# Chromatin Organisation: Molecular Role of H2A Variants

**By**

**Saikat Bhattacharya**

[LIFE09201004024]

**Tata Memorial Centre**

**Mumbai**

*A thesis submitted to the*

*Board of Studies in Life Sciences*

*In partial fulfillment of requirements*

*for the Degree of*

**DOCTOR OF PHILOSOPHY**

***of***

**HOMI BHABHA NATIONAL INSTITUTE**


**September, 2016**

# HOMI BHABHA NATIONAL INSTITUTE

## Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Saikat Bhattacharya entitled **'Chromatin Organisation: Molecular Role of H2A Variants'** and recommend that it may be accepted as fulfilling the dissertation requirements for the Degree of Doctor of Philosophy.

---------- Date: 8/9/2016

**Chairperson:** -- Dr. Shubhada Chiplunkar

---------- Date: 8/9/2016

**Guide/Convenor:** -- Dr. Sanjay Gupta

---------- Date: 8/9/2016

**Member 1:** -- Dr. Ashok Varma

---------- Date: 8/9/16

**Member 2:** -- Dr. Dibyendu Bhattacharyya

---------- Date: 08/9/2016

**External Examiner:** Dr. M.M. Chaturvedi

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 8/9/2016

Place: Navi Mumbai

Dr. Sanjay Gupta
(Guide)

# STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Saikat Bhattacharya

# DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. This work is original and has not been submitted earlier as a whole or in part for a degree/ diploma at this or any other Institution / University.

Saikat Bhattacharya

# LIST OF PUBLICATIONS ARISING FROM THE THESIS

## Journal

1. Bhattacharya, S & and Gupta, S. A Simple Method to Produce Sub-Nucleosome Complexes of High Purity *In Vitro*, *Advances in Biosciences and Biotechnology*, March, 2016, Vol. 7, No. 3.
2. Bhattacharya, S., Reddy, D., Ingle, A., Khade, B., & Gupta, S. Histone H2A mono-ubiquitination and cellular transformation are inversely related in N-nitrosodiethylamine-induced hepatocellular carcinoma. *Experimental Biology and Medicine,* 2016. http://doi.org/10.1177/1535370216649262.
3. Bhattacharya, S., Reddy, D., Reddy, R., Sharda, A., Bose, K., & Gupta, S. Incorporation of a tag helps to overcome expression variability in a recombinant host. *Biotechnology Reports*, 2016, 11, 62–69. http://doi.org/10.1016/j.btre.2016.06.002.
4. Histone H2A.1 Promotes Attainment of Distinct Physiological States by Altering Chromatin Dynamics **(Manuscript Communicated).**

## Chapter in books

1. Monica Tyagi, Shafqat A Khan, Saikat Bhattacharya, Divya Reddy, Ajit K Sharma, Bharat Khade, Sanjay Gupta. Techniques to Access Histone Modifications and Variants in Cancer. 'Cancer Epigenetics: Risk Assessment, Diagnosis, Treatment, and Prognosis', *Methods in Molecular Biology Series* (Humana Press/Springer) 2015. 251-272.

## Conferences

1. Presented poster entitled "A Structural and Functional Study of Histone H2A.1 Isoform, an emerging regulator of the epigenetic landscape" at the Chromatin Structure

and Function conference held at Les Diablerets, Switzerland, 2016.

2. Presented poster entitled “Histone H2A.1 isoform: the emerging regulator of the epigenetic landscape” at the TMC Platinum Jubilee Conference held at NCPA, Mumbai, India, 2016.
3. Presented poster entitled “A Structural and Functional Study of the Role of H2A Isoforms H2A.1 and H2A.2 in Cancer” at the Chromatin and Epigenetics conference held at EMBL, Heidelberg, Germany, 2015.
4. Presented poster entitled “A Comprehensive Study of the Role of H2A Isoforms in Carcinogenesis” at the Carcinogenesis 2015 conference held at ACTREC, Navi Mumbai, India, 2015.
5. Presented poster entitled “Histone Isoforms H2A.1 Forms Stabler Nucleosome Compared to H2A.2 and Promotes Proliferation” at the 5th Meeting of the Asian Forum of Chromosome and Chromatin Biology held at JNCASR, Bangalore, India, 2015.
6. Presented poster entitled “Role of Histone Isoforms H2A.1 and H2A.2: A Structural and Functional Study” at the 4th Asian Conference on Environmental Mutagens held at IICB, Kolkata, India, 2014.
7. Presented poster entitled “Effect of H2A Isoforms on Chromatin Structure and Stability” at the Biopaschim conference held at ACTREC, Navi Mumbai, India, 2014.
8. Presented a poster entitled “Deciphering the Molecular Role of Histone H2A Variants, H2A.1 and H2A.2” at the Frontiers in Modern Biology conference organized at Indian Institute of Science, Bangalore, India in 2013.

Saikat Bhattacharya

# ACKNOWLEDGEMENTS

The 5-6 years of my PhD has been eventful. Like any other period of life, good people and friends are needed to help a person go through the ups and downs. Dr. Sanjay Gupta has not only given me the opportunity to work under his guidance but allowed me to explore myself as a researcher, which I now realize is the most important aspect of learning to do science. The way he has bestowed confidence in me in the initial years of my PhD is remarkable considering that none of my experiments were working out at that time and it is easy to lose faith in an amateur young man who is struggling to find his feet in the uncertainties of research. His trust not only helped me to stay motivated and focused through my struggling years, but also made me more responsible. It always was in my subconscious mind that I need to prove Sir's confidence in me right which made me push myself harder that extra bit. I thank you Sir for being so supportive and helping me become the researcher that I am today.

It does not happen many times in one's life that he/she is worried that words may not do justice to the sacrifice done by another person. If anybody has the equal right to celebrate the success of my PhD, then it is Divya Velga Reddy. She was a constant reminder that I am not in this struggle alone. In her I found a junior and a friend on whom I knew I can always count on professionally as well as personally. She has been my companion through this demanding journey of PhD. Divya, this thesis belongs to you as much as it does to me.

To come through the tough periods in life, people need inspiration. *Bordi* (Ila Mukherjee), you have been my constant inspiration for the last 14 years to contribute to the knowledge of cancer. I remember how much happy you were when you got the news that I have been selected at ACTREC. I wish you stayed with us long enough to witness the day on which I become a Doctorate. I know that your blessings will always be there with me and your absence will be a reminder for me of the importance of the work that I am doing.

I would like to thank my seniors Ajit, Monica and Shafqat for helping me learn my first steps of research work. Ram and Asmita, thank you for being such good human beings. Thank you Sanket and Mudasir for being such great juniors. My sincere appreciation to all my trainees (Vignesh, Dhanashree, Rakesh, Madhuli, Sanidhi, Krupali, Sneha, Riddhi, Sujay) for being part of my every happiness and disappointment during their stay in ACTREC. My deepest appreciation to Bharat, Santosh and Arun ji for supporting the research work and maintaining an amiable environment in the lab.

I am thankful to Dr. S. Chiplunkar (Director, ACTREC), Dr. R. Sarin (Ex-Director, ACTREC) and Dr. S. Zingde (Ex-Deputy Director, ACTREC) for providing the infrastructure and constant encouragement to carry out research in the premier cancer centre, ACTREC, Tata Memorial Centre. I am indebted to CSIR, New Delhi for funding my research fellowship.

I would like to thank my present Doctoral Committee Members, Dr. S. Chiplunkar, Dr. Ashok Varma and Dr. D. Bhattacharyya and earlier members, Late Dr. R. Kalraiya, Dr. G. B. Maru and Dr. R. Makde for their critical analysis and comments on data, helpful suggestions and cooperativity towards the progress of work. Special thanks to Mr. Nikhil Gadewal, BTIS, ACTREC and CDAC, Pune for their bioinformatics support and Dr. Arvind Ingle from Laboratory Animal Facility, ACTREC. My heartfelt gratitude to Dr. Sanjeev Waghmare and Dr. Kakoli Bose for their constant support and encouragement.

At last I would like to thank the most important part of my life, my family and friends. I have been a selfish individual who has bothered only about his PhD. I have not maintained contacts, not attended ceremonies, not responded to their calls. But my friends and family have never held that against me and were always there for me even if I ignored them. I have not been able to take care of my old parents. Sorry *ma* and *baba*. I will never be able to compensate for my mistakes but I hope I have made you proud. Thank you *dada*. You have been the greatest pillar

of strength in my life. You do not realize the feeling of security you give to me because I know you are there to take care of everything.

I hope I will someday be able to contribute to the development of my country and the society after all that you people have done for me.

Saikat Bhattacharya

# Contents

# SYNOPSIS

# Homi Bhabha National Institute

**Ph. D. PROGRAMME**

**PROGRESS REVIEW REPORT**

1. **Name of the Student: Saikat Bhattacharya**
2. **Name of the Constituent Institution: Tata Memorial Centre, ACTREC**
3. **Enrolment No.: LIFE09201004024, 1st September 2010.**
4. **Title of the Thesis: Chromatin Organisation: Molecular Role of H2A Variants**
5. **Board of Studies: Life Sciences**

## Synopsis

### 1. Introduction

Epigenetics is the study of the change in gene expression pattern which does not involve change in the nucleotide sequence of DNA. The term epigenetics was coined by Waddington in 1942 and was derived from the Greek word "epigenesis" which originally described the influence of genetic processes on development. Epigenetic studies were initially focused on DNA methylation, however, subsequent studies elucidated the role of histones in regulating gene expression pattern.

Histones are a class of highly conserved basic proteins and packaging the genome was the primary function previously attributed to them. The core histones comprise of H2A, H2B, H3 and H4, form an octameric protein core around which ~147bp of DNA is wrapped to form the nucleosome core particle (NCP) [1]. Further compaction is achieved with the aid of linker histone H1 [2]. The canonical histone genes are synthesized during the S-phase and to meet up with their high demand during DNA replication, their encoding genes are present in cluster. For instance, there are three clusters of canonical histone genes present in humans at

chromosome number 1 and 6. Interestingly, differences in the primary sequence are observed amongst the histone H2A proteins encoded by these genes. These are termed as isoforms. In humans, there are 16 genes coding for H2A giving rise to 12 isoforms [3].

H2A isoforms, especially two isoforms H2A.1 (encoded by HIST1H2AH, accession number: NM_001315492.1) and H2A.2 (encoded by HIST2H2AA3, accession number: NM_001315493.1) have been reported to be differentially expressed in a variety of physiological states [4][5]. The two isoforms differ in the identity of only three residues: T16S, L51M and R99K. Earlier studies from our lab have shown the differential expression of these two major H2A isoforms in rat hepatocellular carcinoma model system [6]. In humans, the expression of H2A1C isoform was reported to be downregulated in chronic lymphocytic leukaemia (CLL) and gall bladder cancer cell line [7][8]. Interestingly, later on with a larger cohort of samples, H2A1C expression was conversely reported to be up regulated in CLL [9]. Also, H2A1C was found to be upregulated in non-small cell lung carcinoma [10]. The levels of H2A1C alters in diseases including human papillomaviruses hyperplasia, AIDS and multiple sclerosis [11,12]. Collectively, these reports suggest that the expression of the H2A.1 isoform in different states is intricately regulated in the cells. However, whether the observed changes are a consequence of the given state or H2A.1/H2A1C contributes to the attainment of such states has not been addressed in depth. Considering that in humans there are 12 such H2A isoforms, the possibility that these might have non-redundant functions needs to be investigated as it may contribute to the complexity necessary for generating the diversity in the epigenome to achieve cell type specific gene expression.

## 2. Objectives

I. To study alteration in organization of nucleosomal structure with H2A.1 and H2A.2.

II. To study the functional significance of H2A.1 and H2A.2.

## 3. Work Plan

**I. To study alteration in organization of nucleosomal structure with H2A.1 and H2A.2.**

1. *Purification of recombinant histones and their mutants using bacterial expression and purification system for reconstitution of histone oligomeric complexes.*
2. *Reconstitution of H2A.1/H2B and H2A.2/H2B dimers and their mutants using the purified recombinant histones for comparing the stability of the two complexes.*
3. *Circular dichroism and fluorescence spectroscopy of reconstituted H2A/H2B dimers and their different mutants for secondary and tertiary structure validation.*
4. *Equilibrium unfolding of dimers and their different mutants to obtain biophysical parameters to compare stability of H2A.1/H2B dimer with H2A.2/H2B dimer.*
5. *Salt dissociation and Fluorescence Recovery after Photobleaching (FRAP) assay for assessing the dynamics of H2A.1 and H2A.2 and their different mutants in cells.*
6. *Micrococcal nuclease digestion assay to investigate the effect of H2A.1 and H2A.2 on chromatin organisation.*
7. *Molecular Dynamic Simulation to investigate the stability and understand the basis of the altered stability of H2A.1 and H2A.2 containing nucleosomes.*

**II. To study the functional significance of H2A.1 and H2A.2.**

1. *RT-PCR and Real-time PCR to investigate the changes in expression level of H2A.1 (H2A1C) and H2A.2 isoforms in different cell lines and organs.*
2. *Cell cycle synchronization and release followed by flow cytometry analysis to investigate the effect of overexpression of H2A.1 and H2A.2 isoforms on cell cycle profile.*
3. *Cell proliferation and colony formation assay to see the effect of overexpression of H2A.1 and H2A.2 isoforms and their different mutants on cell proliferation.*
4. *Wound healing assay to investigate the effect of H2A.1 and H2A.2 isoforms on the migration of cells.*

5. *Partial hepatectomy followed by histone analysis on AUT-PAGE to investigate the expression level of H2A.1 and H2A.2 isoforms during the process of liver regeneration.*

## 4. Results

### 4.1. Global decrease in H2Aub is observed in hepatocellular carcinoma

We implanted $3mm^2$ tumor from Sprague-Dawley rat into NOD-SCID mice. Successful development of tumor demonstrates that NDEA leads to stable transformation of cells. Along with the alteration in the intensity of bands on AUT-PAGE corresponding to H2A.1 and H2A.2, we observed a marked decrease in the intensity of a band in tumor in the H2A region. We identified that band as H2A by mass spectrometry, although it did not correspond to any one isoform. Immunoblotting confirmed that the band corresponds to H2Aub119. We found Usp21, which is a known deubiquitinase for H2Aub, to be upregulated in cancer. Also, this correlated with the increase in expression of growth factor gene lipocalin 2. These observations were also validated in CL44 (pre-neoplastic) and CL38 (neoplastic) liver cell lines. Further, an inverse correlation was observed between H2Aub and proliferation. There was an increase in H2Aub with concomitant decrease in H3S10p and H4ac on arresting cell proliferation be serum starvation or HDAC inhibitor (TSA) treatment.

### 4.2. H2A.1/H2A1C expression level varies in cancer cell lines

H2A1C in humans is the closest counterpart of H2A.1 in rat, differing in only the S16T substitution. The altered expression level of H2A1C has been reported to alter in certain human cancers [7–10,13–15]. Our observations in rat hepatocellular carcinoma prompted us to investigate the expression level of H2A1C and H2A.2 isoforms in human transformed cell lines. We included cell lines of liver (HepG2), colon (A431) and stomach (KATOIII, AGS) origin and their immortalised untransformed counterparts, that is, HHL5 (liver), HaCaT (colon) and stomach (HFE145) in our study, as the expression level of H2A1C in these cell lines has not been previously reported. An increase in the relative expression of H2A1C was

observed in HepG2 and A431. We did not find any significant changes in both the transformed cell lines of stomach. We also found upregulation of H2A1C in MCF7 consistent with a previously published report [14].

### 4.3. H2A.1 isoform is functionally non-redundant from the H2A.2 isoform

The varied expression levels of H2A.1 in different scenarios prompted us to investigate whether H2A.1 has functionally non-redundant roles. A marked increase in proliferation was observed in CL38 cells on H2A.1 overexpression. Similar effects were reflected in the colony formation assay, with H2A.1 over-expressing colonies substantially larger. To see the effect of the isoforms overexpression on cycling of cells, we studied the cell cycle profile of G1-enriched H2A.1/H2A.2 CL38 cells post 72hours serum release. Overexpression of H2A.1 led to discernible increase in the mitotic cell population (12%) compared to vector control (4%). No significant difference in the closure of the wound in scratch assays on H2A.1 overexpression was perceived. Notably, we did not observe increased proliferation of CL44 cells upon H2A.1 overexpression. Taken together, these data suggest that H2A.1 lead to proliferation in context of cancer.

### 4.4. Leu51 and Arg99 are important in conferring non-redundant functionality to H2A.1 isoform

The H2A isoforms, H2A.1 and H2A.2, differ in three residues in their primary amino acid sequence. We substituted the residues of H2A.1 to the corresponding ones of H2A.2 to identify the residue having the most significant impact on the non-redundant functionality of H2A.1. MTT assays suggested that mutating residue $99^{th}$ of H2A.1 drastically reduced the pro-proliferative effect observed on H2A.1 overexpression. Mutating residue $51^{st}$ also negatively affected proliferation and had a synergistic effect with residue $99^{th}$.

### 4.5. Recombinant expression and purification of histones

The histones were expressed with N-terminal 6xHis tag in the *E. coli* strain BL21 (DE3) pLysS. The expressed histones were completely part of insoluble fraction. 6M urea was used to solubilize the proteins which were then purified using Ni-NTA beads. All the core histones could be very efficiently purified and were used to reconstitute H2A/H2B dimers

### 4.6. H2A.1/H2B dimer is less stable than H2A.2/H2B dimer

To investigate how the histone isoforms, impart their non-redundant functional effects, we started off by comparing the *in vitro* stability of H2A.1/H2B and H2A.2/H2B dimer with purified recombinant histones. Equilibrium unfolding of reconstituted full length H2A/H2B dimers, which was previously described [16], was conducted to conduct the stability analysis of H2A.1/H2B and H2A.2/H2B dimers. The temperature of melting (Tm) for H2A.1/H2B dimer was determined to be $50.04^0$C whereas that of H2A.2/H2B dimer was found to be higher by ~$3^0$C at $53.31^0$C, suggesting the former is less stable. The CD and fluorescence data plotted in response to increasing chemical denaturant corroborated these observations the $[urea]_{1/2}$ for H2A.1/H2B obtained lower than H2A.2/H2B. Further, studies with the mutants suggest that the alteration of the $51^{st}$ residue had the biggest impact on the stability of the dimers. Mutating $51^{st}$ residue of H2A.1 into Met increased the Tm from $50.04^0$C to $52.3^0$C and that of H2A.2 (H2A.1 $16^{th}$ + $99^{th}$) into Leu decreased the stability by $2.1^0$C.

### 4.7. Histone H2A.1 isoform containing-nucleosomes are more stable

To understand the importance of alteration in dimer stability in context of chromatin, we carried out FRAP assay. The percentage recovery of H2A.1 after 1hour was markedly less (44.14%) compared to H2A.2 (64.7%) in the FRAP assay, suggesting that H2A.1 is less dynamic than H2A.2 isoform. Further, we found the R99K substitution, which is involved in the interaction with H4 tails in the NCP, independently brought about the most drastic increase (20%) in the dynamics of H2A.1 followed by L51M (12%). Mutating both the $51^{st}$ and $99^{th}$ residue together, of H2A.1 into the corresponding residues of H2A.2 led to almost

similar dynamics as observed for H2A.2. Molecular Dynamic Simulation corroborated the *in vitro* experiments as the RMSD of H2A.1 containing nucleosomes was found to be lower than H2A.2 containing ones. Further, Arg99 of H2A.1 could form more number of hydrogen bonds.

### 4.8. Incorporation of H2A.1 isoform does not impart structural alterations to chromatin

To see whether there is any global change in nucleosome spacing or chromatin accessibility, the chromatin of CL44 and CL38 cell lines was subjected to MNase digestion. No structural alterations were discernible on carrying out the assay. Similarly, the digestion profile was virtually identical for chromatin isolated from H2A.1/H2A.2 overexpressing CL38 cell lines. Further, on overlapping mononucleosomal DNA from the simulated structures, again, no differences were observed. Taken together, the chromatin structure and accessibility does not alter significantly on incorporation of H2A.1 and H2A.2 isoforms.

## 5. Discussion

We for the first time elucidate that even the homomorphous sequence divergent forms of histones can alter nucleosome dynamics. We found the L51M alteration to have the most significant impact on H2A/H2B dimer stability. The difference observed between H2A.1/H2B and H2A.2/H2B dimer stability is subtle compared to the change brought about by histone variants like H2A.Z [16]. Possibly, the abundance of H2A isoforms in the genome, as opposed to variants, makes this difference significant to induce alteration in epigenetic regulation. Further, we found that H2A.1 isoform gives rise to more stable nucleosomes although the H2A.1/H2B dimers were less stable. This is consistent thermodynamically as a less stable dimer would favour a more stable nucleosome. Similar to our observations, for H2A.Z variant it was reported that H2A.Z/H2B dimer was unstable as compared to the canonical H2A/H2B [16], however, the nucleosome was found to be more stable [17]. A

more stable nucleosome is expected to cause hindrance to chromatin-mediated processes like transcription, replication and repair.

We also found a decrease in transcription repressive mark H2Aub119 in cancer. The deubiquitinase Usp21 is profoundly upregulated in HCC. However, role of other ubiquitinases and deubiquitinases cannot be ignored. We also observed a decrease in H4ac and increase in H3S10p. Our data thus suggests for the first time an inverse correlation between the global levels of cellular transformation mark H3S10p and H2Aub *in vivo*.

One very important aspect that collectively emerges from our study and the earlier reports is that the functional effects exhibited by H2A isoforms might be context dependent, both in terms of extent and the effect itself. For example, the pro-proliferative effect conferred by H2A.1 was not observed in the pre-neoplastic CL44 cells. Notably, human H2A1C isoform, which was initially reported to be downregulated in CLL, was shown to exhibit anti-proliferative effects [7]. However, in a later study with higher number of samples H2A1C levels were found to be higher in CLL patients compared to samples from healthy individuals [9]. Further, the high expression of H2A.1 seen in brain, which are terminally differentiated, suggests that the actual functional effect of H2A.1 may be context dependent. Consistent with this, high expression of H2A1C has been seen in chemo-resistance in pancreatic cancer cell lines [18], which is again a different context.

In summary, we here show that NDEA administration leads to stable transformation of liver to HCC. At the molecular level, there is a marked decrease in the level of H2Aub in HCC with an inverse correlationship with H3S10p and H4ac. Further, there is a concomitant decrease in the expression of H2A.1 isoform. We demonstrated that H2A.1-containing nucleosomes are more stable owing to M51L and K99R substitutions that have the most prominent effect on cell proliferation, suggesting, that the nucleosome stability is intimately linked with the physiological effects observed. Possibly, the increased nucleosome stability

resulting from H2A.1 incorporation, contributes to the contextual alteration in global gene expression pattern, which collectively promotes the attainment of different physiological states. This possibility of the non-redundant function, when extended to the plethora of the histone isoforms encoded by multiple genes (H2A, 16 genes; H2B, 22 genes; H3, 14 genes; H4, 14 genes; and H1, 6 genes), truly increases the complexity of the epigenome by many folds.

**References**


1. Luger K, Mäder a W, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**: 251–260.
2. Finch JT, Klug A (1976) Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A* **73**: 1897–1901.
3. Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ (2002) The Human and Mouse Replication-Dependent Histone Genes. *Genomics* **80**: 487–498.
4. Vassilev a P, Rasmussen HH, Christensen EI, Nielsen S, Celis JE (1995) The levels of ubiquitinated histone H2A are highly upregulated in transformed human cells: partial colocalization of uH2A clusters and PCNA/cyclin foci in a fraction of cells in S-phase. *J Cell Sci* **108 ( Pt 3**: 1205–1215.
5. Tyagi M, Khade B, Khan SA, Ingle A, Gupta S (2014) Expression of histone variant, H2A.1 is associated with the undifferentiated state of hepatocyte. *Exp Biol Med (Maywood)* 1535370214531869.
6. Khare SP, Sharma A, Deodhar KK, Gupta S (2011) Overexpression of histone variant H2A.1 and cellular transformation are related in N-nitrosodiethylamine-induced sequential hepatocarcinogenesis. *Exp Biol Med (Maywood)* **236**: 30–35.
7. Singh R, Mortazavi A, Telu KH, Nagarajan P, Lucas DM, Thomas-Ahner JM, Clinton SK, Byrd JC, Freitas MA, Parthun MR (2013) Increasing the complexity of chromatin:

functionally distinct roles for replication-dependent histone H2A isoforms in cell proliferation and carcinogenesis. *Nucleic Acids Res* **41**: 9284–9295.

8. Su X, Lucas DM, Zhang L, Xu H, Zabrouskov V, Davis ME, Knapp AR, Young DC, Payne PRO, Parthun MR, et al. (2009) Validation of an LC-MS based approach for profiling histones in chronic lymphocytic leukemia. *Proteomics* **9**: 1197–1206.
9. Singh R, Harshman SW, Ruppert AS, Mortazavi A, Lucas DM, Thomas-Ahner JM, Clinton SK, Byrd JC, Freitas MA, Parthun MR (2015) Proteomic profiling identifies specific histone species associated with leukemic and cancer cells. *Clin Proteomics* **12**: 22.
10. Zhang H-H, Zhang Z-Y, Che C-L, Mei Y-F, Shi Y-Z (2013) Array analysis for potential biomarker of gemcitabine identification in non-small cell lung cancer cell lines. *Int J Clin Exp Pathol* **6**: 1734–1746.
11. Olivares I, Ballester A, Lombardia L, Dominguez O, López-Galíndez C (2009) Human immunodeficiency virus type 1 chronic infection is associated with different gene expression in MT-4, H9 and U937 cell lines. *Virus Res* **139**: 22–31.
12. Singh MK, Scott TF, LaFramboise WA, Hu FZ, Post JC, Ehrlich GD (2007) Gene expression changes in peripheral blood mononuclear cells from multiple sclerosis patients undergoing β-interferon therapy. *J Neurol Sci* **258**: 52–59.
13. Zhan F, Barlogie B, Arzoumanian V, Huang Y, Williams DR, Hollmig K, Pineda-Roman M, Tricot G, Van Rhee F, Zangari M, et al. (2007) Gene-expression signature of benign monoclonal gammopathy evident in multiple myeloma is linked to good prognosis. *Blood* **109**: 1692–1700.
14. Su CH, Tzeng TY, Cheng C, Hsu MT (2014) An H2A histone isotype regulates estrogen receptor target genes by mediating enhancer-promoter-3'-UTR interactions in breast cancer cells. *Nucleic Acids Res* **42**: 3073–3088.

15. Kavak E, Unlü M, Nistér M, Koman A (2010) Meta-analysis of cancer gene expression signatures reveals new cancer genes, SAGE tags and tumor associated regions of co-regulation. *Nucleic Acids Res* **38**: 7008–7021.

16. Placek BJ, Harrison LN, Villers BM, Gloss LM (2005) The H2A.Z/H2B dimer is unstable compared to the dimer containing the major H2A isoform. *Protein Sci* **14**: 514–522.

17. Hoch D a., Stratton JJ, Gloss LM (2007) Protein-Protein Förster Resonance Energy Transfer Analysis of Nucleosome Core Particles Containing H2A and H2A.Z. *J Mol Biol* **371**: 971–988.

18. Yi Zhang, Xin Shi, Department of General Surgery, Zhongda Hospital of Southeast University, Nanjing 210009, Jiangsu Province C (2009) Expression of HIST1H2AC and HIST1H2BC genes in pancreatic cancer cell line Puta8988 treated with pemetrexed. *World Chinese J Dig* **14.**:

**Publications in Refereed Journal:**

a. Published: Saikat Bhattacharya and Sanjay Gupta. A Simple Method to Produce Sub-Nucleosome Complexes of High Purity In Vitro, Advances in Biosciences and Biotechnology, 2016, March, 2016, Vol. 7, No. 3.

b. Accepted:

c. Communicated:

- Protein Degradation is One of the Key Determinant for Accumulation of Recombinant Protein in Bacterial Host.
- Histone H2A.1 Promotes Attainment of Distinct Physiological States by Altering Chromatin Dynamics.

- Histone H2A Mono-ubiquitination and Cellular Transformation are Inversely Related in N-nitrosodiethylamine-induced Hepatocellular Carcinoma.

d. Other Publications:

- Monica Tyagi, Shafqat A Khan, Saikat Bhattacharya, Divya Reddy, Ajit K Sharma, Bharat Khade, Sanjay Gupta. Techniques to Access Histone Modifications and Variants in Cancer. 'Cancer Epigenetics: Risk Assessment, Diagnosis, Treatment, and Prognosis', Methods in Molecular Biology Series (Humana Press/Springer) 2015.
- Best poster award for poster entitled "A Comprehensive Study of the Role of H2A Isoforms in Carcinogenesis" at the Carcinogenesis 2015, ACTREC, Navi Mumbai,
- Presented poster entitled "A Structural and Functional Study of the Role of H2A Isoforms H2A.1 and H2A.2 in Cancer" at the Chromatin and Epigenetics conference held at EMBL, Heidelberg, Germany, 2015.
- Presented poster entitled "Histone Isoforms H2A.1 Forms Stabler Nucleosome Compared to H2A.2 and Promotes Proliferation" at the 5th Meeting of the Asian Forum of Chromosome and Chromatin Biology held at JNCASR, Bangalore, India, 2015.
- Presented poster entitled "Role of Histone Isoforms H2A.1 and H2A.2: A Structural and Functional Study" at the 4th Asian Conference on Environmental Mutagens held at IICB, Kolkata, India, 2014.
- Presented a poster entitled "Hereditary Cancer in Post Genomic Era" at the 1st Global Cancer Genomics Consortium (GCGC) Symposium organized at Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre (TMC), Navi Mumbai, India in 2011.
- Presented poster entitled "Effect of H2A Isoforms on Chromatin Structure and Stability" at the Biopaschim conference held at ACTREC, Navi Mumbai, India, 2014.

- Presented a poster entitled "Deciphering the Molecular Role of Histone H2A Variants, H2A.1 and H2A.2" at the Frontiers in Modern Biology conference organized at Indian Institute of Science, Bangalore, India in 2013.

Signature of Student: dBhattacharya

Date: 23/02/16

**Doctoral committee**

| S. No. | Name | Designation | Signature | Date |
|---|---|---|---|---|
| 1. | Dr. S.V. Chiplunkar | Chairman | | 9/3/16 |
| 2. | Dr. Sanjay Gupta | Guide & Convener | | 23/2/16 |
| 3. | Dr. Ashok Varma | Member | | 23/2/16 |
| 4. | Dr. Dibyendu Bhattacharyya | Member | | 23/2/16 |

Forwarded Through:

Dr. S.V. Chiplunkar
Director, ACTREC
Chairperson,
Academic & training Program, ACTREC

Dr. S. V. Chiplunkar
Director
Advanced Centre for Treatment, Research & Education in Cancer (ACTREC)
Tata Memorial Centre
Kharghar, Navi Mumbai 410210.

Prof. K. Sharma
Director, Academics
T.M.C.

Prof. K.S. Sharma
DIRECTOR - ACADEMICS, TMC
Mumbai - 400 012

# *List of Abbreviations Used*

| | |
|---|---|
| **APS** | ***A***mmonium ***P***er***s***ulfate |
| **ATM** | ***A***taxia ***T***elangiectasia ***M***utated |
| **ATR** | ***A***taxia ***T***elangiectasia and ***R***ad3-related |
| **ATRX** | ***A***lpha ***T***halassemia/Mental ***R***etardation Syndrome ***X***-Linked |
| **AUT** | ***A***cetic Acid, ***U***rea, ***T***riton-X |
| **BPB** | ***B***romo***p***henol ***B***lue |
| **CAI** | ***C***odon ***A***daptation ***I***ndex |
| **CD** | ***C***ircular ***D***ichroism |
| **CLL** | ***C***hronic ***L***ymphocytic ***L***eukemia |
| **CRAS** | ***C***oding ***R***egion ***A***ctivator ***S***equence |
| **CTCF** | CC***CTC***-binding ***F***actor |
| **DAXX** | Death-Domain Associated Protein |
| **DDR** | ***D***NA-***D***amage ***R***esponse |
| **DNMT** | ***DN***A ***M***ethyl***t***ransferase |
| **DTT** | ***Dt***hio***t***hreitol |
| **ERE** | ***E***strogen ***R***esponsive ***E***lement |
| **DUSP1** | ***D***ual ***S***pecificity ***P***hosphatase ***1*** |
| **ES Cells** | ***E***mbryonic ***S***tem ***C***ells |
| **FRAP** | ***F***luorescence ***R***ecovery ***a***fter ***P***hotobleaching |
| **GFP** | ***G***reen ***F***luorescent ***P***rotein |
| **GST** | ***G***lutathione-***S***-***T***ransferase |
| **Gdn-Chl** | ***G***uani***di******n***e-***Chl***oride |
| **H2ABbd** | ***H2A B***arr ***B***ody ***D***eficient |
| **HAR Domain** | ***H***istone H2***A*** ***R***epression ***Domain*** |
| **HAT** | ***H***istone ***A***cetyl***t***ransferase |
| **HCC** | ***H***epato***c***ellular ***C***arcinoma |
| **HDAC** | ***H***istone ***Deac***etylase |
| **HDE** | ***H***istone ***D***ownstream ***E***lement |
| **HMG** | ***H***igh ***M***obility ***G***roup |
| **IPTG** | ***I***so***p***ropyl β-D-1-***t***hio***g***alactopyranoside |
| **MALDI** | ***M***atrix-***A***ssisted ***L***aser ***D***esorption/***I***onization |
| **MDS** | ***M***olecular ***D***ynamic ***S***imulation |
| **MNase** | ***M***icrococcal ***N***uclease |
| **MSK1** | ***M***itogen and ***S***tress ***A***ctivated ***K***inase ***1*** |
| **MTT** | 3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide |
| **NAP1** | ***N***ucleosome ***A***ssembly ***P***rotein ***1*** |
| **NCBI** | ***N***ational ***C***entre for ***B***iotechnology ***I***nformation |
| **NCP** | ***N***ucleosome ***C***ore ***P***article |
| **NDEA** | ***N***itroso***die***thyl***a***mine |
| **NOD-SCID** | ***N***on ***O***bese ***D***iabetic-***S***evere ***C***ombined ***I***mmuno ***D***eficiency |
| **NPC** | ***N***uclear ***P***ore ***C***omplex |
| **NTA** | ***N***itrilo***t***riacetic ***A***cid |
| **PBS** | ***P***hosphate ***B***uffered ***S***aline |
| **PCA** | ***P***rincipal ***C***omponent ***A***nalysis |
| **PCNA** | ***P***roliferating ***C***ell ***N***uclear ***A***ntigen |
| **PMSF** | ***P***henyl***m***ethyl***s***ulfonyl ***f***luoride |
| **PP1** | ***P***rotein ***P***hosphatase ***1*** |
| **PRC1** | ***P***olycomb ***R***epressive ***C***omplex ***1*** |
| **Pre-RC** | ***Pre-R***eplication ***C***omplex |

| | |
|---|---|
| **PTM** | ***P**ost **T**ranslational **M**odification* |
| **PVDF** | ***P**oly**v**inyli**d**ene **F**luoride* |
| **RBS** | ***R**ibosome-**B**inding **S**ite* |
| **RMSD** | ***R**oot **M**ean **S**quare **D**eviation* |
| **RT-PCR** | ***R**everse **T**ranscriptase - **P**olymerase **C**hain **R**eaction* |
| **SAM** | ***S**-**A**denosyl **M**ethionine* |
| **SDS** | ***S**odium **D**odeocyl **S**ulfate* |
| **SET** | ***S**u(var)3-9, **E**nhancer of Zeste, **T**rithorax* |
| **SLBP** | ***S**tem-**L**oop **B**inding **P**rotein* |
| **TEMED** | ***Te**tra**me**thylethylene**d**iamine* |
| **TEV** | ***T**obacco **E**tch **V**irus* |
| **TFA** | ***T**ri**f**luoroacetic **A**cid* |
| **TMAO** | ***T**ri**m**ethyl**a**mine **O**xide* |
| **TSA** | ***T**richo**s**tatin **A*** |
| **TSS** | ***T**ranscription **S**tart **S**ite* |
| **VPA** | ***V**al**p**roic **A**cid* |

# *List of Tables*



# *List of Figures*

# I. INTRODUCTION

## Chapter 1

*An overview of epigenetics and review of literature on histones, their post-translational modifications, variants and isoforms*

## 1.1. Overview

Histones are highly basic proteins involved in packaging the chromatin efficiently. This is achieved by formation of nucleosomes that are the basic repeating units of the chromatin and consists of a protein component, comprised of an octameric core of histones and a nucleic acid component, which is ~147bp of DNA[1]. Although, the histone genes are highly conserved, sequence divergence is observed within the genes present in these clusters. These genes, although code for the same core histone, have minor variations in their primary sequence and are sometimes, referred to as histone isoforms[2].

The histone isoforms have been reported to be differentially expressed in various patho-physiological states[2,3,4,5]. The distinct functional effects of replication-dependent histone H2A isoforms have been demonstrated, however, the mechanistic basis of the non-redundancy remains unclear. The work done in this dissertation investigated the specific functional contribution of H2A.1 isoform, which differs from the more prevalent isoform, H2A.2 in identity of only three amino acids. The work shows that H2A.1 exhibits drastically altered expression pattern in different normal tissues and human cancer cell lines (H2A1C in humans) and promotes cell proliferation in a context dependent manner. Further, the work done elucidates that the M51L alteration at the H2A-H2B dimer interface decreases the temperature of melting of H2A.1-H2B as compared to H2A.2-H2B dimer. This difference in dimer stability is also reflected in chromatin dynamics as H2A.1 containing nucleosomes are more stable owing to M51L and K99R substitutions. Interestingly, the same two substitutions were found to have the most prominent effect on cell proliferation, suggesting, that the nucleosome stability is intimately linked with the physiological effects observed. Another interesting observation that emerged from the studies is the marked decrease in H2Aub119 in the hepatocellular carcinoma model system. Though the H2Aub119 occurs on both H2A.1 and H2A.2 isoforms,

considering that this modification shows an inverse correlation with cell proliferation and H2A.1 expression (which promotes proliferation), the role of this modification will be interesting to investigate in future. The work done in the dissertation not only demonstrates the non-redundant functions of H2A isoforms but also shows that even the highly similar H2A isoforms can modulate chromatin dynamics. In addition, on attempting expression of histone in bacterial host for carrying out *in vitro* experiments, it was observed that some histone proteins do not express in bacteria. For obtaining expression of the desired histones, the effect of multiple parameters affecting heterologous expression was investigated. To arrive at a more generalized and unbiased conclusion, few non-histone proteins were also included in the study. This part of the thesis demonstrates that possibly degradation by the host proteases of unstable recombinant proteins results in the lack of recombinant protein accumulation in bacterial host. Further, a simple protocol to purify and reconstitute histone oligomeric complexes is described.

## 1.2. Epigenetics

Epigenetics is the study of the change in gene expression pattern which does not involve change in the nucleotide sequence of DNA. The term 'epi' means above. Therefore, the literal meaning of 'epigenetics' is 'above genetics'. The term epigenetics was coined by C.H. Waddington in 1942 and was derived from the Greek word "epigenesis" which originally described the influence of genetic processes on development[6].

Importance of epigenetics was realized when the study of gene regulatory mechanisms elucidated that not only the inherent nucleotide sequence governs the transcription of genes, but the chemical modification of nucleotides comprising the DNA molecule, like methylation of cytosine residues, are very important. The covalent addition of the methyl group at the 5-carbon of the cytosine ring results in formation of 5-methylcytosine (5-mC). In human DNA,

5-methylcytosine, which is brought about by the enzymes DNA methyltransferases (DNMT's), is found in approximately 1.5% of genomic DNA[7]. In somatic cells, 5-mC occurs in the context of paired symmetrical methylation of a CpG site, in which a cytosine nucleotide is located next to a guanidine nucleotide. However, in embryonic stem (ES) cells a substantial amount of 5-mC is also observed in non-CpG contexts[8]. DNA methylation, that occurs on the stretch of

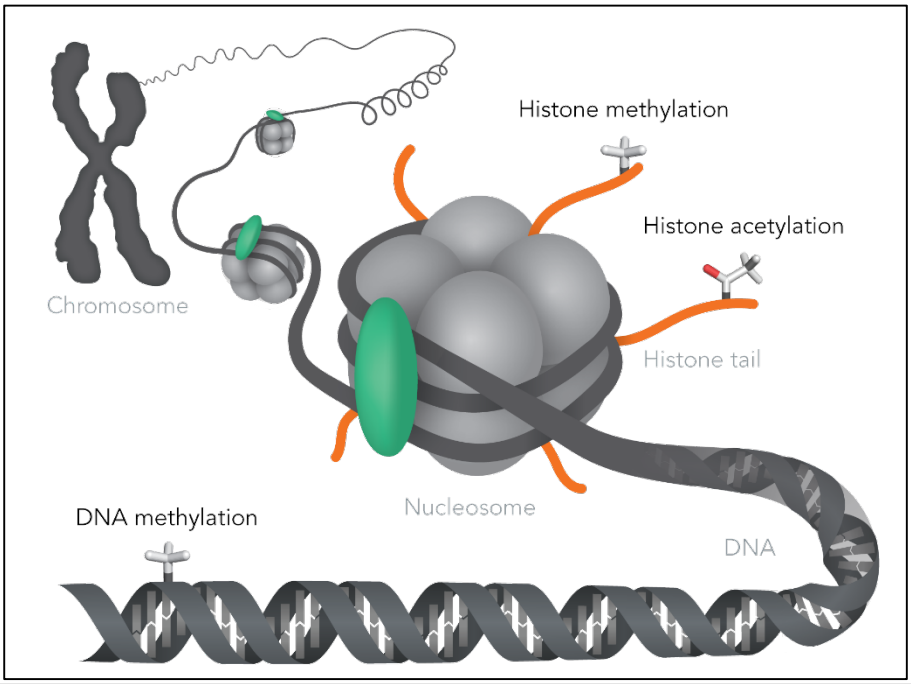



**Figure 1.1. Epigenetic Regulators.** Covalent modifications of DNA (e.g. cytosine methylation) or of histone proteins (e.g. lysine acetylation, lysine and arginine methylation, serine and threonine phosphorylation, and lysine ubiquitination) play central roles in epigenetic regulation.
*Adapted from:*
*https://www.horizondiscovery.com/cell-lines/all-products/explore-by-your-research-area/epigenetics*

CpG dinucleotides, referred to as CpG islands, on the promoter of certain genes, mediates silencing of genes[9]. It is one of the most broadly studied and well-characterized epigenetic modification dating back to studies done by Griffith and Mahler in 1969 which suggested that DNA methylation may be important in long term memory function[10]. DNA methylation was soon studied in context of disease as well and it was unraveled that in many cancers global hypo-methylation and the hyper-methylation of tumor suppressor genes occurs[11]. In fact, cancer was the first human disease to be linked to epigenetics. Studies performed by Feinberg

and Vogelstein in 1983, using primary human tumor tissues, uncovered that genes of colorectal cancer cells were substantially hypomethylated compared with normal tissues[12]. These observations, along with the increasing realization that besides the genome of an individual, environment has an important role in governing phenotype, brought the field of epigenetics into limelight.

For many decades DNA methylation was the focus of investigators to uncover the epigenetic regulatory mechanisms used by cells. However, it was suspected that the condensation of DNA by histones may result in repression of processes such as DNA replication, transcription and repair by limiting access of the underlying DNA sequence to the cellular apparatus. This idea was proved correct by *in vitro* experiments demonstrating that transcription was inhibited with nucleosomal DNA templates but not naked DNA. This, along with the identification of various post translational modifications (PTMs) that histones undergo, ignited the investigation of histones as influential players in contributing to epigenetic regulation of chromatin mediated processes. Continued research has advanced our knowledge of the role of DNA methylation and histone PTMs in governing cellular phenotypes in different contexts [**Figure 1.1**]. Additionally, it also has led to the identification of new epigenetic regulators like the sequence divergent form of the canonical histones, referred to as histone variants and also microRNAs.

## 1.3. Chromatin

The DNA of eukaryotic cells is tightly bound to small basic proteins histones that package the DNA in an orderly fashion in the nucleus [**Figure 1.2**]. The complex between eukaryotic DNA and proteins is known as chromatin, which typically contains about twice as much protein as DNA. Chromatin, in addition to histones and DNA, contains an approximately equal mass of a variety of non-histone proteins like High Mobility Group (HMG) proteins. There are more

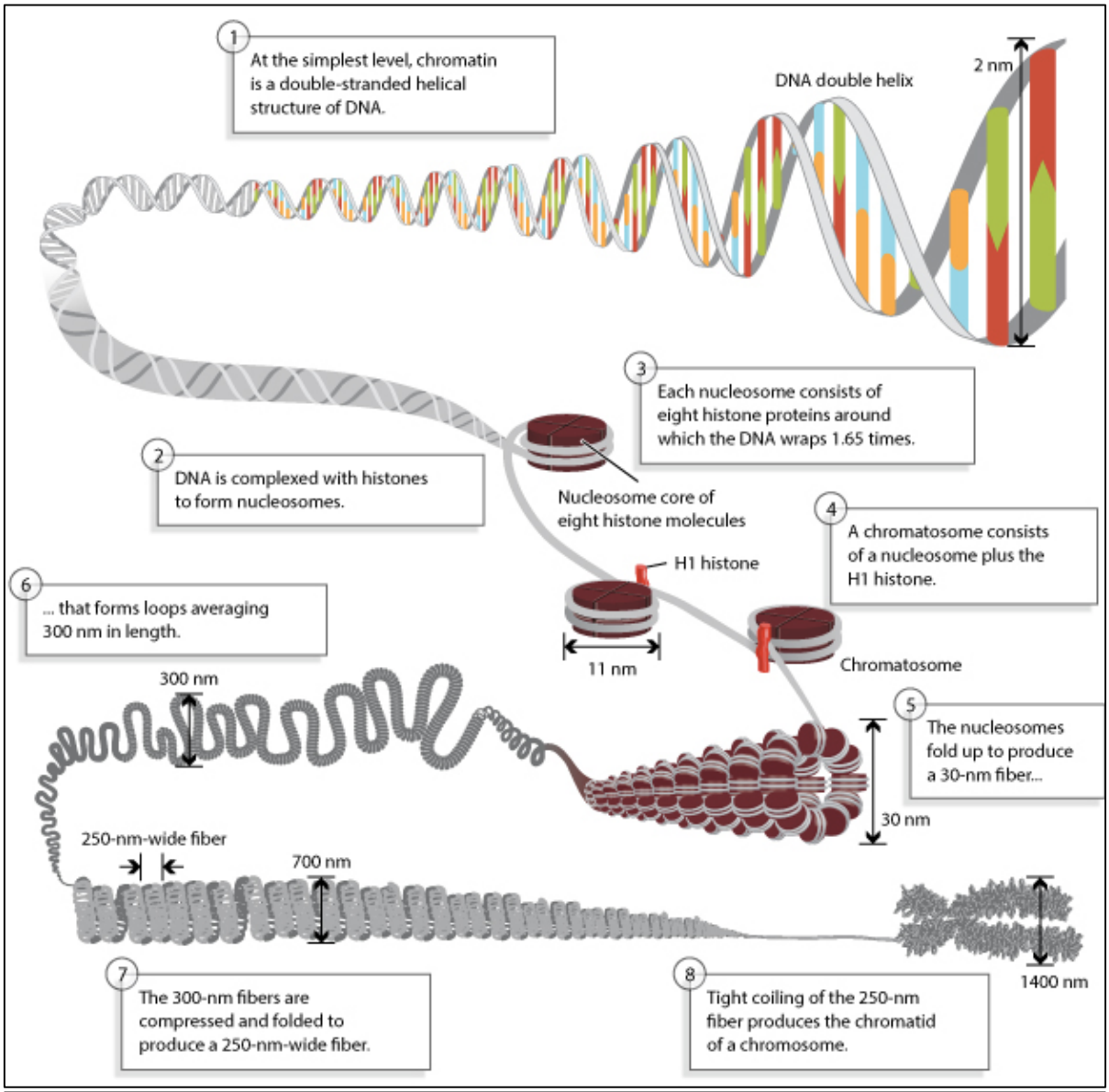



**Figure 1.2. Hierarchical organisation of DNA inside nucleus.** Chromosomal DNA that is more than 6 feet in length is packaged inside a nucleus which is only a few microns in diameter. This is brought about with the aid of highly basic proteins called histones. With the aid of histones, the naked DNA is organized in form of nucleosome which is the fundamental repeating unit of the chromatin. Nucleosomes further fold up to form a 30-nanometer chromatin fibre, which forms loops averaging 300 nanometres in length. The 300 nm fibres are compressed and folded to produce a 250 nm-wide fibre, which is tightly coiled into the chromatid of a chromosome.
*Adapted from: Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed.*

than a thousand different types of proteins, which are involved in a range of activities involving chromatin mediated processes[13].

