
hH1.0	-----SSAKRAGKKK-----	195
hH1t	EVNVRKA-----TSKK-----	207
hH1.1	VTKPKTA-----KP-----KKAAPKKK-----	215
hH1.5	AAKPKAA-----KP-----KAAAKKAAAKKK-----	226
hH1.3	SGKPKVT-----KA-----KKAAPKKK-----	221
hH1.2	VVKPKKA-----AP-----KKK-----	213
hH1.4	TAKPKAA-----KP-----KKAAPKKK-----	219
hH1LS	-----	231
hH1T2	-----	255
hH1oo	IKASSSKVSSQRAEA	346
hH1x	-----	213
hH1.0	-----	195
hH1t	-----	207
hH1.1	-----	215
hH1.5	-----	226
hH1.3	-----	221
hH1.2	-----	213
hH1.4	-----	219

Figure A 1 Clustal omega multiple sequence alignment of full-length human linker histone proteins.

Molecular weight estimation by Mass spectrometry for some of the recombinant human linker histone variants protein preparations

1. Human H1t- Expected MW is 22019Da. We obtain 22024Da.

2. Human H1.5- Expected MW is 22580.1Da. Observed molecular mass is 22577 Da.

3. Human H1t2 – Expected MW of HT-hH1t2 is 28053.27 Da. The observed molecular weight is slightly higher as 29726 Da. The DNA sequence has been verified and the nucleosome binding ability is confirmed by other means. We attribute this discrepancy to some modifications or attachment of small peptides contaminants to the highly charged regions.

Qualitative Analysis Report

Data Filename	hH1T2.d	Sample Name	Sample31
Sample Type	Sample	Position	Vial 1
Instrument Name	ESI TOF	User Name	
Acq Method	Protein18min4degrCLR.m	Acquired Time	3/13/2017 4:08:35 PM
IRM Calibration Status	All Ions Missed	DA Method	autoreportMS11-13kDa_LR.m
Comment			
Sample Group	Info.		

User Chromatograms

User Spectra

Peak List

m/z	z	Abund
100.07121		1903.7
101.05637		1827.5
105.06075		2460.2
107.07634		2915.3
109.06968	1	6645.7
111.07135	1	5233.3
113.06302		2470.9
115.06911		3813.1
119.07846		2088.8
121.08857		7937.9
122.08997		29885.5
123.0772	1	20056.8
124.07709	1	2356.8
125.07427	1	2014.9
127.07037		1747.5
131.1026		2875.3

Qualitative Analysis Report

132.06055	2051.4
136.10284	3111.4
137.08923	5324.1
138.08435	5306.9
139.08317	3599.4
141.09089	1897.8
151.09772	1876.7
153.08389	2281
155.09288	1889.3
167.0947	1893.9
169.07834	5266.9
179.10124	2258.1
192.0918	4380.2
193.10759	2002.6

Peak List

<i>m/z</i>	Abund	Score (DB)
14287.3353	5675	3
14529.1144	22405.1	4
14891.8103	9026.6	6
18916.5182	3926.5	4
27732.5139	4871.1	8
28095.2595	6957.1	8
28575.2703	9718.2	8
29060.3153	6390.9	8
29726.5259	48999	8
29744.9111	12653.7	8
29766.1197	9535.8	8
29829.7321	4228.7	8
29907.4944	3035.3	9
30394.1837	4540.9	9
30514.1512	6353.7	9
30638.5645	3466.2	9
30759.8684	3138.4	9
36871.2203	3796.1	9
42357.2954	3237.8	9
42861.8691	5349.3	9
43702.8188	4306.8	9
45168.2053	5788.2	9
45285.5381	4045.3	9
45527.5582	3675.5	9
45778.3901	3582.8	9
46389.6813	3729.9	9

4. Human Hils – The molecular mass was higher than that of the expected 28054 Da. The observed is 29057 Da, by MALDI-TOF and is 29727/29831 by LC/MS. However, the DNA sequencing results of the plasmid extracted from the cells harvested for protein purification was verified to be exact. We attribute this

discrepancy to some modifications or attachment of small peptides contaminants to the highly charged regions.