Partial digestion of chromatin with the enzyme micrococcal nuclease was found to yield DNA fragments approximately 200 base pairs long in contrast to a continuous smear obtained with

naked DNA. Further, electron microscopy revealed that chromatin fibers have a "beads-on-a-string" appearance, with the beads spaced at intervals of approximately 200 base pairs. Both these observations led Roger Kornberg to propose the nucleosomal model of chromatin[14]. The formation of nucleosomes yields a chromatin fiber approximately 10nm in diameter that is composed of "chromatosomes" separated by linker DNA segments averaging about 80 base pairs in length. The chromatin can then be further condensed by coiling into 30-nm fibers[15], the structure of which still remains to be determined. Two models have been proposed for the structure of the 30nm chromatin fiber- the solenoid model and the zigzag model, although, the recent findings favors the existence of zigzag model[16,17].

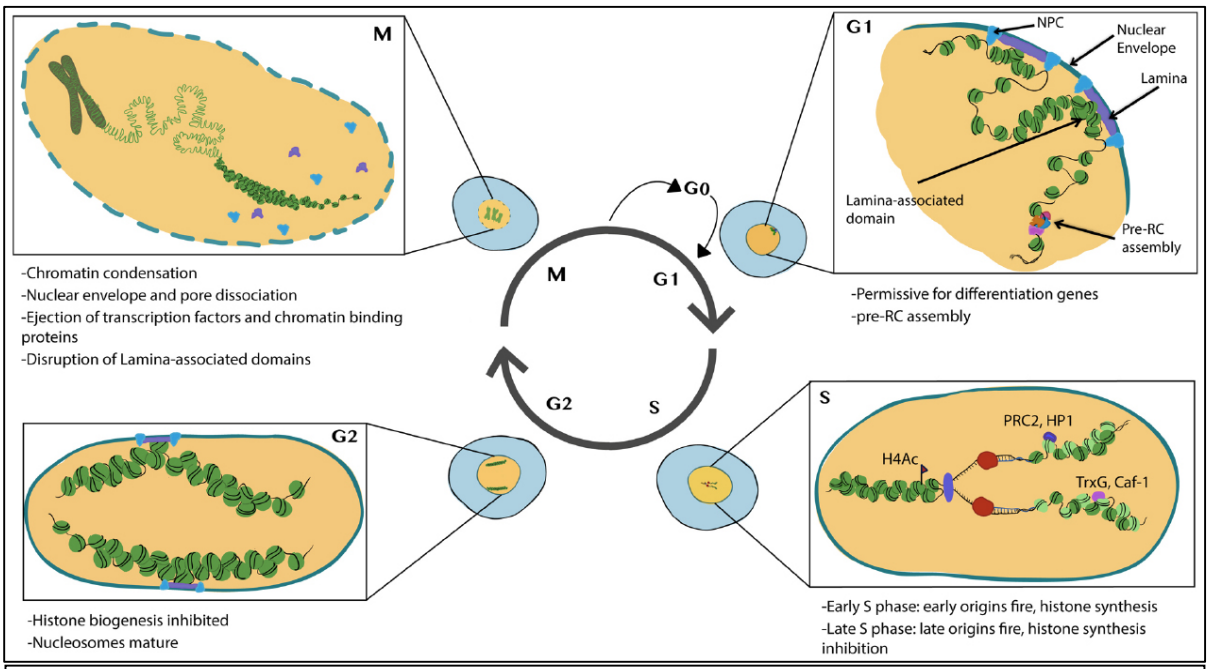



**Figure 1.3. Changes in chromatin during different phases of the cell cycle.** Cells in G1 phase exhibit subnuclear domains with some regions associated with nuclear pore complexes (NPCs) and nuclear lamina. Pre-RCs (Pre- Replication Complexes) preferentially form at accessible chromatin. During S-phase histones are transcribed and synthesized, DNA is replicated and new (light green) and recycled (dark green) nucleosomes assemble to form nascent chromatin. PTM writers and readers also associate with nascent chromatin. During G2 phase, nucleosomes mature and histone biogenesis is inhibited. During mitosis, chromosomes condense and many transcription factors and chromatin binding proteins are ejected from the chromatin. The nuclear envelope breaks down disrupting nuclear lamina associated domains.
*Adapted from: Ma, Yiqin, Kiriaki Kanakousaki, and Laura Buttitta. "How the cell cycle impacts chromatin architecture and influences cell fate." Frontiers in genetics 6 (2015).* Illustration by Nicole Ethen.

The extent of global chromatin condensation varies during the life cycle of a cell [**Figure 1.3**].

In interphase cells, most of the chromatin is relatively decondensed (called euchromatin) and distributed throughout the nucleus[18]. This is consistent with the fact that during this period of the cell cycle, genes are transcribed and the DNA is replicated in preparation for cell division. Most of the euchromatin in interphase nuclei appears to be in the form of 30-nm fibers, organized into large loops containing approximately 50 to 100 kb of DNA. About 10% of interphase chromatin is in a very highly condensed state (called heterochromatin) that is transcriptionally inactive and contains highly repeated DNA sequences[19]. During the G2/M phase of the cell cycle the chromatin condenses in the form of metaphase chromosomes to facilitate their proper distribution to daughter cells.

## 1.4. Histones

The major protein component of chromatin are the histones- small proteins containing a high proportion of basic amino acids, arginine and lysine. The basic residues impart a positive charge to the histone molecules that facilitate their binding to the negatively charged DNA. There are four major types of core histones—H2A, H2B, H3, and H4—and the linker histone H1. During the formation of nucleosome [**Figure 1.4**], H2A and H2B together form a hetero-dimer. One molecule each of H3 and H4 first forms a dimer. Two such dimers then form tetramers with their tetramerisation domains[1]. One H3-H4 tetramer is deposited onto the naked DNA by the highly specialized class of proteins known as histone chaperones[20]. This leads to the bending of DNA around the partial left-handed super-helical ramp formed by the H3-H4 tetramer. This process is known as nucleosomal positioning and it marks the site where the formation of nucleosome has to take place[21]. Subsequently, two molecules of H2A-H2B dimers are deposited onto the tetramer to complete the octameric core around which approximately 147bp of DNA is wrapped in 1.65 left-handed super-helical turn to complete the nucleosome.

The linker histone H1 binds to the nucleosome at the entry and exit sites of the DNA, thus locking the DNA into place and allowing the formation of higher order structure.

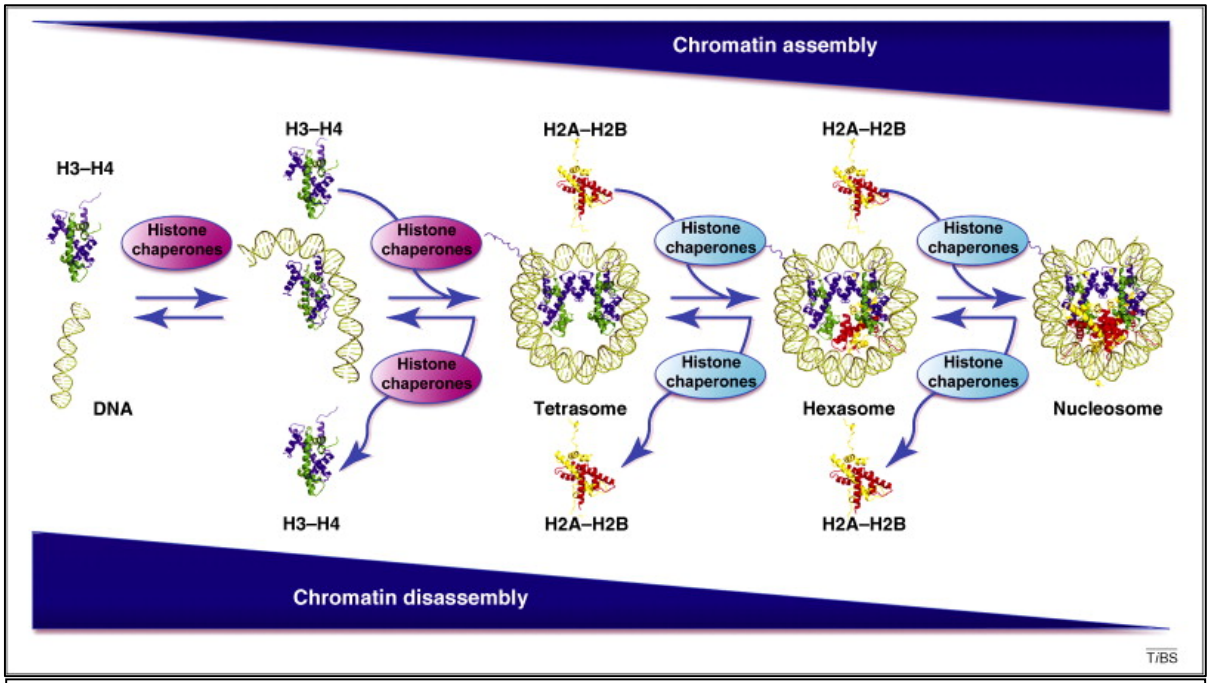



**Figure 1.4. Assembly and disassembly of nucleosome.** Stepwise assembly and disassembly of nucleosomes mediated by histone chaperones. DNA is wrapped around two H3–H4 dimers and two H2A–H2B dimers to form the nucleosome core particle. The stepwise assembly, along with the different possible intermediates including the tetrasome and hexasome, are represented here. Histone chaperones mediate each step of the assembly/disassembly process. Histone H2A is depicted in yellow, H2B in red, H3 in blue, and H4 in green.
*Adapted from:*
*Das, C., Tyler, J. K., & Churchill, M. E. A. (2016). The histone shuffle: histone chaperones in an energetic dance. Trends in Biochemical Sciences, 35(9), 476–489. http://doi.org/10.1016/j.tibs.2010.04.001*

Consistent with their role in DNA packaging, the synthesis of histones is concomitant with that of DNA synthesis. In accordance with this, there is a marked up-regulation in synthesis of histones when the cells enter the S-phase of cell cycle[22]. Further, for quick synthesis of large amount of histones that are needed to package the newly synthesized DNA, the histone genes are iterated and are present in clusters in the genome[23].

The histones have been generally conserved in evolution[24]. H2A and H2B are not nearly as conservative in sequence as H3 and H4, and H1 is quite a variable molecule as compared to many other proteins[25]. The evolutionary changes are unevenly distributed within the sequence

of each histone. In H2A, H2B and H1, the "core" is highly conserved, however, lysine-arginine-rich terminal domains show the evolutionary divergence. Histone H3 and H4 divergence show a much more uniform distribution of amino acid substitutions[25]. Further, wide variations in N and C domains are seen in H2A and H2B. High conservation of these regions in H3 and H4 suggests that they have a much more defined role in these proteins.

| Histone Type | Molecular Weight | Number of Amino Acids | Content of Basic Amino Acids |
|---|---|---|---|
| **H1** | 17000 - 28000 | 200 - 265 | 27% Lys, 2% Arg |
| **H2A** | 13900 | 129 - 155 | 11% Lys, 9% Arg |
| **H2B** | 13800 | 121 - 148 | 16% Lys, 6% Arg |
| **H3** | 15300 | 135 | 10% Lys, 15% Arg |
| **H4** | 11300 | 102 | 11% Lys, 4% Arg |

**Table 1.1. Composition of histones.** Approximate molecular weight, number of residues and content of basic amino acids of the linker histone H1 and the core histones H2A, H2B, H3 and H4.

As previously discussed, all histones contain relatively large amounts of lysine and arginine [**Table 1.1**]. The lysine/arginine ratio varies considerably, from about 20 in the "lysine-rich" histones such as H1 to less than 1 in the "arginine-rich" H3 and H4[25]. The positive charge is distributed quite unevenly in the histone sequences[25]. The N-terminal region contains a high concentration of basic residues. Histone H1 has a very long positively charged C-terminal domain in addition to a lysine-rich N-terminal region[25].

## 1.5. Structure of Nucleosome and its Subcomplexes

### 1.5.1. Electron Microscope Structure of Nucleosome

The octameric histone core of the nucleosome can exist as an intact entity even in the absence of DNA or cross-linking agents at high ionic strengths[26]. This allowed the crystallization of histone complex. The crystal structure of the histone octamer at 3.1-A° resolution was a landmark in the field of chromatin structure as it revealed for the first time the tertiary structure of all the core histones. This elucidated that the four core histones share a common structural motif, termed the histone fold[27]. Each core histone has a structural motif of an N-terminal tail,

three α-helices connected by two loops (α1-L1-α2-L2-α3), and a C-terminal tail[28]. The histone fold consists of a long central helix flanked on either side by a loop and a short helix. The fold has an approximate two-fold axis of symmetry. The central helices of H2A and H4 appear to be slightly bent[1]. Apart from the central histone-fold region, histones H2A and H3 contain additional helices at the N- terminus, whereas histone H2B contains an additional helix at the C-terminus. The N-terminal tails of histones are not visible in the octamer structure, suggesting that they are disordered in the absence of DNA.

## 1.5.2. Dimer and Tetramer of Histones

Each of the core histones exists as part of a dimer, in which the two monomers are closely associated in a head-to-tail manner[29]. The basis for the formation of dimer and tetramer is burial

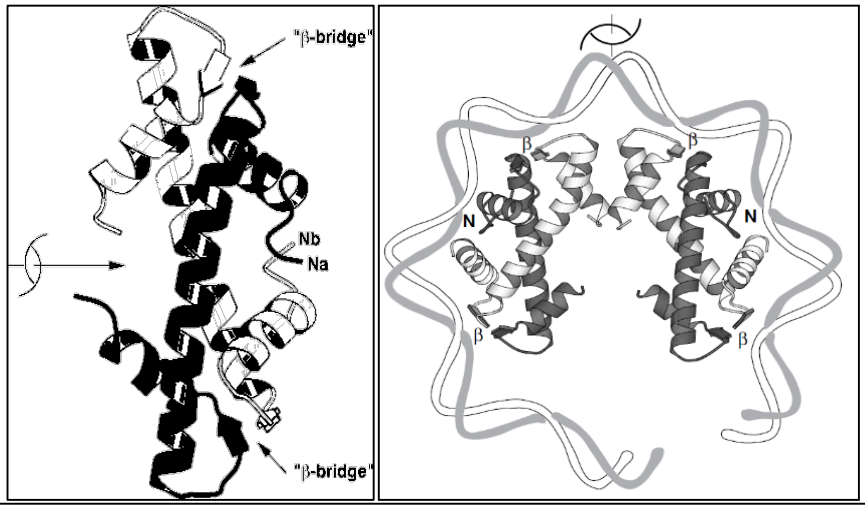



**Figure 1.5. Structure of histone dimer and tetramer.** The dimer *(left)* has a twofold axis of symmetry. The strand from the N-terminal HSH motif of one member pairs with the C-terminal strand of its partner to form a β-bridge, and the ends of the N-terminal helices are also close together to form another paired element. Tetramerisation *(right)* occurs via interactions between the C-terminal halves of two H3. A schematic path of the DNA illustrates how the repeated paired elements of β-bridges and N-terminal ends form a superhelical ramp that interacts with minor grooves of the DNA.
*Adapted from: Ramakrishnan, V. (1997) Histone structure and the organization of the nucleosome. Annu. Rev. Biophys. Biomol. Struct., 26, 83–112.*

of the hydrophobic residues from the solvent which increases stability. The histone-fold portion of each interaction span the entire length of the histone-fold regions of each monomer. The long central helices associate in an antiparallel manner [**Figure 1.5**]. The strand in the HSH

motif in the N-terminal half of a monomer pairs with the strand in the HSH motif of the C-terminal half of its partner to form a short two-stranded β-sheet or β-bridge[29]. A similar sheet is found at the other end in the symmetric position. The solvent-exposed surfaces of these sheets consist of conserved basic residues that are likely to be involved in binding DNA[29]. The (H3-H4) dimer further associates to form a tetramer. In the case of the $(H3\text{-}H4)_2$ tetramer, the interaction is between the C- terminal HSH motif of H3 [**Figure 1.5**]. Unlike the case of the (H3-H4) heterodimers, the (H2A-H2B) heterodimer does not show a marked tendency to associate further into a tetramer[26]. Although some of the residues involved in tetramerisation are also found in H2A and H2B, presumably, not enough of the interactions that are found in the H3-H4 tetramer interface can be formed by either H2A or H2B to form a stable tetramer[29].

### 1.5.3. Octamer of Histones and Nucleosome

An H2A-H2B dimer associates on each side of the H3-H4 tetramer to form the histone octamer. This association occurs primarily via an interaction between the C-terminal halves of H2B and

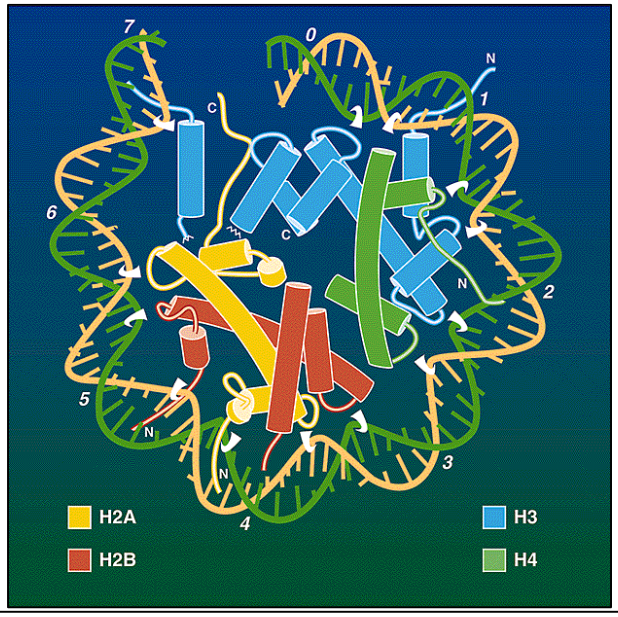



**Figure 1.6. The nucleosome.** One turn (73bp) of the DNA superhelix is depicted, and the view is down the superhelix axis. The core histones are indicated by different colours, and helical regions in the proteins are represented by cylinders. The center of the nucleosome (the dyad position) is indicated by 0 above the DNA superhelix, and it terminates at site 7. Points of protein-DNA contacts are indicated by white hooks. *Adapted from: http://www.nature.com/nature/journal/v389/n6648/full/389231a0.html*

H4. The binding of H2A-H2B dimers to opposite sides of the tetramer results in a tripartite structure for the octamer [**Figure 1.6**]. The association of H2A-H2B to the H3-H4 tetramer is weaker than the association of the two halves of the tetramer, however, the interface between each H2A-H2B dimer and the H3-H4 tetramer is larger than that between the two halves of the tetramer[29]. Electron microscopy had also detected the presence of a left-handed superhelical ramp on the surface of the octamer[30], which presumably is the path of interaction with DNA. The nucleosome core has a dyad symmetry axis that passes between the two superhelical turns of DNA wound around the octamer of core histones[31].

### 1.5.4. H1: The Linker Histone

The linker histone H1 binds to the nucleosome and promotes the organization of nucleosomes into a higher-order structure[32] [**Figure 1.7**]. The linker histones, which do not contain the histone fold motif, are critical to the higher-order compaction of chromatin, because they bind to internucleosomal DNA and facilitate interactions between individual nucleosomes. Nearly all members of the linker histone family consist of a globular domain that is flanked by N- and C-terminal tails that are relatively unstructured in the free histones in solution[29]. The tails are highly basic, and in particular, over half the residues in the C-terminal tail consist of lysines and arginines[29]. The globular domain of H1 is the nucleosome-binding domain, whereas the C-terminal tail binds to the linker DNA between nucleosomes.

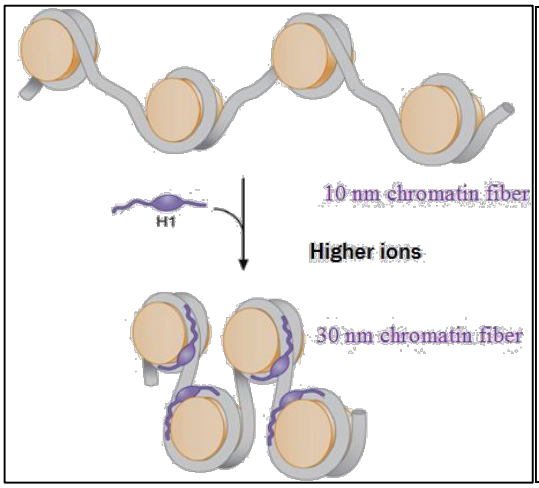



**Figure 1.7. Role of linker histone.** Histone H1, also known as linker histone, has a globular domain and a large C-terminal tail. H1 seals off the nucleosome by sitting at the entry and exit site of DNA on the nucleosome. The long C-terminal tail of H1 is thought to mediate interactions between neighbouring nucleosomes and helps to bring them closer for compaction of chromatin.
*Adapted from:*
*http://biologyforhighschool.net/?p=431*

## 1.6. Histone Post-Translational Modifications

Histones are modified extensively by posttranslational modifications that alter chromatin structure and serve to recruit protein complexes from DNA[33]. Histone tails provide sites for a variety of posttranslational modifications, like acetylation, phosphorylation, methylation, ADP-ribosylation and ubiquitination [**Figure 1.8**]. The modifications play a variety of important role in epigenetic processes bringing about regulation of transcription, mitosis, DNA repair etc.

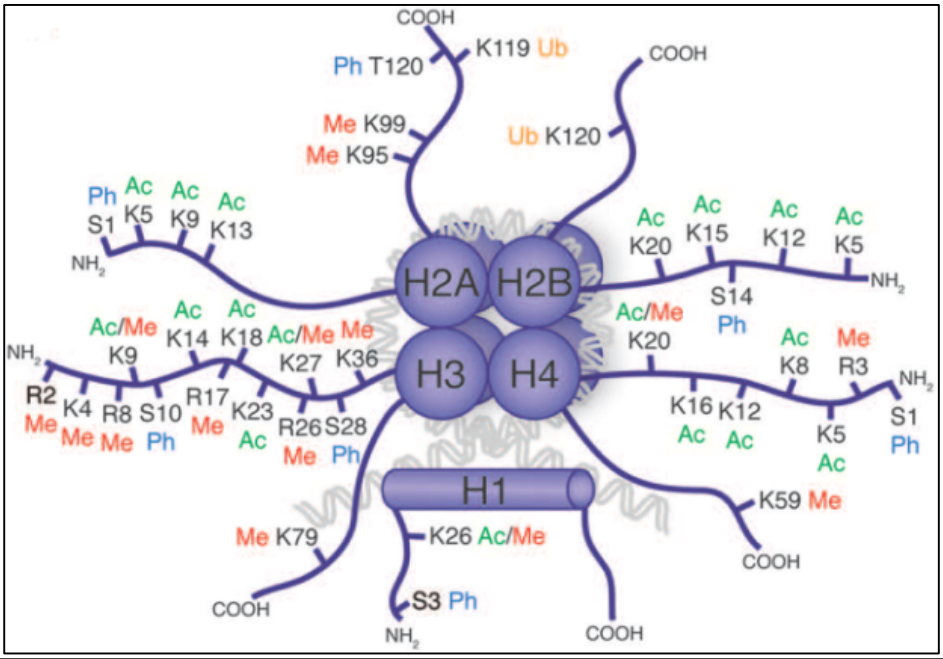



**Figure 1.8. Histone post-translational modifications.** Schematic drawing of a nucleosome with the four canonical histones (H2A, H2B, H3 and H4) and the linker histone H1. The covalent PTMs [methylation (Me), acetylation (Ac), ubiquitination (Ub), and phosphorylation (Ph)] are highlighted on the N- and C-terminal tails of each histone. The modifications which occur in the core body of histones like H3K56Ac are not depicted in the drawing.
*Adapted from: https://www.scienceopen.com/document/vid?id=e595120a-1921-4dfc-96c0-27361577f165*

### 1.6.1. Histone Acetylation

Histone acetylation is a reversible modification that occurs on the ε-amino groups of lysine residues generally at the N-terminal tails of core histones. These reactions are typically catalyzed by enzymes with "histone acetyltransferase" (HAT) or "histone deacetylase"

(HDAC) activity[34] [**Figure 1.9**]. Acetylation removes the positive charge on the histones, thereby decreasing the interaction of the N termini of histones with the negatively charged sugar phosphate backbone of the DNA. This leads to opening up of chromatin and favors transcription of genes. This can be reversed by HDAC activity. Histone acetylation has been linked to cancer and certain types of HDAC inhibitors are already being used to treat certain forms of cancer[35].

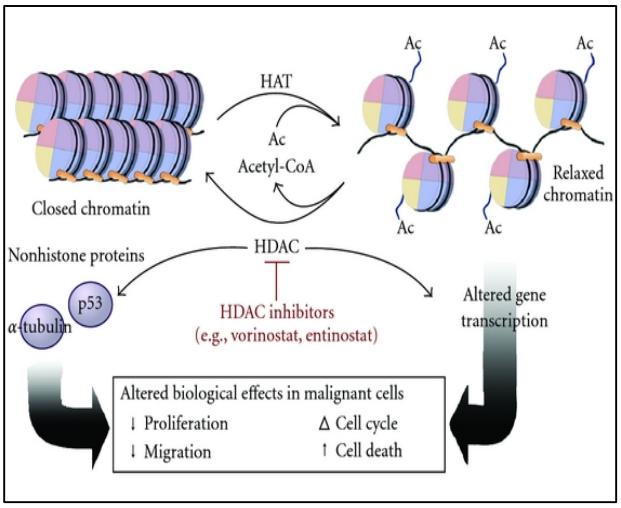



**Figure 1.9. Histone acetylation.** Shown in this illustration, the dynamic state of histone acetylation/deacetylation regulated by HAT and HDAC enzymes. Acetylation of histones increases accessibility of chromatin and gene transcription. Treatment with HDACi alters gene transcription leading to altered downstream biological effects. *Adapted from: https://www.scienceopen.com/document/vid?id=e595120a-1921-4dfc-96c0-27361577f165*

## 1.6.2. Histone Methylation

Histone methylation is a process by which methyl groups are transferred to lysine or arginine residues of histone proteins[36]. Methylation of histones can either increase or decrease transcription of genes, depending on which amino acids in the histones are methylated, and how many methyl groups are attached. For example, trimethylation of histone H3 at lysine 4 (H3K4me3) is an active mark for transcription[37] whereas, trimethylation of histone H3 at lysine 9 (H3K9me3) is a signal for transcriptional silencing[38][**Figure 1.10**].

Histone methyltransferases (HMT) catalyze the transfer of one, two, or three methyl groups to lysine and arginine residues of histone proteins. Two major types of histone methyltransferases exist, lysine-specific (which can be SET (Su(var)3-9, Enhancer of Zeste, Trithorax) domain

containing or non-SET domain containing) and arginine-specific[39,40]. In both types of histone methyltransferases, S-Adenosyl methionine (SAM) serves as a cofactor and methyl donor group[36].

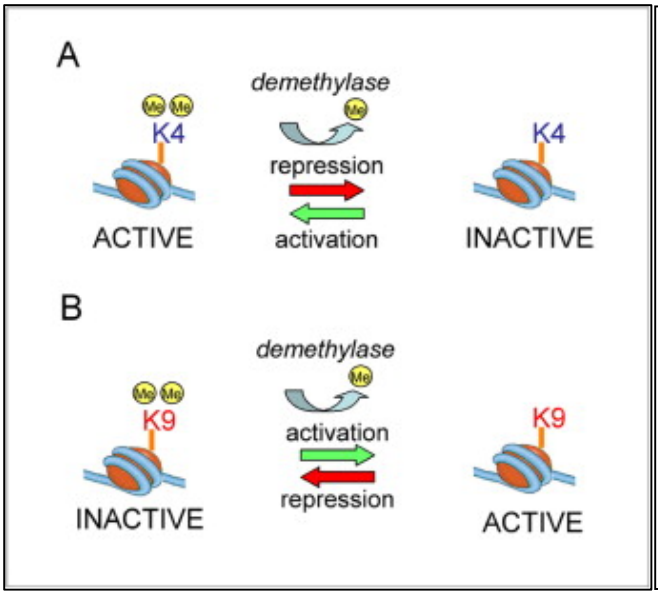



**Figure 1.10. Histone methylation.** The methylation of histone lysines is reversible and regulates gene expression. (a) Methylation of H3K4 is often associated with active genes, and conversely, its demethylation accompanies gene repression. (b) In contrast, methylation of H3K9 is often associated with silenced genes, hence removal of H3K9 methyl marks coincides with gene activation.
*Adapted from:*
*Wysocka J, Milne TA, Allis CD. Taking LSD1 to a new high. Cell. 2005;122(5):654–658.*

## 1.6.3. Histone Phosphorylation

Histone phosphorylation can occur on serine, threonine and tyrosine residues [**Figure 1.11**]. A number of proteins containing phospho-binding modules such as 14-3-3 and BRCT domains that can recognize phosphorylated histones have been identified and characterized as downstream effectors[41,33]. A wide array of kinases (like Aurora B, MSK1, ATM etc.) and phosphatases (like DUSP1, PP1 etc.) participate in modulating the levels of phosphorylation mark on histones in different cell cycle stages and contexts[42]. Characterization of the biological functions of the histone phosphorylation marks is an area of intensive investigation.

One well studied phosphorylation mark is the phosphorylation of H2A variant H2AX. This modification demarcates large chromatin domains around the site of DNA breakage is one the earliest events in the DNA damage response pathway[43]. This modification takes place on serine 139 in mammals and serine 129 in yeasts of the variant histone H2AX, and is commonly

referred to as γH2AX. Protein kinases ATM and ATR in mammals carry out this phosphorylation[44]. This wide distribution of γH2AX around the break is thought to create a specific signaling platform for recruitment and/or retention of DNA damage repair and signaling factors[45].

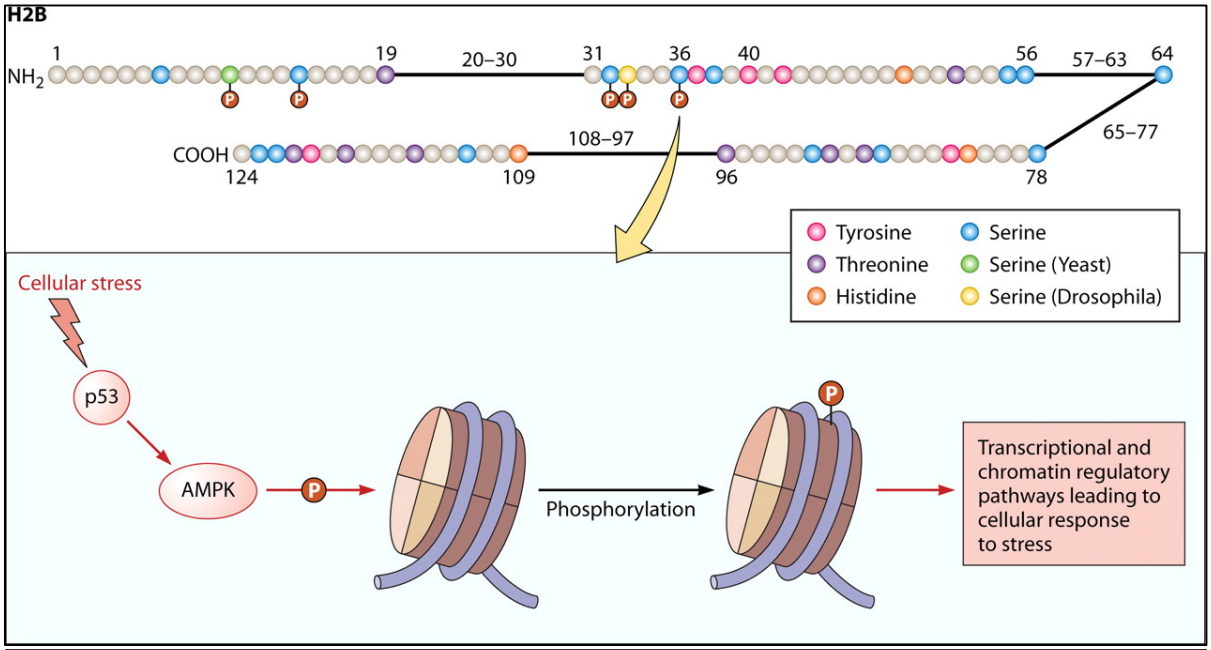



**Figure 1.11. Histone phosphorylation.** Cartoon of the human H2B histone showing all the serine, threonine, tyrosine, and histidine residues. Residues that have been reported to be phosphorylated are shown with a red "P." Yeast has a serine at residue 10 (green), and Drosophila has a serine at residue 33 (yellow), both of which are also phosphorylated. (Bottom) Schematic of the events occurring at S36.
*Adapted from: Banerjee T, Chakravarti D. A peek into the complex realm of histone phosphorylation. Molecular and cellular biology. 2011;31(24):4858–4873.*

Apart from the role of phosphorylated H2AX, phosphorylated histone residues are associated with gene expression. For example, H3S10p and H2BS32p have been linked to the expression of proto-oncogenes such as c-fos, c-jun and c-myc[46,47,48]. Furthermore, targeting H3S28 phosphorylation to promoters of genes such as c-fos and α-globin was shown to control their activation[49]. Phosphorylation of H3 on T11 and T6 has also been implicated in transcription regulation in response to androgen stimulation[50]. Under metabolic stress, an AMP-activated kinase phosphorylates H2BK36 to regulate transcriptional elongation [**Figure 1.11**].

## 1.7. Histone Variants

Apart from the classical arrangement of histone genes in clusters, which give rise to the canonical histones, there are other coding genes for histones as well, which are not present in cluster or in multiple copies [**Figure 1.12**]. The histone variants are generally classified in two ways. They are grouped as replication-dependent, if they are expressed only during the S-phase or are categorized as replication-independent, if they are expressed throughout the cell cycle.

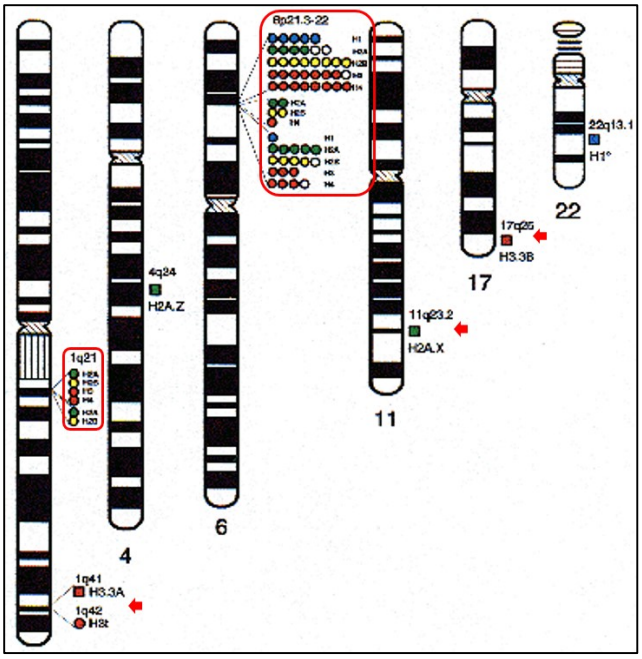



**Figure 1.12. Histone genes in humans.** The ideogram depicts the location of the histone genes in human chromosomes. The clusters indicated in red box shows the location of replication-dependent canonical histone genes. The two major clusters are present on chromosome 1 and 6, with the one on chromosome 6 being the largest one. The genes indicated with red arrows marks the gene for replication-independent histone variants like H3.3, H2A.Z, H2A.X etc. Unlike the genes for canonical histones, genes for variants are not present in multiple copies.

For example, the histone variants H2A.Z and H3.3 are replication-independent histone variants. In addition, there are variants like H2A.X which exhibit both replication-dependent as well as independent expression. Another way of classifying histones is on the basis of sequence heterogeneity. In this way of categorization, histones are classified as homomorphous, if the sequence divergence is of few amino acids. For example, the histone variant H3.3 is a homomorphous variant differing from the canonical histone H3.1 in only 5 amino acids. If the sequence divergence is more, then the variant is categorized as heteromorphous, for example, the variant H2A.Z falls into this category as it has an extended C-terminal tail. Thus, two variants may fall under the same or different categories based on the mode of classification.

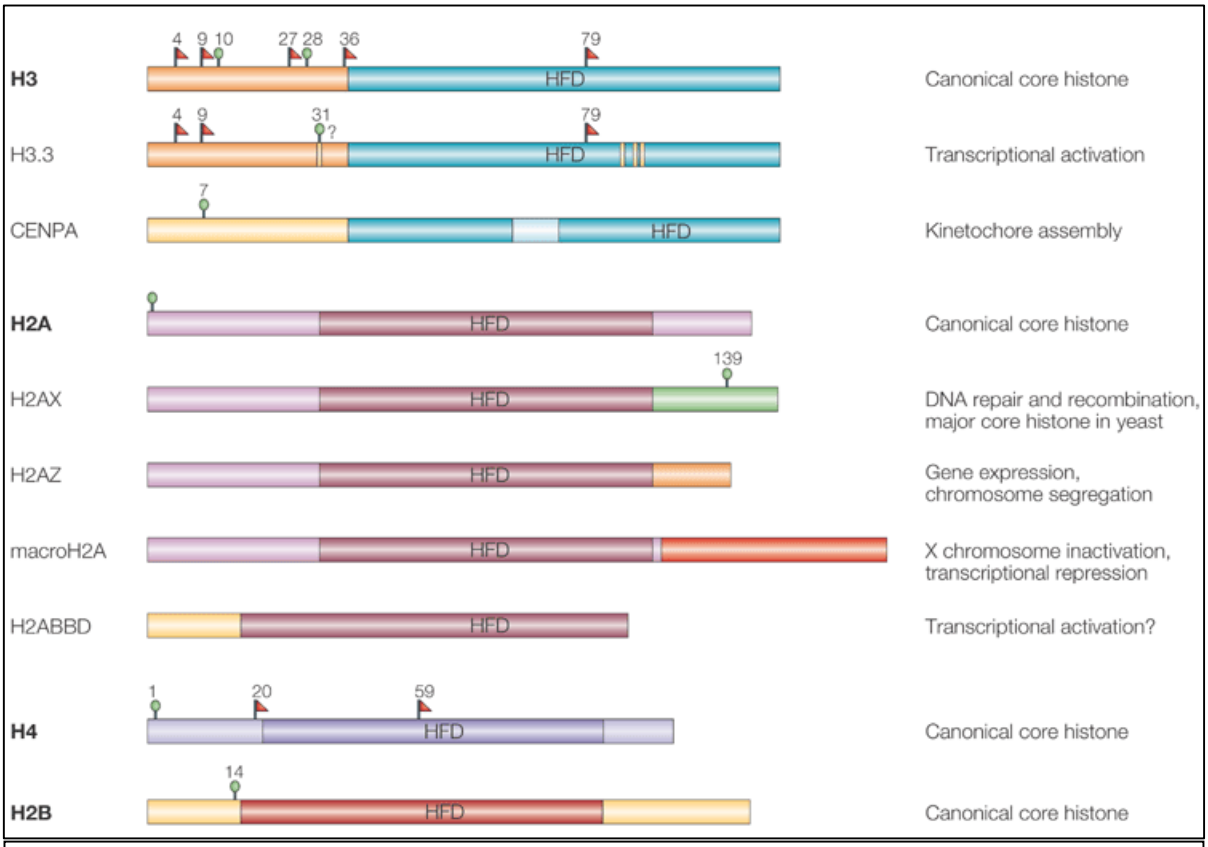



**Figure 1.13. Histone variants and their functions.** The major core histones contain a conserved histone-fold domain (HFD). In addition, they harbour sites for various post-translational modifications. Only well-established sites for lysine methylation (red flags) and serine phosphorylation (green circles) are shown. In the histone H3.3 variant, the residues that differ from the major histone H3 (also known as H3.1) are highlighted in yellow. This N-terminal residue (Ser31) is a site for phosphorylation on H3.3. The centromeric histone CENPA has a unique N terminus, which does not resemble other core histones. Two sites of phosphorylation have been identified in this region, of which Ser7 phosphorylation has been shown to be essential for completion of cytokinesis. The region in the globular domain that is required for targeting CENPA to the centromere is highlighted in light blue. Histone H2A variants differ significantly from the major core H2A in their C terminus. The C terminus of H2AX harbours a conserved serine residue (Ser139), the phosphorylation of which is an early event in response to DNA double-strand breaks. A short region in the C terminus of H2AZ is essential for viability in Drosophila melanogaster. MacroH2A has an extended C-terminal macro domain, the function of which is unknown. Finally, the H2ABBD is the smallest of the H2A variants and contains a distinct N terminus, which lacks all of the conserved modification sites that are present in H2A. The C terminus is also truncated and lacks the docking domain that is found in other H2A species. The histones H4 and H2B are also shown, including their known methylation and phosphorylation sites. The proposed functions of the variants are listed. *Adapted from: Sarma, K., Reinberg, D., 2005. Histone variants meet their match. Nat Rev Mol Cell Biol 6, 139–149.*

The fact that certain histone proteins are synthesized outside of S-phase, suggested to the researchers that they might have some additional role to play in chromatin other than simply packaging the genetic material. These histones were later discovered to carry out specific

functions of the chromatin metabolism [**Figure 1.13**]. Not only do the histone variants perform specialized functions but their localization and recruitment onto the chromatin is very specifically regulated by specialized proteins called histone chaperones [**Figure 1.14**].

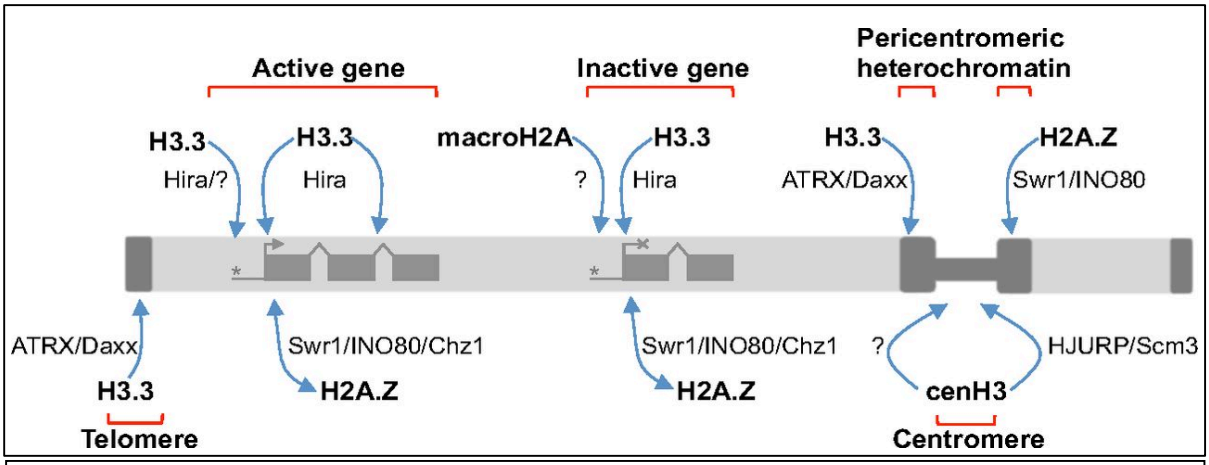



**Figure 1.14. Genomic locations at which histone variants are enriched.** Some histone variants are preferentially deposited at certain regions of the genome, including telomeres, centromeres and pericentromeric heterochromatin, and at both active and inactive genes; deposition at gene enhancers (asterisks) is also observed. The histone chaperones and chromatin remodelers that are known to be involved in this deposition are indicated. Question marks indicate unknown deposition pathways. This is an overview of multiple species. *Adapted from: Skene, P.J., Henikoff, S., 2013. Histone variants in pluripotency and disease. Development 140, 2513–24. doi:10.1242/dev.091439*

Some prominent histone variants are discussed below.

## 1.7.1. H3.3

H3.3 (136 amino acids) has only four amino acid differences with H3.2 (at positions 31, 87, 89 and 90) and five with H3.1 (with an additional difference at amino acid 96). These residues have been proposed to account for particular properties of histone H3.3. The residues 87, 89 and 90 are thought to participate in the regulation of histone-histone interaction stability[51]. In support of this, H3.3 containing nucleosomes are highly sensitive to salt-dependent disruption[52].

Apart from the fact that serine 31, which is specifically found in H3.3, can be phosphorylated[53], H3.3 relate to the increased proportion of PTMs associated with active chromatin such as

acetylation and H3K4 methylation[54,55,56]. Consistent with this, chromatin immunoprecipitation has revealed specific enrichment of H3.3 throughout the gene body of transcribed genes as well as at the promoter regions[57,58]. Notably, H3.3 enrichment at promoters has been observed not only at active but also at inactive genes, possibly accounting for a poised state of these genes[59]. As might be expected H3.3 has different set of proteins which deposit it in different chromatin territories. The H3.3 chaperone HIRA is responsible for the deposition of H3.3 at active chromatin regions[60,61] whereas the DAXX/ATRX complex deposits H3.3 at pericentric heterochromatin and telomeric regions[62,63].

### 1.7.2. H2A.Z

H2A.Z is essential for early embryonic development and is the only histone variant that has been shown to be indispensable for survival[64]. It's a 128 amino acid containing protein and its homology with canonical H2A is less than 60%, however, it is highly conserved from yeast to human[65,66]. H2A.Z has been implicated to participate in many diverse biological processes such as transcription regulation[67], DNA repair[68], heterochromatin formation[69] and chromosome segregation[70]. H2A.Z preferentially localizes at the transcription start site (TSS) where it frequently flanks nucleosome-deficient regions[71,72,67]. Notably, the TSS of an active gene comprises a labile histone H3.3–H2A.Z double-variant nucleosome[73] which could only be purified under conditions of very low ionic strength. These double-variant nucleosomes also marks CTCF insulator binding factor sites[73]. The presence of H2A.Z at the −1 and +1 nucleosomes adjacent to the transcription start site (TSS) has been linked to dynamic changes in gene expression[74,75]. H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin[76]. In H2A.Z$^{-/-}$ cells, Sir2 and Sir3 spread into flanking euchromatic regions, producing changes in histone H4 acetylation and H3 4-methylation indicative of ectopic heterochromatin formation.

### 1.7.3. macroH2A

Histone variant macroH2A are characterized by their large size (368 amino acids), which is three times that of canonical histone H2A[77]. There are two macroH2A genes in the mouse[78,77], termed macroH2A.1 and macroH2A.2[79]. MacroH2A is unique among variants in that it possesses multiple domains, including the histone domain, a 38 residue linker sequence, and a large "macro-domain". MacroH2A incorporate into chromatin via their histone H2A- like region. This is attached via a linker region to the macro domain that projects out of the nucleosome[80]. MacroH2A is generally associated with heterochromatin and transcriptional repression[81]. Consistent with this, macroH2A exists in large populations in the inactive X chromosome of females but is scarcely found in active genes[82]. On its own, the histone domain is sufficient for reducing transcriptional activity *in vivo* and increasing the stability of the nucleosome complex[83]. MacroH2A has been demonstrated to interfere with transcription factor binding and SWI/SNF nucleosome remodeling[83].

### 1.7.4. H2ABBD

H2ABBD (H2A Barr Body Deficient)(115 amino acids) is a highly divergent H2A variant (shares about 48% sequence identity with the canonical H2A) and found only in mammals[84]. This variant is excluded from the inactive X chromosome (hence the name). H2ABBD lacks both the typical C-terminus of the H2A family and the very last sequence of the docking domain[84]. Because of its unique structure, it has been proposed play important role in transcriptional activation[85]. Consistent with this, it colocalizes with regions having high levels of acetylated H4[84]. Role of H2ABBD in transcription activation is further supported by the fact that H2ABBD destabilizes the nucleosome core particles (NCP) causing a more relaxed NCP conformation[86,87]. H2ABBD not only confers lower stability to nucleosomes[88] but NCP containing this variant organized only 118bp of DNA[89]. Consistent with its association with

transcription activation, mouse H2ABBD homologue H2A.Lap1 is localized at the transcription start sites of active genes[90]. Interestingly, the chromatin remodeling complexes SWI/SNF and ACF are unable to mobilize the variant H2ABBD nucleosome[91]. Other noted roles of H2ABBD include its role in splicing. Ectopically expressed human H2ABBD in HeLa cells was deposited at intron-exon boundaries. This is functionally important as H2ABBD knockdown cells show splicing defects[92].

## 1.8. Nucleosome Structure/Stability on Incorporation of Histone Variants

One of the primary means by which histone variants carry out their specialized functions is by modulating chromatin dynamics that occurs due to the change in the biochemical composition of the nucleosome brought about by incorporation of the variant. Some of the well-studied histone variants in this context are discussed below.

### 1.8.1. CENP-A

CENP-A consists of a highly variable N-terminal sequence (which varies inter-species in both length and composition)[93]. This is followed by a histone-fold domain, which in the case of human CENP-A, has 62% identity to the canonical H3[94]. Mutagenic analysis in yeast has demonstrated that both these domains perform essential roles[95]. *In vitro* reconstitution experiments with purified CENP-A were able to produce NCPs (Nucleosome Core Particles) with DNase I footprints and AFM images almost identical to those of native NCPs[96]. However, using deuterium exchange/mass spectrometry and hydrodynamic analysis, it has been shown that CENP-A-H4 tetramers are more compact and rigid than H3-H4 tetramer[97]. This difference has been attributed to the synergistic action of the first loop (L1) and second α-helix (α-2) of the histone-fold of CENP-A. Interestingly, the same region has been suggested to be

responsible for targeting the CENP-A–H4 complexes to centromeres[97]. In addition, this region of CENP-A been postulated to bind to the narrow groove of DNA[98]. Notably, the N-terminal tail of H3 interacts with the linker DNA and is involved in the maintenance of the higher order structure of the chromatin fibre[99,100]. Hence, it is possible that this region of CENP-A plays a critical role in defining the topological characteristics of centromeric chromatin. This is supported by the fact that centromeric chromatin is more compact than pericentric or bulk chromatin[101].

### 1.8.2. H2A.Z

*In vitro* recombinant H2A.Z failed to form dimer with H2B[102]. Only under the presence of high concentration of protein stabilizing osmolyte TMAO, H2A.Z-H2B dimer could be reconstituted[102]. Furthermore, by equilibrium unfolding experiments H2A.Z-H2B dimer was found to be less stable than H2A-H2B dimer[102]. These differences in dimer stability as compared to canonical H2A-H2B dimer is also reflected in the nucleosome stability, although, contradictory results have been reported. The crystallographic structure of H2A.Z-containing NCPs suggested subtle destabilization of the interaction between the (H2A.Z–H2B) dimer and the (H3–H4) tetramer[103]. In agreement with this, it was shown that NCPs reconstituted with recombinant- H2A.Z–H2B dimers were less stable at higher ionic strength as analyzed by the analytical ultracentrifuge[104]. Corroboratively, H2A.Z-containing NCPs appear to have an enhanced thermal- dependent mobility[105]. In contrast, a more recent characterization of the salt-dependent stability using FRET indicated that H2A.Z stabilizes the histone octamer within the NCP[106]. This is consistent thermodynamically with the *in vitro* stability data of H2A.Z-H2B dimer[102], as an unstable dimer should favor a more stable octamer. It is worth noticing that the splice variants of histone H2A.Z exist. Hence, it would be important to clearly specify the isoform while reporting such studies to avoid ambiguity.

### 1.8.3. macroH2A

The primary sequence of macroH2A is ~65% identical to canonical H2A. Sedimentation analysis using sucrose gradients revealed that macroH2A stabilizes the NCP[107]. The crystal structure of an NCP containing the macro domain shows that variant incorporation causes only minor NCP rearrangements[108]. It mainly differs from H2A in two important regions: the L1 loops and the docking domain. The canonical 38NYAE41 L1 loop possesses a net negative charge, while in contrast the macroH2A L1 38HPKY41 sequence has a net positive charge and an increased hydrophobicity. Notably, substitutions of the L1 loops in canonical H2A with a macroH2A sequence creates nucleosomes with *in vitro* stabilities and *in vivo* enrichments that are nearly identical to NCPs containing the complete macroH2A histone domain[108]. Therefore, the L1 loops appear to be very important for the functional abilities of macroH2A. Changes to the docking domains, although increases *in vivo* enrichment, it has little effect on *in vitro* stability[108].

### 1.8.4. H2ABBD

Incorporation of H2ABBD has a marked effect on nucleosome architecture which is reflected in its ability to organize only 118bp of DNA[87]. Recombinant H2ABBD does not form stable octamers although its equimolar mixture with H3-H4-H2B is able to form NCPs in the presence of DNA[89]. The nucleosomes that are formed exhibit a relaxed conformation compared with the native NCPs under a broad range of salt concentrations as determined by analytical ultracentrifugation and by FRET[89,86]. This is most likely the result of the flanking DNA being released from the NCP constraints[89,86] with the whole particle adopting a more open structure. FRAP experiments showed that H2ABBD exchanges more rapidly than canonical H2A in chromatin[86]. In addition, the high dynamic flexibility of NCPs consisting of H2ABBD is reflected in the ease with which the former can be assembled and disassembled by nucleosome

assembly protein (NAP1)[109]. Notably, the assembly was more efficient when the NCPs contained H3.3 than when the NCPs contained the canonical H3. It would be interesting to investigate the *in vivo* association of H2ABBD and H3.3.

## 1.9. Histone Variants in Cancer

Consistent with their role in governing gene expression and other chromatin mediated processes, some of the histone variants have been found to be aberrantly expressed in cancer [**Table 1.2**]. Emerging evidence points towards the role for histone variants in contributing to tumor progression. This is further emphasized with the recent findings of cancer-associated mutation in H3.3 variant gene. Some of the reports linking histone variants and cancer are discussed below.

### 1.9.1. H3.3

A breakthrough finding which reemphasizes the role of histone variants in development of cancer is the driver mutations in H3.3 identified in pediatric gliobastoma[110]. Somatic mutations in the H3.3-ATRX-DAXX chromatin remodeling pathway were identified in 44% of tumors (21/48). Recurrent mutations in H3F3A, which encodes the replication-independent histone variant H3.3, were observed in 31% of tumors, and led to amino acid substitutions at two critical positions within the histone tail (K27M, G34R/G34V) involved in key regulatory post-translational modifications[110]. Mutations in ATRX and DAXX, encoding two subunits of a chromatin remodeling complex required for H3.3 incorporation at pericentric heterochromatin and telomeres, were identified in 31% of samples overall, and in 100% of tumors harboring a G34R or G34V H3.3 mutation. Furthermore, the presence of H3F3/ATRX -DAXX/TP53 mutations were strongly associated with alternative lengthening of telomeres and specific gene expression profiles[110].

| Variant | Cancer type | Dysfunction | Proposed Mechanism | References |
|---|---|---|---|---|
| H2A.Z (H2A.Z.1) | Colorectal cancer | Overexpression | N.D. | 121 |
| | Undifferentiated cancers | Overexpression | N.D. | 122 |
| | Metastatic breast carcinoma | Overexpression | N.D. | 123 |
| | Primary breast cancer | Overexpression | ER alpha- and Myc-dependent upregulation | 112 |
| | Breast cancer (MCF7 cell line) | Overexpression | Increased proliferation | 113 |
| | Breast cancer | Overexpression | H2A.Z recruited at promoters of ER alpha target genes | 119 |
| | Prostate cancer (LNCaP cell line) | Overexpression | Myc-mediated H2AFZ upregulation | 116 |
| | Prostate cancer (LNCap xenograft) | N.D. | H2A.Zub evicted from PSA promoter/enhancer upon activation | 124 |
| | Prostate cancer (LNCaP cell line) | N.D. | H2A.Zac associated with oncogene activation, unmodified H2A.Z with tumor-suppressor silencing | 115 |
| H2A.X | B-CLL and T-PLL | Translocations and deletions chr 11q23 | Increased genome instability | 125,126 |
| | Head and neck squamous cell carcinoma | Gene deletion | Increased genome instability | 127 |
| | Non-Hodgkin lymphoma | Gene mutation | Increased genome instability | 128 |
| | Gastrointestinal stromal tumor | Upregulation | Promote apoptosis upon the treatment with a kinase inhibitor | 129 |
| | Breast cancer | Gene deletion | Increased genome instability | 130 |
| | Lung cancer | Reduced protein levels/splicing defects | Suppression of cell proliferation via reduced PARP-1 | 131 |
| mH2A.1 | Breast cancer | Reduced protein levels/splicing defects | N.D. | 132 |
| | Melanoma | Transcriptional downregulation | Upregulation of CDK8 | 133 |
| | Testicular, bladder, ovarian, cervical, endometrial cancers | Splicing defects | N.D. | 132 |
| | Colon cancer | Reduced protein levels and splicing defects | N.D. | 132 |
| mH2A.2 | Lung | Reduced protein levels | N.D. | 131 |
| | Melanoma | Transcriptional downregulation | Upregulation of CDK8 | 133 |
| CENP-A | Colorectal cancer | Overexpression | Aneuploidy | 134 |
| | Invasive testicular germ cell tumors | Overexpression | N.D. | 135 |
| | HCC | Overexpression | Deregulation of cell cycle and apoptotic genes | 136 |
| | Breast cancer | Overexpression | N.D. | 137 |
| | Lung adenocarcinoma | Overexpression | N.D. | 138 |

| | | | | |
|---|---|---|---|---|
| H3.3 | Carcinoma of the esophagus | Overexpression | N.D. | 139 |
| | GBM | Mutation (K27M, G34V/R) | N.D. | 140 |

**Table 1.2. Histone variants in cancer.** Cancer associated variants and the proposed mechanism. *Adapted from Angelica, M.D., Fong, Y., 2008. NIH Public Access. October 141, 520–529. doi:10.1016/j.surg.2006*

## 1.9.2. H2A.Z

Overexpression and redistribution of H2A.Z have been reported in cancer[111,112,113]. H2A.Z is upregulated in bladder cancer and stimulates cell proliferation[114]. Further, H2A.Z redistribution correlates with increased expression of proliferation related genes[114]. In prostate cancer cells there is an H2A.Z reorganization that poises the oncogene promoters for activation. Furthermore, the overall levels of H2A.Z decrease at the transcription start sites (TSSs) of such promoters upon activation which is accompanied by a gain of acetylated H2A.Z[115]. In this way, in the presence of androgen H2A.Z operates as a facilitator of transcription that is subject to a rapid dynamic turnover once the gene is undergoing cycles of transcription[116].

High expression of H2A.Z is ubiquitously detected in the progression of breast cancer[113]. Importantly, H2A.Z expression is significantly associated with lymph node metastasis and patient survival. This makes H2A.Z an excellent target for diagnostic and therapeutic interventions. Expression of TFF1 has been used as a marker for breast cancer and ER- positive tumors[117]. Interestingly, H2A.Z is also associated with estrogen responsive element (ERE) of the TFF1 promoter. Upon stimulation by estrogen, H2A.Z disassociates from the ERE regions and deposits at the proximal regions of promoter leading to poising of gene for expression[118]. Similarly, H2A.Z has been shown to incorporate within the promoter of the cathepsin D gene that upregulates its expression upon estrogen stimulation[119]. Further, aberrant expression of H2A.Z facilitates abnormal cellular proliferation through c-Myc, p21/p53, and estrogen target genes expression. Collectively, these findings suggest that H2A.Z overexpression causes

increased breast cancer proliferation. To restore H2A.Z function, two potential drug treatments for cancer therapies has been suggested: targeting c-Myc transcription factor (using inhibitor 10058-F4) and the H2A.Z-remodelling enzyme, p400/Tip60 (using anacardic acid)[120].

### 1.9.3. macroH2A

Evidences suggest that macroH2A acts as a tumor suppressor as it is frequently found to be downregulated in a variety of cancers. First report to suggest involvement of macroH2A in cancer progression came from human breast and lung tumor biopsies where macroH2A1.1 and macroH2A2 expression inversely correlated with proliferation[121]. Subsequently, macroH2A was found to be downregulated in melanoma and acts as a tumor suppressor by binding to CDK8 promoter and suppresses tumor progression of malignant melanoma[122]. Loss of macroH2A1.2 isoforms positively correlated with increasing malignant phenotype of melanoma cells in culture and human tissue samples. Knockdown of macroH2A isoforms in melanoma cells of low malignancy results in significantly increased proliferation and migration.

However, the role of macroH2A1.2 appears to be more complex than generally being a tumor suppressor. As discussed, in favor of a tumor suppressive function, macroH2A1.2 was found to reduce metastatic potential of melanoma cells[122]. In contrast, in breast cancer cells macroH2A1.2 increased migration, invasion and growth in attachment-free conditions[123]. Thus, macroH2A1.2 function and gene targets may have context dependence or cell type specificity.

## 1.10. Histone Isoforms

Apart from the sequence divergence observed between the canonical histones and the variants, sequence divergence also exists amongst the canonical histones arising from cluster [**Table**

**1.3**]. These are often termed as "histone isoforms". There are three clusters of canonical histone gene present in humans at chromosome number 1 (cluster 2 and 3) and 6 (cluster 1). There are three clusters in rat at chromosome 2, 10 and 17. Similarly, mice have three clusters coding for histones. These histone isoform genes are replication dependent.

| Human gene | Accession | Human gene | Accession | Human gene | Accession | Human gene | Accession |
|---|---|---|---|---|---|---|---|
| HIST1H2AA | AY131982 | HIST1H3A | AF531274 | HIST1H1E | AF531302 | HIST1H2BO | AF531298 |
| HIST1H2AB | AY131983 | HIST1H3B | AF531275 | HIST1H1T | AF531301 | HIST2H2BE | AY131979 |
| HIST1H2AC | AY131984 | HIST1H3C | AF531276 | HIST1H2BA | AF531284 | HIST3H2BB | AY131981 |
| HIST1H2AD | AY131985 | HIST1H3D | AF531277 | HIST1H2BB | AF531285 | HIST1H4A | AY128654 |
| HIST1H2AE | AY131986 | HIST1H3E | AF531278 | HIST1H2BC | AF531286 | HIST1H4B | AY128655 |
| HIST1H2AG | AY131987 | HIST1H3F | AF531279 | HIST1H2BD | AF531287 | HIST1H4C | AY128656 |
| HIST1H2AH | AY131988 | HIST1H3G | AF531280 | HIST1H2BE | AF531288 | HIST1H4D | AY128657 |
| HIST1H2AI | AY131989 | HIST1H3H | AF531281 | HIST1H2BF | AF531289 | HIST1H4E | AY128658 |
| HIST1H2AJ | AY131990 | HIST1H3I | AF531282 | HIST1H2BG | AF531290 | HIST1H4F | AY128659 |
| HIST1H2AK | AY131991 | HIST1H3J | AF531283 | HIST1H2BH | AF531291 | HIST1H4G | AY128660 |
| HIST1H2AL | AY131992 | HIST2H3C | AF531307 | HIST1H2BI | AF531292 | HIST1H4H | AY128661 |
| HIST1H2AM | AY131993 | HIST3H3 | AF531308 | HIST1H2BJ | AF531293 | HIST1H4I | AY128662 |
| HIST2H2AA | AY131971 | HIST1H1A | AF531299 | HIST1H2BK | AF531294 | HIST1H4J | AY128663 |
| HIST2H2AB | AY131972 | HIST1H1B | AF531304 | HIST1H2BL | AF531295 | HIST1H4K | AY128664 |
| HIST2H2AC | AY131973 | HIST1H1C | AF531300 | HIST1H2BM | AF531296 | HIST1H4L | AY128665 |
| HIST3H2A | AY131974 | HIST1H1D | AF531303 | HIST1H2BN | AF531297 | HIST2H4 | AF525682 |
| | | | | | | HIST4H4 | AY128653 |

**Table 1.3. Human histone isoforms.** Genes encoding human histone isoforms along with their accession number. *Data taken from: Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J., & Maltais, L. J. (2002). The Human and Mouse Replication-Dependent Histone Genes. Genomics, 80(5), 487–498. http://doi.org/10.1006/geno.2002.6850*

The replication-dependent histone genes encode a unique class of transcripts present only in metazoans. These genes encode the only mRNAs in eukaryotes that lack a poly(A) tail, ending instead in a stem-loop sequence that has been conserved in all metazoans[124]. These genes also lack introns. The 3' end of histone mRNA is formed by endonucleolytic cleavage that is directed by a purine-rich sequence known as the histone downstream element (HDE)[125] [**Figure 1.15**]. Two trans-acting factors have been identified that bind to the cis-acting elements at the 3'end of histone transcripts and are required for histone pre-mRNA processing: the stem-loop binding protein (SLBP), which binds to the stem-loop[126], and the U7 snRNP, which binds to the HDE[127]. Each time a cell divides it must rapidly synthesize large amounts of histones to package the newly synthesized DNA molecule during the brief S-phase. Therefore, the histone mRNAs are highly cell-cycle regulated, with their abundance increasing 35-fold as cells enter S-phase and decreasing again at the end of S-phase[128].

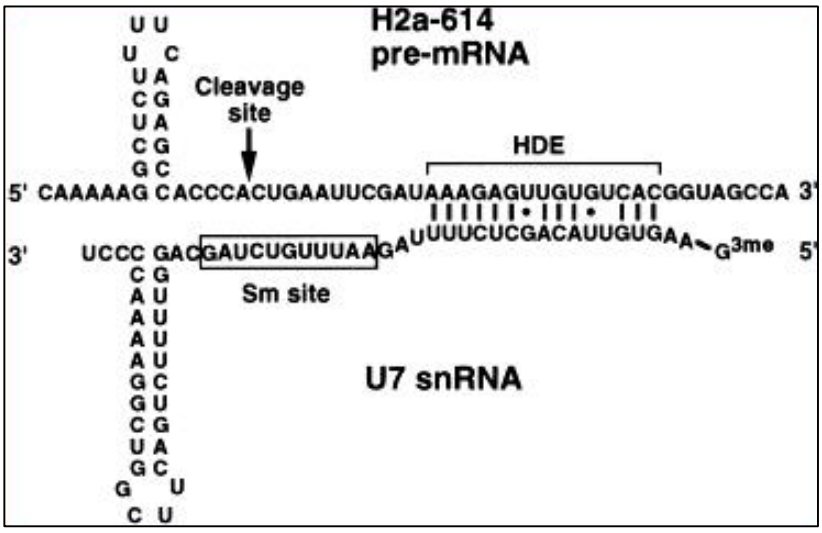



**Figure 1.15. Histone 3' UTR.** Representation of the processing reaction occurring at the 3' end of replication-dependent histone precursor mRNA. The endonucleolytic cleavage occurs between the highly conserved hairpin and the purine-rich sequence HDE. Trans-acting factors are hairpin binding protein (HBP), the U7 snRNP and zinc finger protein 100. Possible alignment of structural elements of the U7 snRNA and the histone pre-mRNA resulting from formation of a duplex between the U7 snRNA 5' end and the histone downstream element (HDE). Sequences of the mouse H2a-614 pre-mRNA near the processing site and mouse U7 snRNA are shown. The possible Watson–Crick base pairs (vertical lines) and GU base pairs (dots) between the 5' end of U7 snRNA and the HDE are indicated.
*Adapted from: Dominski, Z., & Marzluff, W. F. (1999). Formation of the 3' end of histone mRNA. Gene, 239(1), 1–14. http://doi.org/http://dx.doi.org/10.1016/S0378-1119(99)00367-4.*

## 1.11. Differential Expression of H2A Isoforms

There are three clusters in rat at chromosome 2, 10 and 17. Cluster 1 (chromosome 17) has 9 H2A genes, cluster 2 (chromosome 2) has 3 genes and cluster 3 (chromosome 1) has only 1 gene [**Figure 1.16**]. In total there are 13 H2A genes that putatively code for 10 different isoforms (most are with "predicted status" in NCBI database). In humans, there are 16 genes coding for H2A giving rise to 12 isoforms[23]. Histone isoforms, like the variants, have been reported to be differentially expressed in a variety of physiological states. For instance, the proportion of two isoforms H2A.1/H2A.2 has been shown to decrease during development and differentiation of rat brain cortical neurons[129] and during *in vitro* differentiation and aging[130]. H2A.1 and H2A.2 genes are present in histone cluster of chromosome 17 and 2, respectively, in rats. During the course of liver development as well similar changes are observed[5].

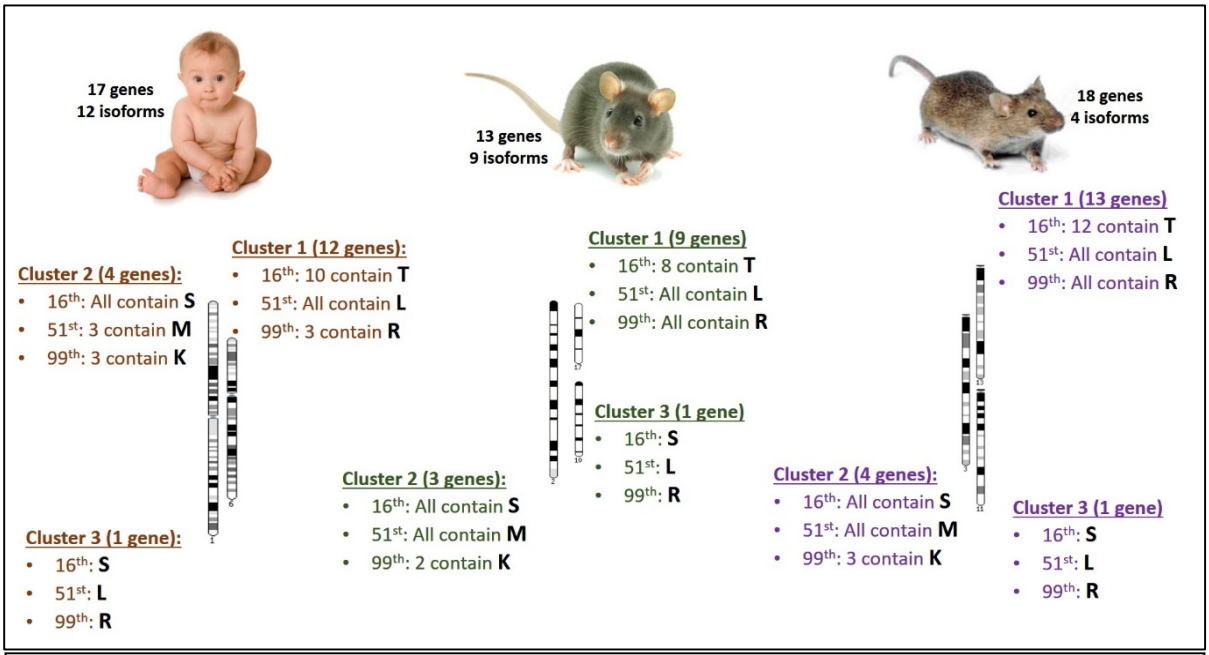



**Figure 1.16. Histone H2A genes in mammals.** Information of the histone H2A coding genes of human, rat and mice are depicted. Humans have 17 H2A genes which code for 12 isoforms, rat 13 genes give rise to 9 isoforms and 18 H2A genes of mouse code for 4 isoforms. Majority of H2A isoforms coded from Cluster 1 of all the three species, like H2A.1 has threonine, leucine and arginine at 16$^{th}$, 51$^{st}$ and 99$^{th}$ positions, respectively. Likewise, majority of H2A isoforms coded from Cluster 2 of all the three species, like H2A.2 has serine, methionine and lysine at 16$^{th}$, 51$^{st}$ and 99$^{th}$ positions, respectively.

In one study, there was a decrease in Hist1H2AC levels on treatment of HCT116 cells with Nutlin-3a, which increases the levels of the p53[131], indicating an inverse correlation of H2A1C isoform with proliferation. Consistent with this the levels of Hist1H2AC have also been shown to decline following a cell cycle arrest[131]. The levels of Hist1H2AC specifically have been reported to change in diseases including human papillomaviruses hyperplasia, AIDS and multiple sclerosis[132,133.]

## 1.12. H2A Isoforms in Cancer

In cancer the expression level of H2A isoforms have been reported to alter. Earlier studies from our lab have shown the differential expression of two major H2A isoforms, H2A.1 and H2A.2, in rat hepatocellular carcinoma model system[3]. The pattern of variants of histones was investigated during N- nitrosodiethylamine (NDEA)-induced hepatocarcinogenesis. The study showed for the first time *in vivo* overexpression of a major histone H2A isoform H2A.1 and a decrease in H2A.2 at protein and mRNA in HCC. H2A.1 and H2A.2 are highly homologous, replication- dependent, non-allelic variants of histone H2A differing at only three amino acid positions.

In human cancers as well, the expression of H2A1C isoform is downregulated in chronic lymphocytic leukemia and gall bladder cancer[2,134]. In 4 normal individuals and 40 CLL patients, a significant decrease in the relative abundance of histone H2A1C was observed in primary CLL cells as compared to normal B cells[134]. Similar changes in the abundance of H2A isoforms are also associated with the proliferation and tumorigenicity of bladder cancer cells[2]. The regulation of replication-dependent histone H2A expression was also reported to occur on a gene-specific level[2]. Interestingly, later on with a larger cohort of samples, H2A1C

expression was conversely reported to be up regulated in CLL (10). In breast cancer H2A1C is upregulated and controls ER target genes in ER positive breast cancer cell lines[4].

Collectively, these reports suggest that not only the proportion of isoforms alter in different pathological states but the cell intricately maintains the balance of the H2A isoforms during different processes. However, whether this change in the balance of the H2A isoforms is a consequence of the pathological state or these isoforms contribute to the attainment of such states has not been addressed in depth.

# II. AIMS AND OBJECTIVES

## Chapter 2

*Experiments carried out to address the objectives and the key outcomes of the subsequent chapters are briefly described.*

## 2.1. Statement of the Problem

It is not known whether two major H2A isoforms, H2A.1 and H2A.2, that are known to exhibit altered expression under different physiological conditions have non-redundant function. Further, although, the distinct functional effects of H2A isoforms in humans has now been reported, the mechanistic basis of the non-redundancy still remains elusive. In addition, whether the histone isoforms may modulate chromatin structure and stability has not been investigated in depth. Hence, it is important to investigate the possible functional significance of the altered expression level of H2A.1 and H2A.2 isoforms, which is observed during development and disease like cancer.

## 2.2. Hypothesis

We hypothesize that the differential incorporation of H2A.1 and H2A.2 may alter chromatin dynamics which in turn may influence the gene expression and phenotypic behaviour of the cell.

## 2.3. Objectives

I. To study alteration in organization of nucleosomal structure with H2A.1 and H2A.2.

II. To study the functional significance of H2A.1 and H2A.2.

## 2.4. Experiments Conducted

The following experiments were conducted in order to address the objectives:

### *I. To study alteration in organization of nucleosomal structure with H2A.1 and H2A.2.*

1. Expression and purification of histones using recombinant methods
2. Reconstitution of H2A.1-H2B and H2A.2-H2B dimer and their structural validation
3. Equilibrium unfolding of H2A.1-H2B and H2A.2-H2B dimer by thermal and chemical denaturation monitored by circular dichroism spectroscopy
4. Equilibrium unfolding of H2A.1-H2B and H2A.2-H2B dimer by chemical denaturation monitored by fluorescence spectroscopy
5. Mutational studies with H2A.1-H2B dimer by thermal denaturation monitored by circular dichroism spectroscopy
6. Energy minimisation of H2A.1-H2B and H2A.2-H2B dimers with analysis of non-covalent bonds
7. Salt dissociation experiment for assessing the stability of H2A.1 with H2A.2 containing nucleosomal complexes
8. Fluorescence Recovery after Photobleaching (FRAP) assay for comparing the dynamics of H2A.1 and H2A.2 in live cells
9. FRAP assay with H2A.1 mutants for identifying residue(s) having the most significant impact on the dynamics of H2A isoforms
10. Molecular dynamic simulation studies to compare the stability of H2A.1 with H2A.2 containing nucleosomal complexes
11. Micrococcal nuclease digestion assay to assess the effect of H2A.1 and H2A.2 on chromatin organisation

## *II. To study the functional significance of H2A.1 and H2A.2.*

1. Comparative analysis of expression level of H2A.1 and H2A.2 in different organs of rat

2. Comparative analysis of expression level of H2A1C and H2AA3 in different human transformed cell lines with respect to their immortalised counterparts
3. Determination of subnuclear distribution of H2A.1 and H2A.2
4. Cell based assays, namely, MTT, colony formation, cell cycle progression and migration assays for determining the effect of overexpression of H2A.1 and H2A.2 isoforms
5. Determination of the residue important for the pro-proliferative effect of H2A.1 by carrying out mutational analysis followed by MTT assay
6. Comparative analysis of H2Aub119 level and expression level of major ubiquitinase and deubiquitinase in normal vs tumour tissue and pre-neoplastic vs neoplastic cell lines
7. Analysis of correlation between H2Aub119 level and proliferation of cells
8. Investigating the effect of HDAC inhibitor and H2Aub level in cells

The remaining chapters will describe the results of my work that has been carried out to fulfill the above mentioned objectives. Each of the chapters consists of the subsections *Introduction*, *Methods*, *Results*, *Discussion* and *Conclusion*. **Chapter III** "Heterologous Expression in Bacteria" will describe the effects of various expression parameters on the accumulation of recombinant histones. It was found that some of the histone proteins, especially H2A.2 and H4, do not express robustly in recombinant host. A thorough investigation suggests that H2A.2, H2B.1, H3.2 and H4 are probably unstable upto varying degrees leading to their disappearance from the host in a time dependent manner. Destabilization could be prevented by incorporation of the commonly used purification tags, like GST, of appropriate size and position leading to accumulation. Our results provide compelling evidence that the proposed parameters which modulate recombinant expression like codon bias usage, mRNA secondary structure and GC content of the transcript modulate this process possibly by influencing protein folding. This

part of the work was published in the journal *Advances in Biotechnology Reports* with the title "**Incorporation of a tag helps to overcome expression variability in a recombinant host**".

**Chapter IV** "Reconstitution of Histone Oligomers" describes the methodology used for purification of bacterially expressed histones and the steps performed to reconstitute the histone oligomers for carrying out *in vitro* experiments. We describe a protocol to obtain large quantities of highly pure histones using bacterial expression system for GST pull-down and reconstitution experiments. In addition, we describe methods to quickly reconstitute and purify H2A-H2B dimers, H3-H4 tetramers and histone octamers for *in vitro* experiments. We demonstrate that these sub-complexes are properly folded and are hence, true representatives of the actual substrates *in vivo*. We also show that histones have a propensity to be non-specifically cleaved by proteases. Our results suggest that TEV protease is the most suitable protease while working with histones. The methodology described here should allow researchers to purify histone complexes in three days enabling functional and structural analyses of histone variants, mutants and post-translational modifications. This part of the work was published in the journal *Advances in Biosciences and Biotechnology* with the title "**A Simple Method to Produce Sub-Nucleosome Complexes of High Purity *In Vitro***".

**Chapter V** "Physiological Effect of H2A.1 Isoform Overexpression" compares the non-redundant functional effects of H2A.1 and H2A.2 isoforms from the physiological aspect. We found that H2A.1 exhibits drastically altered expression pattern in different normal tissues and human cancer cell lines (H2A1C in humans). H2A.1 promotes cell proliferation of neoplastic CL38 cells. Interestingly, similar effects are not observed in pre-neoplastic CL44 cells suggesting that the non-redundant functionality might be context dependent. Studies with mutant demonstrate that K99R substitution is majorly responsible for the pro-proliferative effect observed on H2A.1 overexpression with a synergistic effect of M51L alteration. Notably,

S16T, the only substitution between rat H2A.1 and human H2A1C does not appear to have any effect on the non-redundant functionality of H2A.1 in the assays that we have performed.

**Chapter VI** "Non Redundant Stability and Structural Contribution of H2A Isoforms" investigates the effect of incorporation of H2A.1 isoform on chromatin dynamics to uncover the molecular basis of the non-redundancy. We carried out equilibrium unfolding of recombinant H2A.1-H2B dimer. We found the M51L alteration at the H2A-H2B dimer interface decreases the temperature of melting of H2A.1-H2B by ~$3^0$C as compared to H2A.2-H2B dimer. This difference in dimer stability is also reflected in chromatin dynamics as H2A.1 containing nucleosomes are more stable owing to M51L and K99R substitutions. Molecular Dynamic Simulation suggests that these substitutions increase the number of hydrogen bonds and hydrophobic interactions enabling H2A.1 to form more stable nucleosomes. Interestingly, the same two substitutions have the most prominent effect on cell proliferation, suggesting, that the nucleosome stability is intimately linked with the physiological effects observed.

Exploration of the context in which H2A.1 imparts its non-redundant functional effect led to the work described in **Chapter VII** "H2AUb119 in Hepatocellular Carcinoma". In this study, we report that H2Aub markedly decreases in hepatocellular carcinoma (HCC). Usp21, a H2A deubiquitinase, is probably responsible for decrease in H2Aub. In addition, the H2Aub levels showed an inverse correlation with H3S10 phosphorylation (H3S10p) and the proliferative state of the cells. Down regulation of H2Aub is also associated with increased expression of growth factor gene lipocalin 2. Interestingly, we show that treatment of cells with histone deacetylase inhibitor trichostatin A results in increase of H2Aub and decrease in H3S10p. Our work for the first time suggests the *in vivo* association of H3S10p, H4ac and H2A119ub with cellular transformation. This part of the work was published in the journal *Experimental Biology and Medicine* with the title *"***Histone H2A Mono-ubiquitination and Cellular**

**Transformation are Inversely Related in N-nitrosodiethylamine-induced Hepatocellular Carcinoma***"*.

**Chapter VIII** discusses the overall outcome of the thesis. The thesis ends with the salient findings of the work described in the thesis and some future directions.

# III. HETEROLOGOUS EXPRESSION OF HISTONES IN BACTERIA

## Chapter 3

*Work carried out for optimising expression conditions for histones in E. coli for obtaining protein to carry out in vitro assays*

**This work was published in the journal "Biotechnology Reports"**

## 3.1. Introduction

With the identification of increasing number of sequence divergent forms of histones having functionally important and non-redundant roles it is of interest to understand whether the changes they bring about in the biochemical composition of the nucleosome and its sub-complexes is reflected in terms of stability. Large quantity of histones can be purified by isolating them from cells[135], however, they are highly heterogeneous in terms of variants and PTM's[136,137]. This does not allow the investigation of the possible contribution of a particular histone variant or isoform on stability and structural organization of chromatin. In such scenario use of recombinant histones becomes the method of choice to study the effect of histone variants or modified histones on the structure and stability of chromatin.

Despite the added advantages of well-understood genetics and economically favoured nature of using *E. coli* as a host for carrying out heterologous expression, at times, there is complete absence of recombinant protein expression in bacteria. It is intriguing that the lack of recombinant protein expression is a, rather, commonly encountered phenomenon considering that a DNA sequence with the required elements like host compatible promoter, should be transcribed and expressed. It is puzzling that two proteins of similar nature are sometimes expressed at varied levels in the recombinant host. The classic example is presented by core histones, the expression levels of which, vary drastically on recombinant expression in *E. coli* [26,138].

There is no definitive explanation as to why some proteins do not express in bacteria. The different hypothesis, which have been put forth to account for these observations, suggests that inherent property of transcript that hinders translation is the root cause for this [139,140,141,142]. In this regard, the codon usage bias hypothesis is the most prominent and accepted one. Codon

usage bias refers to the differences in the frequency of occurrence of synonymous codons in DNA and is organism specific [reviewed in 143]. It is speculated that these differences in codon bias may play a role in determining levels of protein expressed in heterologous hosts and sometime, may even lead to the complete absence of expression. Overcoming codon bias may lead to an increment in the yield of recombinant protein. Although, codon optimization proves to be useful in some cases, however, at times, the lack of expression persists[144,138].

To account for the different observations pertaining to recombinant protein expression, the codon bias hypothesis has evolved with time and different rare codon parameters like clusters[145], their number in the first six codons[146] and the ones coding for arginine[147] are now considered as important determinants. The arginine codons AGA and AGG particularly have been shown to be responsible for lower levels of protein expression in *E. coli*[148]. Likewise, high frequency of arginine residues is suspected to negatively affect expression of histone H1 and H5[149]. However, more studies are needed to ascertain these proposed correlations.

As an alternative explanation for this, rather mysterious problem, the importance of factors like strong mRNA secondary structure and the 'GC' content of the transcript have been emphasized in relation to transcription and translation[139,142,149]. It has even been claimed that amino terminal encoded rare codons reduce RNA secondary structure, and thereby increase expression[150]. However, the ribosome possesses intrinsic helicase activity and may allow translation through strong secondary structures[151]. Further, in the case of prokaryotic systems, as the mRNA is mostly covered with ribosomes, the structures anticipated by the prediction tools may not represent the actual mRNA structures *in vivo*[152]. Hence, the possible role of mRNA secondary structure in protein expression is somewhat conflicting and needs more validation. The different parameters discussed above, the diverse hypotheses and the conflicting reports pertaining to lack of heterologous expression are very well reviewed in the

article by Welch *et al*, 2009[153]. In brief, the numerous factors that have been proposed to explain lack of recombinant protein expression often does not lead to a predictable expression outcome.

Many comprehensive studies have been performed to understand the effect of these factors in an almost isolated way but very few have focused on uncovering the possible correlation between the most emphasized parameters and the reason behind such association. Hence, for the better understanding of the effect of different expression parameters on heterologous expression and for obtaining histones in large quantities for our *in vitro* experiments, we conducted a systematic investigation with focus on validating the effect of different proposed elements on the expression of histones. Also, we have emphasized on finding any possible correlation amongst the different factors which has been reported to improve yield of recombinant proteins in certain cases.

## 3.2. Methods

### *3.2.1 Construct preparation*

The coding sequences of the genes used in the experiments were amplified using Taq Polymerase (NEB) from the cDNA synthesised from RNA (treated with DNaseI) isolated from cell lines. The amplicons were cloned into pTZRT57 vector according to manufacturer's (Thermo scientific) instructions. The cloned fragments were further sequenced. For subsequent cloning into different expression vectors, the coding sequences were amplified with primers incorporating the appropriate restriction sites (please see *Annexure III* in *Appendix* for primer sequences) and were sub cloned, maintaining the correct reading frame which was verified by sequencing. Please see please see *Annexure V* in *Appendix* for few of the representative construct maps.

### 3.2.2 *Plasmid Isolation (Mini-prep)*

The protocol for plasmid isolation is as described earlier[154]. 10ml of LB media, containing suitable antibiotic, was inoculated with single colony from the transformed plate and incubated at $37^0$C overnight. 1.5ml of the overnight grown culture were transferred to an Eppendorf tube and the bacterial cells were pelleted down by centrifuging at 10000g for 5min at $4^0$C. The supernatant was discarded and the pellet was resuspended uniformly without any clumps in 150µl of TELT buffer. 5µl of lysozyme (50mg/ml) was added to the tube. The contents were mixed and kept on ice for 1min. The tube was then placed in a boiling water bath for 1min and immediately chilled on ice for 10min. The tube was then centrifuged at 12000g for 10min at $4^0$C. The pellet was discarded with the help of a micro-tip. 330µl of chilled absolute alcohol was added to the supernatant and placed at $-20^0$C for 30min or $-70^0$C for 5min. The tube was then centrifuged at 12000g for 10min at $4^0$C. The supernatant was discarded and the pellet was washed with 70% ethanol. The supernatant was discarded and the pellet was air dried. The air dried pellet was dissolved in 20µl of nuclease free water.

### 3.2.3 *Preparation of Competent Cells (Calcium Chloride Method)*

The protocol for preparation of competent cells is as described earlier[155]. 100ml of LB media was inoculated with 1ml of overnight grown culture that was obtained by inoculating a single colony of bacteria from streaked plate incubated at $37^0$C overnight. The cells were grown for approximately 2hrs ($OD_{600}$ = 0.4) in a shaker incubator. The culture was chilled on ice for 30min. The cells were harvested by centrifugation by centrifuging at 10000g for 5min at $4^0$C and the supernatant was discarded. The cells were washed with chilled 100mM $CaCl_2$ and centrifuged at 8000g for 5min at $4^0$C. The cells were resuspended in 20ml of chilled 100mM $CaCl_2$ and kept on ice for 45min. The cells were again harvested by centrifugation by centrifuging at 8000g for 5min at $4^0$C and the supernatant was discarded. The cells were

resuspended in ice-cold sterile solution of 100mM $CaCl_2$ + 15% glycerol and kept on ice for 10min. 200µl aliquots were made in pre-chilled tubes and stored at -80$^0$C after freezing in liquid nitrogen.

### *3.2.4 Preparation of Ultra Competent Cells*

The protocol for preparation of ultra-competent cells is as described earlier[156]. DH5α was streaked onto LB agar plate and incubated overnight at 37$^0$C. 100ml of SOB in 1000ml flask was inoculated with a single colony from the plate. The flask was incubated at 18$^0$C with shaking at 200rpm. Once the $OD_{600}$ reached 0.3- 0.4, the culture was kept on ice for 10min. The cells were harvested by centrifugation by centrifuging at 10000g for 5min at 4$^0$C and the supernatant was discarded. The cell pellet was resuspended in 32ml of ice cold transforming buffer and kept on ice for 10min. The cells were harvested by centrifugation by centrifuging at 10000g for 5min at 4$^0$C and the supernatant was discarded. The cell pellet was resuspended in 8ml of ice cold transforming buffer and 600µl DMSO is added. Cells were kept on ice for 10min. 100µl aliquots were made in pre-chilled tubes and stored at -80$^0$C after freezing in liquid nitrogen.

### *3.2.5 Growth and IPTG induction of transformed bacterial expression hosts*

The protocol for recombinant histone expression is as described earlier[26] with slight modifications. The plasmids (100ng) containing the gene of interest was transformed in either competent BL21 (DE3) pLysS and Rosetta (DE3) pLysS and plated onto LB agar plates containing the appropriate antibiotic (ampicillin or kanamycin) for selection of transformed bacteria. In addition, chloramphenicol (25µg/ml) was incorporated into media to maintain selection pressure for the T7 lysozyme-containing phage present in the bacterial strains used for expression studies. The plates were incubated at 37$^o$C for 16hrs in a temperature controlled

incubator. A single colony was inoculated from the plates of transformed bacteria in 5mL or 20mL LB media containing the appropriate antibiotic and incubated at 37°C with constant shaking until the $OD_{600}$ reached between 0.3 and 0.6. Induction of recombinant protein expression were carried out by adding IPTG to a final concentration of 0.2mM. When lack of expression was encountered, titration of IPTG concentration was carried out ranging 0.2mM-2mM. The cultures were induced for 3hrs at 37°C or overnight at 18°C. To investigate the effect of trimethylamine oxide (TMAO), it was added at a working concentration of 60mM, 1.5hrs post IPTG induction and the cells were harvested 1.5hrs post-addition. One tube was kept as uninduced as a negative control.

Post induction the bacteria were harvested by centrifugation at 8000g for 10mins at 4°C and processed. The soluble and the insoluble fractions of proteins were separated resuspending the cell pellet in 1ml of buffer containing 50mM Tris-Cl pH 8.0, 0.5% Triton X-100 and 100µg/ml lysozyme followed by three rounds of sonication, each for 30 seconds at 10% amplitude. The lysate was then centrifuged at 27000g for 30mins at 4°C. The supernatant and pellet, thus obtained, contains the soluble proteins and the insoluble proteins respectively. Sample was prepared by vortexing followed by boiling the lysate for 10mins in a boiling water bath after adding equal amount of 2x Loading Dye. The sample was then cooled by keeping at 4°C and then the proteins were resolved by loading the samples on 10 or 18% SDS-PAGE followed by Coomassie staining (Brilliant Blue R250).

### *3.2.6 Affinity purification of GST fusion proteins*

The protocol for GST based affinity purification is as described earlier[157]. 100µl of 50% slurry of glutathione-sepharose beads (Novagen) was taken in a fresh Eppendorf tube and equilibrated with buffer containing 20mM Tris-Cl pH 8.0, 200mM NaCl, 1mM EDTA pH 8.0, 0.5% Nonidet P-40, 2µg/ml aprotinin, 1µg/ml leupeptin, 0.7µg/ml pepstatin and 25µg/ml PMSF.

500µl of the buffer was added to the tube and the beads were tapped gently for uniformly resuspending them. After 2-3min incubation the tubes were centrifuge at 750g for 1 minute at 4°C to pellet the beads. The supernatant is discarded and the process is repeated at least thrice. Meanwhile, the supernatant containing the GST-fusion protein was kept at 4°C. Post equilibration the fraction consisting of the soluble proteins was added onto the beads in the tube and incubated for 2 hours at 4°C on a rotating mixer to prevent the beads from settling. After incubation, the beads were washed in the same way as they were equilibrated with the same buffer with which they were initially equilibrated. The proteins bound to the beads were eluted with the elution buffer containing 50mM Tris-Cl pH 8.0 and 10mM reduced glutathione (Sigma-Aldrich). The eluted proteins were analysed by preparing the sample as previously described and resolving onto 10% SDS-PAGE.

### *3.2.7 cDNA synthesis, RT-PCR and Real-time PCR*

Total RNA was extracted from previously collected bacterial cells as per the manufacturer's (Macherey-Nagel) instructions. Extracted RNA was further treated with DNaseI (Fermentas) for 60mins at 37°C to degrade any possible DNA contaminant. DNaseI was inactivated by incubating the samples at 72°C for 30mins. RNA was visualized by ethidium bromide staining after electrophoresis on a 1% MOPS denaturing agarose gel. RNA (2µg) was subjected to reverse transcription using M-MLV Reverse Transcriptase and random hexamer primers according to the manufacturer's (Thermo-Scientific) instructions. cDNAs were then amplified with the corresponding gene-specific primer sets (please see please see *Annexure III* in *Appendix* for primer sequences), designed to amplify the total coding sequence, by PCR for 30 cycles using the condition of 30s at 94°C, 1 min at 58°C, and 1 min at 72°C. The PCR products were analysed on a 1% agarose gels containing 0.5 µg/ml ethidium bromide. The cDNA synthesized was further used for Real-time PCR experiments with Syber green dye (Applied

Biosystems). The expression levels were plotted as relative fold change with respect to GAPDH.

### *3.2.8 Agarose gel electrophoresis*

Agarose gel electrophoresis was carried out to analyse levels of PCR and RT-PCR products in different samples[155]. Adequate 1X TAE electrophoresis buffer (40mM Tris-acetate and 2mM $Na_2EDTA$) was prepared to fill the tank of electrophoresis chamber (Techno Source, India) and to prepare the gel. For preparation of gel, agarose (1.5% w/v in TAE) was melted in a microwave oven. Ethidium bromide was added to melted agarose solution which was cooled down to ~55°C at a concentration of 0.5µg/ml. Melted agarose was poured into the gel casting platform with prefixed comb. The comb was removed from the hardened gel. Ethidium bromide was added to tank buffer (0.5µg/ml in 1X TAE) and gel casting platform was placed into the electrophoresis chamber. DNA samples were prepared in appropriate amount of 6X loading buffer (12% Ficoll 400, 60mM EDTA pH 8.0, 0.6% w/v SDS, 0.15% w/v bromophenol blue and 0.15% w/v xylene cyanol). Hundred base pair DNA ladder (NEB, cat#3231) was loaded as molecular weight marker. The gel was electrophoresed at 5V/cm of gel and the progress of separation was monitored by migration of dyes in the loading buffer. After separation of tracking dyes, the gel was examined on UV trans-illuminator to visualize DNA. Band intensities were analysed by ImageJ and Graphpad prism software. Band intensities of histone variants and Ki67 were normalized to those of 18S rRNA.