Qualitative Analysis Report

Data Filename	hHills.d	Sample Name	Sample27
Sample Type	Sample	Position	P1-A1
Instrument Name	ESI TOF	User Name	
Acq Method	Protein18min4degrCLR.m	Acquired Time	3/13/2017 2:04:11 PM
IRM Calibration Status	All Ions Missed	DA Method	autoreportMS11-13kDa_LR.m
Comment			

Sample Group Info.

User Chromatograms

User Spectra

Peak List

m/z	z	Abund
110.06786	1	17025
120.95805		16856.4
129.09727		13686.1
138.96776		24961.2
249.04309	1	11530.3
635.17357		14649.7
642.32156	1	15042.1
649.08985		13503.6
663.8229		15334
678.66053		16681.6
683.67398		15200.7
693.48804		17662.4
710.90532		19067.8
728.19441		18955.1
740.53012		15322.7
747.14749		20327.3

Qualitative Analysis Report

765.51885	19152.5
785.45932	18656.2
807.22219	18875.9
817.24469	12107.1
829.23313	17843.8
852.53645	17527.3
877.86035	16183.9
904.4656	16280.7
932.73195	15459.1
962.70818	14924.4
994.67784	14985.4
1029.07649	13453.8
1065.85279	13455
1105.27278	11960.2

Peak List

<i>m/z</i>	Abund	Score (DB)
14897.1428	12204.8	0
14915.5645	14496.3	3
14961.8929	13871	0
17750.174	76716.3	5
17957.762	101036.4	5
26623.9637	17142.8	8
26936.54	23697.5	8
29727.2677	465955	9
29743.3965	187553.2	9
29759.9091	126223.4	9
29774.5568	147216	9
29794.7306	127042	9
29809.2951	110335.3	9
29831.0695	576619.3	9
29846.5973	305916.7	9
29863.7398	193646.3	9
29879.331	185061.5	9
29894.8893	152566.5	9
29934.6775	91394.6	9
29950.4904	91192.7	9
29965.8608	60377.7	9
29985.0436	32677.3	8
29999.97	17463.8	8
35499.4817	14488.1	9
35917.0392	13864	9
43765.4818	10805.2	9

Secondary Structure Prediction of human H1T2

Figure A 2 Secondary sequence element prediction of human H1T2 protein using GOR tool from Exspasy

'c' represent coils, 'h' represent helices and 'e' indicate extended strands. About 70 residues are shown in each row marked by the numbers above and accordingly residue number can be deduced.

3D Structure prediction of human H1T2

Figure A 3: A Pymol representation of a 3D structure prediction of human H1T2 protein using CASP (Moult et al., 2014)

A 3D cartoon representation, colored as spectrum, of the CASP model of human H1T2. The ends of the different alpha helices are labeled using residue numbers.

Evaluating the appropriateness of molecular replacement using model 1

A model of paired nucleosome derived from the asymmetric unit of the H1x chromatosome structure was used as a search model for performing molecular replacement using Phaser.

Figure A 4: Close view of the Electron density map to validate the appropriateness of using a paired-nucleosome model

A. The model stems from molecular replacement with Model 1 of figure 4.17 and 4.18, followed by rigid body and restrained refinement. Panels B and C highlight some example regions of DNA and core histones, respectively. In panels B and C, 2Fo-Fc electron density map in blue, contoured at 1.10σ , is shown superimposed. Ovals or circles are used to exemplify the regions with a poor fit to the electron density.

Evaluating the appropriateness of molecular replacement using model 3

A model of individual nucleosomes, derived from the H1x-chromatosome structure was used to search for 2 copies of nucleosomes to perform molecular replacement using Phaser.

Molecular replacement and refinement statistics:
Search performed for 2 copies of one nucleosome
of the H1x-chromatosome structure.