### *3.2.9 SDS PAGE analysis*

The protocol for SDS PAGE analysis of proteins is as described earlier[155]. In brief, glass plate sandwich was assembled using 0.1cm thick spacers. Resolving gel solution (*Annexure IV*) was prepared and poured into the glass plate sandwich and allowed to polymerize for approximately

30mins. During the polymerisation process a thin layer of water was overlaid on top of the resolving gel to prevent contact with oxygen. Once the resolving gel polymerised the overlaid water was decanted. Stacking gel solution (*Annexure IV*) was prepared and poured into the glass plate sandwich on top of the polymerised resolving gel. in similar manner. A 0.1cm thick Teflon comb was inserted and gel was allowed to polymerize. Histone samples to be analysed were diluted 1:1 (v/v) with 2X SDS sample buffer and incubated for 5min in boiling water. Teflon comb was removed, sandwich was attached to the electrophoresis chamber and filled with electrophoresis buffer (*Annexure IV*). Samples were loaded into the wells formed by comb. The gel was run at 20mA of constant current until the BPB tracking dye entered the separating gel and then at 30mA until the BPB dye reached the bottom of the gel. The power supply was then disconnected and gel was subjected to Coomassie staining or western blot analysis. For Coomassie staining gel was transferred to tray containing Coomassie Brilliant Blue R-250 (CBBR) staining solution (0.1% w/v CBBR, 50% methanol and 10% acetic acid in water). The gel was stained for ~3hrs on shaker. Gel was then transferred to destaining solution (50% methanol and 10% acetic acid in water). The gel was destained on shaker with several changes in destaining solution until protein bands appeared on clear gel.

### *3.2.10 Agarose formaldehyde gel electrophoresis*

The protocol for denaturing agarose gel electrophoresis is as described earlier[155]. Agarose was dissolved in water and cooled to ~60°C to prepare 1% agarose formaldehyde gel. After cooling, 5ml of 10X MOPS running buffer (0.2M MOPS pH7.0, 0.5M sodium acetate and 0.01M EDTA) and 9ml of 12.3M formaldehyde were added. The gel was poured into electrophoresis tray with comb and allowed to set. Comb was removed and gel was placed in gel tank. Gel tank was filled with 1X MOPS running buffer. For electrophoresis 2µg RNA was loaded per lane. RNA volume was increase to 11µl by water and 5µl of 10X MOPS buffer, 9µl of 12.3M

formaldehyde and 25µl of formamide were added and sample was incubated for 15min at 55°C. To this mixture 10µl formaldehyde loading buffer (1mM EDTA pH8.0, 0.25% w/v BPB, 0.25% w/v xylene cyanol, 50% v/v glycerol) was added and loaded onto the gel. The gel was run at 5V/cm until dye migrated one-third to two-third length of the gel. The gel was removed, transferred to RNase free glass dish with water and soaked twice for 20min each. After sufficient removal of formaldehyde, gel was soaked in 0.5µg/ml ethidium bromide and allowed to stain for 40min. The gel was destained in water for 1hr and examined on a UV trans-illuminator at a wavelength of 365nm to visualize RNA.

### *3.2.11 Rare codon, CAI and RNA secondary structure prediction*

Rare codons in the coding region of the transcript were predicted using the Caltor Prediction tool (*http://people.mbi.ucla.edu/sumchan/caltor.html*). CAI (Codon Adaptation Index) for the coding regions of the transcripts was calculated using CAIcal server (*http://genomes.urv.es/CAIcal/*). In the above two servers the coding sequence of the gene was provided as input and the analysis was performed and tabulated. RNA structures of the transcripts were predicted using Vienna RNA Secondary Structure Prediction program on the web at *http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi*. The sequence from ribosomal binding site (RBS) to the seventh codon of the transcript was provided as input.

### *3.2.12 GC content and instability index analysis*

The GC content analysis of the transcripts was carried out using the OligoCalc tool (*http://basic.northwestern.edu/biotools/OligoCalc.html*). Either the sequence of the entire transcript was provided as input or only the sequence of the first 6 codons was considered for calculating the GC content. The instability index of the proteins was calculated using the

ExPASy ProtParam tool (*http://web.expasy.org/protparam/*). The amino acid sequence of the proteins was provided as input.

### *3.2.13 Accession number*

H2A.1 (GenBank: JX661508.1), H2A.2 (GenBank: JX661509.1), H2B.1 63.1 (XM_002725263.1), H2B.1 68.1(XM_002725268.1 updated to XM_002725268.2), H4 (NM_001123469.1), H3.3 (GenBank: BC006497.2).

## 3.3. Results

### 3.3.1 Inconsistent expression of histones in bacteria

Expression was obtained for H2A.1 (lane 4), H2B.1 63.1(lane 8), H3.2 (lane 12) and H3.3 (lane

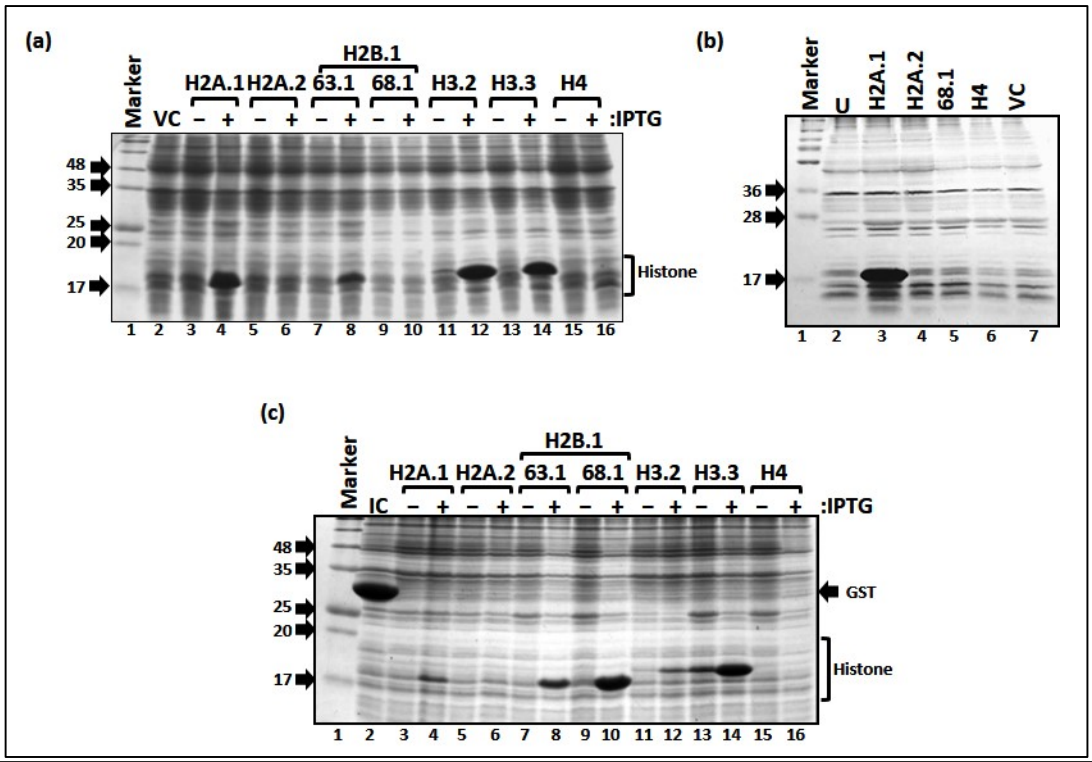



**Figure 3.1: Expression of recombinant histones in *E. coli*.** (a) SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21(DE3) pLysS harbouring the pET3a-histone constructs, induced with 0.2mM IPTG at $37^0$C for 2hrs. The region marked as "Histone" spans the molecular weight range of the histones being expressed. (b) SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS harbouring pET3a -histone constructs, induced with 2mM IPTG at $18^0$C overnight. (c) Expression analysis of pET3a-histone constructs in Rosetta (DE3) pLysS. U- Uninduced, VC- Vector Control, IC- Induction Control.

14) proteins. However, during our attempt to purify rat histone proteins using recombinant methods, we encountered lack of expression of few of the transcripts. There was lack of accumulation of proteins H2A.2 (lane 6), H2B.1 68.1 (lane 10) and H4 (lane 16) even though we used the *E. coli* BL21 (DE3) pLysS strain, which is deficient in *OmpT* and *Lon* protease and especially suited to carry out expression of toxic proteins [**Figure 3.1a**]. These proteins failed to express even at lower growth temperatures post-induction [**Figure 3.1b**] and at varied IPTG concentrations. The presence of rare codons, as proposed earlier, might interfere with translation. Expression of the pET3a H2B.1 68.1 (lane 10) construct was observed in the codon optimised strain Rosetta (DE3) pLysS strain but not of H2A.2 (lane 6) and H4 (lane 16) [**Figure 3.1c**].

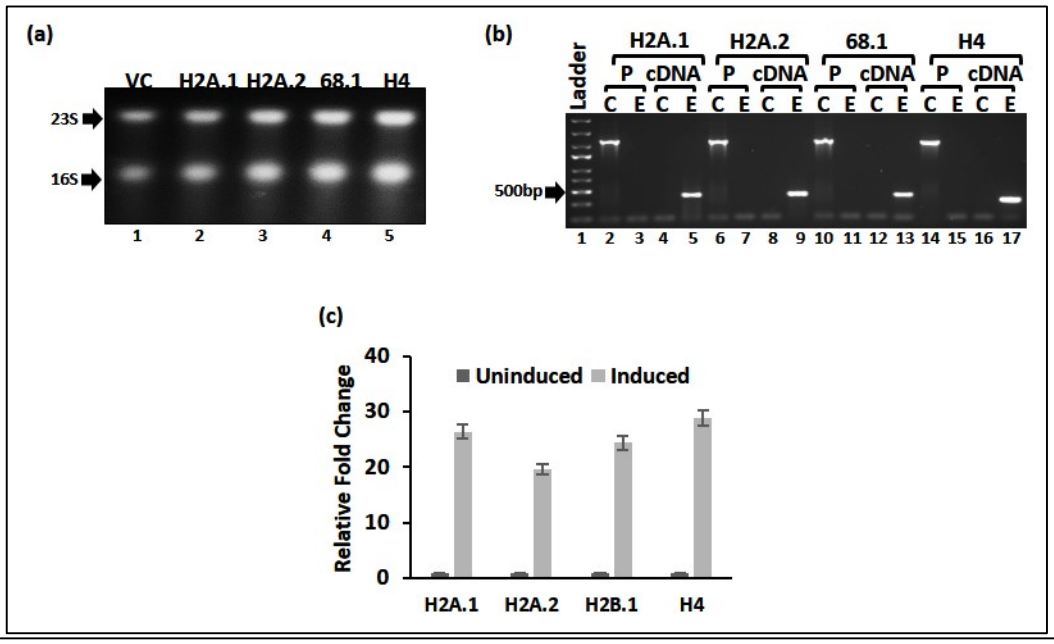



**Figure 3.2: Transcription of recombinant histones in *E. coli*.** (a) MOPS electrophoresis showing the integrity of RNA used to carry out RT-PCR and Real Time PCR. Two distinct bands corresponding to 23S and 16S rRNA shows the integrity of RNA is intact. For checking expression status at the transcript level (b) semi-quantitative RT-PCR was done. 'cDNA' denotes cDNA synthesized by RT-PCR using random hexamer primers followed by PCR amplification with respective gene specific primers. For control (C) cDNA from bacteria having empty pET3a vector was amplified using corresponding gene specific primers and for the experimental sets (E) cDNA from bacteria harbouring pET3a histone constructs was amplified using respective histone primers. H2A.1 was used as a positive control. 'P' denotes experiment to validate DNA free preparation of RNA, in which C is control for amplification where respective plasmids were used as template and amplified using T7 forward and reverse primers. E is the experimental set in which DNA removal by DNaseI from RNA preparation was validated. (c) Real time PCR for respective constructs.

To validate whether the lack of expression is only at the protein level or also at RNA level, purified RNA from transformed bacteria was subjected to RT-PCR. RT-PCR [**Figure 3.2b**] *(see **figure 3.2a** for RNA integrity)* confirmed that the transcripts for recombinant proteins were being produced. Further, real time PCR data demonstrates that H4, was rather, transcribed in relatively more abundance than H2A.1 [**Figure 3.2c**]. The relative transcript levels of H2A.2 and H2B.1 68.1 also were comparable to H2A.1 [**Figure 3.2c**]. These results were particularly interesting considering that there was a remarkable difference in the way the different transcripts (H2B.1 63.1 and 68.1) coding for the same protein expressed in bacteria.

### 3.3.2 Lack of recombinant expression does not stringently correlate with rare codon parameters

| Amino Acid | Rare Codon | H2A.1 | H2A.2 | 63.1 | 68.1 | H3.2 | H3.3 | H4 |
|---|---|---|---|---|---|---|---|---|
| Arg | CGA | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| | CGG | 2 | 2 | 1 | 1 | 5 | 3 | 2 |
| | AGG | 2 | 0 | 0 | 0 | 1 | 5 | 0 |
| | AGA | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| Gly | GGA | 1 | 1 | 3 | 1 | 0 | 1 | 2 |
| | GGG | 0 | 1 | 1 | 1 | 3 | 3 | 1 |
| Iso | AUA | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Leu | CUA | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Pro | CCC | 3 | 2 | 3 | 3 | 3 | 5 | 1 |
| Thr | ACG | 1 | 2 | 2 | 2 | 4 | 1 | 0 |
| | Consecutive | 0 | 0 | 2 | 2 | 2 | 4 | 2 |
| | First 6 Codons | 1 | 0 | 1 | 1 | 0 | 1 | 2 |
| | Frequency | 7.69% | 6.15% | 7.93% | 6.34% | 11.76% | 16.91% | 8.73% |
| | CAI | 0.673 | 0.706 | 0.625 | 0.654 | 0.666 | 0.644 | 0.617 |
| | Total | 10 | 8 | 10 | 8 | 16 | 23 | 9 |
| | Expression | √ | X | √ | X | √ | √ | X |

**Table 3.1: Rare codon content analysis of histone proteins.** The rare codon parameters for the histone coding sequences tabulated. "√" and "X" mark indicates that expression was discernible or non-discernible, respectively, on SDS-PAGE when compared to the "uninduced" control.

Analysis of the rare codon parameters of the transcripts led to the observation of the presence of several inconsistences with the codon bias hypothesis. For instance, the transcript for H3.3 having the highest percentage (16.91%) and number of rare codons in a cluster (four) [**Table**

**3.1**], expressed efficiently, whereas, transcript for H2A.2 having the least percentage (6.15%) and number of rare codons in a cluster (zero), did not express in BL21 (DE3) pLysS [**Table 3.1**]. Likewise, H3.3 transcript having 12 rare codons for arginine, comprising of all of its degenerate rare codons, expressed robustly unlike the transcript of H2B.1 68.1 having only one rare codon for arginine. In addition, the presence of rare codons at the 5'-end of a transcript is also believed to influence mRNA structure and stability and thus, the relative kinetics of translation[158,159]. In this regard, no correlation with expression levels was observed for our transcripts ["First 6 codons", **Table 3.1**]. Next, codon adaptation indexes (CAI) were calculated, which is a measure of the relative adaptiveness of the codon usage of a gene towards the codons used for highly expressed genes. CAI values range from 0 to 1, with higher values indicating a greater proportion of the most abundant codons[160]. Contradictorily, H2A.2 having the highest CAI of 0.706 amongst the histone transcripts [**Table 3.1**] didn't express in either of the strains [**Figure 3.1**].

### 3.3.3 mRNA secondary structure and GC content of transcripts does not always correlate with protein expression levels

Next, we investigated the possible correlation of mRNA secondary structure and GC content of the transcript on protein expression. To maintain consistency with the previously published literature[141] the Vienna RNA web suite[161] was used to predict a possible long-range secondary structure spanning the ribosome-binding site (RBS), translation initiator AUG and the first seven codons. Again, a general pattern did not emerge from our analysis, considering the complexity of the overall secondary structure and the structure near the ATG translation initiation codon [**Figure 3.3**] (free energy of the histone transcripts are tabulated **table 3.2**). For instance, the mRNA secondary structure near the start site is very similar for H2A.1 (expresses robustly) and H4 transcript (no expression).

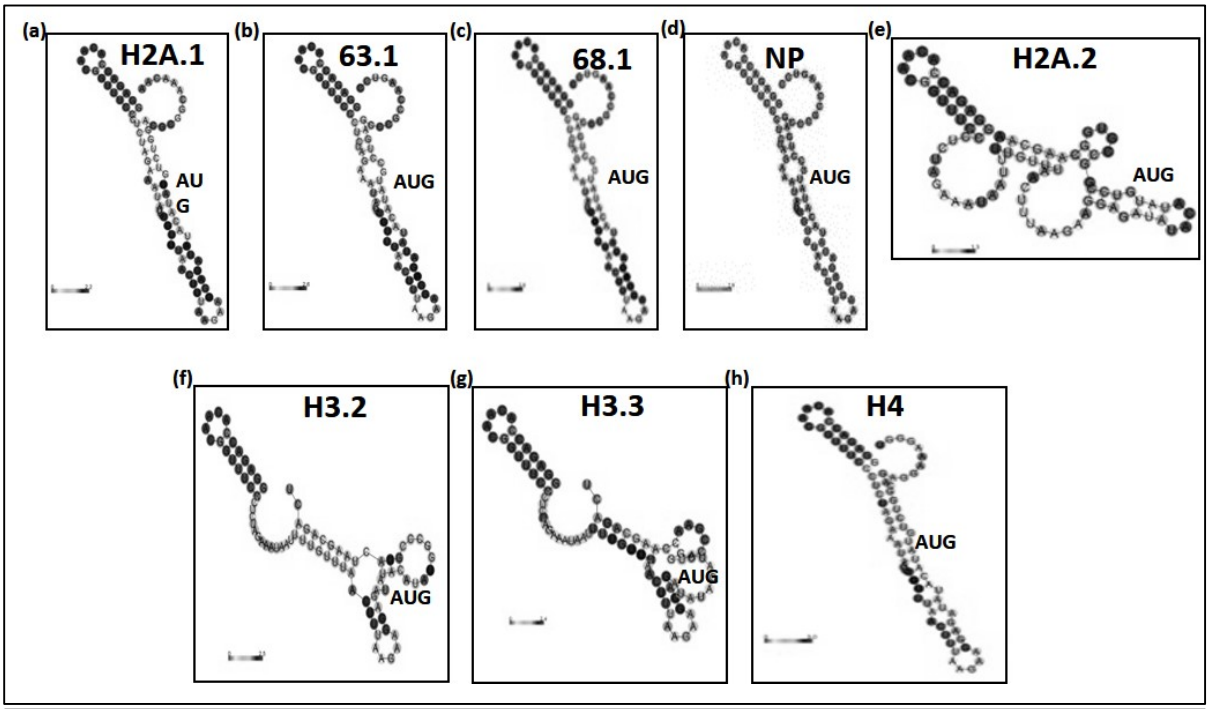



**Figure 3.3: mRNA secondary structure prediction of recombinant transcripts.** mRNA secondary structures for the transcripts used in the study arising from pET3a are depicted in the figure which were predicted using Vienna RNA web-suite. The location of the initiator start codon has been marked with AUG. Residues involved in strong secondary structure formation are depicted as darker circles.

Importantly, the mRNA secondary structure and the free-energy values were identical (-14.20 kcal/mol) in the case of the three transcripts of H2B.1 [**Figure 3.3 (b-d)**]. Thus, their differential expression pattern strongly suggests that the mRNA secondary structure may not be the primary cause behind the results obtained. Finally, the differential expression pattern observed for the same transcript in different strains cannot be explained based on mRNA secondary structure differences. Although, the mRNA structures used in the study were predicted, however, given that the correlation between protein expression and mRNA structure, which were put forward earlier, were also with predicted structure, it is reasonable to be able to contradict the hypothesis using the anticipated structures.

Further, the % GC of the full-length mRNA transcripts as well as the first six codons was calculated **(Table 3.3)**. A pattern appeared to emerge from the GC content of the non-histone transcripts with overall GC content more than 60% inversely correlating with expression and

| Transcript | ΔG (kcal/mol) | Expression |
|---|---|---|
| pET H2A.1 | -17.80 | √ |
| pET H2A.2 | -19.90 | X |
| pET H2B.1 63.1 | -14.20 | √ |
| pET H2B.1 68.1 | -14.20 | X |
| pET H2B.1 NP | -14.20 | √ |
| pET H3.2 | -14.40 | √ |
| pET H3.3 | -15.00 | √ |
| pET H4 | -17.40 | X |

**Table 3.2: Free energy change of mRNA transcripts.** Free energy change for the formation of secondary structure of the histone transcripts arising from pET3a is tabulated. Prediction was performed using Vienna RNA web-suite. The column "Expression" depicts presence (√) or absence (X) of expression on attempting expression in BL21 (DE3) pLysS.

vice-versa. However, the histone transcripts do not follow this pattern. For instance, H3.2 and H3.3 transcripts having GC content of 56% and 61%, respectively, expressed consistently, whereas H4 having a much lesser GC content of 50% did not express.

Also, H2B.1 63.1 and H2B.1 68.1 having identical GC contents express differentially. Further, H4 having the least instability index (42.95), amongst the transcripts analysed was not accumulating, suggesting, that it may not be responsible for the observed accumulation pattern of recombinant proteins.

| Gene | %GC | %GC of 6 codons | Instability Index | Expression |
|---|---|---|---|---|
| H2A.1 | 65 | 56 | 49.18 | **√** |
| H2A.2 | 68 | 67 | 53.68 | × |
| H2B.1 63.1 | 62 | 67 | 47.33 | **√** |
| H2B.1 68.1 | 62 | 67 | 47.33 | × |
| H3.2 | 68 | 56 | 47.19 | **√** |
| H3.3 | 59 | 61 | 44.24 | **√** |
| H4 | 62 | 50 | 42.95 | × |

**Table 3.3: GC content and instability index of histones.** GC content of the histone transcripts used in the study arising from pET3a are tabulated along with the instability index. "√" and "X" mark indicates that expression was discernible or non-discernible, respectively, on SDS PAGE when compared to the "uninduced" control.

### 3.3.4 Rare codons may not always be detrimental to expression

Interestingly, for few transcripts, like H3.2 and H3.3, a positive correlation between rare codon content and expression was found. Also, the expression levels of H2A.1 (lane 4) and H3.2 (lane 12) in the codon-optimised Rosetta (DE3) pLysS strain [**Figure 3.1c**] were significantly lower as compared to that in BL21 (DE3) pLysS strain [**Figure 3.1**a]. These two observations suggested that altering codon bias related parameters may have a context dependent effect on recombinant expression *[see **Figure 3.9e** for quantified data]*.

```
63.1  ATGCCTGAGCCCGCCAAGTCCGCTCCCGCCCCGAAGAAGGGCTCCAAGAAGGCTGTGACC 60
NP    ATGCCTGAGCCCGCCAAGTCCGCTCCCGCCCCGAAGAAGGGCTCCAAGAAGGCCGTGACC 60
68.1  ATGCCTGAGCCCGCCAAGTCCGCTCCCGCCCCGAAGAAGGGCTCCAAGAAGGCCGTTACC 60
      ************************************************** ** ***

63.1  AAGGCCCAGAAGAAGGACGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTACTCGGTG 120
NP    AAGGCGCAGAAGAAGGACGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTACTCGGTG 120
68.1  AAGGCCCAGAAGAAAGACGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCGTG 120
      ***** ******** ****************************************** ** ***

63.1  TACGTGTACAAGGTGCTGAAGCAAGTGCACCCGGATACCGGAATCTCTTCCAAGGCCATG 180
NP    TACGTGTACAAGGTGCTGAAGCAAGTGCACCCGGACACTGGCATCTCTTCTAAGGCCATG 180
68.1  TACGTGTACAAGGTGCTGAAGCAAGTGCACCCGGATACCGGCATCTCTTCCAAGGCCATG 180
      ************************************ ** ** ******** *********

63.1  GGCATCATGAACTCGTTCGTGAACGACATCTTCGAGCGCATCGCGGGAGAGGCGTCGCGC 240
NP    GGCATCATGAACTCGTTCGTGAACGACATCTTCGAGCGCATCGCGGGAGAGGCGTCGCGC 240
68.1  GGCATCATGAACTCGTTCGTGAACGACATCTTCGAACGCATCGCGGGCGAGGCGTCGCGC 240
      ************************************ *********** ************

63.1  CTGGCGCATTACAACAAGCGCTCGACCATCACGTCCCGGGAGATCCAGACGGCCGTGCGC 300
NP    CTGGCGCATTACAACAAGCGCTCGACCATCACGTCCCGGGAGATCCAGACGGCCGTGCGC 300
68.1  CTGGCGCATTACAACAAGCGCTCGACCATCACGTCCCGGGAGATCCAGACGGCCGTGCGC 300
      ************************************************************

63.1  CTGCTGCTGCCCGGGGAGCTGGCCAAGCACGCGGTGTCCGAGGGCACCAAGGCAGTCACC 360
NP    CTGCTGCTGCCCGGGGAGCTGGCCAAGCACGCGGTGTCCGAGGGCACCAAGGCGGTCACC 360
68.1  CTGCTGCTGCCCGGGGAGCTGGCCAAGCACGCGGTGTCCGAGGGCACCAAGGCGGTCACC 360
      ************************************************** ******

63.1  AAGTACACCAGCTCCAAG 378
NP    AAGTACACCAGCTCCAAG 378
68.1  AAGTACACCAGCTCCAAG 378
      ****************
```

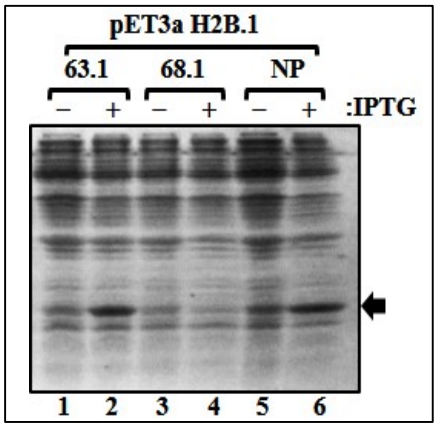



**Figure 3.4: Expression of recombinant histone H2B.1 in *E. coli*.** Multiple alignment of three histone H2B.1 coding sequences used in the study *(left)*. The identical nucleotides are marked with asterix. SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS the pET3a- H2B.1 constructs, 63.1, 68.1, and NP, induced with 0.2mM IPTG at $37^{0}$C for 2hrs *(right)*. Arrow marks the induced protein.

To test this proposition, we focussed on the two histone transcripts of H2B.1, namely 63.1 and 68.1 The only difference in rare codons between the two H2B transcripts are for the rare codons of glycine at position number 158 and 226 in H2B.1 63.1 transcript. As both the transcripts code for the same protein, it was noteworthy that expression was obtained from the transcript

having a greater number of rare codons [**Figure 3.1a**]. By error prone PCR, we were able to obtain another transcript variant (H2B.1 NP) of H2B.1 68.1 in which the rare codon at position number 226 is present. Multiple alignment of the three H2B.1 transcript is shown in ***Figure 3.4***. However, the robust expression of H2B.1 NP transcript, scored by the presence of an intense band at ~17kDa in the induced sample (lane 6), demonstrates that, indeed, rare codons may be favourable to protein expression and not always detrimental [**Figure 3.4**].

### 3.3.5 Incorporation of N-terminus tag positively affect protein accumulation

As tags have been frequently used in purification of heterologous proteins, we wanted to test the effect of tags on recombinant protein expression. A 6xHis tag was incorporated at the N-terminal end of protein to understand its importance in translation kinetics. Interestingly, with

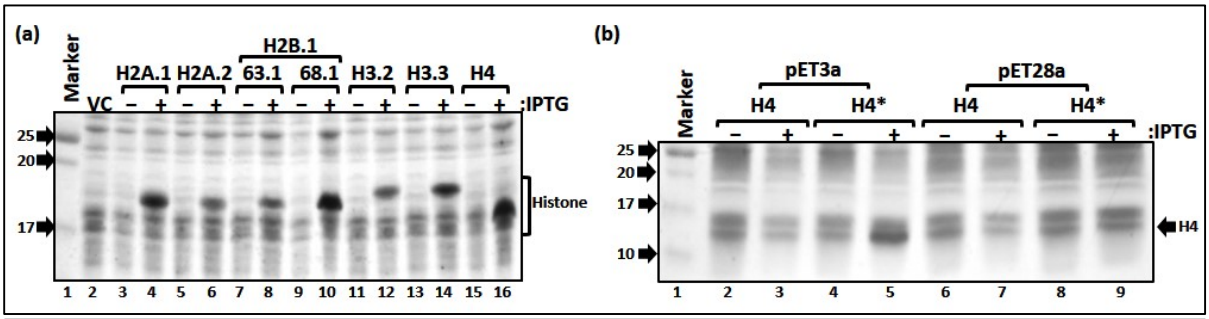



**Figure 3.5: Expression with 6xHis tag.** (a) SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS harbouring the pET28a-histone constructs, induced with 0.2mM IPTG at $37^0$C for 2hrs. (b) 18%, coomassie stained SDS gel shows the expression status of H4 and H4* (codon optimized) in pET3a and pET28a vector backbone in BL21 (DE3) pLysS induced with 0.2mM IPTG at $37^0$C for 2hrs.

the 6xHis tag accumulation of H2A.2 protein (lane 6) was seen even without using a codon optimized strain [**Figure 3.5a**]. This further strengthens our data on lack of direct correlation between rare codons and protein accumulation. Expression was also seen in case of pET28a H2B.1 68.1 (lane 10) and H4 (lane 16) but the results were "inconsistent". In Rosetta (DE3) pLysS consistent levels of expression from pET28a H2B.1 68.1 construct was achieved; however, for pET28a H4 construct the inconsistency persisted without change in the growth of

*E. coli*. Importantly, this suggested to us the possibility of involvement of protein degradation behind the varied accumulation pattern observed for different proteins with changing parameters.

### 3.3.6 Degradation might be the underlying cause behind lack of accumulation of most recombinant proteins in *E. coli*

We speculated that altering the transcript parameters like rare codons, secondary structure, GC content or presence/absence of tag affects protein accumulation either by altering folding and/or degradation. We found that histone H4 expresses most inconsistently in *E. coli,* an observation supported by other reports[26] and hence, tested the occurrence of destabilisation with H4 by western blotting. We speculated that altering rare codons should theoretically alter degradation, hence, we incorporated a codon optimized H4 transcript as well, in our studies ($H4^{*}$). Initially, the expression of $H4^{*}$ was tested both in the BL21 (DE3) pLysS and the Rosetta (DE3) pLysS strain. Consistent with the previous reports, pET3a $H4^{*}$ expressed robustly in bacteria, however, expression was not obtained with pET28a vector [**Figure 3.5b**].

A time-chase experiment was carried out post-induction to monitor H4 degradation with non-codon optimized H4 in both pET28a [**Figure 3.6a**] and pET3a [**Figure 3.6b**] vector backbone. A decrease in the intensity of band in the immunoblot analysis suggested the destabilisation of synthesized protein with time. Notably, we did not observe any smear generally associated with degradation on using antibodies against the histone proteins. This is in agreement with the previous report where decrease in intensity, without any smear, was observed for histone H4[162]. Such degradation was not observed with the $H4^{*}$ transcript [**Figure 3.6f**]. Similar observations were made for H2A.2, H2B.1 68.1 and H3.2 proteins as well [**Figure 3.6c, d, e**], however, with varying kinetics of degradation.

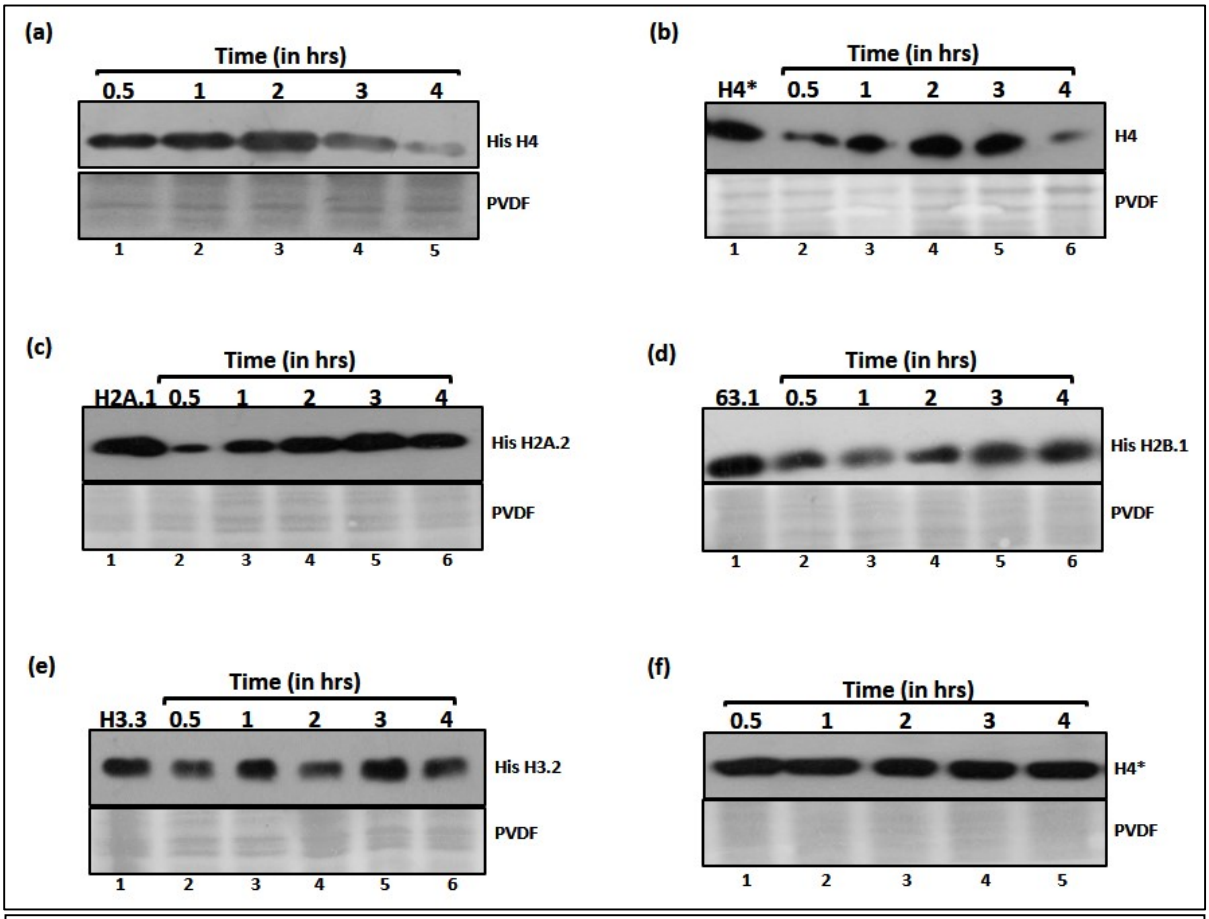



**Figure 3.6: Evidence of degradation of recombinant proteins in *E. coli*.** (a) Western blot of total cell lysates of BL21 (DE3) pLysS harbouring pET3a H4 construct at different time points post-induction, with anti-H4 antibody. (b)(c)(d) and (e) Western blot of total cell lysates of BL21 (DE3) pLysS harbouring pET28a H4, H2A.2, H2B.1 and H3.2 construct respectively at different time points post-induction, with anti-His antibody. (f) Western blotting analysis of total cell lysates of BL21(DE3) pLysS harbouring pET3a H4* construct at different time points post-induction, with anti-H4 antibody. H4* was selected as a negative control for degradation experiment, as it does not undergo degradation. PVDF membrane was stained with Fast green.

The bacterial metabolism and hence, the translation kinetics should depend on growth medium. When the expression of H4 constructs in different expression systems was tested in LB and the much more nutrient enriched 2xTY media [**Figure 3.7a, b, c**], the expression levels were dissimilar, however, no stringent correlation between the nature of media and yield of protein was seen. Incorporation of osmolytes like trimethylamine oxide (TMAO) that has been previously used to stabilise proteins[102] and glycerol in growth media did not enhance protein accumulation [**Figure 3.7d**]. This might possibly be due to the fact that both TMAO and glycerol can themselves be metabolised by bacteria.

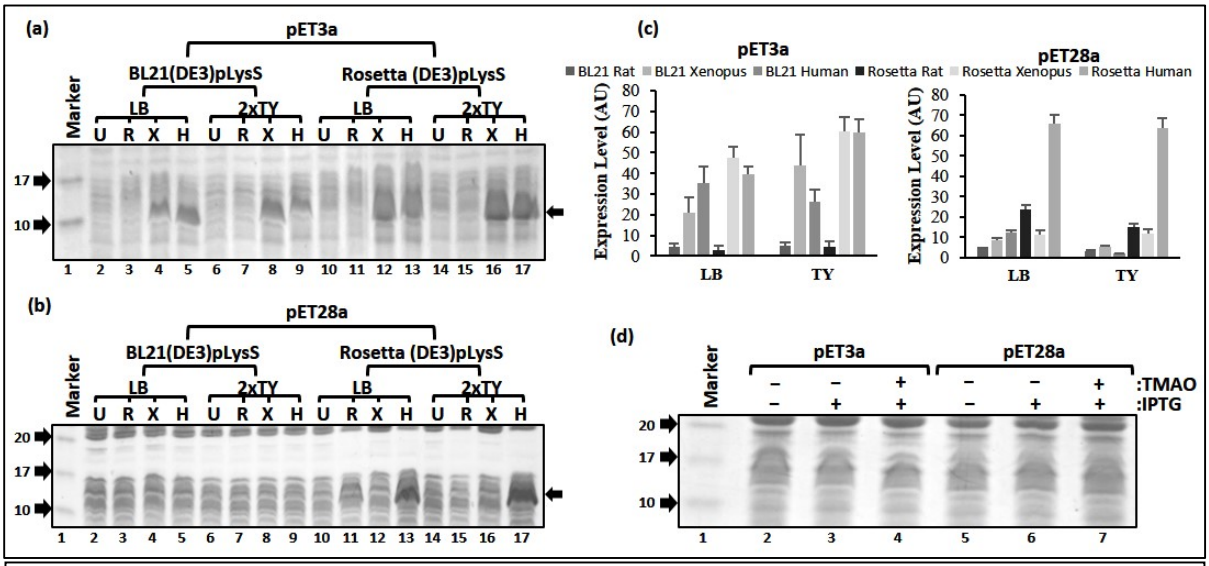



**Figure 3.7: Expression of histones in response to different media, vectors and solutes.** SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS and Rosetta (DE3) pLysS harbouring (a) pET3a H4, (b) pET28a H4 construct, induced with 0.2mM IPTG at $37^0$C for 2hrs. (c) Quantitative representation of three independent experiments depicted in (a) and (b). (d) Effect of TMAO on expression level of pET3a H4 and pET28a H4 depicted as SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS harbouring histone constructs U-uninduced, R-Rat H4 induced, X- Xenopus H4 induced, H- Human H4 induced.

## 3.3.7 Incorporation of GST tag prevents the recombinant protein from destabilization

Our studies with His-tag suggested that a fusion partner may lead to the stabilization of the downstream unit. Strains like BL21 (DE3) pLysS and Rosetta (DE3) pLysS, that are deficient in *OmpT,* can be particularly useful in this regard as they spare GST fusion proteins from degradation[163]. In order to see the effect of GST on accumulation of protein, initially, expression of H2A.1 and H2A.2 were attempted with pGEx2T vector. For both the proteins, accumulation was obtained [**Figure 3.8**]. The expression of H2A.2 was consistently obtained with N-terminal GST tag. To rule out the possible contribution of difference in vector backbone and copy number towards expression level, GST was cloned upstream of the histones already cloned in pET3a (*please see Annexure V in Appendix for details pertaining to construct preparation*). Indeed, consistent levels of expression were seen for all the constructs, including

H2B.1 68.1 and H4 which were most inconsistent, without tag as judged by the presence of an intense band around ~35kD in the induced samples in coomassie stained gels [**Figure 3.9a**].

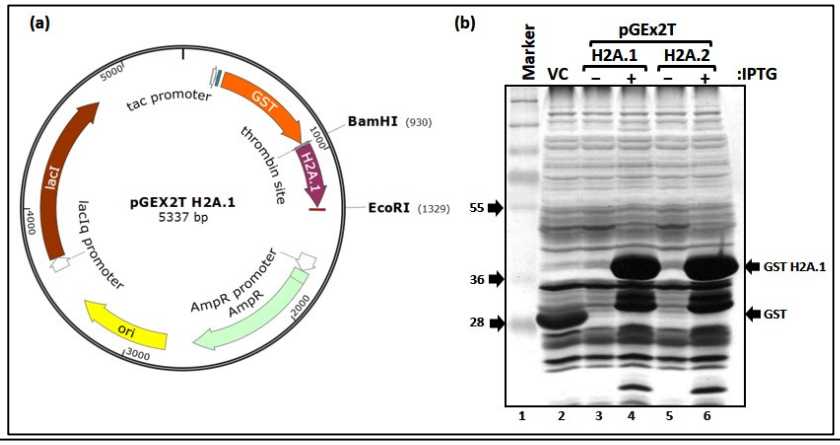



**Figure 3.8: Expression of pGEx2T Histone constructs.** (a) Cartoon showing H2A.1 cloned in pGEX2T vector. (b) SDS-PAGE analysis (18% coomassie stained) of whole cell lysates of BL21 (DE3) pLysS harbouring pGEX2T H2A.1 and H2A.2 constructs induced with 0.2mM IPTG at 37$^0$C for 2hrs. VC- Vector Control.

### 3.3.8 N-terminal tag is better suited to prevent destabilization than a C-terminal one

To test whether the location of the tag is also a determinant, the GST tag was cloned at the C-terminus of histones. Apart from H2A.2 (lane 5), all constructs expressed robustly, although, there was a considerable difference in expression levels of different proteins [**Figure 3.9b (i)**]. Importantly, this rules out the possibility that the expression achieved was due to altered stability of the 5' end mRNA or 5' end amino acid residues or enhanced translation initiation resulting from tag incorporation. Interestingly, at instances the immuno-blot data showed a smear suggestive of degradation of protein with C-terminal GST tag [**Figure 3.9c**], which might be the possible reason for differential protein levels. This was not observed with an N-terminus tag. Further, the binding of the fusion proteins to the glutathione conjugated beads suggested that the GST conformation is maintained when expressed as a C-terminal fusion [**Figure 3.9b (ii)**].

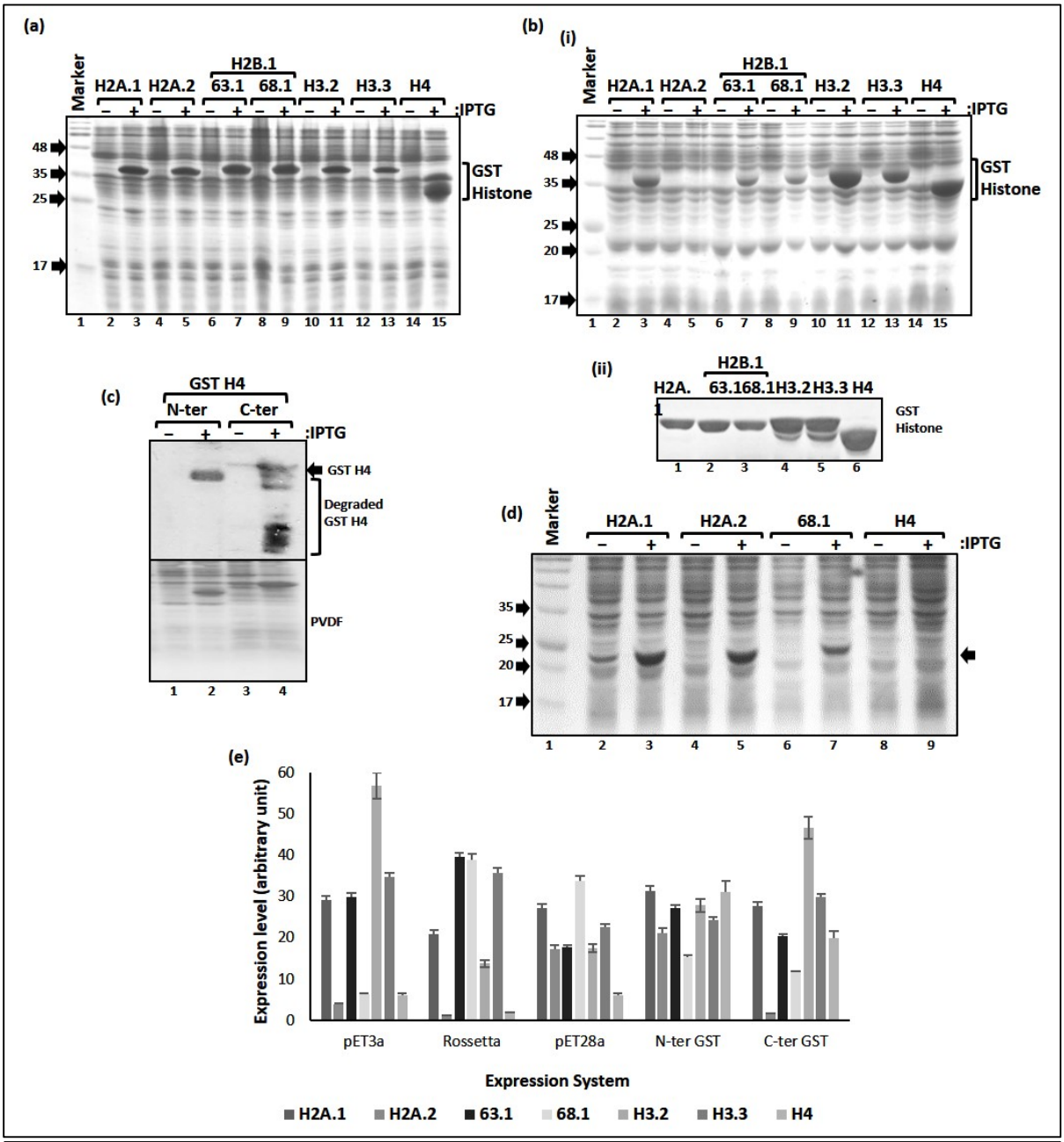



**Figure 3.9: Expression of histones as GST fusions.** (a) SDS-PAGE analysis (10%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS harbouring pET GST (N-terminal) - histone constructs induced with 0.2mM IPTG at 37$^0$C for 2hrs. (b)(i) SDS-PAGE analysis (10% coomassie stained) of whole cell lysates of BL21 (DE3) pLysS harbouring pET GST (N-terminal) - histone constructs induced with 0.2mM IPTG at 37$^0$C for 2hrs. (ii) SDS-PAGE gel (10%, coomassie stained) showing affinity purified C-terminal GST fused Histones. (c) Western blot analysis of N-terminal and C-terminal GST fused H4 with anti-H4 antibody. Region marked spans the degradation of C-terminal GST H4. (d) SDS-PAGE analysis (10%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS harbouring constructs for histone fused with N-terminal domain of GST induced with 0.2mM IPTG at 37$^0$C for 2hrs. (e) Quantitative data showing comparative expression levels of all histone constructs in various vectors.

### 3.3.9 Size of the tag is an important determinant of expression levels

Next, we asked whether the size of the tag influences accumulation levels. To address this, the N-terminal domain of GST comprising of 86 amino acid residues was cloned upstream of the histones in pET3a vector and their expression was attempted. The results were similar to that obtained with the 6xHis tag with stable expression of H2A.2 and H2B.1 and the H4 protein being expressed inconsistently [**Figure 3.9d**]. This clearly suggested that the size of the tag with respect to that of protein of interest is an important determinant.

### 3.3.10 Solubility of proteins doesn't necessarily correlate with accumulation

Proteins in the insoluble aggregates would decrease accessibility to host proteases preventing degradation. However, both GST and MBP tags are known to solubilize proteins and hence, the improved accumulation observed should not be because of the above mentioned reason. However, to rule out this possibility the correlation of protein solubility was investigated with expression level. The solubility profile of histones expressed using pET28a vector in BL21 (DE3) pLysS strain at $37^0$C is depicted in **Figure 3.10a**. All the histones were mostly insoluble.

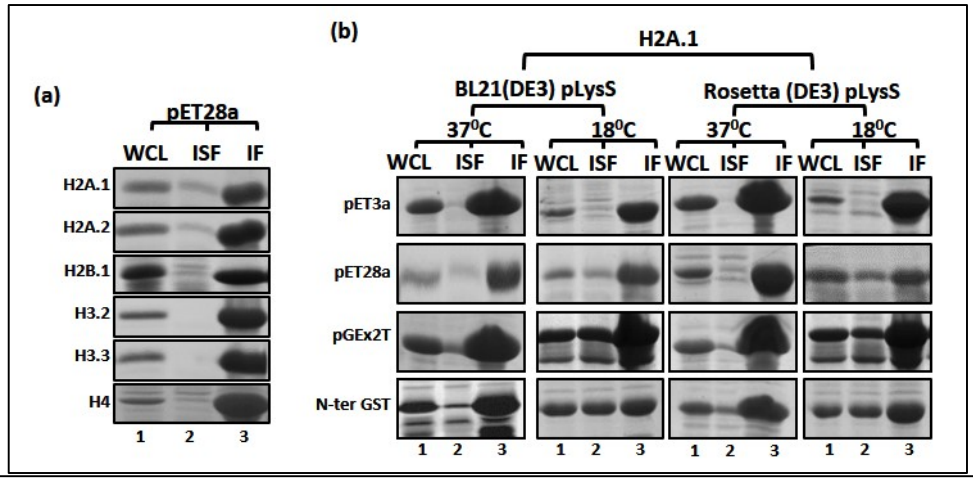



**Figure 3.10: Effect of induction temperature and purification tag on solubility of histones.** (a) Coomassie stained gel showing solubility profile of all histones in pET28a expressed in BL21 (DE3) pLysS. (b) SDS PAGE analysis for solubility profile of H2A.1 in pET3a, pET28a, pGEX2T and with N-ter domain GST expressed in BL21 (DE3) pLysS and Rosetta (DE3) pLysS (see Materials and Methods for induction conditions).

Further, we tested the solubility of H2A.1 with different tags. The experiments were carried

out at $18^0$C, as lower temperature may lead to lower metabolic rate allowing better folding and solubility. The codon optimised Rosetta (DE3) pLysS strain was also used as codon optimisation have been shown to alter protein solubility[164]. We did find altered solubility for 6x His-tagged H2A.1 at $18^0$C in Rosetta (DE3) pLysS strain. As expected the major alteration in solubility profile was observed with GST tag [**Figure 3.10b**]. However, we didn't find any correlation between protein solubility and expression, as irrespective of solubility we could achieve robust accumulation.

## 3.4. Discussion

Heterologous expression is a very powerful tool for producing recombinant proteins. Unfortunately, the lack of heterologous expression is a rather commonly encountered phenomenon and systematic studies are to be done for building strategies to circumvent this problem. Many reports provide guidelines to optimise recombinant protein expression, however, not many studies address why many a times the proposed hypotheses fail to explain the observed results. This might be due to lack of fundamental understanding which dictates recombinant protein accumulation in bacteria and the proposed factors like rare codons, mRNA secondary structure and GC content probably are contributory factors to the fundamental cause. Our results that the rare codons need not be detrimental to protein accumulation and codon-optimisation, at times, may be unfavourable for protein expression, are indicative that this indeed might be true. Possibly, rather than a barrier to the translation machinery, the various transcript parameters like rare codons or mRNA secondary structure (to which GC content also contributes), are modulators of translation kinetics. The overall translation speed can be modulated through tRNA recycling at the ribosome[165] and attenuation of *trp* operon presents a classic example[166] [**Figure 3.11 (2)**].

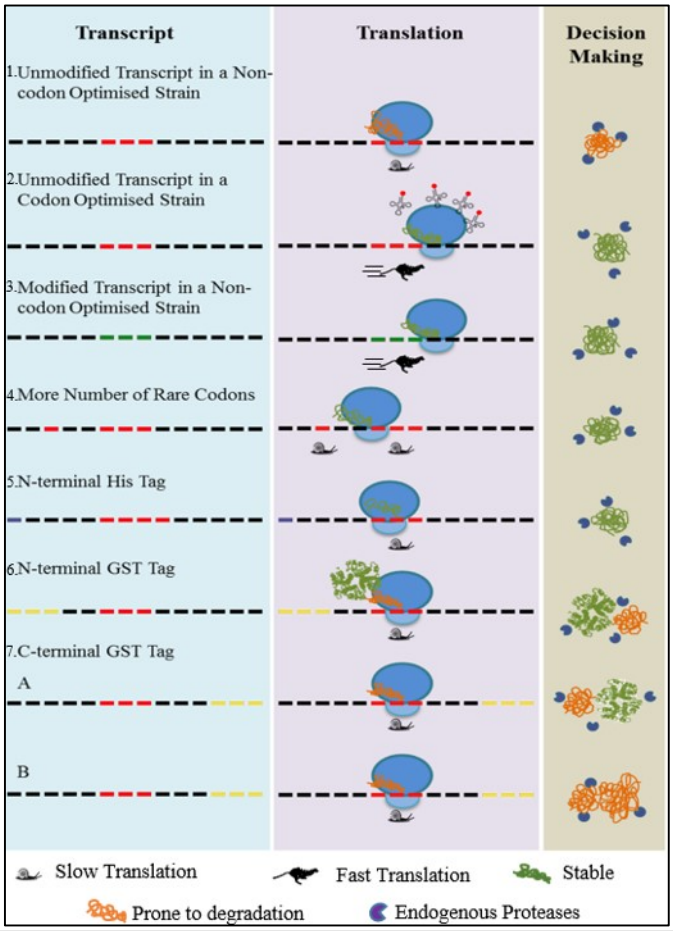



**Figure 3.11: Proposed model for factors influencing recombinant protein accumulation.** The red dotted lines indicate the region of the transcript which are either rich in rare codon, have a strong secondary structure or a high GC content. The green dotted lines indicate the modified region of the transcript having reduced rare codons, secondary structure or GC content. The violet dotted line indicates 6xHis tag whereas the yellow ones indicate the GST tag.

In such scenario, the question arises that why sometimes there is complete lack of expression? One probable reason, might be the fact that transcription, translation and protein folding are coupled in bacteria. Presence of rare codons can affect the rate of translation and would influence the folding, thus local protein structure[167]. The local protein structure in turn may define the response of targeting and degradation by the host protease. Therefore, altering the "translation kinetics code" inherent in the sequence of the transcript can lead to complete lack of expression owing to degradation at times or may result in improved accumulation. This also emphasizes that the position of a rare codon should be more important than its absolute number. Subsequently, coding sequences having more rare-codons, but at positions which promote

"better" folding may lead to better accumulation [**Figure 3.11 (4)**]. This hypothesis is consistent with the proposition that rare codons may provide the genetic instruction for the regulation of rate of protein synthesis to allow proper folding of the nascent polypeptide[168].

Though the strains used for carrying out expression were deficient in the protease *OmpT* and *Lon,* however, other host protease involved in the normal turnover of proteins like ClpXP, ClpAP, HslUV, etc. might be involved in the degradation of heterologous proteins[169]. It is in agreement with an earlier report in which human H4 was shown to be degraded 2hr post-induction [170]. Martinez *et al.* also showed by immuno-blotting that the human phenylalanine hydroxylase yield was greatly reduced because of proteolytic degradation[171]. In our experiments also western blotting was preferred over pulse labelling technique because use of specific antibody does not lead to the background observed with pulse labelling technique. Western blotting has been previously used and recommended to monitor degradation of protein[172,162]. Misfolding of the protein is likely to be responsible for this degradation observed. It is quite expected in bacteria, that eukaryotic proteins won't be properly folded due to the lack of specific chaperones and/or post-translational modification required for folding. This in turn leads to misfolded proteins which are targeted by the bacterial host proteases.

The challenge remains to establish a predictable correlation between any of the proposed parameters and expression, which can help us to engineer means of circumventing destabilisation. Unfortunately, it appears it is not going to be very straight forward. An earlier study showed that for the same transcript replacing the rare codons with the major ones or changing the GC content can independently bring about improved expression[173]. Clearly, not only are these parameters independently important, but correlations amongst secondary structure, the GC content and the codon usage of the transcript are significant[174]. As an

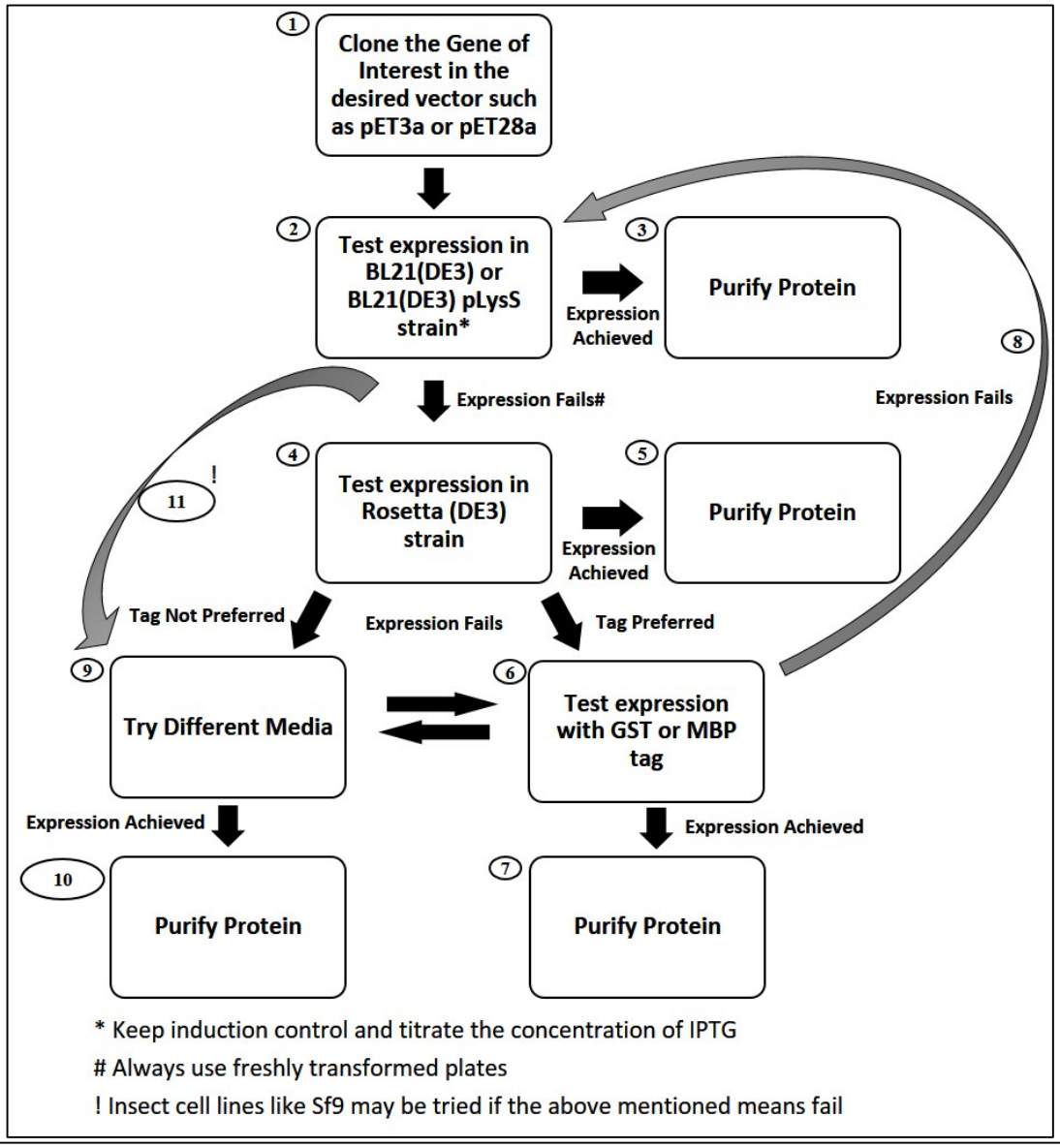



**Figure 3.12: Flow chart for obtaining recombinant protein expression.** A general guideline for obtaining recombinant expression of protein with minimal number of troubleshooting steps.

alternative, we found that incorporation of a tag leads to much more consistent and predictable outcomes of protein expression and the size of the tag with respect to that of proteins is also important. The above hypothesis also gives us the reason as to why some proteins, which do not efficiently express without a tag, are sometimes robustly expressed as fusion with known tags like GST and MBP. Importantly, an N-terminus tag appears to be better suited. Probably, at the N-terminus of a fusion, the folding of GST/MBP precedes that of the downstream unit.

However, when GST is present downstream of a protein, its folding might be influenced by its fusion partner leading to the attainment of a misfolded form [**Figure 3.11 (7B)**].

## 3.5. Conclusions

To summarise, the destabilisation of the protein might be the fundamental cause behind the lack of accumulation of most recombinant proteins and this may be averted by incorporating an N-terminal tag of suitable size. The model presented in **Figure 3.11** depicts the different parameters affecting protein accumulation and the probable mechanism (see legend for details). Finally, the steps depicted in **Figure 3.12** should allow a researcher to troubleshoot expression problems in minimum time with maximum probability of success. Although, affinity tags have been used for a long time to facilitate expression and purification of proteins, we are able to propose the mechanism behind the observed results.

# IV. RECONSTITUTION OF HISTONE OLIGOMERS

## Chapter 4

*Optimization of purification conditions of recombinant histones and reconstitution of histone oligomers to carry out biophysical assays*

**This work was published in the journal "Advances in Biosciences and Biotechnology"**

## 4.1. Introduction

Histones being the primary structural scaffold of chromatin, even minor differences in their primary sequence, can modulate chromatin stability and accessibility. Histones exist as a part of oligomeric complex inside the cells in the form of nucleosomes. H2A is incorporated into this complex in a very precise order in which it first interacts with monomeric H2B to form a dimer and then is incorporated into the octamer around which DNA is wrapped [**Figure 4.1**]. Hence, subtle changes in stability of each of this sub-complexes may manifest its effect at the level of nucleosome stability ultimately leading to altered chromatin dynamics and is an interesting aspect to investigate.

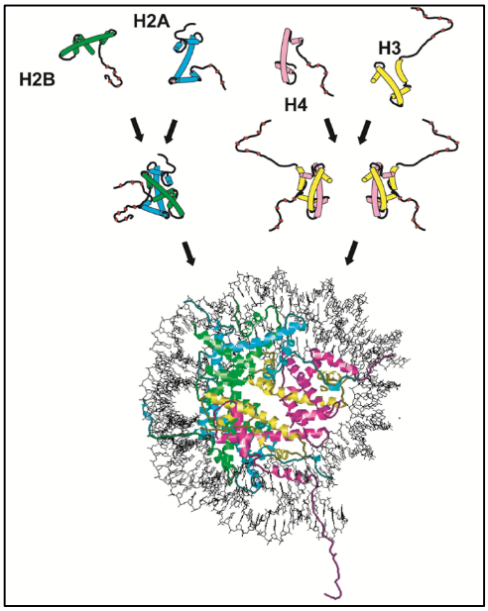



**Figure 4.1. Formation of the nucleosome core particle.** Individual core histones H2A (blue), H2B (green), H3 (yellow) and H4 (magenta) first hetero-dimerize to form the H2A-H2B and the H3-H4 complexes. The different complexes can either under different stringencies or with the help of histone chaperones associate together to form the nucleosome composed of a central tetramer of H3-H4 flanked by two heterodimers of H2A-H2B, and wrapped by 146 base pairs of DNA.
*Adapted from: http://www.intechopen.com/source/html/43165/media/image1.png*

Histones when expressed in bacteria form inclusion bodies and hence, their purification first requires solubilizing them with denaturing agents like urea or guanidine chloride followed by ion exchange chromatography and gel filtration[26]. Requirement of using multiple chromatography steps makes it technically and practically challenging for a non-specialist lab to employ these methods to purify recombinant histones. In addition to the technical challenges faced for purifying recombinant histones, the use of full length histone proteins for carrying

out in vitro assays is limited largely due to their unfolded confirmation in hydrophilic environments and also, due to protein aggregation. Nucleosome sub-complexes like histone H2A-H2B dimer, H3-H4 tetramers and H2A-H2B-H3-H4 octamers should circumvent these issues and in addition, would fare as better representatives of *in vivo* substrates in *in vitro* assays. Although, the production of nucleosome complexes with purified histones has been described earlier[175] it requires considerable expertise and resources to successfully execute the protocol. Further, a major challenge faced for the use of full length histone proteins for carrying out *in vitro* assays and reconstitution is their aggregation in commonly used buffers.

In the present study, various methods to purify histone proteins were evaluated to come up with a simple protocol to enable quick reconstitution of H2A-H2B dimers, H3-H4 tetramers and histone octamers for *in vitro* experiments. The method was subsequently used to reconstitute H2A.1-H2B and H2A.2-H2B dimer to compare their *in vitro* stability.

## 4.2. Methods

### *4.2.1 Growth and IPTG induction of transformed bacterial expression hosts*

Same as described in chapter III (page number 79).

### *4.2.2 Affinity purification of GST fusion proteins*

Same as described in chapter III (page number 80).

### *4.2.3 Ni-NTA affinity purification*

The protocol for Ni-NTA affinity purification of recombinant histone is as described earlier[176] with certain modifications. The insoluble pellet containing histone were dissolved in 50mM Tris–HCl buffer (pH 8.0) containing 500mM NaCl, 5% glycerol, and varying concentration of urea [**Figure 4.5**c] for more details). Pellet obtained from 1000ml of bacterial culture was

resuspended in 10ml of buffer. After the pellet was dissolved the samples were centrifuged at 27,000g for 20mins at $4^0$C. The supernatant was transferred to a fresh tube. The insoluble pellet obtained was again resuspended in 10ml of the same buffer for re-extraction of histones. Samples were again centrifuged at 27,000g for 20mins at $4^0$C. The supernatant was pooled with the previously collected supernatant and the insoluble pellet obtained was discarded. 2ml of 50% slurry of nickel–nitrilotriacetic acid (Ni–NTA) agarose beads (Qiagen) were equilibrated with buffer containing 500mM NaCl, 5% glycerol, and 6M urea. The 50% slurry was added onto a PD10 column (GE) with a stopper. 5 times bed volume buffer was added and allowed to flow through. This was repeated at least thrice. The stopper was attached to prevent the liquid from flowing through and the supernatant containing the 6xHis-tagged histones were then added onto the column. The column was kept on a rocker at $4^0$C for allowing efficient and uniform binding of the proteins to the beads. After a 60min incubation, the stopper was opened and the flow-through was collected and labelled as unbound. The beads were washed with 50mM Tris–HCl buffer (pH 8.0) containing 500mM NaCl, 5% glycerol, 6 M urea, and 10mM imidazole in a way similar to what was described for equilibration. The 6xHis-tagged histones were eluted by a linear gradient of imidazole from 5 to 300mM in buffer containing 500mM NaCl, 5% glycerol and 6 M urea. The eluted fractions were analysed by preparing the sample as described earlier and loading on 18% SDS-PAGE. The histones obtained were dialysed against 2 litre 10mM HCl to remove urea and salts with the help of D-tube Dialyser 6-8KDa cut-off (Millipore). The histones were processed immediately for reconstitution experiments or were stored at $-80^0$C post lyophilisation.

### *4.2.4 Reconstitution of H2A-H2B dimer and H3-H4 tetramer*

To reconstitute the H2A-H2B dimer, the His6-tagged histones H2A and H2B or H3 and H4 in 10mM HCl were mixed in 1:1 stoichiometry. The concentration was determined for the

histones by using their molar extinction coefficient. Further, the quantification was validated by resolving the histones on an 18% SDS-PAGE followed by coomassie staining. After adding the histones, the mixture was rapidly diluted three folds by adding 50mM Tris–HCl (pH 8.0) containing 10mM DTT and 2mM EDTA with continuous stirring. The diluted mixture was then dialyzed (cut-off 10kDa) against 1 litre of 50mM Tris–HCl (pH 8.0) containing 10mM DTT, 2mM EDTA and 2M NaCl for 1 h at room temperature, and the dialysis was continued overnight at $4^0$C. After an overnight dialysis, the impurities and excess histones precipitated. The insoluble histones were removed by centrifugation at 27,000g for 20 mins at $4^0$C. The dialysate was then concentrated using a 10kDa cut-off protein concentrator (Amicon Ultra-15, Millipore). The concentrated proteins were loaded onto equilibrated Superdex-200 HiLoad 16/60 (GE) gel filtration column and the peak fractions were collected. The elution of the protein was monitored by absorbance at 280nm. The salt concentration was reduced by stepwise dialysis for downstream assays.

### *4.2.5 Reconstitution of histone octamer*

The reconstituted H2A-H2B dimer and H3-H4 tetramer were mixed in 1:1 stoichiometry and mixed properly with the help of a magnetic stir bar. The mix was then put into a dialysis tubing (cut-off 10kDa) and dialysed overnight with changes against 2 litre of 50mM Tris–HCl (pH 8.0) containing 10mM DTT, 2mM EDTA and 2M NaCl at $4^0$C. The dialysate was then centrifuged at 13000g for 20min at $4^0$C to remove any precipitate. The clear supernatant was concentrated, if required, using a 10kDa cut-off protein concentrator (Amicon Ultra-15, Millipore). The concentrated protein was analysed on an 18% SDS-PAGE to confirm the presence of all the four core histones. The protein was then injected into equilibrated Superdex-200 HiLoad 16/60 (GE) gel filtration column (flow rate 0.5ml/min) and the peak fraction was collected.

# 4.3. Results

## 4.3.1 Purification of recombinant histones using affinity purification

Histones are highly insoluble when expressed in bacteria. We tried to use the commonly used GST tag as it often solubilizes protein and also enables use of affinity purification. We found

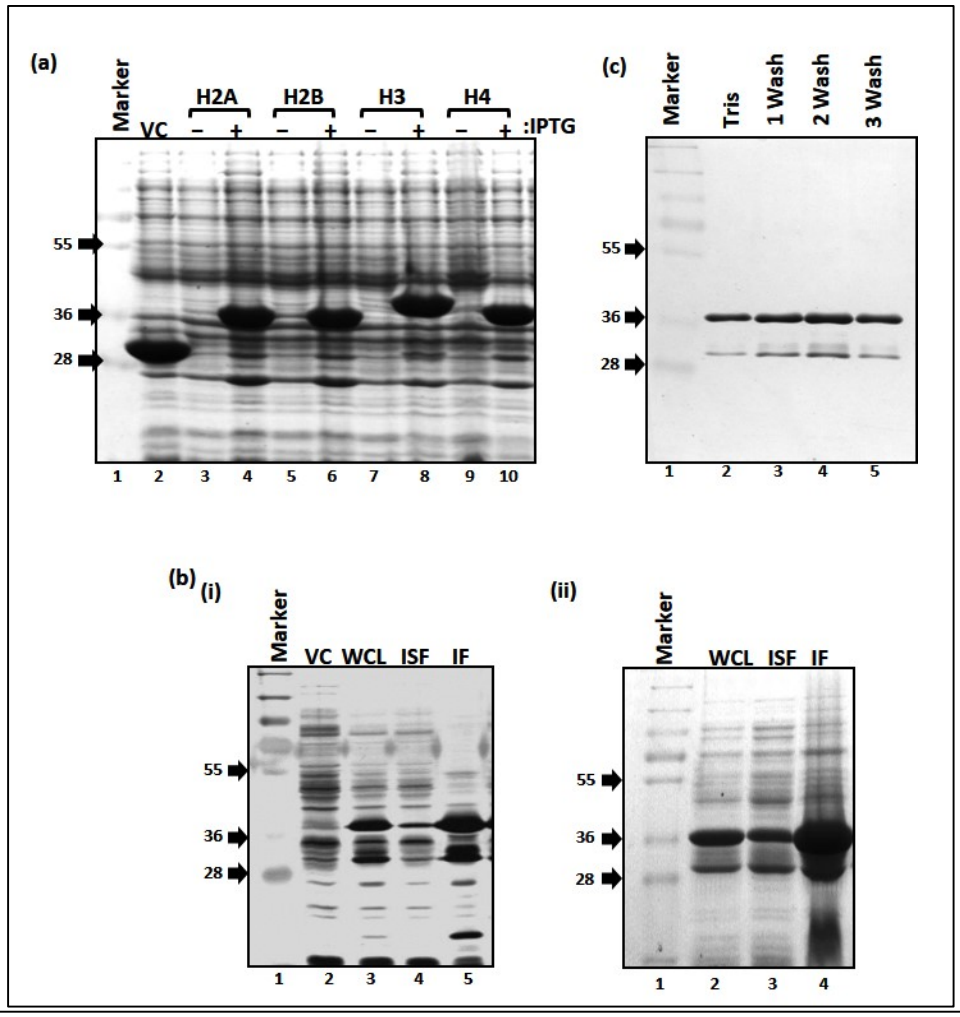


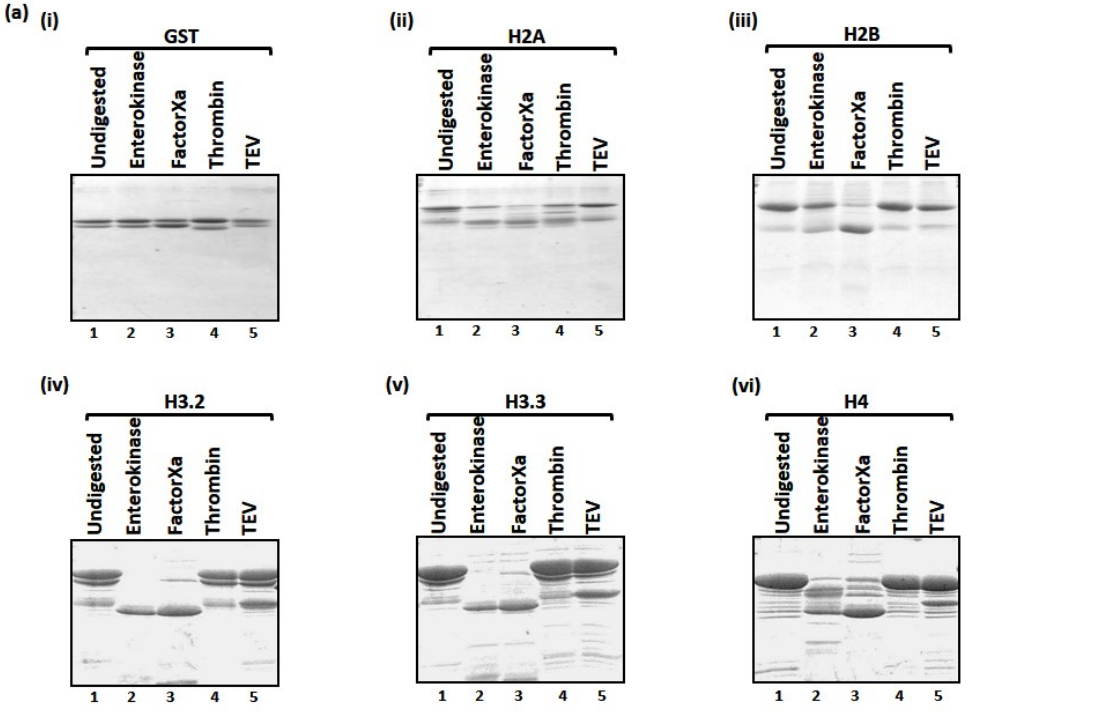

**Figure 4.2: Expression and purification of GST tagged histones.** (a) SDS-PAGE analyses (10%, coomassie stained) of whole cell lysates of BL21(DE3) pLysS harbouring the pGEX2T-histone constructs, induced with 0.2mM IPTG at $37^0$C for 2hrs. (b) SDS-PAGE analyses (10%, coomassie stained) of cell lysates of BL21(DE3) pLysS harbouring the pGEX2T-H2A (i) at $37^0$C, induced with 0.2mM IPTG for 2hrs (ii) at $18^0$C, induced with 0.2mM IPTG overnight. The cells were harvested, lysed post induction and the different fractions were separated by centrifugation prior to loading on the gel. VC- Vector Control, WCL- Whole Cell Lysate, ISF- Induced Soluble Fraction, IF- Insoluble Fraction. (c) SDS PAGE (10%, coomassie stained) showing GST-histone fusions bound to glutathione conjugated beads. The beads were washed with binding buffer increasing number of times to remove non-specifically bound proteins.

the histones expressed robustly with the GST tag [**Figure 4.2a**], however, the expressed proteins were largely insoluble when growth temperature post-induction of recombinant

protein expression was maintained at $37^0$C. Similar results were obtained for all the core histones [**Figure 4.2b (i)**]. On attempting expression at $18^0$C, approximately 10% of the protein was recovered in the soluble fraction [**Figure 4.2b (ii)**]. Proteins were purified by affinity purification by binding to glutathione conjugated beads and washing increasing number of times to remove non-specifically bound proteins [**Figure 4.2c**].

## 4.3.2 Removal of GST tag from recombinant proteins

GST being of 26kDa size and a homodimer itself is expected to interfere with the hetero-

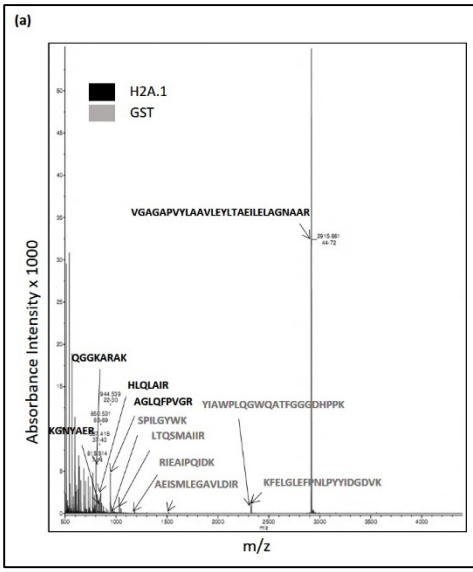



(b)

| | GST | | H2A | | H2B | | H3.2 | | H3.3 | | H4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E | O | E | O | E | O | E | O | E | O | E | O |
| Enterokinase | X | X | X | √ | X | √ | X | √ | X | √ | X | √ |
| Factor Xa | X | X | X | √ | X | √ | X | √ | X | √ | X | √ |
| Thrombin | X | X | √ | √ | X | X | X | X | X | X | √ | - |
| TEV | X | X | X | X | X | X | X | X | X | X | X | X |

**Figure 4.3: Protease cleavage of GST-histone fusions.** (a) SDS-PAGE analyses (10%, coomassie stained) of (i) GST and (ii)-(vi) GST-histone fusion proteins for removal of GST tag. with all the used enzymes. 50µg of the purified protein was incubated with 50ng of enzyme for 1hr at $25^0$C. For buffer condition please see *Annexure IV* in *Appendix*. (b) Table summarising the expected (E) and observed (O) protease activity of the enzymes used.

dimerization of histones. This makes cleavage of the GST tag essential for carrying out H2A-H2B dimerization or H3-H4 tetramerisation. We tested the efficacy of some commonly used proteases thrombin, factorXa, TEV and enterokinase for histone purification. Incubation of only GST protein did not lead to any cleaved products [**Figure 4.3a (i)**] as expected. However, to our surprise, we found proteases like factorXa and enterokinase, non-specifically cleave the histones as can be seen with the disappearance of the band corresponding to GST fusion and appearance of band corresponding to the GST protein [**Figure 4.3a (ii)-(vi)**]. By mass spectrometry we confirmed that the protein we purified was indeed GST-Histone fusion [**Figure 4.4**]. On screening for different proteases which are continually used in




(b)

Histone H2A type 1-C OS=Rattus norvegicus (Score 49.60)

| Meas. M/z | Calc. MH+ | Intensity | Dev. (Da) | Dev. (ppm) | Range | Sequence |
|---|---|---|---|---|---|---|
| 815.514 | 815.485 | 1952.353 | 0.029 | 35.759 | 7 - 14 | QGGKARAK |
| 837.418 | 837.421 | 1070.318 | -0.003 | -3.794 | 37 - 43 | KGNYAER |
| 850.531 | 850.526 | 2149.802 | 0.005 | 5.775 | 83 - 89 | HLQLAIR |
| 944.539 | 944.531 | 4418.682 | 0.008 | 8.663 | 22 - 30 | AGLQFPVGR |
| 2915.661 | 2915.588 | 30719.427 | 0.073 | 25.073 | 44 - 72 | VGAGAPVYLAAVLEY LTAEILELAGNAAR |

Glutathione S-transferase class-mu 26 kDa isozyme OS=Schistosoma japonicum (Score 57.50)

| Meas. M/z | Calc. MH+ | Intensity | Dev. (Da) | Dev. (ppm) | Range | Sequence |
|---|---|---|---|---|---|---|
| 963.557 | 963.530 | 213.090 | 0.027 | 28.289 | 2 - 9 | SPILGYWK |
| 1032.597 | 1032.587 | 808.265 | 0.010 | 10.076 | 65 - 73 | LTQSMAIIR |
| 1048.591 | 1048.582 | 520.564 | 0.009 | 8.852 | 65 - 73 | LTQSMAIIR 5: Oxidation (M) |
| 1182.701 | 1182.684 | 228.793 | 0.017 | 14.105 | 182 - 191 | RIEAIPQIDK |
| 1516.829 | 1516.804 | 125.200 | 0.025 | 16.357 | 90 - 103 | AEISMLEGAVLDIR |
| 2326.189 | 2326.140 | 874.676 | 0.049 | 20.945 | 198 - 218 | YIAWPLQGWQATFGG GDHPPK |
| 2357.212 | 2357.206 | 36.508 | 0.005 | 2.257 | 45 - 64 | KFELGLEFPNLPYYID GDVK |

**Figure 4.4: Mass spectrometric analysis of GST H2A.** (a) Plot of mass/charge ratio vs absorbance intensity of peptides of GST and H2A.1 marked in different colours. (b) Tabulated list of the peptide sequences of H2A.1 and GST identified by MALDI.

recombinant protein purification, TEV protease was the only one, found not to cleave histones non-specifically. The expected and observed pattern of protease activity is tabulated in **Figure 4.3b**. Although, thrombin is expected to cleave H4 protein (as per ExPASy Peptide Cutter) but the cleavage site is just after four amino acids. This makes the band shift difficult to be perceived on SDS-PAGE [**Figure 4.3a (vi)**]. Post GST removal the monomeric histones were found to precipitate limiting their use.

### 4.3.3 6xHis tag is better suited for affinity purification of histones

Due to the above mentioned limitation encountered, we further explored other affinity based purification strategies for histones. We opted to use N-terminal 6xHis tag for histone purification keeping in consideration that a small tag like 6xHis is not expected to interfere in the oligomerization of histones (N-terminal tails of histones protrude out of the nucleosomes).

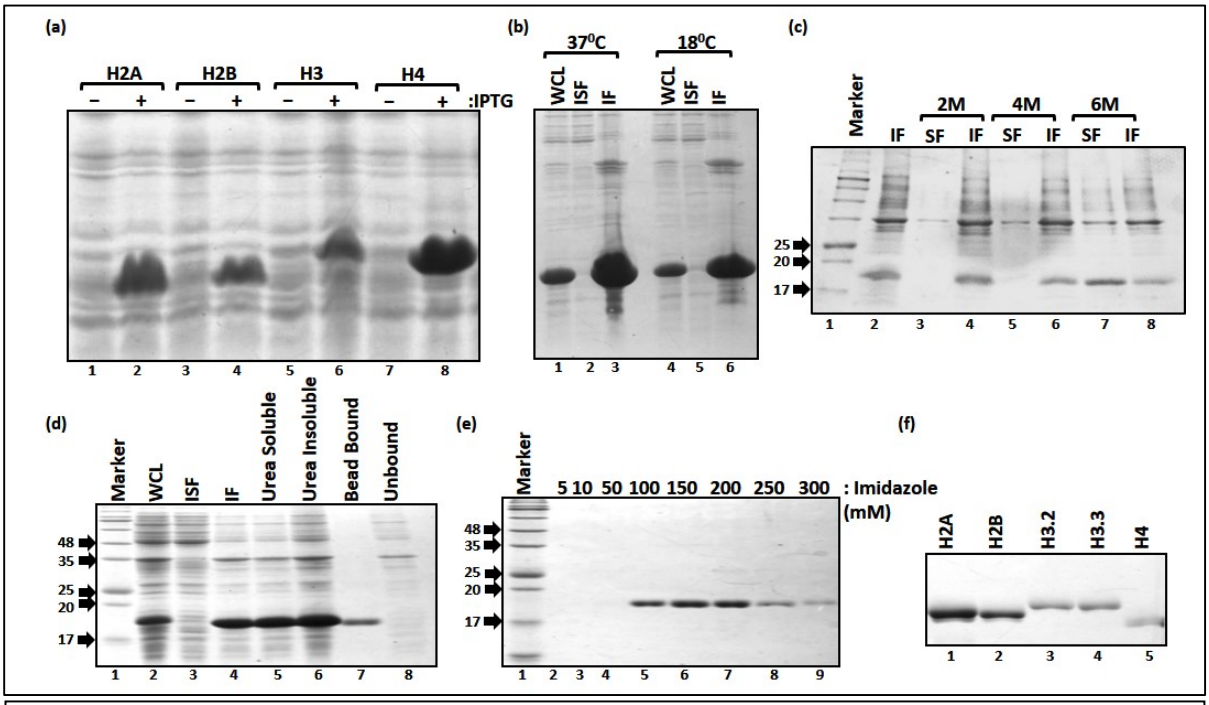



**Figure 4.5: Expression and purification of 6XHis tagged histones.** (a) SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21(DE3) pLysS harbouring the pET28a-histone constructs, induced with 0.2mM IPTG at $37^0$C for 2hrs. (b) SDS-PAGE analyses (18%, coomassie stained) of cell lysates of BL21 (DE3) pLysS harbouring the pET28a-H2A at $37^0$C, induced with 0.2mM IPTG for 2hrs and at $18^0$C, induced with 0.2mM IPTG overnight. The cells were harvested, lysed post induction and the different fractions were separated by centrifugation prior to loading on the gel. Similar results were obtained for all the core histones. WCL- Whole Cell Lysate, ISF- Induced Soluble Fraction, IF- Insoluble Fraction. (c) Solubilisation of histones from the IF with different urea concentrations. Lanes 2, 4, 6, 8 correspond to insoluble fractions and lanes 3, 5, 7 to fractions solubilized with 2M, 4M and 6M urea. (d) SDS-PAGE analyses (18%, coomassie stained) of different fractions obtained at different stages of Ni-NTA based purification of 6xHis tagged H2A. Similar results were obtained for all the core histones. (e) Elution profile of 6xHis tagged H2A on elution with different concentrations of imidazole. (f) SDS-PAGE (18%, coomassie stained) showing the purified core histones.

We used pET28a vector to express histones in BL21 (DE3) pLysS. The recombinant protein expression was robust [**Figure 4.5a**]. On analyzing different fractions, the histones were found to be completely part of insoluble fraction at both $37^0$C and $18^0$C [**Figure 4.5b**]. Similar results

were obtained for all the core histones. Unlike GST, the 6xHis tag can be used for affinity purification even in denaturing conditions, hence, different solubilisation strategies were employed. Urea was the preferred denaturant to solubilize proteins from the insoluble fraction as Gdn-Cl doesn't allow solubilized 6xHis proteins to bind to Ni-NTA column. Upon further titration with different urea concentrations (2M, 4M, 6M), 6M urea was found as the optimal denaturant concentration to solubilize these proteins [**Figure 4.5c**]. The solubilized proteins robustly bound to Ni-NTA beads [**Figure 4.5d**]. Elution with buffers containing increasing concentration of imidazole (5, 10, 50, 100, 150, 200, 250, 300mM) suggested that histones could be very efficiently eluted with 150mM imidazole concentration with almost none of non-specific proteins [**Figure 4.5e**]. Thus, all the core histones could be very efficiently and conveniently purified with one step purification process [**Figure 4.5f**].

### 4.3.4 Histone monomers are not properly folded in solution

To remove denaturant from the purified proteins, the eluted fraction was overnight dialyzed against 50mM phosphate buffer. However, often it led to the precipitation of majority of histones. Histones being highly basic proteins the precipitation could be circumvented by

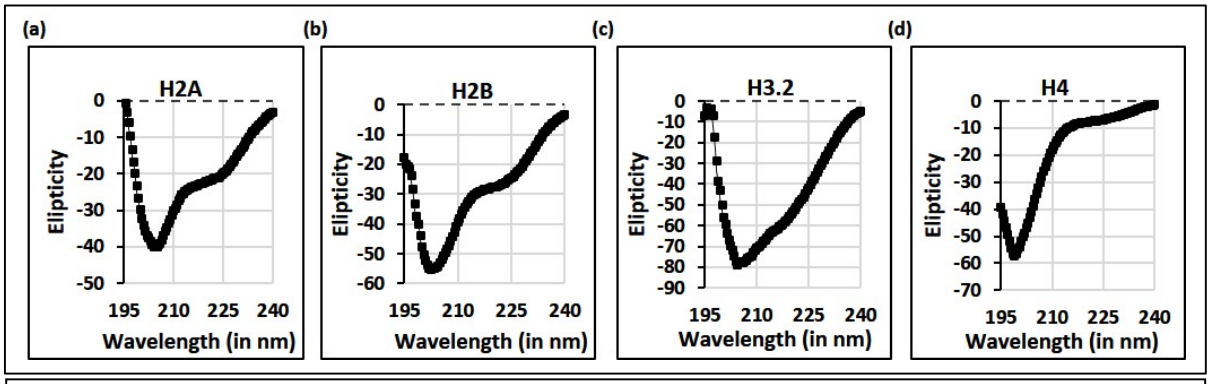



**Figure 4.6: Secondary structure analysis of histone monomer.** Far-UV CD spectra of core histone monomers obtained post affinity purification. Spectra of protein with ~10-20µM concentration was recorded. The lack of dip at 222nm and predominant dip at 205nm suggests that the monomers are in an unfolded state in the solution.

dialyzing histones against 10mM HCL irrespective of the concentration. For carrying out subsequent analysis the concentrated histones could be diluted in phosphate (pH 7.4) or tris

(pH 7.4) buffer without precipitation. Afterwards, to see whether the histone monomers attain their native confirmation on removal of the denaturant, circular dichroism (CD) spectra was taken. Spectra showed a dip at around 200nm for all the histone monomers suggesting that the monomers were not folded properly and were largely in random coil [**Figure 4.6a-d**].

### 4.3.5 Histone oligomers can be conveniently reconstituted from affinity purified histones

To reconstitute H2A-H2B dimers and H3-H4 tetramers as 1) they would be better substrates for carrying out *in vitro* enzymatic assays, 2) would be properly folded owing to the burial of hydrophobic residues and 3) would be easier to use for octamer formation.

In this regard, the protocol of rapid dilution of histones for forming oligomers was used[177]. Post rapid dilution protocol involves binding of the mix to heparin column and subsequent elution[177]. It was found that even H3-H4 tetramers could be *in vitro* reconstituted directly from 10mM HCl without urea denaturation by this method which was not previously reported. Also, dialyzing the rapid dilution mix against 2M NaCl leads to precipitation of excess histone monomers leaving the properly folded dimers in soluble form thus eliminating the requirement of heparin column. In ideal scenario, with usage of equimolar histones, there should be no or minimum precipitate post dialysis. The precipitate (if formed) were removed by centrifuging the sample at 10000g for 20mins. The dimers were purified by size-exclusion chromatography using HiLoad 16/60 Superdex-200 gel filtration column. The dimer eluted at a volume of 82mL [**Figure 4.7a (i)**]. The molecular weight of H2A and H2B being very similar (H2A- 14119.5Da, H2B- 13906.1Da) [**Figure 4.7a (ii)**], the peak fractions from gel filtration were resolved onto AUT-PAGE, which separates proteins on the basis of hydrophobicity, to confirm the formation

of heterodimers [**Figure 4.7a (iii)**]. The presence of two bands confirmed that the peak obtained from gel filtration column consists of two proteins and is thus a heterodimer.

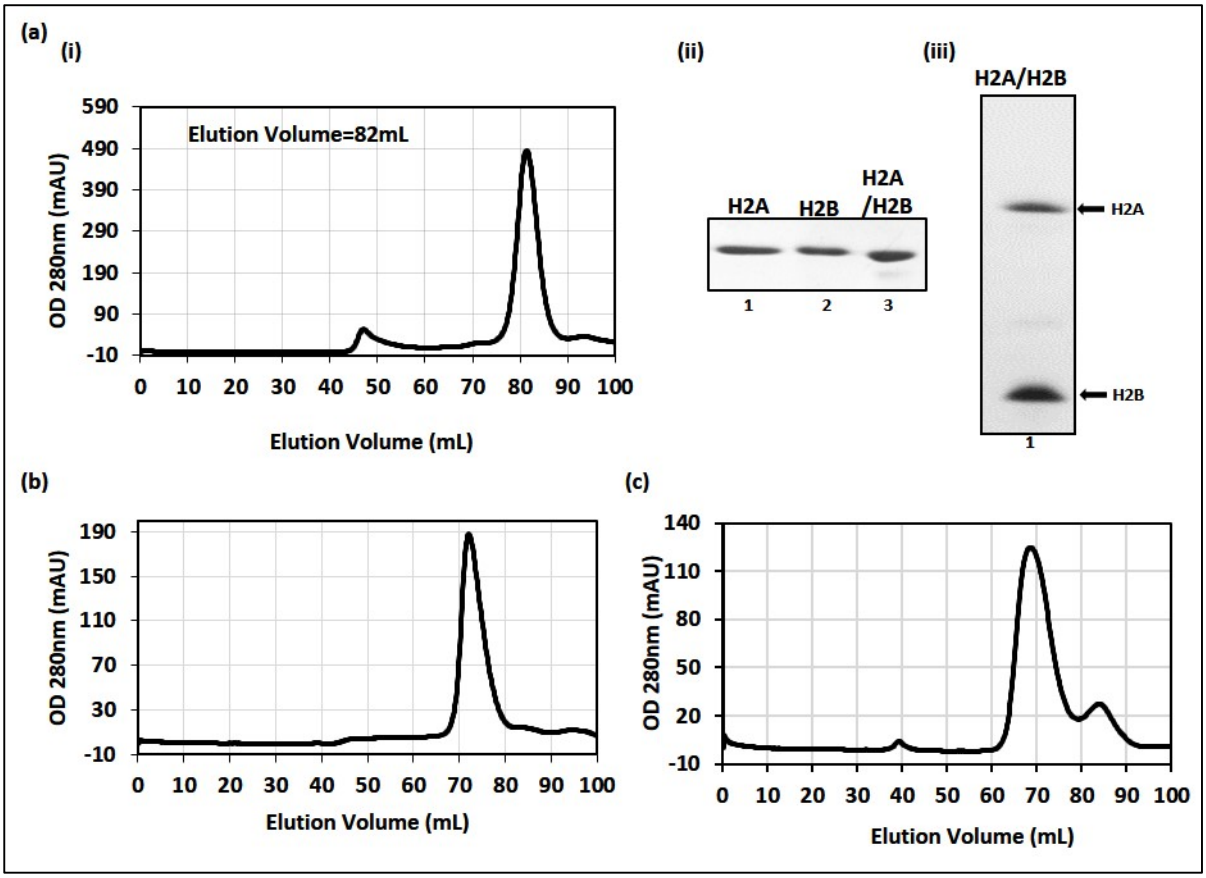



**Figure 4.7: Reconstitution of histone dimer, tetramer and octamer.** (a)(i) Elution profile of reconstituted H2A-H2B dimer with elution volume of 82ml on a Superdex 200 HiLoad 16/60 (GE) gel filtration column. (ii) SDS PAGE analysis of the peak fraction of dimer collected. (iii) AUT PAGE analysis of the peak fraction of dimer collected, showing distinctly resolved H2A and H2B. (b) Elution profile of reconstituted H3-H4 tetramer with elution volume of 72ml on a Superdex 200 HiLoad 16/60 (GE) gel filtration column. (c) Elution profile of reconstituted histone octamer with elution volume of 65ml on a Superdex 200 HiLoad 16/60 (GE) gel filtration column. The smaller shoulder peak is of excess H2A-H2B dimer.