	Initial	Final
R factor	0.2775	0.2652
R free	0.2652	0.2681
Rms BondLength	0.0139	0.0092
Rms BondAngle	1.6318	1.3574
Rms ChirVolume	0.1048	0.0819

Figure A 5: Close view of the Electron density map to validate the appropriateness of using one nucleosome model

A. The model stems from molecular replacement with one nucleosome of Model 1 of figure 4.17 and 4.18, followed by rigid body and restrained refinement. Panels B and C highlight some example regions of DNA and core histones, respectively. In panels B and C, 2Fo-Fc electron density map in blue, contoured at 1.50σ , is shown superimposed. Ovals or circles are used to exemplify the regions with a poor fit to the electron density.

Evaluating the appropriateness of molecular replacement using model 2

A paired-nucleosome model, similar to the asymmetric unit of the H1x-chromatosome was generated using the search performed using model 3 depicted in Figure A2. This model, referred to as model 2 was used to search for 1 copy using molecular replacement. The model was further refined after fitting DNA ends into density.

Figure A 6: Close view of the Electron density map to validate the appropriateness of a modified paired-nucleosome model 2

A. The model stems from molecular replacement with Model 2 of figure 4.17 and 4.18, followed by rigid body and restrained refinement. In B, C and D, 2Fo-Fc electron density map in blue, contoured at 1.10σ , is shown superimposed. B is a representative of the DNA regions fitting to the map density. C and D are representative images of some core histone regions fitting into the electron density. Ovals or circles are used to exemplify the regions with a good fit to the electron density.

Bacterial Strains

BL21 (DE3)	Novagen, USA
BL21 (DE3) pLysS	Invitrogen, USA
HB101	Promega, USA
Rosetta (2) DE3	Novagen, USA

Addgene plasmids for tagged constructs

29706	pET His6 MBP N10 TEV LIC cloning vector (2C-T)
29707	pET His6 GST TEV LIC cloning vector (2G-T)
29708	pET His6 MBP TEV LIC cloning vector (2M-T)
29709	pET His6 NusA TEV LIC cloning vector (2N-T)
29712	pET His6 Thioredoxin TEV LIC cloning vector (2T-T)

Materials

24-well crystallization plates	Hampton Research, USA
26/60 Sephacryl S-200 HR column	GE Healthcare, USA
AKTA FPLC	GE Healthcare, USA
Amicon concentrator device, 10 kDa molecular weight cut off	Millipore, USA
DyeEx 2.0 Spin Kit	QIAGEN, Germany
G:Box machine	Syngene, UK
Gel plates	Bio-Rad, USA
High vacuum grease	Dow Corning, USA
HiLoad 16/60 Superdex 200	GE Healthcare, USA
HiTrap HP (5 ml)	GE Healthcare, USA
HiTrap CM FF (1 ml)	GE Healthcare, USA
HiTrap DEAE FF (1 ml)	GE Healthcare, USA
HiTrap Heparin HP (5 ml)	GE Healthcare, USA
Imaging Screen-K	Bio-Rad, USA
Minisart® syringe filters, 0.2 µm	Sartorius Stedim Biotech, France
Molecular Imager FX	Bio-Rad, USA
MonoQ 5/50 GL column	GE Healthcare, USA
Nanodrop 200 spectrophotometer	Thermo Fisher Scientific, USA
Prep cell, Model 491	Bio-Rad, USA
Quartz SUPRASIL precision cell, 10 mm light path, 15 mm center	Hellma, Germany
Resource S (6 ml)	GE Healthcare, USA
Rigaku R-Axis IV++ image plate	Rigaku, Japan
Siliconized glass coverslips	Hampton Research, USA
Slab Gel Drier Model GD2000	Hoefer Inc., USA
Sonics Vibra Cell Sonicator	Sonics & Materials, USA
Whatman chromatography paper, 3 mm	Sigma-Aldrich, USA