### 4.3.6 H2A-H2B dimers and H3-H4 tetramers are better substrates for octamer reconstitution than histone monomers

The properly folded H2A-H2B dimers and H3-H4 tetramers were next used to give rise to histone octamer. Equimolar ratio of H2A-H2B dimer and H3-H4 tetramer were mixed and were dialysed against 2M NaCl. Under such high NaCl concentrations, the histone octameric core particle was formed which could be confirmed and purified by loading onto size exclusion

chromatography column with elution at an expected volume ~65ml[26] [**Figure 4.7c**]. The excess H2A-H2B dimer eluted at 82mL.

### 4.3.7 Histone oligomers are properly folded

Post reconstitution and purification, secondary structure validation of the H2A-H2B dimer and H3-H4 tetramer were carried out by far-UV CD. The spectra dip at 222 and 208nm, characteristic of helices confirmed that the histone oligomers, which has a high proportion (40%) of helical structure [**Figure 4.8a**], is properly folded [**Figure 4.8b (i)** for H2A-H2B

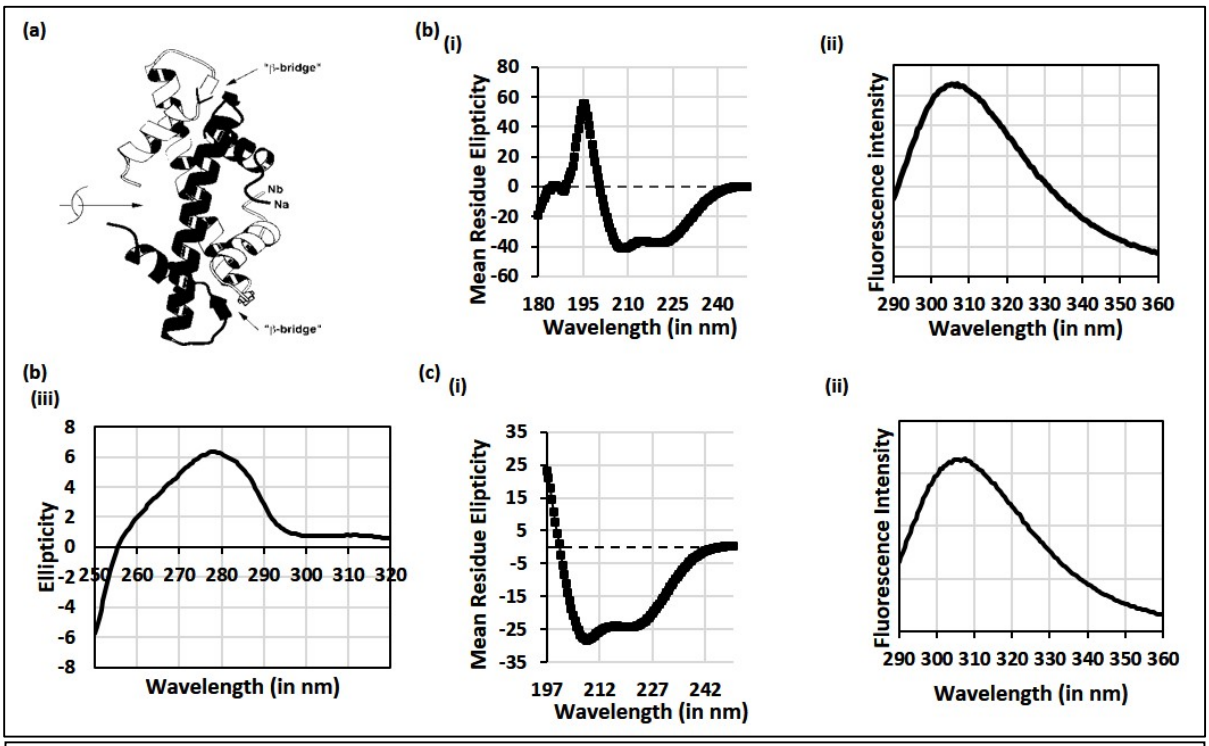



**Figure 4.8: Secondary structure analysis of histone dimer and tetramer.** (a) Cartoon depicting the structure of histone H2A-H2B dimer. (b) (i) Far UV CD spectra of reconstituted and purified H2A-H2B dimer. The dip at 208 and 222nm suggests that the dimer is folded in its native confirmation. The peak at 195nm further corroborates that the dimer has high helical content. (ii) Fluorescence spectral scan of reconstituted and purified H2A-H2B dimer. The peak at 305nm suggests that the dimer has proper tertiary structure with buried tyrosine residue (iii) Near UV CD spectra for tertiary structure validation of dimers. The bell shaped curve with peak near 278nm in the near-UV CD spectra is in agreement with signature tyrosine peak between 275 and 282nm (c) (i) Far UV CD spectra of reconstituted and purified H3-H4 tetramer. The dip at 208 and 222nm suggests that the dimer is folded in its native confirmation. (ii) Fluorescence spectral scan of H3-H4 tetramer. The peak at 305nm suggests that the tetramer has proper tertiary structure with buried tyrosine residue.

dimer, **c (i)** for H3-H4 tetramer]. The greater dip at 208nm compared to 222nm highlights the greater proportion (~60%) of randomly coiled structure of the dimer and tetramer. The tertiary

structure of the H2A-H2B dimer and H3-H4 tetramer was validated by near UV CD and fluorescence spectroscopy. H2A has three tyrosine residues at position 39, 50 and 57 and H2B have five tyrosine residues at position 37, 40, 42, 83 and 121. Histones do not contain tryptophan, hence, the fluorescence of histone proteins is majorly attributed to the tyrosine. The bell shaped curve with peak near 278nm in the near-UV CD spectra is in agreement with signature tyrosine peak between 275 and 282nm [**Figure 4.8b (iii)**]. The dimers exhibited a fluorescence emission maximum at 305nm [**Figure 4.8b (ii)**] which is expected for tyrosine. Similar pattern was observed for the H3-H4 tetramers [**Figure 4.8c (ii)**].

## 4.4. Discussion

We have described a very convenient and robust method to express, purify and reconstitute sub-nucleosomal complexes of histones for carrying out *in vitro* studies. Histones are often insoluble in bacteria. One way of circumventing this issue is by co-expressing histones, for example H2A and H2B together in tandem, in bacteria which at times makes them soluble[178]. However, this method does not ensure that all histones expressed will be soluble[178]. As an alternative approach, we tried to use the commonly used GST tag as it often solubilizes protein and also enables use of affinity purification. Affinity purification may yield ~90% pure proteins after only one round of purification and is technically less challenging than ion-exchange chromatography. GST tag could be used to solubilise and purify histones, however, due to the precipitation of monomeric histones post GST tag removal the strategy was not useful especially considering that only 10% of the protein is solubilised on GST tag incorporation. Nonetheless, the strategy can be useful to carry out some *in vitro* GST pull down assays. Another striking thing which emerged from our data is the propensity of proteases to non-specifically cleave histone proteins. FactorXa and Enterokinase, which have propensity to non-specifically cleave at basic residues. Hence, care is to be taken while choosing the protease to

remove tags from fusion proteins consisting of histones. TEV protease was found to be the best suited for this purpose. For producing recombinant histones in bulk amount, Ni-NTA affinity purification under denaturing conditions was very convenient and robust technique. In addition, 6xHis tag being small and as N-terminus of histones are not involved in formation of the complexes, the His-tagged proteins could be directly used in reconstitution experiments without removal of tag. Further, we documented that the reconstituted complexes are properly folded which makes them useful for carrying out *in vitro* studies pertaining to histone variants/mutants and modifications. Notably, histone monomers are not properly folded in solution as suggested by the CD spectra. This might be attributed to the fact that the alpha 2 helix of histones is largely hydrophobic in nature.

Histones lack tryptophan and contain very less tyrosines and phenylalanines. The scarcity of aromatic residues means that quantification methods based on the presence of these residues like Folin-Ciocalteu and absorbance at 280nm are error prone for histones. During *in vitro* reconstitution experiments the histones have to be added at an equimolar ratio and any error in quantification can lead to large amount of precipitation. In this regard using reconstituted H2A-H2B dimers and H3-H4 tetramers for histone octamer reconstitution is more convenient rather than adding all the four histone monomers stoichiometrically. In our experience using the reconstituted dimers and tetramers for octamerisation gives more consistent results than using the individual monomers. The histone octamer elutes at 65mL. The H3-H4 tetramer elutes at 72mL, whereas the H2A-H2B dimer elutes at 84mL. Hence, it is advised to use a little excess of H2A-H2B dimer during octamer reconstitution as they are easier to distinguish from the octamer peak during gel filtration as compared to H3-H4 tetramer peak.

## 4.5. Conclusions

We have described a protocol in which histones can be purified and its oligomeric structures reconstituted in very few steps in three days. Histones can be purified in bulk by affinity purification and then via use of rapid dilution method followed by dialysis, H2A-H2B dimers and H3-H4 tetramers can easily be reconstituted. For histone octamer formation, the reconstituted and purified dimers and tetramers are mixed proportionately followed by dialysis. These reconstituted and purified octamers which have their native structure are better substrates for use in *in vitro* assays. We also showed that adding GST tag partially solubilises the otherwise insoluble histones which can be used for *in vitro* assays. Further, histones have a propensity to be non-specifically cleaved by proteases and TEV protease should be opted for while working with histones.

# V. PHYSIOLOGICAL EFFECT OF H2A.1 ISOFORM OVEREXPRESSION

## Chapter 5

*Work carried out to investigate the possible non-redundant functional effect of H2A.1 isoform on cell physiology*

## 5.1. Introduction

The canonical histone genes are synthesized during the S-phase and to meet up with their high demand during DNA replication, their encoding genes are present in cluster. For instance, there are three clusters of canonical histone genes present in humans at chromosome number 1 and 6. Interestingly, differences in the primary sequence is observed amongst the histone H2A

```
Hist1h2aa          MSGRAKQGGKARAKAKSRSFRAGLQFPVGRVHRLLRQGNYAERIGAGTPVYLAAVLEYLT
Hist2h2ab          MSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLT
Hist2h2ac          MSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLT
Hist2h2aa3         MSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLT
Hist1h2af          MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLYKGNYSERVGASAPVYLAAVLEYLT
Hist3h2a           MSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAAVLEYLT
Hist1h2ak          MTGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAAVLEYLT
Hist1h2ac          MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAAVLEYLT
H2ae-like          MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAAVLEYLT
Hist1h2aiVx2       MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAAVLEYLT
Hist1h2aiVx1       MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAAVLEYLT
Hist1h2an          MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAAVLEYLT
Hist1h2ah          MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLT
                   *:**.***********:** *************** :***:**:**.:***:********

Hist1h2aa          AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
Hist2h2ab          AEILELAGNAARDNKKTRIIPRHLQLAVRNDEELNKLLGGVTIAQGGVLPNIQAVLLPKK
Hist2h2ac          AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKK
Hist2h2aa3         AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKK
Hist1h2af          AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
Hist3h2a           AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
Hist1h2ak          AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
Hist1h2ac          AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
H2ae-like          AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
Hist1h2aiVx2       AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
Hist1h2aiVx1       AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
Hist1h2an          AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
Hist1h2ah          AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK
                   **************************:*********** ********************

Hist1h2aa          TESHHKSQTK
Hist2h2ab          TDSHKPGKNK
Hist2h2ac          TESHKAKSK-
Hist2h2aa3         TESHHKAKGK
Hist1h2af          TESHHKSKGK
Hist3h2a           TESHHKAKGK
Hist1h2ak          TESHHKAKGK
Hist1h2ac          TESHHKAKGK
H2ae-like          TESHHKAKGK
Hist1h2aiVx2       TESHHKAKGK
Hist1h2aiVx1       TESHHKAKGK
Hist1h2an          TESHHKAKGK
Hist1h2ah          TESHHKAKGK
                   *:**:  .
```

**Figure 5.1: Protein sequence alignment of rat H2A isoforms.**

proteins encoded by these genes. These are often termed as histone isoforms[2]. In humans, there are 16 genes coding for H2A giving rise to 12 isoforms[23]. Likewise, there are 13 genes which code for 9 isoforms in rats (most are having the "predicted" status) [**Figure 5.1**].

H2A isoforms have been reported to be differentially expressed in a variety of physiological states. For instance, the proportion of two isoforms H2A.1 (encoded by HIST1H2AH, accession number: NM_001315492.1) and H2A.2 (encoded by HIST2H2AA3, accession number: NM_001315493.1) has been shown to decrease in the course of development and differentiation of rat brain cortical neurons[129] and during *in vitro* differentiation and aging of fibroblasts[130]. During the course of liver development as well, similar changes were observed[5]. Earlier studies from our lab have shown the differential expression of these two major H2A isoforms in rat hepatocellular carcinoma model system[3]. In humans, there is no isoform of H2A having the identical sequence as rat H2A.1. The most similar isoform in terms of amino acid sequence is H2A1C (encoded by HIST1H2AC), which differs from H2A.1 at the identity of the 16$^{th}$ residue where it has serine instead of threonine [**Figure 5.2**]. H2A.2 is completely conserved amongst vertebrates. Notably, the expression level of H2A1C in humans have been reported to alter in pathological states. Expression of H2A1C isoform was reported to be downregulated in chronic lymphocytic leukaemia (CLL) and gall bladder cancer cell line[2,134]. Interestingly, later on with a larger cohort of samples, H2A1C expression was conversely reported to be up regulated in CLL[179]. Also, H2A1C was found to be upregulated in non-small cell lung carcinoma[180]. The levels of H2A1C, specifically, have been reported to change in other diseases including human papillomaviruses hyperplasia, AIDS and multiple sclerosis[133,181]. Collectively, these reports suggest that the expression of the H2A.1 isoform in different states is intricately regulated in the cells. However, whether the observed changes are a consequence of the given state or H2A.1/H2A1C contributes to the attainment of such states

has not been addressed in depth. One such report that aims to address this issue, showed that specific knockdown of H2A1C leads to marked increase in cell proliferation. This effect is not observed on knocking down the other abundant isoforms like H2A 1B/E[2]. However, how the histone isoforms impart their non-redundant effects remains unclear. This becomes even more important to address as the levels of H2A1C are upregulated in many cancers, as discussed above, which is inconsistent with its anti-proliferative function exhibited in the study conducted by Singh *et. al*[2].

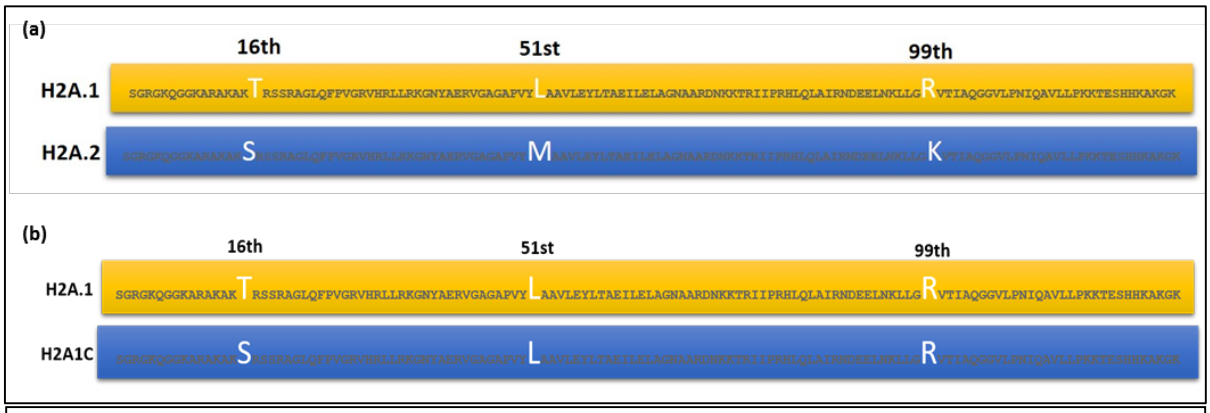



**Figure 5.2: Protein sequence alignment of H2A isoforms.** (a) Alignment of rat H2A.1 and H2A.2. (b) Alignment of rat H2A.1 and human H2A1C.

In the present study, we show that the expression level of H2A.1/H2A1C isoform markedly alters in different tissues as well, in addition to being generally upregulated in many cancer cell lines. We further investigate the effect of overexpression of H2A.1 on cell proliferation and migration.

## 5.2. Methods

### *5.2.1 Animal handling and experiments*

All the experiments were performed on Sprague-Dawley rats (spp. *Rattus norvegicus*). Experiments were performed after approval of Institute Animal Ethics, ACTREC. Procedures for maintenance and conducting experiments were as per Committee for the Purpose of Control

and Supervision on Experiments on Animals (CPCSEA), India, standards. Animals were euthanized in carbon dioxide chamber at the end of each experiment. The sacrificed animals were collected in plastic bags and stored in animal house at -20$^0$C. These bags were collected by municipal garbage vans. Protocol to induce sequential stages liver carcinogenesis is as previously described[3].

For liver tissue transplantation, small piece of ~3mm$^2$ size liver and tumour tissue collected from the donor mice were washed in RPMI medium. Small skin incision was made at the flank region of NOD-SCID mice and liver/tumour tissue was implanted aseptically under the subcutis. For histology analysis liver tissues were excised, washed with ice-cold saline and either fixed in formalin for hematoxylin and eosin (H&E) staining or snap-frozen in liquid nitrogen. Liver histology was evaluated on the basis of vacuolar degeneration, fibrosis, steatosis and inflammation. Further, IHC for PCNA and CycinD1 was done using VECTASTAIN® ABC kit.

Partial hepatectomy was carried out as earlier described[182] with age normalized (22-24 weeks of age) animals. All surgeries were performed under aseptic conditions under general anaesthesia by ketamine. In brief, ketamine was injected at a dose of 10mg/ml per kg of body weight intramuscularly. Each animal's abdomen was shaved and disinfected. A horizontal incision was made, approximately 2-3cm long, just below the xiphoid cartilage. Left lateral and median lobes of liver (50-70% liver) were resected and muscle and skin layers were closed. Animals were provided with an external source of heat during recovery and were maintained post operatively under optimal conditions. Animals were sacrificed at different time points (8, 12, 24, 36hrs) after partial hepatectomy and their liver was excised. The tissue were stored at -80$^0$C. For subsequent analysis histones were isolated from the tissues as described in later. The

histones were resolved on AUT-PAGE and silver staining was performed to visualise the resolved histones.

### *5.2.2 Transmission electron microscopy*

Protocol for specimen preparation for transmission electron microscopy is a previously described[183] with slight modifications. The piece of tissue (normal and tumor) to be analysed was excised. The tissue was cut into thin slices and allowed to harden in the fixative 3% glutaraldehyde for about 10min before further cutting. The tissue was further fixed for 1-2hrs at room temperature. The tissue was post-fixed for 1 hour in 1% osmium tetroxide in distilled water. Further, following incubations were carried out post rinsing 3 times (5min per rinse) in distilled water: 50% ethanol - 10min, 70% ethanol - 10min, 80% ethanol – 10min, 95% ethanol with 2 changes within 10min, 100% ethanol from a newly opened bottle - 3 changes within 15min, propylene oxide - 10min, 1:1 propylene oxide/resin - minimum of 1hr, 1:2 propylene oxide/resin - 1 hour or overnight on a rotator, 100% Epon or Spurr resin - 2 changes over 2-6hrs. Grids were contrasted by use of alcoholic uranyl acetate for 1min and lead citrate for 30sec. The grids were then observed under a Carl Zeiss LIBRA120 EFTEM and images were captured.

### *5.2.3 Total RNA preparation*

All solutions were prepared in 0.1% diethylpyrocarbonate (DEPC) treated water for isolation of RNA. Glassware was baked at $300^0$C for 4hr and compatible plasticware was rinsed with chloroform. Nitrile gloves were used to prevent RNase contamination.

RNA was isolated from liver tissue by guanidine method for total RNA preparation[155]. Frozen liver tissue (0.1g) was powdered in pre-chilled mortar and pestle. The powder was suspended in 1ml denaturing solution (4M guanidine thiocyanate, 25mM sodium citrate, 0.5% sarkosyl

and 0.1M β-ME) in microcentrifuge tube. To this suspension, 0.1ml of 2M sodium acetate pH 4, 1ml of water saturated phenol and 0.2ml of chloroform/isoamyl alcohol (49:1) were added. The suspension was thoroughly mixed and incubated at 0-4$^0$C for 15min. The suspension was then centrifuged for 20min at 10000g (15000rpm in Rota 4R), 4$^0$C. The upper aqueous phase was transferred to fresh microcentrifuge tube.

RNA was precipitated by adding equal volume of 100% isopropanol and incubating for 30min at -20$^0$C. The precipitate was centrifuged for 10min at 10000g (15000rpm in Rota 4R) and 4$^0$C. RNA pellet was dissolved in 0.3ml denaturing solution. RNA was precipitated with 0.3ml of 100% isopropanol for 30min at -20$^0$C and centrifuged for 10min at 10000g (15000rpm in Rota4R) and 4$^0$C. RNA pellet was resuspended in 75% ethanol, vortexed, incubated for 10-15min at room temperature and again centrifuged for 5min at 10000g (15000rpm in Rota4R) and 4$^0$C. RNA pellet was dried for 5min and dissolved in DEPC treated water. RNA was stored at -70$^0$C until required. RNA was quantitated by diluting 5µl in 1ml alkaline water (1mM $Na_2HPO_4$) and reading at A260. Quality of RNA was confirmed by A260/A280 (1.9-2.0), A260/A230 (2.0-2.2) and agarose formaldehyde gel electrophoresis.

### *5.2.4 Cell line maintenance*

CL44, a rat liver non-tumorigenic cell line and CL38, a rat liver tumorigenic cell-line, *in vivo* transformed by NDEA carcinogenesis, was a kind gift from Dr. H. M. Rabes (University of Munich, Germany). Sterile techniques were maintained when working with cell-line. Antibiotics were added to culture medium. Reusable glassware, plasticware and solutions were sterilized by suitable means prior to use. Cell culture work was carried out in laminar flow hood. CL44 and CL38 cells were maintained in Eagle's minimum essential medium supplemented with 10% v/v fetal bovine serum (Invitrogen, cat#10270106) and antibiotic solution (HiMedia, cat#A002) in a humidified 37$^0$C, 5% $CO_2$ incubator. Culture grown to

confluent density was subjected to trypsinization. Medium was removed from culture with a sterile pipette. Adhered cells were washed with PBS, pH 7.2-7.4 and 1ml trypsin/EDTA (0.25% w/v trypsin, 0.2% EDTA in PBS) solution was added. Plates were incubated at $37^0$C until cells detached from surface. Detached cells were suspended in 1ml complete media. Viable cells were counted as described below and plated into fresh culture dishes (~2 X $10^4$ cells/ml).

Number of viable cells were determined by adding 0.5ml of trypan blue (0.4% w/v in PBS), 0.4ml PBS to 0.1ml cell suspension in microcentrifuge tube. Cells were loaded on haemocytometer and viable (unstained) cells were counted. Cells were counted using formula:

Cells/ml = average number of cells per WBC chamber x dilution factor (10) x $10^4$

### 5.2.5 *Synchronisation of cells in G0-G1 phase of the cell cycle*

Exponentially growing CL38 cells were trypsinized and counted as described earlier and plated (2X$10^5$ cells per 100mm X 15mm culture dish) in complete MEM. At 30-40% confluent density, complete MEM was replaced by serum free MEM and cells were incubated for 24, 48, 72 and 96hrs. After completion of incubation, CL38 cells were trypsinized, suspended in 1ml MEM and centrifuged for 10min at 200g (2000rpm in Rota4R). Cell pellet was either used for extraction of histones or processed for flow cytometry analysis.

### 5.2.6 *Flow-cytometry analysis*

Cells were prepared for flow-cytometry analysis as described earlier[184]. Cells were centrifuged for 10min at ~200g (2000rpm in Rota4R) and resuspended in 0.1ml of PBS. Cells were transferred to 0.9ml of 70% ethanol v/v in PBS and stored at -20°C for ≥ 2hr until required. Ethanol-suspended cells were centrifuged for 5min at 200g (2000rpm in Rota4R). Cell pellet was resuspended in 1ml PBS and recentrifuged at 200g (2000rpm in Rota4R). The cell pellet

was resuspended in 1ml PI (1μg/ml)/Triton X100 (0/1%) staining solution with RNase and incubated for 15min at $37^0$C.

Flow cytometer FACS Calibur (Becton Dickinson) was adjusted for excitation with blue light and detection of PI emission at red wavelength. Cell fluorescence was measured in flow cytometer using pulse width-pulse area signal to gate out the cell-doublets and analysed using DNA content frequency histogram deconvolution software (ModFit LT v2.0 for Mac OS 8.6).

### 5.2.7 *MTT assay*

Cell viability was quantified by its ability to reduce tetrazolium salt 3-(4,5-dimethylthiazole-2Ƴ)-2,5-diphenyl tetrasodium bromide (MTT) to coloured formazan products as described earlier[185]. MTT reagent (5mg/ml in PBS) was added to the cells at 1/10th volume of the medium to stain only viable cells and incubated at $37^0$C for 4hrs. MTT solubilisation buffer (0.01M HCl, 10% SDS) of two-fold volume was added to cells, followed by incubation in the dark at $37^0$C for 24hrs. The absorbance was measured at 570nm with Spectrostar Nano-Biotek, Lab Tech plate reader. Cell viability was expressed as the percentage of absorbance obtained in control cultures.

### 5.2.8 *Colony formation assay*

Clonogenic assay was performed as described earlier[186]. The cells (n=1000) were plated in triplicate in 60mm tissue culture plates and they were allowed to grow in a monolayer for 14 days. Cells were incubated in complete culture medium, with media changes after every 2-3 days. After 14 days, the cells were fixed with 4% paraformaldehyde for 1hr. The colonies were stained with 0.5% crystal violet (0.5% in 70% ethanol) for 1hr at room temperature, rinsed and air-dried. Surviving colonies with more than 50 cells were counted and images were captured

using a high-resolution Nikon D70 camera (Nikon, Tokyo, Japan). For quantification of the size of the colonies ImageJ was used.

### 5.2.9 *Histone Isolation*

a) *From tissue:* First, nuclei were isolated from liver tissues by sucrose density gradient centrifugation[187] and then, histones were extracted from isolated nuclei by low concentration acid extraction method. Liver tissues (~3g) excised from euthanized rats were rinsed in excess of ice-cold 'nuclear buffer A' (15mM Tris-Cl pH7.5, 60mM KCl, 15mM NaCl, 2mM EDTA, 0.5mM EGTA, 0.34M sucrose, 0.15mM β-ME, 0.15mM spermine and 0.5mM spermidine). Liver tissues were minced in 30ml fresh nuclear buffer A and homogenized in motor driven glass-Teflon homogenizer. Homogenate was filtered through 8 layers of cheesecloth to remove large solid impurities. The filtrate was centrifuged for 15min at 1500g (5000rpm in Rota 6R-V/Fm, Plasto Crafts) and $4^0$C. The crude nuclear pellet obtained was resuspended in fresh 30ml of nuclear buffer A. The suspension was layered on 10ml of 'nuclear buffer B' (15mM Tris-Cl pH7.5, 60mM KCl, 15mM NaCl, 0.1mM EDTA, 0.1mM EGTA, 1.8M sucrose, 0.15mM β-ME, 0.15mM spermine and 0.5mM spermidine) and centrifuged for 90min at 100000g (26000rpm in AH629 swinging bucket rotor) and $4^0$C. The nuclear pellet was resuspended in 3ml nuclear buffer A. The nuclei were either used for histone extraction as earlier described[188].

Nuclei were resuspended in 0.2M $H_2SO_4$ at a concentration of ~2mg DNA/ml and incubated overnight at $4^0$C. For determining concentration, the nuclei were ruptured in 2M Urea- 5M NaCl and the absorbance was measured at 260nm. The suspension was centrifuged for 10min at ~12000g (15000rpm in Rota4R V/Fm, Plasto Crafts) and $4^0$C. The supernatant comprising of acid soluble nuclear proteins was transferred to fresh clean microcentrifuge tube. Nuclear proteins (predominantly histones) in the supernatant were precipitated by adding 4 volumes of acetone and stored overnight at $-20^0$C. The mixture was centrifuged for 10min at ~12000g and

$4^0$C as mentioned before to pellet the proteins. The pellet was washed once in chilled acidified acetone (0.05M HCl in 100% acetone) and once in chilled 100% acetone. Protein pellet was dried in vacuum centrifuge for 15min. The pellet was resuspended in 0.1% β-ME in water at a concentration of ~5µg/µl and stored at -$20^0$C. Histones resolved on 18% SDS-polyacrylamide gel were transferred to PVDF membrane, probed with site-specific modified histones antibodies and signals were detected by using ECL plus detection kit (Millipore; Catalogue no. WBKLS0500).

b) *From cell line:* First nuclei were isolated from cells. Cell pellet was resuspended in 0.1ml PBS in a microcentrifuge tube. To this suspension 0.9ml lysis solution (250mM sucrose, 50mM Tris-Cl pH7.5, 25mM KCl, 5mM $MgCl_2$, 0.2mM PMSF, 50mM $NaHSO_3$, 45mM sodium butyrate, 10mM β-ME and 0.2% v/v Triton X-100) was added. Tube was inverted several times and centrifuged for 15min at 800g (4000rpm in Rota4R) and $4^0$C. The nuclear pellet obtained was subjected to histone extraction by low concentration acid extraction method by adding 0.3ml of 0.2M $H_2SO_4$ as described previously.

### *5.2.10 Analysis of Histones*

Extracted histones were estimated and subjected to SDS-PAGE for western blot analysis or processed further for AUT-PAGE.

a) *Protein estimation*: Histone concentrations in various samples were determined by Bradford method of protein estimation[189]. Protein standards were prepared containing a range of 0 to 100µg of Bovine Serum Albumin in 5ml of 1X Bradford reagent. Histone samples were also prepared similarly. Samples were vortexed and incubated at room temperature for 5min. Absorbance was measured at 595nm and the blank was adjusted. Histone samples were estimated for protein concentration by plotting standard curve.

b) *SDS-PAGE of extracted histones*: Same as described in chapter III (page number 82). After the run was complete, the power supply was then disconnected and gel was subjected to Coomassie staining for protein visualization or subjected to western blot analysis.

c) *Coomassie staining:* After the electrophoresis gel was transferred to tray containing Coomassie Brilliant Blue R-250 (CBBR) staining solution (0.1% w/v CBBR, 50% methanol and 10% acetic acid in water). The gel was stained for ~3hrs on shaker. Gel was then transferred to destaining solution (50% methanol and 10% acetic acid in water). The gel was destained on shaker with several changes in destaining solution until protein bands appeared on clear gel. The gel was documented as .tiff image and gel bands were quantified by densitometry using ImageJ software (v1.42q, NIH). Individual band intensity levels were normalized to total histone intensity on the gel and analysed using GraphPad Prism software.

### *5.2.11 Western Blot Analysis*

Levels of particular proteins or their post-translational modifications were assessed by western blot analysis using the appropriate antibody. See *Annexure I* in *Appendix* for the details of the antibodies used in the study.

a) *Electroblotting from SDS-PAGE:* Histones (5-10µg) were electroblotted from SDS-PAGE gels to PVDF membranes for western blot analysis. The transfer tank of electroblotting apparatus (Trans-Blot Cell, Bio-Rad) was filled with 1x transfer buffer (see *Annexure IV* for composition). PVDF membrane was activated in 100% methanol for 5sec. The activated membrane and SDS-PAGE gel were equilibrated in 1x transfer buffer. The gel membrane transfer sandwich was prepared and inserted into the transfer tank with gel on cathode side and membrane on anode side. Transfer was conducted at a constant current of 300mA for 200min. Proteins transferred onto the membrane were detected by staining with Ponceau S (0.5% w/v Ponceau S in 1% v/v acetic acid) and destaining with several changes of water.

b) *Immunoblot detection:* Proteins transferred onto PVDF membrane were probed with antibodies. In general, membrane with transferred proteins was incubated in 'blocking buffer' i.e. 5% BSA in Tween20/Tris-buffered saline (TTBS, 100mM Tris-Cl pH7.5, 0.9% w/v NaCl and 0.1% v/v Tween20) for 1hr at room temperature on orbital shaker. Blocking buffer was then replaced by recommended dilutions of primary antibodies in TTBS and incubated for 1hr at room temperature in orbital shaker. The membrane was vigorously washed four times with TTBS for 15min each at room temperature. Further the membrane was incubated in recommended concentrations of HRPO labelled secondary antibodies in TTBS for 1hr at room temperature on orbital shaker. The membrane was again washed vigorously four times with TTBS at room temperature and developed using Immobilon Western (Millipore, cat#P90719). The membrane was exposed to X-ray film in dark room and developed using Optimax X-ray film processor (Protec). Band intensities were analysed using ImageJ and GraphPad prism software.

### *5.2.12 AUT-PAGE Analysis*

AUT-PAGE analysis of histones was done as described earlier[190]. For preparation of AUT-PAGE (15cm long) gel, glass-plate sandwich was assembled using 0.15cm thick spacers. Separating gel solution was prepared and poured into glass plate sandwich and was allowed to polymerize in light. Stacking gel solution was then prepared, poured and polymerized into the glass plate sandwich. See *Annexure IV* in *Appendix* for details of chemicals required. The sandwich was attached to electrophoresis chamber and filled with electrophoresis buffer (1M acetic acid and 0.1M glycine). The gel was pre-electrophoresed for 2hrs at constant voltage of 200V. The buffer was then replaced and histone samples (see Annexure IV in Appendix for sample preparation) were loaded in the wells. AUT gel was washed twice in ten gel volumes of buffer A (0.05M acetic acid, 0.5% w/v SDS), 30min each and once with modified buffer of

O'Farrell (0.0625M Tris-Cl pH 6.8, 2.3% w/v SDS, 1% β-ME). The gel was treated with three washes of 50% methanol of 1hr each for rapid fixation of proteins. For silver staining the protocol described earlier was used[191]. The gel was immersed in ten gel volumes of freshly prepared ammoniacal silver nitrate solution for 30min. The gel was washed thrice with water each time for 2min. The gel was then developed by using developing solution (0.01% w/v citric acid monohydrate and 0.1% w/v formaldehyde). The development was stopped when protein spots appeared on clear gel using stopping solution (40% methanol and 10% acetic acid).

## 5.3. Results

### 5.3.1 Nitrosodiethylamine leads to stable transformation of liver cells with upregulation of H2A.1 expression

Nitrosodiethylamine (NDEA) administration for four months led to development of HCC as confirmed by H&E staining, with cancer cells exhibiting evident atypia and large nuclei [**Figure 5.3a (i)**]. Nuclear contours are irregular and large nucleoli are notable. In addition, electron microscopy revealed several indentations giving rise to a more irregular shaped nucleus [**Figure 5.3a (ii)**]. This was further corroborated by high Cyclin D1 [**Figure 5.3a (iii)**] and PCNA [**Figure 5.3a (iv)**] expression in tumour sections by immuno-histochemistry analysis and also their marked upregulation at the transcript level [**Figure 5.3b**]. We wanted to see whether the increased expression of H2A.1 in hepatocellular carcinoma (HCC), which we have previously demonstrated[3], persists even without the influence of the chemical carcinogen nitrosodiethylamine (NDEA). To address this, $3mm^2$ tissue was excised from the tumour developed in the liver of Sprague-Dawley rat after NDEA administration and subcutaneously implanted into a NOD-SCID mice. Two weeks' post-implantation tumours of ~3cm size

developed in only in the mice with tumour tissue [**Figure 5.3c (ii)**]. These results demonstrate the stable transformation of liver cells by administration of NDEA.

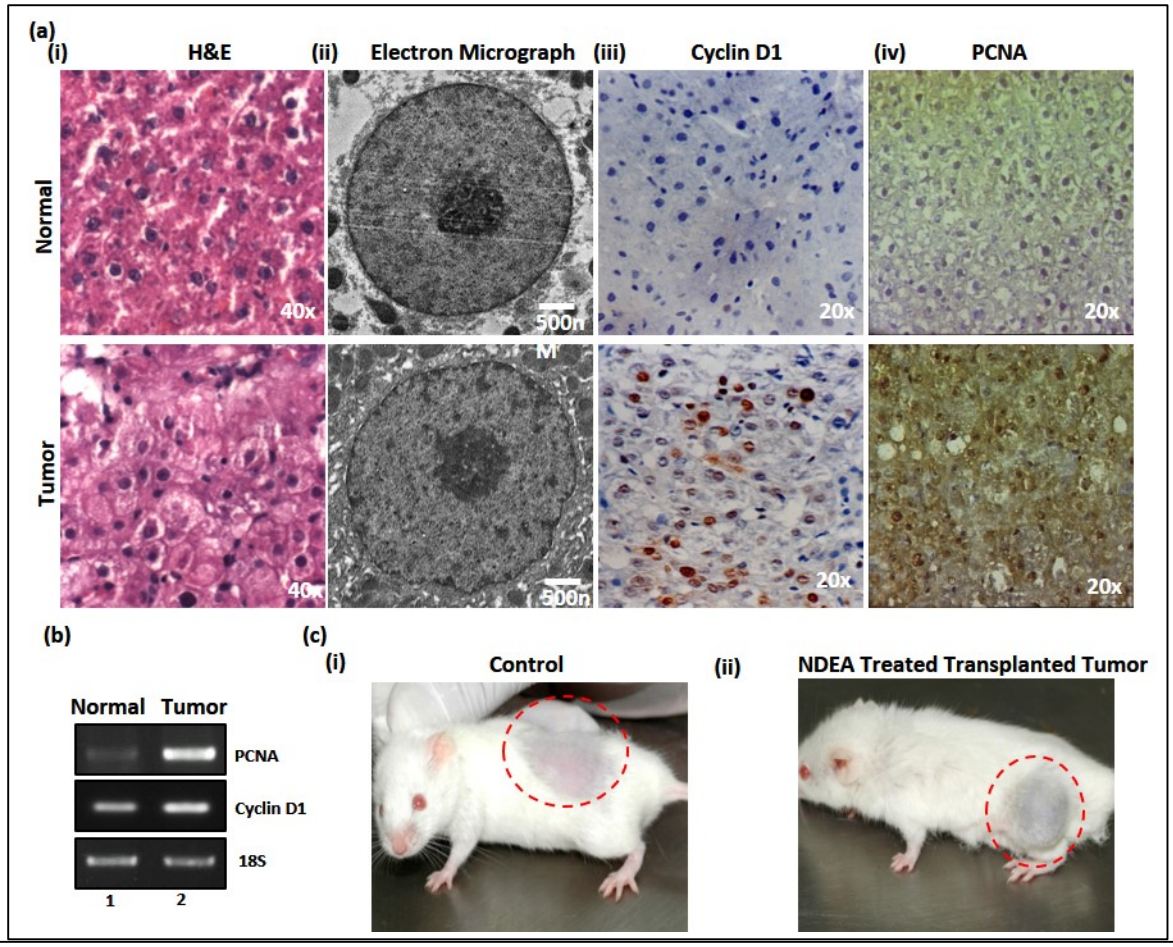



**Figure 5.3: NDEA mediated liver cell transformation.** (a) (i) Comparative H&E, (ii) Electronmicrographs, (iii) IHC of Cyclin D1 and (iv) IHC of PCNA of normal and four month NDEA treated liver tissues suggesting development of HCC. (b) (i) Semi-quantitative RT-PCR of Cyclin D1 and PCNA of the normal and tumor (4 month NDEA treated) tissues. 18S was used as loading control (c) Pictures depicting development of tumor after two weeks of subcutaneously implanting liver tissue of (i) normal and (ii) NDEA-treated rat in NOD- SCID mice. Tumor is observed only in the mice in which liver tumor tissue developed in NDEA-treated rat was used for transplantation. NDEA- N-nitrosodiethylamine, H and E- hematoxylin and eosin, HCC- hepatocellular carcinoma, IHC- Immunohistochemistry.

Histones isolated from normal and dysplastic liver as well as tumour and transplantable HCC were resolved on AUT-PAGE [**Figure 5.4a**]. In dysplastic liver, significant changes in H2A isoform profile was not observed. However, H2A.1 was significantly upregulated in tumour and transplantable HCC compared to normal liver [**Figure 5.4b**]. This suggests that H2A.1 upregulation is a stable alteration that occurs during the process of transformation of liver cells.

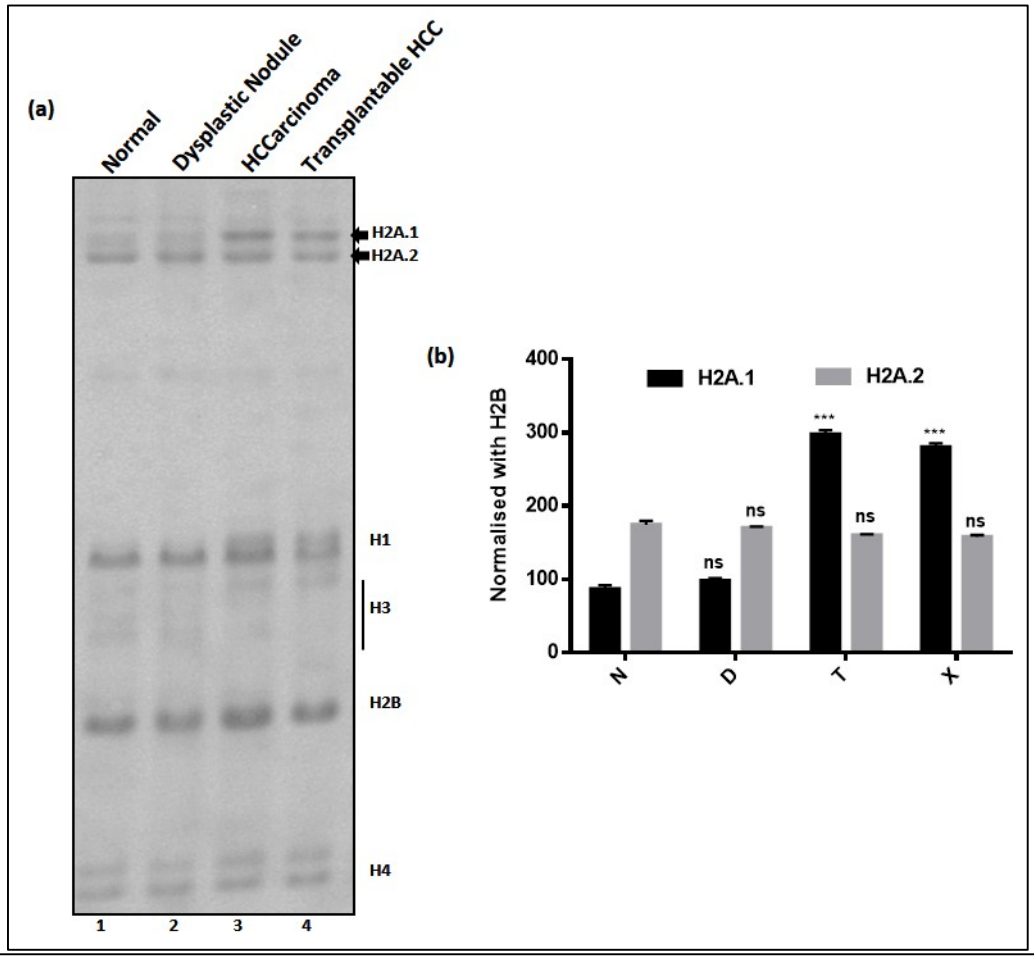



**Figure 5.4: H2A.1 and H2A.2 expression level in transplantable HCC.** (a) AUT-PAGE analysis of histones isolated from xenograft tumour (marked Transplantable HCC) along with histones from different stages of NDEA-induced liver cancer in rat. (b) Densitometric analysis using GelAnalyser software of the bands corresponding to H2A.1 and H2A.2 in the depicted AUT-PAGE. The bands of H2A.1 and H2A.2 were normalised with band of H2B. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$). N-Normal, D-Dysplastic, T-HCCarcinoma, X- Transplantable HCC. The significance was calculated with respect to Normal.

## 5.3.2 H2A.1/H2A1C expression level varies in human cancer cell lines and amongst different rat normal tissues

H2A1C in humans is the closest counterpart of H2A.1 in rat, differing in only the S16T substitution. The altered expression level of H2A1C has been reported to alter in certain human cancers[2,4,134,179,180]. Our observations in rat hepatocellular carcinoma prompted us to investigate the expression level of H2A1C and H2A.2 isoforms in human transformed cell lines of liver (HepG2). We also included cell lines of colon (A431) and stomach (KATOIII, AGS) origin

and their immortalised untransformed counterparts, that is, HHL5 (liver), HaCaT (skin) and HFE145 (stomach) in our study [**Figure 5.5a**]. H2A1C was found to be significantly overexpressed in MCF7 [**Figure 5.5a (iii)**], consistent with previously published report[4] and in HepG2 and A431 [**Figure 5.5a (i)**]. We did not find significant changes in the expression of the two H2A isoforms in transformed cell lines of stomach [**Figure 5.5a (ii), (iv)**].

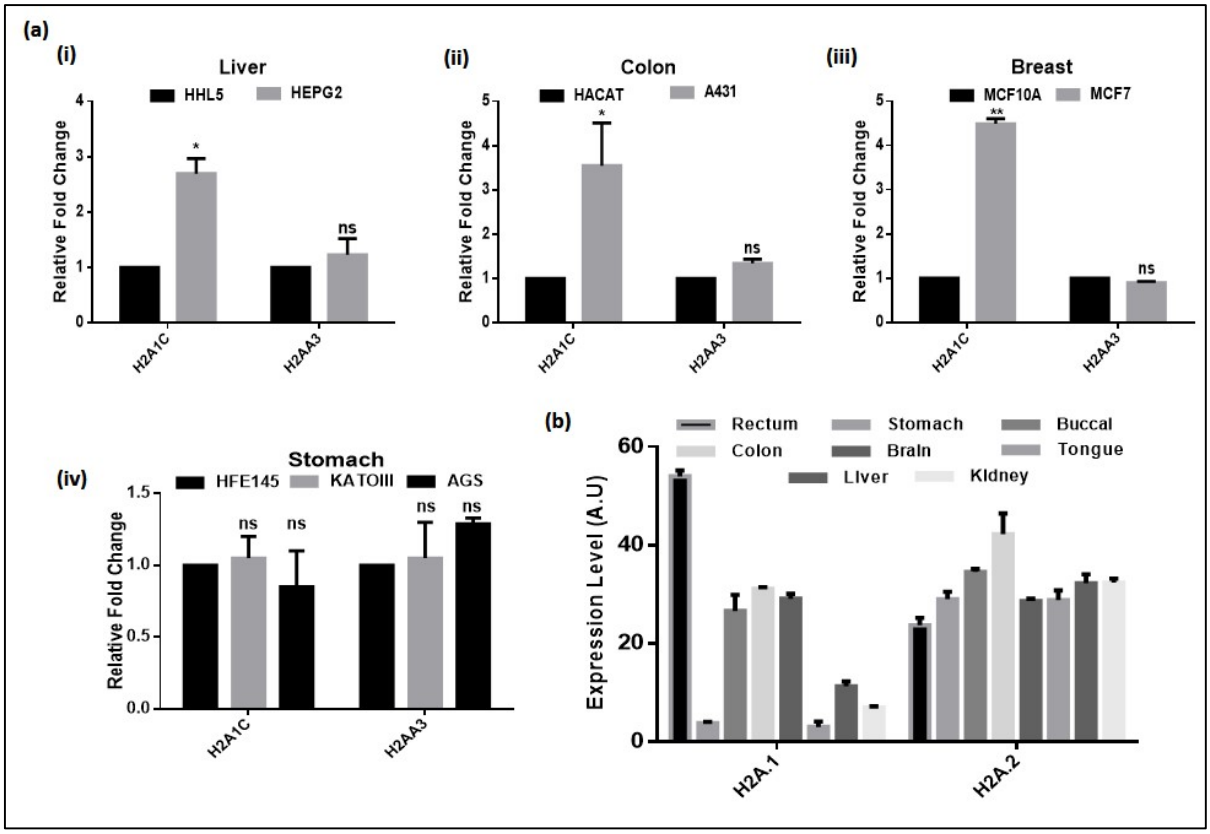



**Figure 5.5: H2A.1/H2A1C expression in different cell lines and tissues.** (a)(i)(ii)(iii)(iv) Quantitative real time PCR data showing the relative expression levels of H2A1C and H2AA3 with respect to GAPDH in different human cell lines (see text for more details). (b) Quantitative data showing the relative levels of H2A.1 and H2A.2 in various normal rat tissues, monitored at transcript level by semi-quantitative RT-PCR. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$).

In addition to the altered expression in cancer cell lines, a marked variation in transcript levels of H2A.1 was observed in various organs. A very high level of H2A.1 expression was observed in rectum [**Figure 5.5b**], whereas, in stomach and tongue, the expression levels were

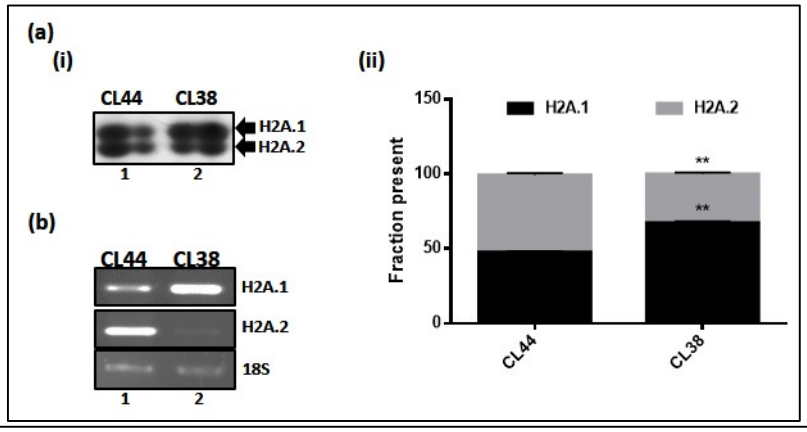



**Figure 5.6: Relative levels of H2A.1/H2A.2 in CL38 and CL44.** (a)(i) H2A region of the AUT-PAGE gel showing the relative levels of H2A.1 and H2A.2 (arrows marked) in CL44 and CL38. (ii) Quantitative data of the AUT-PAGE showing the relative levels of H2A.1/H2A.2. (b) Agarose gel showing the expression status of both the isoforms in CL44 and CL38 cell lines, done by semi-quantitative RT-PCR. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$).

particularly low [**Figure 5.5b**]. Altered expression of H2A.2 isoform was also noted, however, the differences were not as pronounced as those observed for H2A.1. These results, along with the earlier published reports, suggests that the H2A.1 expression level in cells is intricately regulated in various patho-physiological conditions. The above described results, along with the earlier published reports, suggests that the H2A.1 expression level in cells is intricately regulated in various patho-physiological conditions.

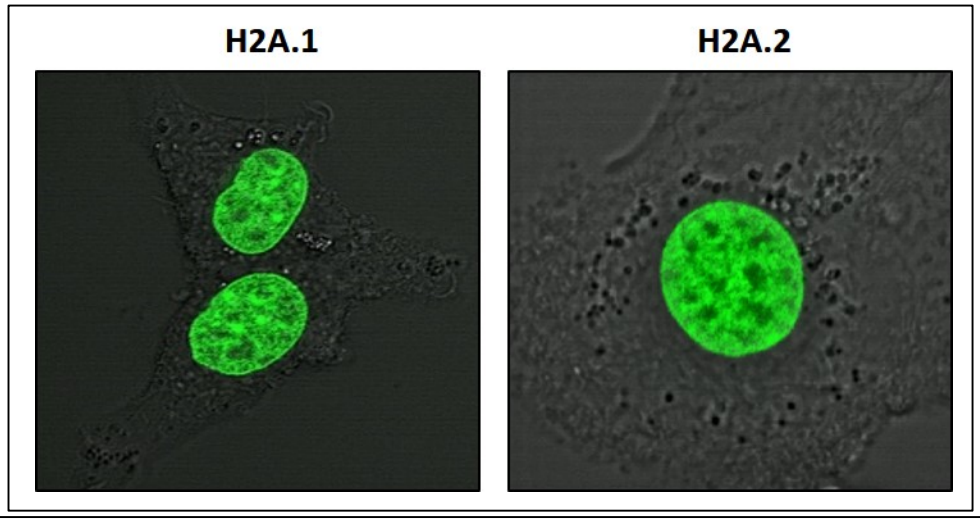



**Figure 5.7: Localization of H2A.1 and H2A.2.** Histone H2A.1 and H2A.2 were tagged with yellow fluorescent protein (YFP) using peYFPn1 vector and transfected in CL38 cells. The image was captured using Olympus LSM510 Scanning Confocal Microscope. The tagged histones are appearing green due to pseudo-colour.

### 5.3.3 H2A.1 isoform is functionally non-redundant from the H2A.2 isoform

The varied expression levels of H2A.1 in different tissues, differentiation status, age and

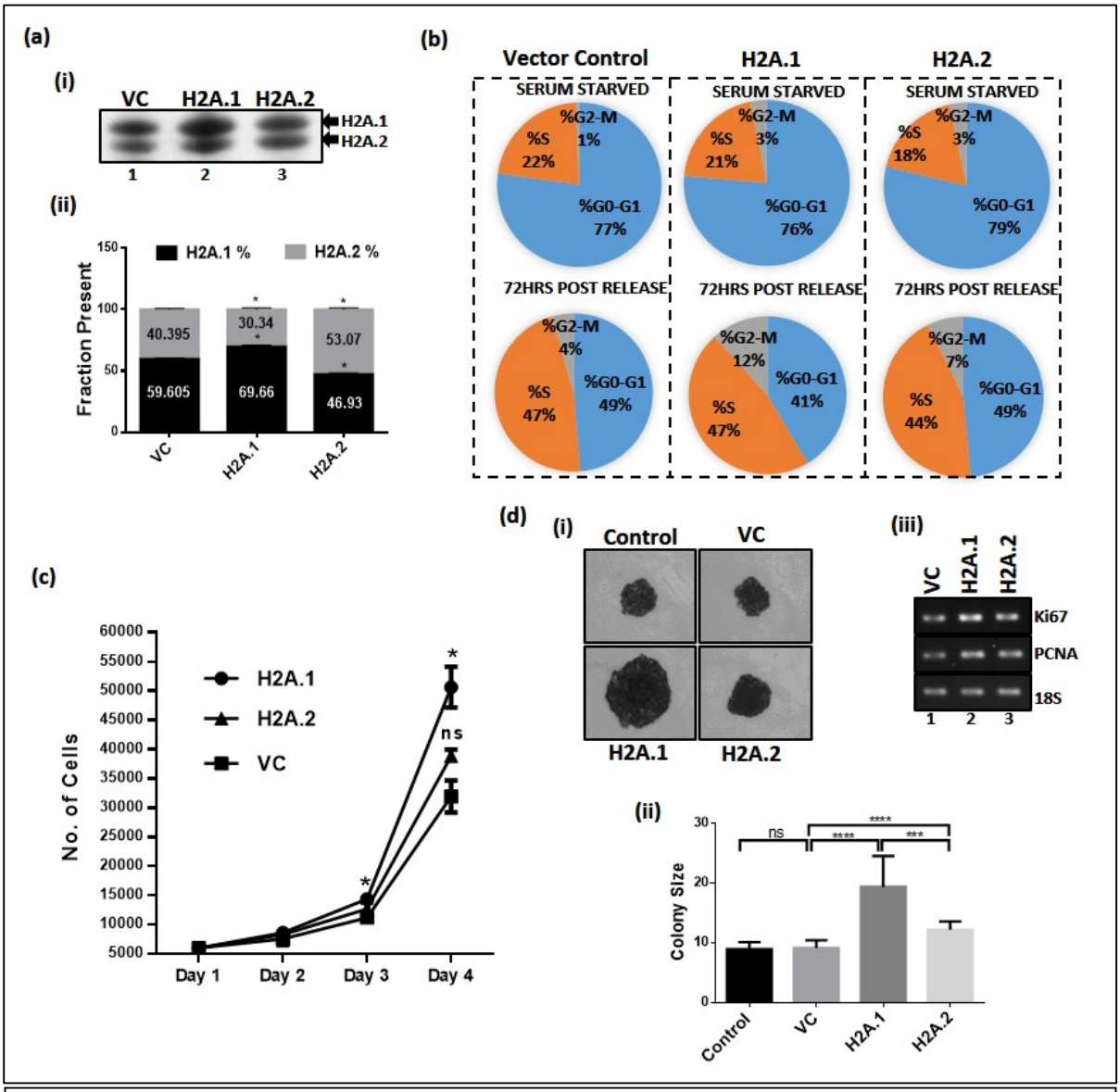



**Figure 5.8: Effect of H2A.1 overexpression on cell proliferation.** (a)(i) AUT PAGE analysis showing the H2A region demonstrating the enrichment of H2A isoforms in chromatin upon their ectopic overexpression using pcDNA 3.1 (+) vector. (ii) Quantitative analysis of isoforms enrichment in chromatin on overexpression. The bands of H2A.1 and H2A.2 were normalised with band of H2B (b) Cell cycle analysis of H2A isoform overexpressing stable cell lines post serum starvation and release. The data was plotted post performing flow cytometry analysis. (c) Cell proliferation curves of H2A.1 and H2A.2 overexpressing cells in comparison to vector control (VC) in CL38 cell line monitored for 4 days through MTT assay. The significance was calculated with respect to VC. (d)(i) Picture depicting the differences in the size of colonies obtained after plating 1000 cells and allowing growth for 14 days. The colonies were fixed and stained with crystal violet stain. (ii) Quantitative analysis of the colony size of 50 colonies of each set measured using ImageJ. (iii) RT PCR for cell proliferation markers Ki67 and PCNA on H2A.1 and H2A.2 overexpression cells in comparison to vector control (VC) in CL38 cell line. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$).

diseases prompted us to investigate whether H2A.1 are functionally non-redundant. Our results show that H2A.2 is expressed abundantly in different rat tissues [**Figure 5.5b**]. Hence, the effects of H2A.1 were compared in reference to H2A.2 isoform. Two cell lines were chosen that are derived from the liver of NDEA-administered Sprague-Dawley rats for the studies; CL44 (pre-neoplastic), with equimolar ratio of H2A.1 and H2A.2, and CL38 (neoplastic), in which H2A.1 is overexpressed [**Figure 5.6**].

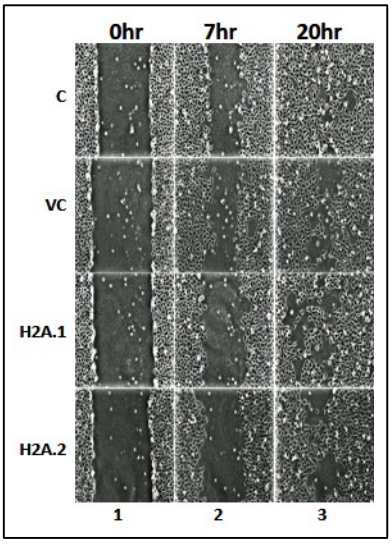



**Figure 5.9: Effect on migration on H2A.1 overexpression.** Wound healing assay for analysis of changes in migration potential of the cell upon overexpression of H2A isoforms in CL38 cell line.

YFP tagged H2A.1/H2A.2 suggested that these isoforms are widely incorporated in the chromatin [**Figure 5.7**]. It was confirmed that overexpression of H2A.1/H2A.2 leads to ~16% increase in its abundance in chromatin [**Figure 5.8a**]. This essentially, can be considered to mimic the patho-physiological condition in cells, where overexpression of H2A.1 leads to its enrichment in chromatin. The cell cycle profile of H2A.1/H2A.2 CL38 cells was checked 72hours post serum release after they were G1-enriched by serum starvation. Ectopically H2A.1 overexpressing cells had a discernible increase in the mitotic cell population (12%) proliferation was observed post 3 days in CL38 cells on H2A.1 overexpression [**Figure 5.8c**]. Similar effects were reflected in the colony formation assay, with H2A.1 over-expressing colonies substantially larger suggesting more rapid proliferation [**Figure 5.8d (i), (ii)**]. Associated up-regulation in proliferation markers, Ki67 and PCNA, was also detected [**Figure 5.8d (iii)**]. No significant difference in the closure of wound in scratch assays was perceived

[**Figure 5.9**] on overexpressing histone isoforms. Notably, we did not observe increased proliferation of CL44 cells upon H2A.1 overexpression [**Figure 5.10a**].

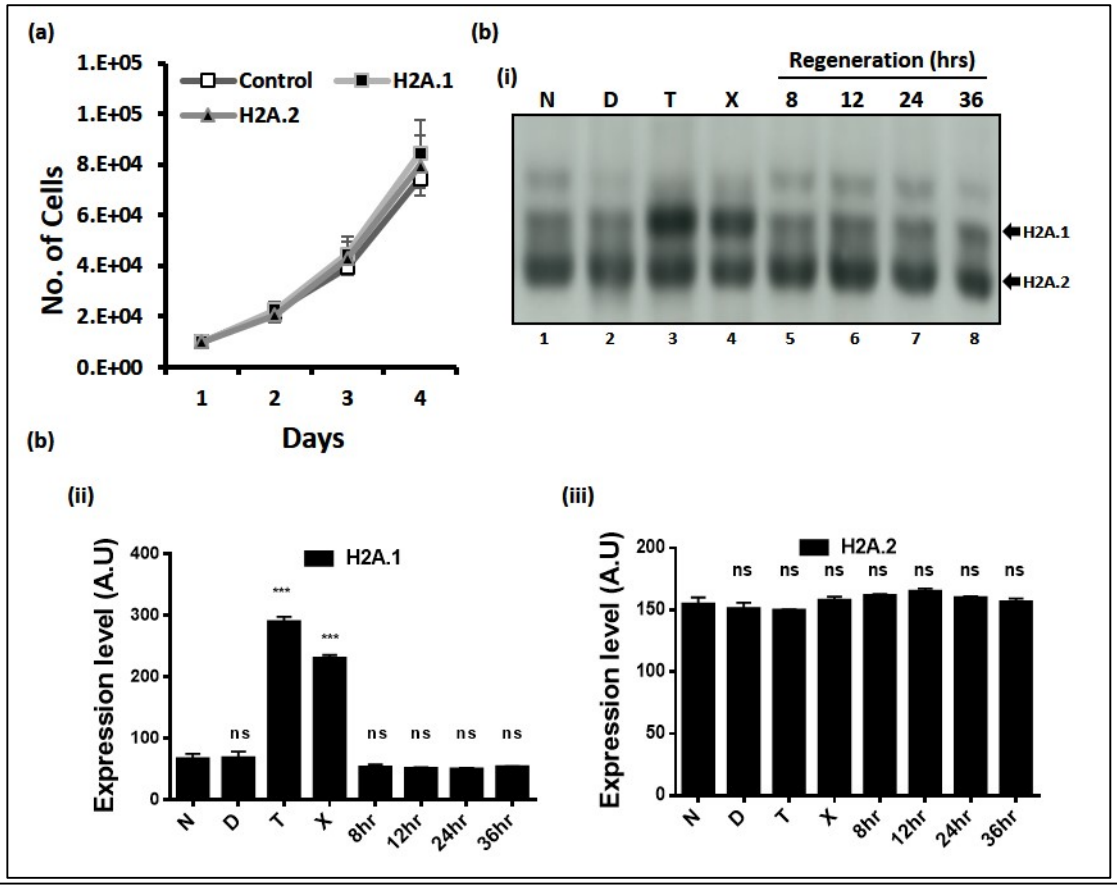



**Figure 5.10: H2A.1 and H2A.2 expression level and cell proliferation.** (a) Cell proliferation curves of H2A.1 and H2A.2 overexpressing cells in comparison to vector control (VC) in CL44 cell line monitored for 4 days through MTT assay. (b) (i) AUT-PAGE depicting the H2A region of the histones isolated from different stages of liver transformation and time points of liver regeneration. Quantification of band corresponding to H2A.1 and H2A.2 from three independent experiment is depicted in b (ii) and b (iii), respectively. N- Normal, D- Dysplastic, T- Tumor, X- Transplantable HCC. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$). The significance was calculated with respect to Normal.

During the process of liver regeneration, rapid cell proliferation occurs. To investigate the correlation between H2A.1 and cell proliferation, regeneration of liver was induced by performing partial hepatectomy and the tissue samples were collected at the indicated time points [**Figure 5.10b (i)**]. However, unlike in HCC, where there is marked upregulation of H2A.1, the level of neither H2A.1 or H2A.2 was found to alter during the regeneration process

[**Figure 5.10b (ii)**]. Taken together, these data suggest that H2A.1 upregulation does not always lead to or positively correlate with proliferation.

### 5.3.4 Leu51 and Arg99 are important in conferring non-redundant functionality to H2A.1 isoform

The H2A isoforms, H2A.1 and H2A.2, differ in three residues in their primary amino acid sequence [**Figure 5.11a**]. The residues of H2A.1 were substituted to the corresponding ones of H2A.2 to identify the residue having the most significant impact on the non-redundant functionality of H2A.1. Mutating residue 99$^{th}$ of H2A.1 drastically reduced the pro-proliferative effect observed on H2A.1 overexpression [**Figure 5.11c**]. Mutating residue 51$^{st}$ also negatively affected proliferation and had a synergistic effect when substituted alongside residue 99$^{th}$ [**Figure 5.11d**]. The assays were performed with clones exhibiting similar levels of over-expressed proteins to rule out variations resulting from differences in expression level

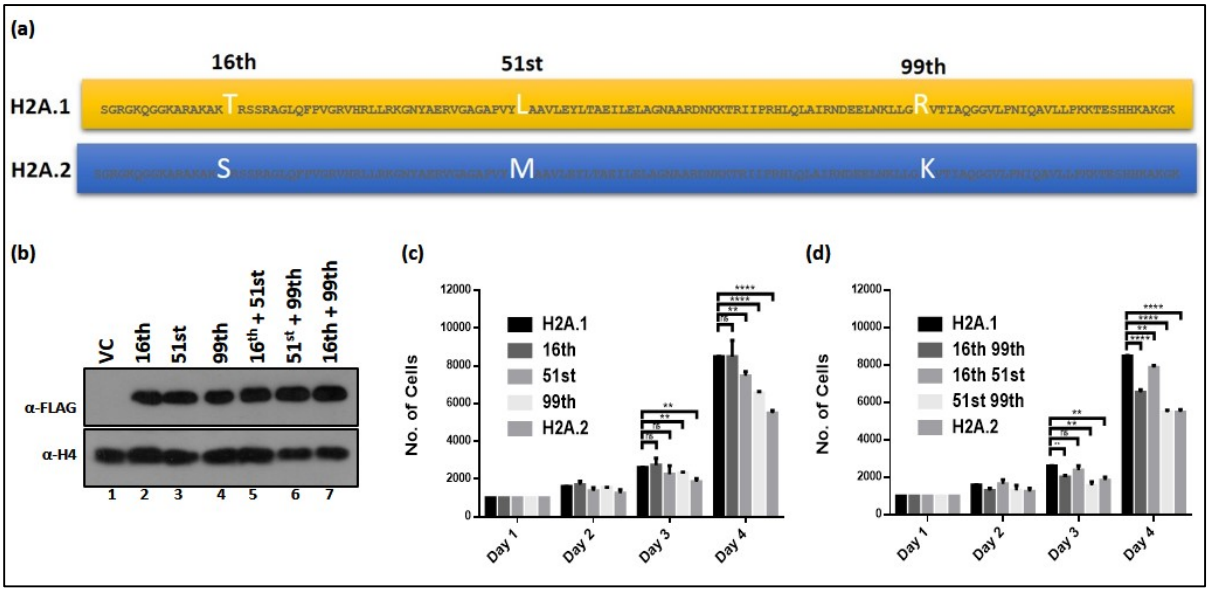



**Figure 5.11: Effect of H2A.1 mutants on cell proliferation.** (a) Alignment of rat histone H2A.1 and H2A.2 isoforms. (b) Western blot analysis depicting the expression level of ectopically expressed H2A.1 and its mutants in the selected stable clones of CL38 for performing proliferation assays. H2A.1 and its mutants were cloned in pcDNA FLAG HA vector, hence, anti-FLAG antibody was used to detect and quantify the expression level. Cell proliferation data of 6 independent experiments of H2A.1 (c) single and (d) double mutants depicted as histogram. MTT assay was performed for 4 days in CL38 cells for comparing the effect of H2A mutants on cell proliferation. VC- Vector control. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$).

[**Figure 5.11b**]. Notably, the 16$^{th}$ residue, in which the rat H2A.1 and human H2A1C differ, did not have any significant effect on the non-redundant effects of H2A.1 in the assays [**Figure 5.11c**].

## 5.4. Discussion

The aberrant expression of histone isoforms in various patho-physiological states has prompted the question for decades whether the H2A isoforms have non-redundant functions in the cell. We provide here definitive evidence of the differential effect of the H2A isoforms on cellular proliferation implicating their non-redundancy in a context dependent manner.

The expression level analysis in different organs also shows that particularly the expression of H2A.1 (or H2A1C) varies in different tissues. Similarly, in humans particularly the expression level of H2A1C alters markedly in different contexts. This suggests that probably H2A.2 plays a "house-keeping" function in the cells whereas H2A.1 or H2A1C helps to bring about specialised non-redundant outcomes. Notably, the 16$^{th}$ residue, in which the rat H2A.1 and human H2A1C differ, did not have any significant effect on the non-redundant effects of H2A.1 in the assays.

Previous studies from our lab has shown that H2A.1 is highly overexpressed in undifferentiated hepatocytes and embryonic stem cells[5]. These observations along with the reports that the proportion of H2A.1 and H2A.2 decreases in the course of development and differentiation of rat brain cortical neurons[129] and *in vitro* differentiation of fibroblasts[130] suggests that along with the proliferation, the differentiation state of the cells may govern the ratio of the two isoforms. This is consistent with the observation that H2A.1 expression gradually increases as the liver sequentially starts to dedifferentiate during the sequential stages of hepatocellular carcinoma[3].

One very important revelation that occurs from our study and the earlier reports is that the functional effects exhibited by H2A isoforms might be context dependent, both in terms of extent and the effect itself. For example, the pro-proliferative effect conferred by H2A.1 was not observed in the pre-neoplastic CL44 cells. Notably, human H2A1C isoform, which was initially reported to be downregulated in CLL, was shown to exhibit anti-proliferative effects[2]. However, in a later study with higher number of samples H2A1C levels were found to be higher in CLL patients compared to samples from healthy individuals[179]. Further, the high expression of H2A.1 seen in brain cells, which are terminally differentiated and do not regenerate, suggests that the actual functional effect of H2A.1 may also be context dependent and need not be always proliferation associated. Consistent with this, the higher expression of H2A1C has been seen in chemo-resistance in pancreatic cancer cell lines[192], which is again a different context.

Based on the discussion above, some of the questions which arises are: what determines the context in which the non-redundant functionality of H2A isoforms are exhibited? And in those contexts, which are the genes which are regulated by the particular isoform? It is reasonable to hypothesize that other factors which contribute to the epigenetic landscape of cells and/or the differential PTM's that the histone isoform itself may undergo, determines the context in which the differential functional effects of H2A isoforms are exhibited. Interestingly, Arg 99 of H2A has been shown to undergo methylation[193]. One study, that has tried to identify the genes in the particular context, shows that H2A1C isoform control ER target genes in ER positive breast cancer cell lines[4]. Interestingly, the deletion of the H2A N-terminal domain (Δ4-20) led to upregulation of only 248 genes and also conferred sensitivity to radiation induced DNA damage[194]. Clearly, much remains to be understood of the correlation between H2A mediated nucleosome stability and how this regulates chromatin mediated processes like gene expression and DNA repair.

## 5.5. Conclusion

In summary, H2A.1 expression varies markedly in different patho-physiological states. Incorporation of H2A.1 into the chromatin leads to promotion of cellular proliferation in neoplastic liver cell line. The M51L and K99R substitutions have the most prominent effect on cell proliferation. The possibility of the non-redundant function, when extended to the plethora of the histone isoforms encoded by multiple genes (H2A, 16 genes; H2B, 22 genes; H3, 14 genes; H4, 14 genes; and H1, 6 genes), truly increases the complexity of the epigenome by many folds. Undoubtedly, such complexity is the necessity for multicellular organisms as the diversity in the epigenome plays central role in cell type specific gene expression. This in turn leads to specialised functions in thousands of cell types with the same genome.

# VI. NON REDUNDANT STABILITY AND STRUCTURAL CONTRIBUTION OF H2A ISOFORMS

## Chapter 6

*Work carried out for investigating the non-redundant stability and structural contribution of H2A.1 isoform*

## 6.1. Introduction

We found the H2A.1 expression varies markedly in different states and it does have context dependent non-redundant functionality compared to H2A.2 owing to M51L and K99R substitutions. Previous studies as well have demonstrated the non-redundant functionality of H2A1C isoform[2,4]. However, how the histone isoforms, impart their non-redundant functional effects remains unclear which we next wanted to focus upon. Histone variants, which are the better studied sequence divergent forms of histone, impart their non-redundant functionality by modulating chromatin dynamics. One primary way by which this is achieved is by changing the biochemical composition of the nucleosome core particle which results in its altered structure and/or stability (for details please refer to section 1.8). Histone variants, which are relatively more studied, have been reported to alter the structure of chromatin. For example, H2ABBD containing nucleosomes are less stable[88] and organises only 118bp of DNA instead of 147bp[87]. Histone H2A.Z and H3.3 hetero-nucleosomes could only be purified under conditions of low ionic strength due to their lesser stability[195]. The nucleosome is an oligomeric complex and the stability of each of the sub complexes may modulate the nucleosome stability. Previously, it was reported that H2A.Z-H2B dimer is less stable compared to canonical H2A-H2B[102]. These reports clearly suggest that histone variants alter chromatin structure and stability. Whether the histone isoforms also modulate chromatin stability has not been reported.

H2A.1 when compared to H2A.2 has the following three substitutions S16T, M51L and K99R. The 16th residue of H2A is involved in interaction with the minor groove of DNA in the NCP, residue 51st lies in the dimer interface with H2B and residue 99th of H2A interacts with the H4 tail in the octamer core [**Figure 6.1**]. It is worth noticing that residue 16$^{th}$ is part of the HAR domain of H2A and has been shown to be a determinant of nucleosome stability and gene repression[194]. The 51st residue of H2A, which lies in the dimer interface with H2B has leucine

(H2A.1) to methionine (H2A.2) alteration. Leucine and methionine are very similar amino acids. However, previously these have been reported to give rise to altered stability in proteins[196,197]. In one study up to 10 adjacent residues within the core of T4 lysozyme were replaced by methionine. Such variants folded cooperatively with progressively reduced stability[196]. Likewise, lysine (H2A.2) to arginine (H2A.1) substitution that occurs at the 99$^{th}$ position, brings about alteration in stability of proteins[198] although both the amino acids are very similar. Lysine to arginine substitution led to new salt bridges and hydrogen bond interactions that help improve the rigidity of the mutant protein. Therefore, the three differences in amino acid residues between H2A.1 and H2A.2 can potentially alter the stability of the nucleosome and its sub complexes.

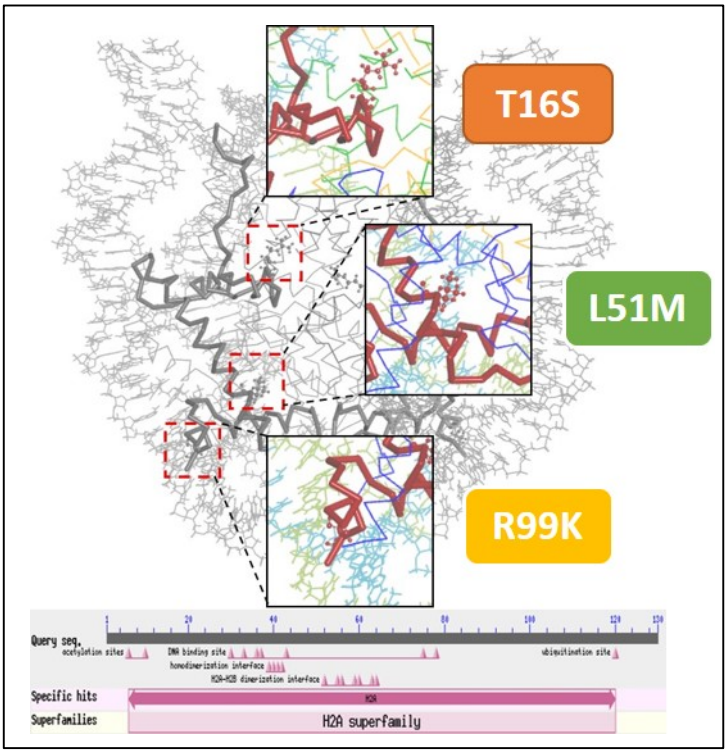



**Figure 6.1: Location of differential residues in nucleosome.** Nucleosome depicting the location of the three differential amino acids between H2A.1 and H2A.2. Below is a representation of the important regions of interaction of H2A.

In this part of the work the possible non-redundant structural/stability contribution of histone

H2A isoforms H2A.1 and H2A.2 have been investigated by using *in vitro*, *in cellulo* and *in silico* approaches.

## 6.2. Methods

### *6.2.1 Reconstitution of H2A-H2B dimer and H3-H4 tetramer*

See under Methods section of chapter IV (page number 108).

### *6.2.2 Equilibrium unfolding*

Equilibrium unfolding as a function of urea and temperature was monitored by CD spectroscopy. Also, equilibrium unfolding as a function of urea and Gdn-Chl (guanidine chloride) was monitored by change in fluorescence. For CD spectroscopy, the protein (25 µg/ml) was incubated in presence of various concentrations of the denaturant at $25^0$C for 24 h. In the thermal unfolding experiments, the spectra were recorded 3min after the desired temperature was attained. Spectra were recorded in the temperature range of 20-$100^0$C. The intrinsic tryptophan fluorescence spectra of the H2A-H2B dimers were recorded on a spectrofluorimeter, equipped with water bath. The protein was excited at 280nm using a cell of 1.0 cm path-length and both excitation and emission slit widths were set at 3nm. CD spectra were recorded on a computer interfaced JASCO spectropolarimeter using a cylindrical quartz cell of 1 mm (190 – 250nm) and 10mm (250 – 300) at a H2A-H2B concentration of 1µg/ml. For each spectrum, 20 successive scans were collected and the averaged spectra were used for further analysis.

### *6.2.3 Data fitting*

The unfolding data was fit into two-state model of unfolding as described previously[199]. The denaturation curves were plotted, with the fluorescence intensities at 305nm of native and denatured H2A-H2B dimer, against denaturant concentration, and further analysis of the data

was performed as described by Pace *et al.*[200]. Similar analysis was performed with the denaturation curves plotted with the drop in CD signal at 222nm with increasing temperature or urea concentration. From the denaturation curves, a two state $\mathrm{F} \leftrightarrow \mathrm{U}$ unfolding mechanism was assumed, and consequently, for any of the points, only the folded and unfolded conformations were present at significant concentrations. Thus, if $f_F$ and $f_U$ represent the fraction of protein present in the folded and unfolded conformations, respectively then

$$f_F + f_U = 1 \tag{1}$$

$f_U$ was calculated using the following equation

$$f_U = (F_F - F_0) / (F_F - F_U) \tag{2}$$

where $F_F$ is the fluorescence intensity of completely folded or native protein, $F_0$ is the observed fluorescence intensity at any point of denaturant concentration or temperature, $F_U$ is fluorescence intensity of the completely denatured or unfolded protein.

For a two state $\mathrm{F} \leftrightarrow \mathrm{U}$ unfolding mechanism, the equilibrium constant $K$ and $\Delta G_U$, the free energy of unfolding was calculated using Equation 4 and 5 respectively.

$$K = f_U / (1 - f_U) \tag{3}$$

$$\Delta G_U = \mathrm{RT} \ln K \tag{4}$$

where R is the gas constant and T is the absolute temperature. It is assumed that the free energy of unfolding, $\Delta G_U$, has a linear dependence on the concentration of the denaturant $[\mathrm{D}]$.

$$\Delta G_U = \Delta G^{H20} + m[\mathrm{D}] \tag{5}$$

$\Delta G^{H20}$ and $m$ are therefore the intercept and the slope respectively, of the plot of $\Delta G_U$ versus [D]. $\Delta G^{H20}$ corresponds to the free energy difference between the folded and unfolded states in the absence of any denaturant and $m$ is a measure of the cooperativity of the unfolding reaction. The concentration of denaturant at which the protein is half unfolded (when $\Delta G_U = 0$) is given by $D_{1/2}$ and from Equation 5, $\Delta G^{H20} = - m D_{1/2}$.

The data from the thermal unfolding curves were obtained under the same conditions as those for denaturant unfolding curves. Values of $f_U$, $K$ and $\Delta G_U$ were calculated using Equations 2, 3 and 4. The midpoint of thermal ($T_m$) denaturation was obtained as the temperature at which $\Delta G_U = 0$ from the plot of $\Delta G_U$ versus T. The slope of such a plot at Tm yielded $\Delta S_m$, the change in entropy. The enthalpy change for unfolding at $T_m$, $\Delta H_m$, was calculated using the equation:

$$\Delta H_m = T_m \Delta S_m \qquad (6)$$

$\Delta C_p$, the change in heat capacity that accompanies protein unfolding was obtained from the slope of the plot of $\Delta H_m$ versus $T_m$, where $T_m$ was varied as a function of Gdn-Chl concentration. $\Delta G_U$ at $25^0$C was calculated using equation:

$$\Delta G\ (T) = \Delta H_m\ (1 - T/\ T_m) - \Delta C_p\ [(T_m - T) + T \ln (T/\ T_m)] \qquad (7)$$

### 6.2.4 *FRAP (Fluorescence Recovery After Photobleaching) assay*

FRAP assay was performed as described earlier[201] with slight modifications. H2A.1 and H2A.2 coding sequences were cloned into peYFPn1 and stable clones were made in CL38 cells. The cells were seeded in 30mm optical plates and incubated to allow the cells to reach confluency.

This helps to minimise movement of cells during acquisition. FRAP was performed on a laser confocal microscope LSM510 Meta (Zeiss) equipped with $CO_2$ and temperature maintenance accessory. The microscope was equipped with a heated environmental chamber set to 37°C and $CO_2$ perfusion set to 5%. For photobleaching experiments, square bleach regions with a size of 5 × 5 µm were positioned on selected cell nuclei within the field of view. Photobleaching was performed using 100 iterations with the 488-nm and the 514-nm laser line set to 100% transmission. To determine long-term recovery kinetics, 3D image stacks of 8-µm height and a z distance between image planes of 1 µm were recorded with an initial speed of 1 frame/30sec for the first 15min, followed by intervals of 5min for the next 45min to minimise photobleaching. Quantitative evaluations were performed with ImageJ (*http://rsb.info.nih.gov/ij/*). Lateral and rotational movements of the cell nucleus were corrected by image registration using the StackReg plug-in of ImageJ. Mean intensities over time were extracted from the total nuclear area (T). The background ROI outside of the cell was defined manually from the initial field of view. The mean gray values over time were measured, background subtracted, and normalized to the respective means of the last prebleach values. The resulting postbleach B values were then divided by the respective T value to correct for the superimposed gain or loss of total fluorescence during postbleach acquisition, potentially caused by newly synthesized GFP-histones, bleaching-by-acquisition, and flux of residual fluorescence from above and below the recorded optical plane.

### *6.2.5 Cell fractionation*

The harvested cells were washed twice with chilled PBS, lysed in MKK lysis buffer (10mM Tris-Cl, pH 7.4, 0.27M Sucrose, 1mM EDTA, 1mM EGTA, 1% Triton X-100) containing protease and phosphatase inhibitors (1mM sodium orthovanadate, 10mM sodium fluoride, 10mM β-glycerophosphate, 10µg/ml leupeptin, 10µg/ml aprotinin, 1mM PMSF) for 30min at

$4^0$C. Following hypotonic lysis, cells were centrifuged at 12,500rpm for 20min at $4^0$C to separate chromatin fraction and total soluble protein fraction.

### 6.2.6 *Salt dissociation experiment*

CL38 cells were transfected with H2A.1 pcDNA FLAG or H2A.2 pcDNA FLAG. The nuclei of the cells were isolated as described in chapter V (page number 130). Pelleted nuclei were resuspended in buffer containing 15mM HEPES [pH 7.9], 60mM KCl, 15mM NaCl, 0.34mM sucrose and 10% glycerol and incubated with different NaCl concentrations (600-1000mM) at 4°C for 30min. Supernatants were collected and the proteins were precipitated with 20% TCA. The histone in the pellet remaining after salt treatment was extracted with 0.2M $H_2SO_4$ as described in chapter V (page number 129). The proteins were resolved on 18% SDS-PAGE and probed with suitable antibodies.

### 6.2.7 *Dimer simulation*

Sequence analysis of rat nucleosome identified human nucleosome PDB ID: 2CV5 as a closet matched based on histone variants. Histone H2B was 100% identical. In histone H2A.1 dimer only one change incorporated at 40th position from Ser40Ala, while for histone H2A.2 dimer additional mutations were introduced at Thr16Ser, Leu51Met and Arg99Lys positions. Both the dimers were subjected to energy minimization using smart minimizer algorithm of Discovery studio 2.5. It performs first steepest descent with RMS gradient tolerance of 3, followed by conjugate gradient minimization. The convergence criterion for minimization is set to gradient tolerance of RMS 0.1.

### 6.2.8 *Nucleosome system building*

Sequence analysis of rat nucleosome identified human nucleosome PDB ID: 2CV5 as a closet matched based on histone variants. Histones H4 and H2B were 100% identical. In histone

H2A.1 only one change was at 40th position from Ser40Ala. In order to convert H3.1 of human to H3.3 of rat, four point mutations Ser87Ala, Val89Ile, Met90Gly and Cys96Ser were introduced. Thus, rat nucleosome consists of Histone (H2A.1, H2B, H4, H3.3) was build labelled as 'system 1'. From system 1, system 2 was build where H2A.1 was converted to H2A.2 by introducing mutations at Thr16Ser, Leu51Met and Arg99Lys positions on both the copies of the histone variants. The rest of the histone variants (H2B, H4 and H3.3) were unchanged. In both the systems the DNA were identical as with PDB ID: 2CV5.

### *6.2.9 Molecular dynamics simulation*

Molecular dynamics (MD) simulations experiments were performed based on principles described earlier[202]. Simulation were performed in triplicates for both the systems in a truncated octahedron box by adding sufficient ions to balance the nucleosome charge. Nucleosomal complex were represented using amber ff03 force filed and the TIP3P model to describe the water molecules. System was energy minimized in two steps, initially using steepest-descent method putting restrain on nucleosomal complex and allowing water to relax. In next minimization step entire system was allowed to relax. Then system was equilibrated first by NVT followed by NPT ensemble.

All the Molecular simulations were carried out using the Gromacs-4.6.5 software, with periodic boundary conditions. The particle mesh Ewald method was used to treat the long-range electrostatics, together with a cut-off of 1.2 nm for the short-range repulsive and attractive dispersion interactions, which were modelled via a Lennard-Jones potential. The Settle algorithm was used to constrain bond lengths and angles of water molecules, and P-Lincs for all other bond lengths. The time step of 2fs was used for entire system. The temperature was kept constant at 300K by using the Nose-Hoover thermostat method. The simulation was carried out for 250ns. To control the pressure at 1 atmosphere, Parrinello-Rahman method was

used. Gromacs tools were used for RMSD and H-bond calculation[203]. Principal component analysis (PCA) was done using prody software. VMD software was used for contacts analysis calculations using TCL script and for image generation.

The simulation work was carried out in collaboration with CDAC, Pune and BTIS, ACTREC.

### *6.2.10 MNase digestion assay*

The assay was performed as described earlier[187]. Nuclei containing 2mM $CaCl_2$ were incubated for 2, 4, 6, 8 and 10min with 5U MNase/mg of DNA at 37°C in MNase digestion buffer (15mM Tris-Cl pH 7.4, 15mM NaCl, 2mM $CaCl_2$, 60mM KCl, 15mM β-ME, 0.5mM spermidine, 0.15mM spermine, 0.2mM PMSF, protease and phosphatase inhibitors). The digestion was stopped by adding 2X lysis buffer (0.6M NaCl, 20mM EDTA, 20mM Tris-Cl pH 7.5, 1% SDS). MNase digested samples were treated with RNaseA (100µg/ml) for 30min at $37^0$C followed by proteinase K (80µg/ml) treatment for 2hrs at $50^0$C. The samples were extracted sequentially with phenol, phenol: chloroform and chloroform followed by ethanol precipitation at $-20^0$C. The precipitated DNA was recovered by centrifugation at 10000g for 20min. The DNA pellet was washed, air dried, dissolved in 50µl of TE buffer and concentration was determined by A260/A280 absorbance. MNase-digested samples were resolved on 1.8% 1XTAE agarose gel electrophoresis with 0.5µg/ml ethidium bromide. The image was scanned with ImageJ software version 1.43u; Java 1.6.0_10 (32-bit).

## 6.3. Results

### 6.3.1 Equilibrium unfolding of H2A-H2B dimer

The stability analysis of H2A-H2B dimer were carried out by reconstitution of H2A.1-H2B and H2A.2-H2B dimers and subjecting them to equilibrium unfolding. Reconstitution was performed as described in chapter IV (page number 108). Subsequently, equilibrium unfolding

was performed which was monitored by observing both secondary and tertiary structure changes. The data for H2A.1-H2B is depicted in figures. Similar unfolding behaviour was obtained for H2A.2-H2B dimer.

*Secondary Structure Changes:* Unfolding was observed in response to thermal and chemical denaturant by circular dichroism.

*Thermal Unfolding:* Unfolding was carried out starting from $20^0$C upto $80^0$C with a $2^0$C increment and an equilibration time of 3min. The CD spectra of only three temperatures are plotted for clarity. Analysis of the thermal unfolding curves suggest that there is a gradual dip at 222nm with the unfolding of dimer suggesting that it can serve as a good spectroscopic probe for monitoring secondary structure unfolding [**Figure 6.2a**]. Further, the unfolding was completely reversible with no protein aggregation as suggested by the completely overlapping unfolding and refolding curves [**Figure 6.2b**]. The data obtained could be fit into two-state unfolding model for dimeric proteins with residual in the range of only ±2 using IgorPro [**Figure 6.2c**].

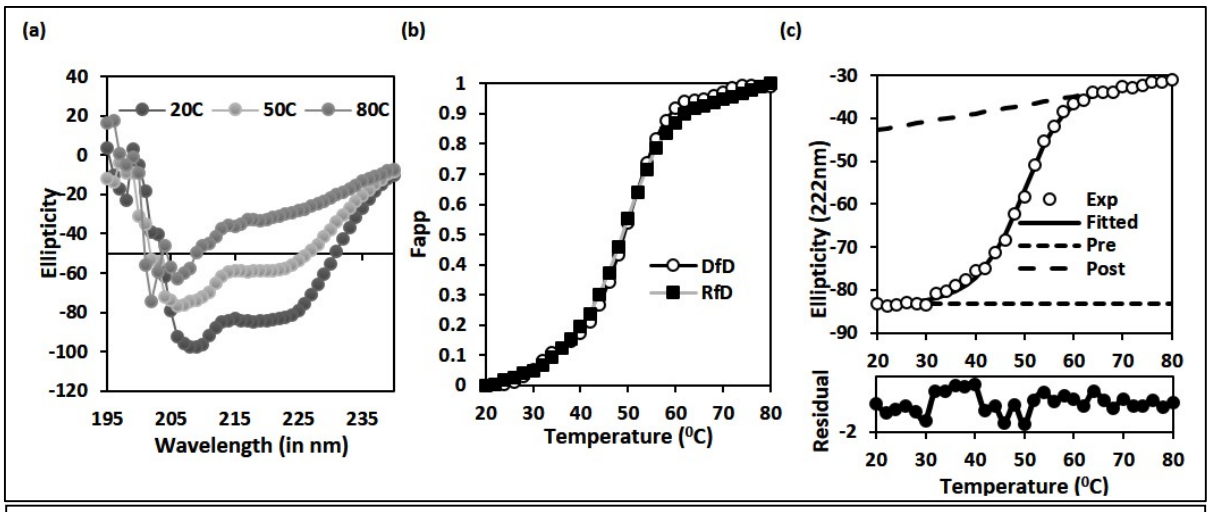



**Figure 6.2: Thermal unfolding monitored by secondary structure changes.** (a) Thermal denaturation of H2A.1-H2B dimer monitored by CD spectroscopy. (b) Fraction unfolded against the denaturant is plotted for both denaturation as well as renaturation experiment. Completely overlapping curves suggests that the denaturation was highly reversible. (c) The data obtained from thermal denaturation was fit into two-state model of unfolding.

*Chemical Unfolding:* Urea induced denaturation was also monitored with CD with an increment of 0.2M urea concentration starting from 0M and similar to thermal denaturation the

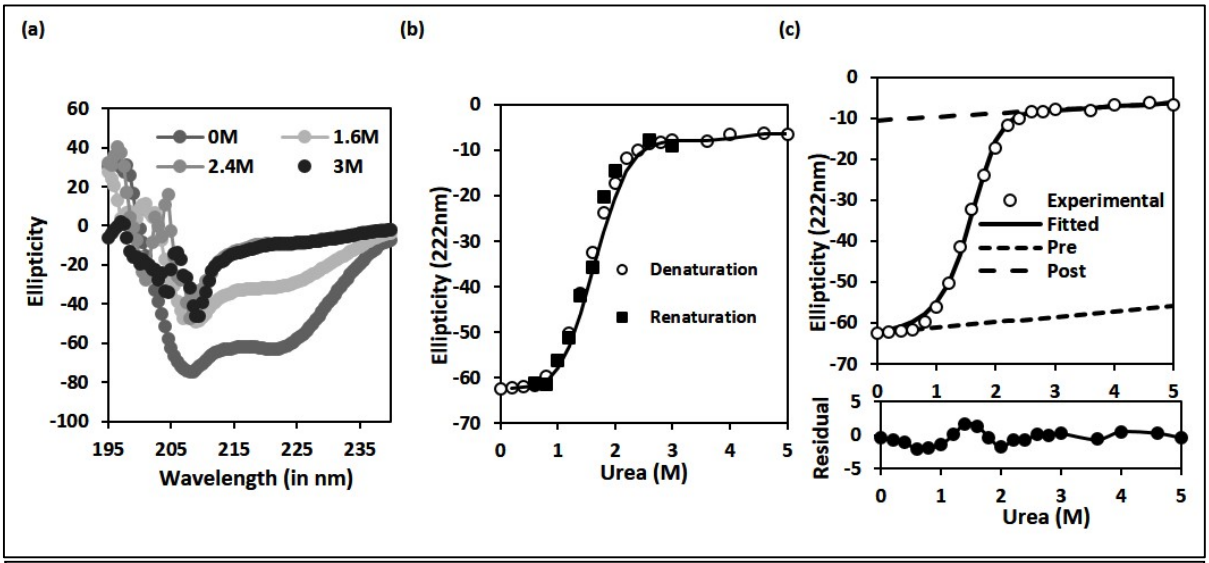



**Figure 6.3: Chemical unfolding monitored by secondary structure changes.** (a) Urea denaturation of H2A.1-H2B dimer monitored by CD spectroscopy. (b) Fraction unfolded against the denaturant is plotted for both denaturation as well as renaturation experiment. Completely overlapping curves suggests that the denaturation was highly reversible. (c) The data obtained from urea denaturation was fit into two-state model of unfolding.

dip at 222nm of CD spectra was used to plot the unfolding [**Figure 6.3a**]. Initially titration upto 8M urea was carried out, however, as the unfolding is complete in 3M urea, subsequent titrations were carried out with upto 5M concentration of urea. The denaturation was completely reversible [**Figure 6.3b**]. Similar to the thermal unfolding data the chemical denaturation data could be fit into two-state unfolding model with a residual between the range of ±5 [**Figure 6.3c**].