Chemicals

2'-deoxythymidine triphosphate (dTTP)	Thermo Fisher Scientific, USA
2-methyl-1,3 propanediol (MPD)	Sigma-Aldrich, USA
Acetic acid	Merck, USA
Acrylamide	Bio-Rad, USA
Acrylamide:Bis, 30% (29:1)	Bio-Rad, USA
Agarose	Affymetrix, USA
Ammonium persulfate (APS)	Acros Organics, Belgium
Ammonium sulfate	Sigma-Aldrich, USA
Ampicillin	Sigma-Aldrich, USA
Bacto Agar	BD Biosciences, USA
Bacto Tryptone	BD Biosciences, USA
Bis	Bio-Rad, USA
Bis-Tris propane	Sigma-Aldrich, USA
Boric acid	Merck, USA
Bovine serum albumin (BSA)	Sigma-Aldrich, USA
Bromophenol blue	Sigma-Aldrich, USA
Chloramphenicol	Acros Organics, Belgium
Chloroform	Fisher Scientific UK Ltd, UK
cOmplete, EDTA-free protease inhibitor cocktail	Roche, USA
Coomassie Brilliantblau G-250	Applichem, Germany
Dihydrogen potassium phosphate (KH ₂ PO ₄)	Merck, USA
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, USA
Dipotassium hydrogen phosphate (K ₂ HPO ₄)	Merck, USA
Dithiothreitol (DTT)	Gold Biotechnology, USA
DNA Loading Dye	Bio-Rad, USA
Ethanol	Merck, USA
Ethidium bromide	Bio-Rad, USA
Ethylenediaminetetraacetic acid (EDTA)	Bio-Rad Laboratories, USA
Formamide	Sigma-Aldrich, USA
Formic acid	Sigma-Aldrich, USA
Glucose	Alfa Aesar, USA
Glycerol	USB Corporation, USA
Glycine	1st Base Pte. Ltd., Singapore
Guanidine hydrochloride	Sigma-Aldrich, USA
Hydrochloric acid (HCl)	Merck, USA
Imidazole	Sigma-Aldrich, USA
InstantBlue	Expedeon, UK
Isoamyl alcohol	Acros Organics, Belgium
Isopropanol	Merck, USA
Isopropyl β -D-1-thiogalactopyranoside (IPTG)	Gold Biotechnology, USA
Loading dye, Blue/Orange, 6x	Promega, USA
Magnesium chloride (MgCl ₂)	Sigma-Aldrich, USA

Magnesium sulfate (MgSO ₄)	Sigma-Aldrich, USA
Manganese chloride (MnCl ₂)	Sigma-Aldrich, USA
Methanol	Merck, USA
Phenol	Sigma-Aldrich, USA
Piperidine	Sigma-Aldrich, USA
Polyethylene glycol, 6 kDa (PEG-6000)	Sigma-Aldrich, USA
Potassium acetate	Alfa Aesar, USA
Potassium cacodylate (K-caco)	Sigma-Aldrich, USA
Potassium chloride (KCl)	Merck, USA
Potassium dihydrogen phosphate (KH ₂ PO ₄)	Sigma-Aldrich, USA
Potassium permanganate (KMnO ₄)	Sigma-Aldrich, USA
Riboadenosine triphosphate (rATP)	Sigma-Aldrich, USA
RNA, sodium salt	Merck, USA
Sodium acetate	Sigma-Aldrich, USA
Sodium chloride (NaCl)	Merck, USA
Sodium dihydrogen phosphate (NaH ₂ PO ₄)	Sigma-Aldrich, USA
Sodium dodecyl sulphate (SDS)	Hoefer, Inc., USA
Sodium hydroxide (NaOH)	Merck, USA
Tetramethylethylenediamine (TEMED)	Bio-Rad, USA
Trehalose	Sigma-Aldrich, USA
Tris	USB Corporation, USA
Triton X-100	Bio-Rad, USA
Tween-20	Bio-Rad, USA
Tween-20	Bio-Rad, USA
Urea	USB Corporation, USA
Xylene cyanol FF	Sigma-Aldrich, USA
Yeast extract	USB Corporation, USA
Zinc chloride (ZnCl ₂)	Sigma-Aldrich, USA
Zinc sulfate (ZnSO ₄)	Sigma-Aldrich, USA
Zinc sulfate (ZnSO ₄)	Sigma-Aldrich, USA
β-mercaptoethanol (BME)	Sigma-Aldrich, USA
γ ³² P-ATP	PerkinElmer Inc., USA

Enzymes and Buffers

Calf intestinal alkaline phosphatase (CIAP)	New England Biolabs, USA
EcoRV	New England Biolabs, USA
HinfI	New England Biolabs, USA
NEBuffer 1	New England Biolabs, USA
PNK buffer	New England Biolabs, USA
RNase A	Bio-Rad Laboratories, USA
T4 DNA ligase	New England Biolabs, USA
T4 DNA ligase buffer	New England Biolabs, USA
T4 polynucleotide kinase (PNK)	New England Biolabs, USA
Thrombin	GE Healthcare, USA

Media and Buffer Recipes

10% polyacrylamide gel	10% (v/v) Acrylamide 0.17% (v/v) Bis 1x TBE buffer 0.1% (v/v) APS 0.1% (v/v) TEMED
15% SDS-PAGE	Resolving gel 15% (v/v) Acrylamide 0.25% (v/v) Bis 375 mM Tris-HCl (pH 8.8) 0.1% (w/v) SDS 0.1% (v/v) APS 0.1% (v/v) TEMED Stacking gel 5% (v/v) Acrylamide 0.25% (v/v) Bis 125 mM Tris-HCl (pH 6.8) 0.1% (w/v) SDS 0.1% (v/v) APS 0.1% (v/v) TEMED
18% SDS-PAGE	Resolving gel 18% (v/v) Acrylamide 0.3% (v/v) Bis 375 mM Tris-HCl (pH 8.8) 0.1% (w/v) SDS 0.1% (v/v) APS 0.1% (v/v) TEMED Stacking gel 5% (v/v) Acrylamide 0.25% (v/v) Bis 125 mM Tris-HCl (pH 6.8) 0.1% (w/v) SDS 0.1% (v/v) APS 0.1% (v/v) TEMED

2xTY media	1.6% (w/v) Bacto Tryptone 1.0% (w/v) Yeast extract 86 mM NaCl
4% polyacrylamide gel	4% (v/v) Acrylamide 0.05% (v/v) Bis 0.25x TBE buffer 0.1% (v/v) APS 0.1% (v/v) TEMED
6% polyacrylamide gel	6% (v/v) Acrylamide 0.15% (v/v) Bis 0.25x TBE buffer 0.1% (v/v) APS 0.1% (v/v) TEMED
Coomassie Blue	40% Methanol 10% Acetic acid 0.1% Coomassie Brilliant Blue G-250
Denaturing DNA loading dye (1x)	95% (v/v) Formamide 0.025% (w/v) Bromophenol blue 0.025% (w/v) Xylene cyanol FF 0.5 mM EDTA
Protein Loading Dye (1x)	15 mM Tris-HCl (pH 6.8) 0.5% (w/v) SDS 5% (v/v) Glycerol 0.00625% (w/v) Bromophenol blue 2.63% (v/v) BME
TB media	12% (w/v) Bacto Tryptone 24% (w/v) Yeast extract 17 mM KH ₂ PO ₄ 72 mM K ₂ HPO ₄ 0.4% (v/v) Glycerol
TBE buffer (1x)	89 mM Tris-HCl (pH 7.5) 89 mM Boric acid 2 mM EDTA
TCS-n buffer	20 mM Tris-HCl (pH 7.5) 1 mM EDTA (pH 8.0) 1 mM DTT n M KCl
TE(10,0.1)	10 mM Tris-HCl (pH 8.0) 0.1 mM EDTA (pH 8.0)
TE(10,50)	10 mM Tris-HCl (pH 8.0) 50 mM EDTA (pH 8.0)
TGS buffer (1x)	25 mM Tris base 200 mM Glycine 0.1% SDS
TYE agar	10% (w/v) Bacto Tryptone 5% (w/v) Yeast extract 137 mM NaCl 15% (w/v) Bacto Agar
Tris-borate (1x)	89 mM Tris-HCl (pH 7.5) 89 mM Boric acid