*Tertiary Structure Changes:* To follow tertiary structure unfolding, urea induced denaturation monitored by fluorescence spectroscopy was performed.

*Chemical Unfolding:* On carrying out urea induced denaturation there was a drop in the fluorescence intensity with the unfolding as expected because of quenching of the fluorescence of tyrosines previously buried in the dimer interface [**Figure 6.4a**]. The drop in the intensity of

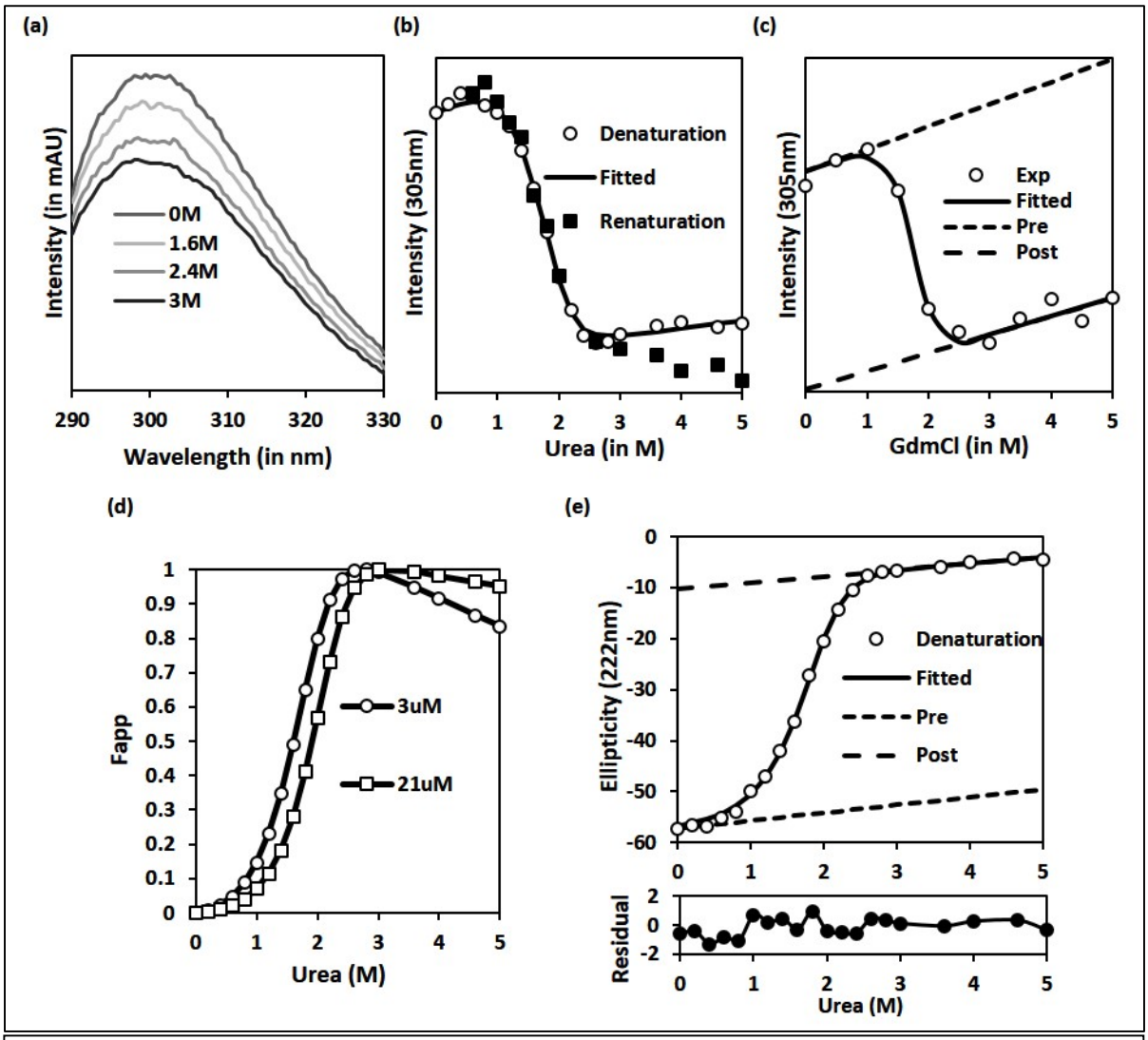


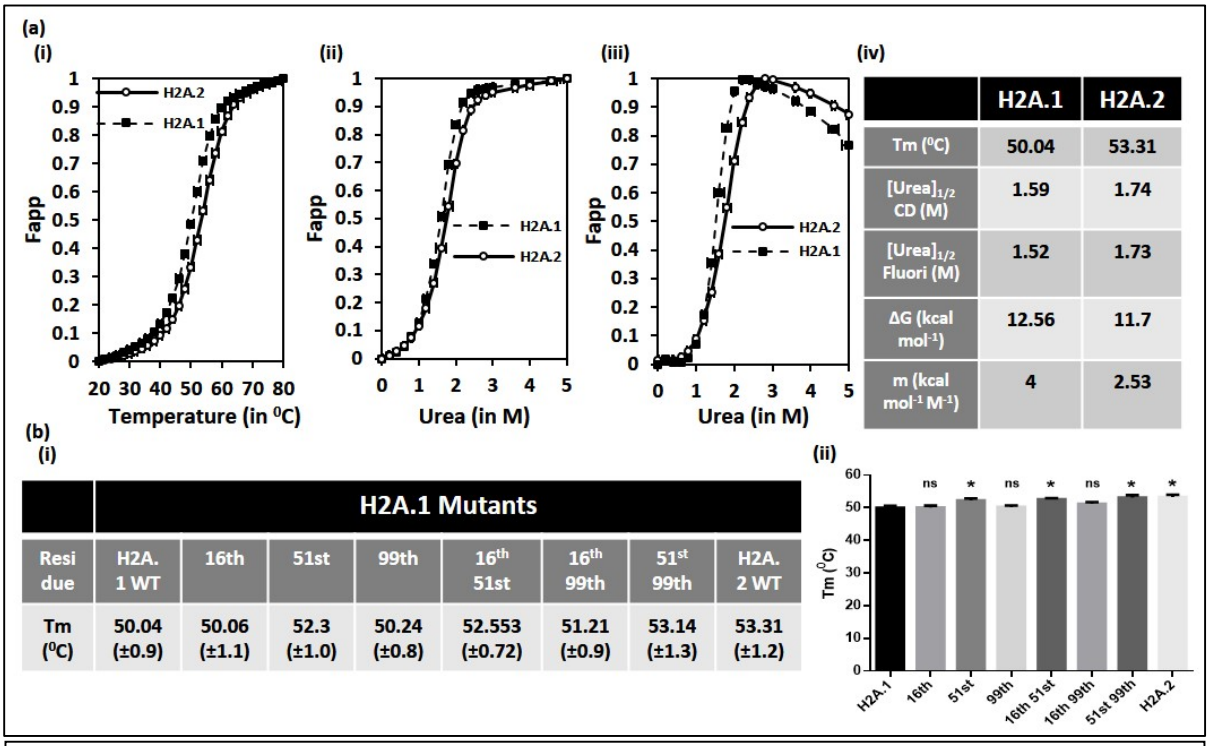

**Figure 6.4: Equilibrium unfolding monitored by tertiary structure changes.** (a) Urea denaturation of H2A.1-H2B dimer monitored by fluorescence spectroscopy to observe tertiary structure unfolding. Spectral scan at different urea concentration is plotted. (b) Fluorescence intensity at 305nm at different urea concentration is plotted for both denaturation as well as renaturation experiment. Highly overlapping curves suggests that the denaturation was reversible. (c) Fluorescence intensity for H2A.1-H2B dimer at 305nm at different guanidine chloride concentration is plotted. (d) Fraction unfolded against the denaturant is plotted for two different concentrations of H2A.1-H2B dimer. (e) The data obtained from urea denaturation was fit into two-state model of unfolding.

emission maxima at 305nm could be used for monitoring and plotting denaturation as there was no apparent red-shift [**Figure 6.4a**]. The folding was reversible [**Figure 6.4b**], however, the pre- and post-transition baseline in urea denaturation curve had a positive slope as also observed in previous reports[102]. However, to ensure that transitions were not missed during the unfolding process, denaturation was carried out with Gdn-Cl as well. Similar pre- and post-transition baselines corroborated the urea denaturation data [**Figure 6.4c**]. The unfolding also

showed concentration dependence as is expected for a dimeric protein [**Figure 6.4d**] and could be fit into two-state model of unfolding [**Figure 6.4e**] substantiating the data obtained for secondary structure unfolding.

### 6.3.2 H2A.1-H2B dimer is less stable than H2A.2-H2B dimer

We initiated the investigation of the possibility discussed above by comparing the *in vitro* stability of H2A.1-H2B and H2A.2-H2B dimer with purified recombinant histones. Equilibrium unfolding of reconstituted full length H2A-H2B dimers, which was previously described [102], was conducted to compare the stability of H2A.1-H2B and H2A.2-H2B dimers. Once the equilibrium unfolding curves for both H2A.1-H2B and H2A.2-H2B dimers were obtained, a comparative analysis of stability was carried out [**Figure 6.5a**]. Co-plotting the Fapp (apparent fraction unfolded) of H2A.1-H2B and H2A.2-H2B dimers against the increasing temperature/denaturant concentration shows a hysteresis that is indicative of difference in response to the denaturant [**Figure 6.5a (i), (ii), (iii)**]. The temperature of melting (Tm) for H2A.1-H2B dimer was determined to be $50.04^0$C whereas that of H2A.2-H2B dimer was found to be higher by ~$3^0$C at $53.31^0$C [**Figure 6.5a (iv)**], suggesting the former is less stable. The CD and fluorescence data plotted in response to increasing chemical denaturant were in pretty good agreement with the $[urea]_{1/2}$ for H2A.1-H2B obtained as 1.59 and 1.52M respectively. The $[urea]_{1/2}$ for H2A.2 was found to be 1.74 and 1.73M with CD and fluorescence spectroscopy, respectively. *m* represents the slope of the plot of ΔG vs T. The *m* value obtained for H2A.1-H2B dimer was 4kcal $mol^{-1}M^{-1}$ and that of H2A.2-H2B dimer was 2.53kcal $mol^{-1}M^{-1}$ [**Figure 6.5a (iv)**] suggesting the higher sensitivity of H2A.1-H2B dimer to denaturant concentration. Our thermal and chemical unfolding data monitored by both fluorescence and circular dichroism spectroscopy data suggests that H2A.1-H2B dimer is less stable than H2A.2-H2B dimer.




| | H2A.1 | H2A.2 |
|---|---|---|
| Tm ($^0$C) | 50.04 | 53.31 |
| $[Urea]_{1/2}$ CD (M) | 1.59 | 1.74 |
| $[Urea]_{1/2}$ Fluori (M) | 1.52 | 1.73 |
| ΔG (kcal mol$^{-1}$) | 12.56 | 11.7 |
| m (kcal mol$^{-1}$ M$^{-1}$) | 4 | 2.53 |

| | H2A.1 Mutants | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Residue | H2A. 1 WT | 16th | 51st | 99th | 16$^{th}$ 51st | 16$^{th}$ 99th | 51$^{st}$ 99th | H2A. 2 WT |
| Tm ($^0$C) | 50.04 (±0.9) | 50.06 (±1.1) | 52.3 (±1.0) | 50.24 (±0.8) | 52.553 (±0.72) | 51.21 (±0.9) | 53.14 (±1.3) | 53.31 (±1.2) |

**Figure 6.5: Comparative stability of H2A.1-H2B and H2A.2-H2B dimer.** (a)(i)(ii) Apparent fraction unfolded (Fapp) obtained from analysis of equilibrium unfolding curves of H2A.1-H2B and H2A.2-H2B monitored through circular dichroism during thermal and urea denaturation. (iii) Apparent fraction unfolded (Fapp) obtained from analysis of equilibrium unfolding curves of H2A.1-H2B and H2A.2-H2B monitored by fluorescence spectroscopy during urea denaturation. (iv) Comparative tabulation of various biophysical parameters obtained by fitting the unfolding curves of H2A.1-H2B and H2A.2-H2B in two state model of unfolding. (b) Temperature of melting (Tm) of various H2A.1 single and double mutants obtained by analysis of thermal denaturation curve monitored through CD spectroscopy. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$). The significance has been calculated with respect to H2A.1.

### 6.3.3 The L51M substitution in H2A at the dimer interface with H2B is primarily responsible for the differential stability

Stability of H2A.1-H2B dimer was determined to be lower than H2A.2-H2B dimer. Subsequently, the contribution of the three residues, in which the two H2A isoforms differ, was investigated on dimer stability by carrying out thermal denaturation with the reconstituted mutant dimers. Tm of the mutant H2A.1-H2B dimers suggest that the L51M alteration had the biggest impact on the stability of the dimers. Mutating 51st residue of H2A.1 into methionine increased the Tm from 50.04$^0$C to 52.30$^0$C and that of H2A.2 (H2A.1 16th + 99th) into leucine decreased the stability by 2.1$^0$C [**Figure 6.5b**].

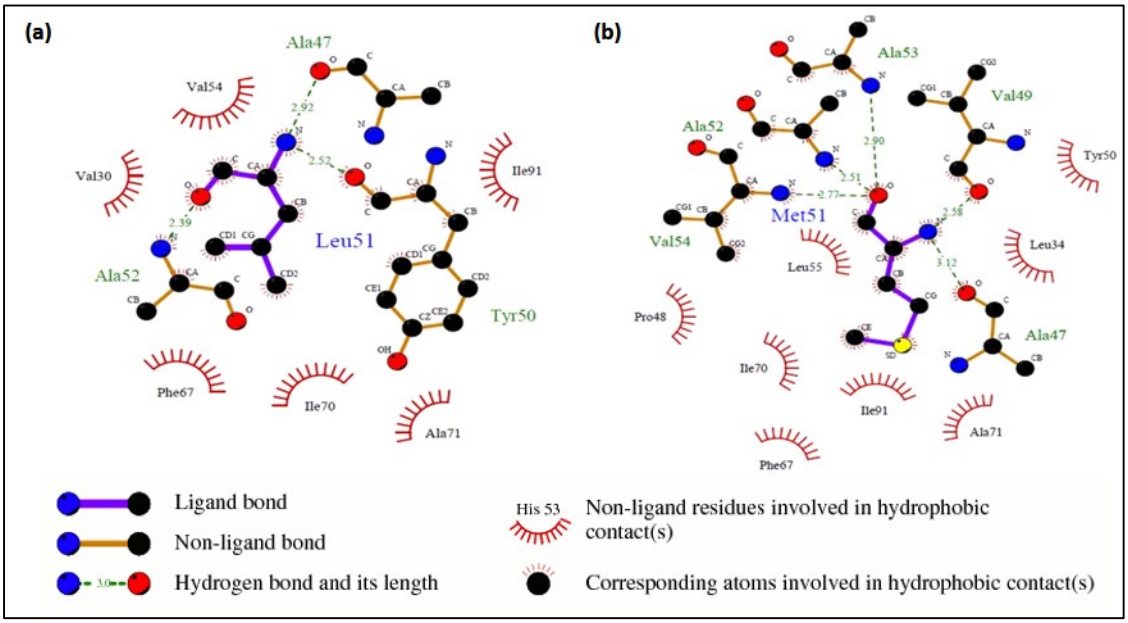



**Figure 6.6: Ligplot depiction of interactions of 51st residue of H2A.** Ligplots depicting the interactions of the 51st residue in the dimer interface of H2A and H2B with (a) leucine and (b) methionine. Methionine 51 is not only able to form more number of hydrophobic interactions but also owing to the orientation of its die chain, it is able to form more number of hydrogen bonds in the dimer interface as compared to leucine.

Leucine and methionine are very similar amino acids. Hence, it was intriguing that why such a substitution would lead to altered stability of H2A-H2B dimer. In order to understand the possible changes on incorporation of methionine which led to increased stability, we carried out *in silico* energy minimisation of H2A.1-H2B and H2A.2-H2B dimer and looked for the differences in interactions. Consistent with its more hydrophobic nature we observed that the presence of methionine at the dimer interface of H2A and H2B led to increased number of hydrophobic interactions [**Figure 6.6**]. Also, presence of methionine results in the formation of higher number of hydrogen bonds due to the favourable orientation of methionine side chain. Altering the 16th and 99th residues of H2A.1 into those of H2A.2 in isolation didn't significantly influence the dimer stability. Although, the R99K substitution, did not contribute significantly towards dimer stability, however, it seemed to have a destabilising effect when mutated along with leucine 51. The R99K alteration increased the stability of H2A-H2B dimer by an additional ~$0.8^0$C.

### 6.3.4 Histone H2A.1 isoform containing nucleosomes are more stable owing to formation of higher number of hydrogen bonds

To understand the importance of alteration in dimer stability in context of chromatin, we investigated the effect of these isoforms on nucleosome stability. The chromatin was incubated in buffers with increasing ionic strength (please see Methods section for more details). Detectable levels of H2A.2 isoform (FLAG tagged) were obtained in the nucleosolic fraction at a lower ionic strength (600mM NaCl) compared to H2A.1 (700mM NaCl) [**Figure 6.7a (i)**].

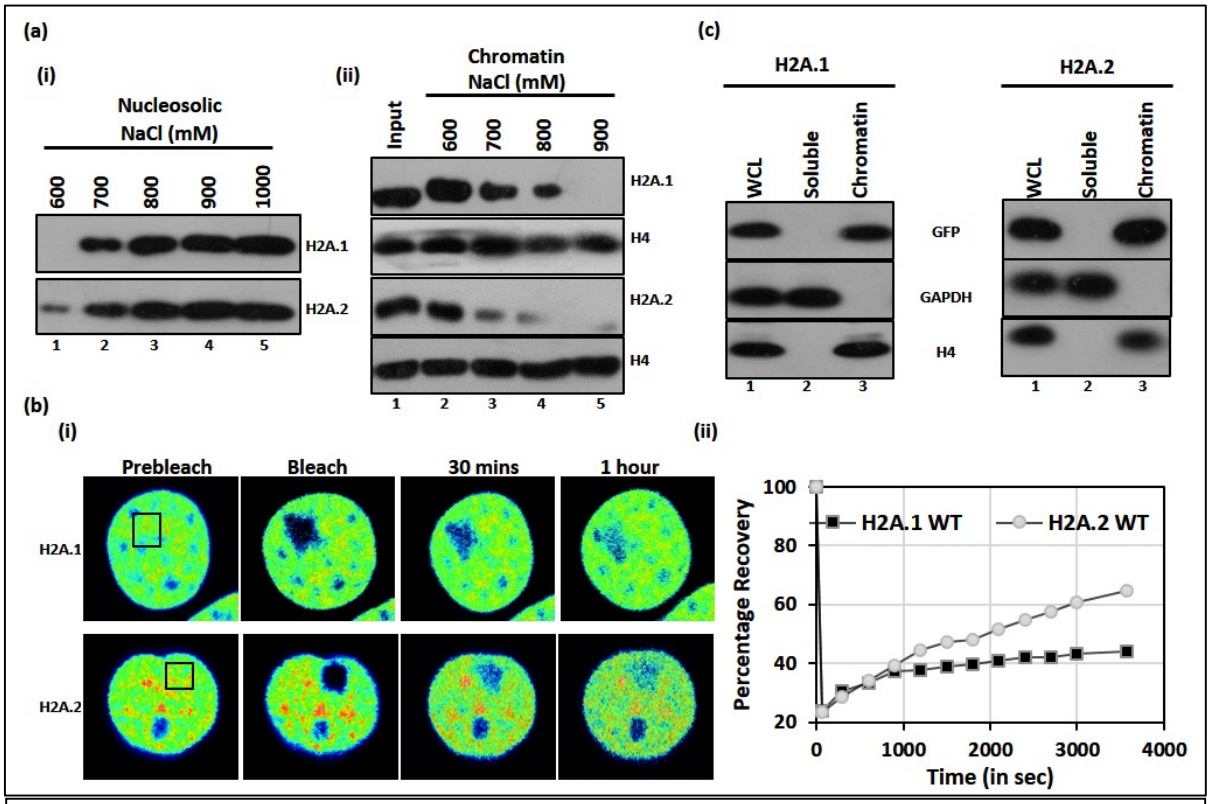



**Figure 6.7: Dynamics of H2A.1 and H2A.2 with chromatin.** (a) Western blot showing the levels of FLAG H2A.1 and H2A.2 in (i) nucleosolic and (ii) chromatin fractions upon incubation of nuclei with increasing ionic strength buffer. (b)(i) FRAP assay performed with CL38 cell line transfected with YFP-tagged H2A.1 and H2A.2. The region indicated with box was bleached with laser. (ii) Graph depicting the percentage recovery of YFP H2A.1 and YFP H2A.2 monitored for a period of 1 hours. (c) Cellular fractionation followed by immunoblotting with the marked antibodies to determine sub cellular distribution of histones.

Analysis of chromatin fraction also indicated that H2A.1 isoform is more resistant to elution from chromatin with increasing ionic strength compared to H2A.2 isoform [**Figure 6.7a (ii)**].

To address whether the more stable association of H2A.1 with chromatin is also reflected in its dynamics, we monitored the recovery of fluorescently tagged H2A.1/H2A.2 in a bleached region of nucleus in a live cell [**Figure 6.7b (i)**]. We documented that the distribution of both the isoforms is similar in soluble and chromatin bound fraction with undetectable levels in soluble fraction [**Figure 6.7c**]. The difference in the dynamics of the two proteins started becoming apparent after 1000sec. The percentage recovery of H2A.1 after 1hour was markedly less (44.14%) compared to H2A.2 (64.7%) [**Figure 6.7b (i)(ii)**] in the FRAP assay, suggesting that H2A.1 is less dynamic than H2A.2 isoform.

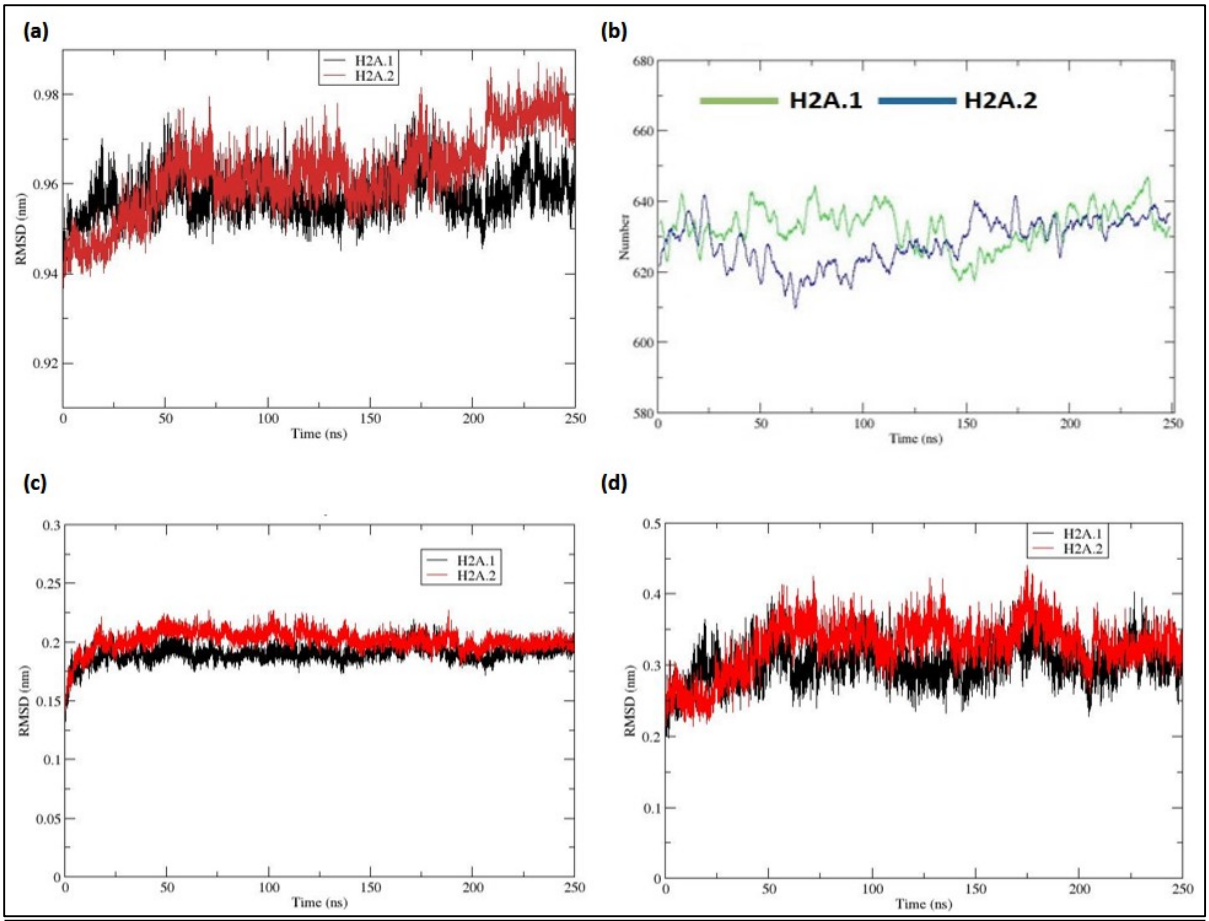


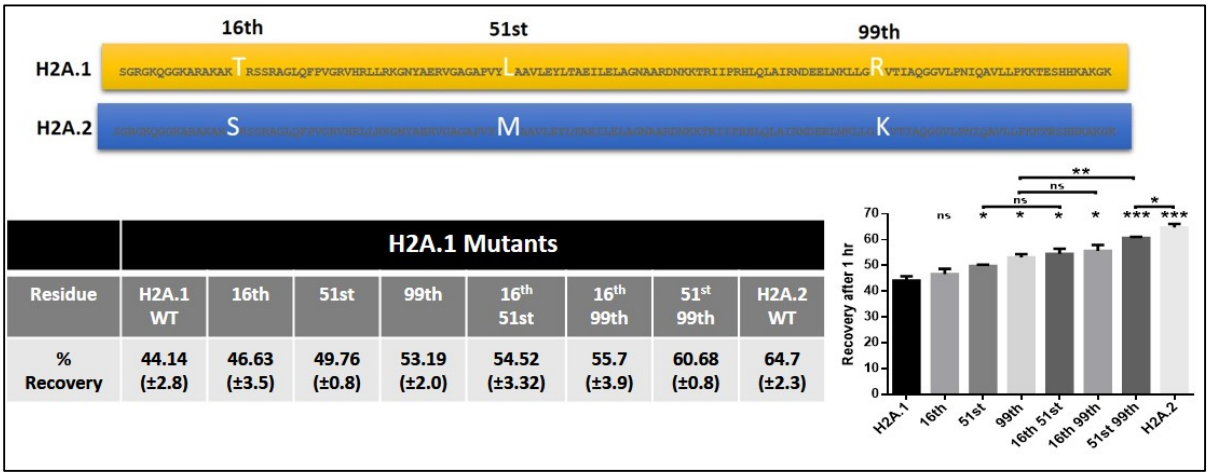

**Figure 6.8: Molecular Dynamic Simulation of H2A.1/H2A.2 containing nucleosomes.** (a) RMSD of H2A.1 (black) and H2A.2 (red) nucleosomes over a span of 250ns. (b) Hydrogen bond analysis of H2A.1 (green) and H2A.2 (blue) containing nucleosome over the span of molecular dynamic simulation (MDS). RMSD of (c) protein and (d) during the course of MDS.

To understand the basis of the increased stability of H2A.1 containing nucleosomes, we carried

out molecular dynamic simulation (MDS). The convergence of the MD simulation in terms of structure was calculated by root mean square deviation (RMSD) with respect to initial structure. The RMSD analysis was in agreement with *in vitro* data with lower RMSD of H2A.1 containing system suggesting that it forms more stable nucleosomes as compared to H2A.2 [**Figure 6.8a**]. Corroboratively, hydrogen bonding analysis shows that during the course of simulation H2A.1 nucleosome has higher number of hydrogen bonds [**Figure 6.8b**]. The RMSD of the octamer and DNA independently showed similar trend [**Figure 6.8c and d**].

### 6.3.5 Leu51 and Arg99 residues leads to increased stability of H2A.1 containing nucleosomes as compared to H2A.2 containing ones

We carried out site directed mutagenesis of H2A.1 followed by FRAP to identify the important alteration(s) out of the three residues in which H2A.1 and H2A.2 differ, that is majorly




| | H2A.1 Mutants | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Residue | H2A.1 WT | 16th | 51st | 99th | $16^{th}$ $51st$ | $16^{th}$ $99th$ | $51^{st}$ $99th$ | H2A.2 WT |
| % Recovery | 44.14 (±2.8) | 46.63 (±3.5) | 49.76 (±0.8) | 53.19 (±2.0) | 54.52 (±3.32) | 55.7 (±3.9) | 60.68 (±0.8) | 64.7 (±2.3) |

**Figure 6.9: FRAP with H2A.1 mutants.** Table (left) and histogram (right) showing the percentage recovery of YFP tagged H2A.1, H2A.2 and mutants of H2A.1 after 1hour. The error bars depicted in the figure are of SEM (standard error of mean). ns (P > 0.05), * (P ≤ 0.05), ** (P ≤ 0.01), *** (P ≤ 0.001), **** (P ≤ 0.0001).

responsible for the difference in dynamics. The R99K substitution, which is involved in the interaction with H4 tails in the NCP, independently brought about the most drastic increase (from 44.14 to 53.19 percent recovery) in the dynamics of H2A.1 followed by L51M (from 44.14 to 49.76 percent recovery) [**Figure 6.9**]. Mutating both the 51st and 99th residue

together, of H2A.1 into the corresponding residues of H2A.2 led to almost similar dynamics as observed for H2A.2. Mutating the 16th residue did not have significant impact on H2A.1 dynamics.

We wanted to understand how substitutions with very similar amino acids brought about changes in nucleosome stability. Analysis of the number of hydrogen bonds formed by the residues at the three positions with nearby residues throughout the simulation time of 250ns were performed for both H2A.1 and H2A.2 containing nucleosomes. The data suggested that the 51st and the 99th residue majorly participate in the formation of hydrogen bonds with very

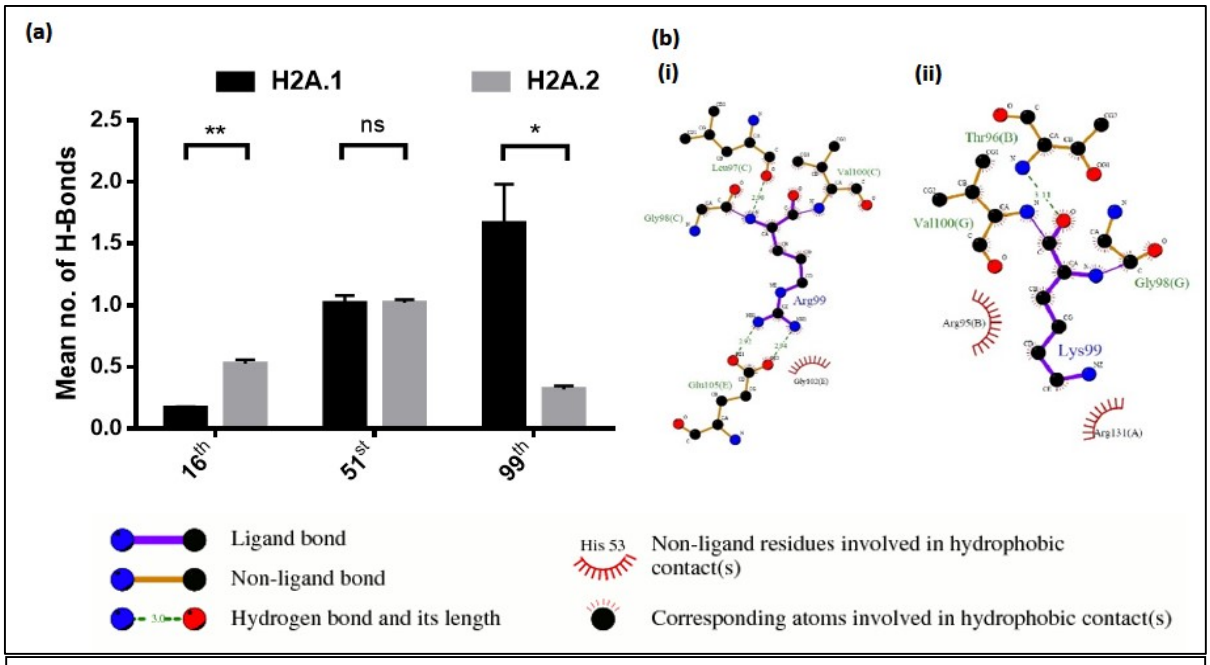



**Figure 6.10: Hydrogen bonding analysis of H2A.1/H2A.2 containing nucleosomes.** (a) Graph depicting number of hydrogen bonds formed by three differential residues (16th,51st and 99th) amongst H2A.1 and H2A.2 during MDS (b) (i) Ligplot depicting the hydrogen and hydrophobic interaction of 99th residue of H2A.1 with the neighbouring residues. (ii) Ligplot depicting the hydrogen and hydrophobic interaction of 99th residue of H2A.2 with the neighbouring residues. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$).

less contribution from the 16th residue [**Figure 6.10a**]. Importantly, arginine at 99th position in H2A.1 system forms more number of hydrogen bonds than lysine [**Figure 6.10a**]. The ligplots depicts the hydrogen and hydrophobic interactions between the 99th and nearby residues [**Figure 6.10b**].

### 6.3.6 Principal component analysis suggests that H2A.1 containing nucleosome structures are better correlated

Principal component analysis (PCA) was carried out to discriminate relevant conformational changes in a protein from the background of atomic fluctuations. **Figure 6.11a (i)** shows cross correlation plot for protein octamer for H2A.1 and H2A.2. In H2A.1 nucleosome, nearby interacting chains shows positive correlation while distant region are showing negative

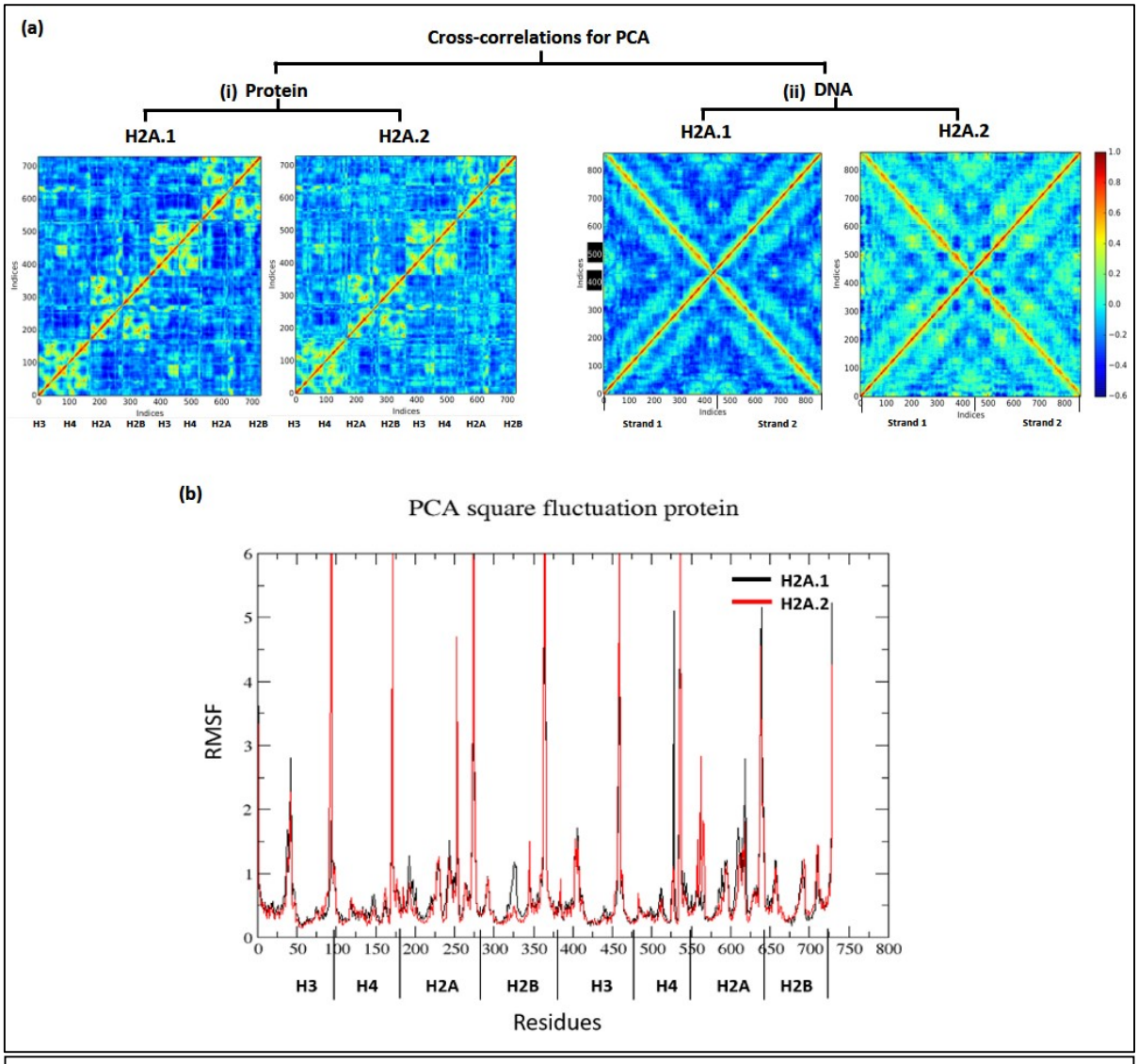



**Figure 6.11: Principle component analysis (PCA) of H2A.1/H2A.2 containing nucleosomes.** (a) Cross correlation plots showing the atomic fluctuations at (i) protein and (ii) DNA for H2A.1 and H2A.2 containing nucleosome. The blue colour indicates negative cross correlation while red colour indicate positive cross correlation. (b) Graph depicting the comparative PCA square fluctuations of DNA of H2A.1 and H2A.2 containing nucleosome.

correlation. Generally, positive correlation is seen in nearby residue with synchronous motion, whereas negative correlation is observed between distantly interacting residues with asynchronous motion. H3 and H4 together forms a dimer, therefore H3 shows a positive correlation for H4 while negative for rest of histone chains. Similarly, H2A shows positive correlation for H2B. The pattern of correlation observed with H2A.1 or H2A.2 containing nucleosome is same for nearby chains, however the correlation between H2A.2 and H2B (system 2) is slightly less positive compared to H2A.1 and H2B containing nucleosome (system 1). Also, in system 2 there is less negative correlation between distant chains. Thus, comparing cross correlation data with PCA square fluctuation [**Figure 6.11b**] it can be seen that negative correlated motion between distant chains are providing a rigidity and stability to the H2A.1 nucleosome. Similarly, the cross-correlation of DNA [**Figure 6.11a (ii)**] follows the same trend.

### 6.3.7 Incorporation of H2A.1 isoform does not impart structural alterations to chromatin

The difference in the cross correlation plot for DNA of H2A.1 and H2A.2 containing system [**Figure 6.11b**] prompted us to investigate whether there might be structural alteration in the DNA on incorporation of H2A isoforms. However, overlaying structures of different time points of simulation suggested that there is no prominent structural alteration [**Figure 6.12a**]. To address whether there is any global change in nucleosome spacing or chromatin accessibility, the chromatin isolated from H2A.1/H2A.2 overexpressing CL38 cell line was subjected to MNase digestion for different time points [**Figure 6.12b**]. MNase cuts at the linker DNA and results in appearance of fragments of approximately 200bp corresponding to mononucleosomes (nucleosome core plus the linker). With increasing incubation, the intensity of the trail below the chromatin increased with appearance of mononucleosomes at 6min time

point [**Figure 6.12b**]. Further digestion (8 and 10min time points) clearly shows the characteristic ladder pattern expected from MNase digestion assay [**Figure 6.12b**]. The digestion pattern was almost identical suggesting that the chromatin structure and accessibility does not alter significantly on incorporation of H2A.1 and H2A.2 isoforms. However, more sensitive experiments are needed to be performed to rule out the possibility of very minute changes that might occur in the chromatin on incorporation of the histone isoforms.

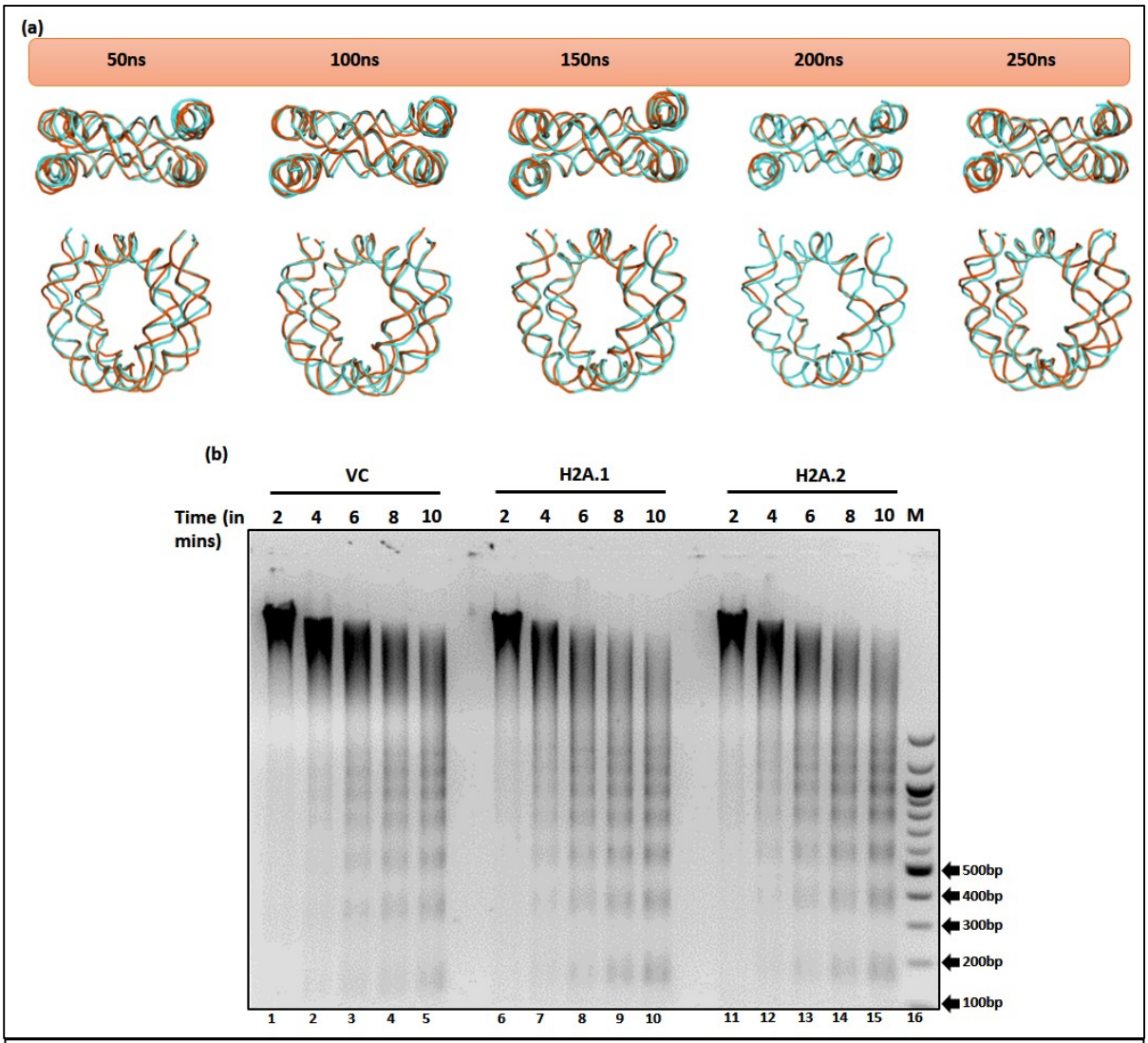



**Figure 6.12: Effect of H2A.1/H2A.2 incorporation on chromatin architecture.** (a) Overlay of DNA of H2A.1 and H2A.2 containing nucleosomes at different time points of MDS. (b) Agarose gel electrophoresis (1.8%) of Micrococcal nuclease digested chromatin isolated from H2A.1 and H2A.2 overexpressing CL38 cells. VC represents vector control.

## 6.4. Discussion

The non-redundancy of histone isoforms has made the understanding of the epigenetic regulations employed by cells more complex, nevertheless, interesting. Previous studies have attempted to elucidate the role of H2A1C isoform in context of cancer[4], however, no report exists which provides insights into the basic non-redundant role of H2A isoforms which may contribute to the attainment or persistence of a particular physiological or pathological state. We for the first time elucidate that even the homomorphous sequence divergent forms of histones can alter nucleosome dynamics without major alterations in the chromatin organisation.

We found the L51M alteration to have the most significant impact on H2A-H2B dimer stability. The difference observed between H2A.1-H2B and H2A.2-H2B dimer stability is subtle compared to the change brought about by histone variants like H2A.Z[102]. This is consistent with previous reports where L to M replacements altered the protein stability only by 0.4 to 1.9 kcal/mol[196]. Possibly, the ubiquitous abundance of H2A isoforms in the genome, as opposed to variants, makes this difference significant to induce alteration in epigenetic regulation. Probably, the cell uses the histone variants to bring about major changes in gene regulation and has evolved histone isoforms for subtle modulations of chromatin mediated processes at nucleosomal level.

Stability alteration of proteins on leucine to methionine substitution or vice-versa has been suggested to be context dependent[197]. Although, the van der Waals volume occupied by leucine is the same as that for methionine, two opposing forces are at play when leucine to methionine substitution occurs. The substitution of methionine with leucine within the interior of a protein is expected to increase stability both because of a more favourable solvent transfer term as well

as the reduced entropic cost of holding a leucine side chain in a defined position. At the same time, this expected beneficial effect may be offset by steric factors due to differences in the shape of leucine and methionine[197]. Our simulation studies suggest that presence of methionine in the dimer interface not only leads to increased hydrophobic interactions but also results in the formation of higher number of hydrogen bonds due to the favourable orientation of methionine side chain.

Interestingly, besides the involvement of the 51st residue in determining H2A-H2B dimer stability, a trend of synergistic effect was seen when residue 99th is mutated along with the 51st. This was intriguing as residue number 99th is not present in the dimer interface. Arginine (in H2A.1) and lysine (in H2A.2) are positively charged residues and play important roles in stabilising proteins by forming ionic interactions and hydrogen bonds in the protein as well as with water[198]. Notably, the guanidine group in arginine allows interactions in three possible directions through its three asymmetrical nitrogen atoms in contrast to only one direction of interaction allowed for lysine. Owing to this difference in geometry of the two amino acids, arginine might have a more stabilising effect on proteins over lysine which is corroborated by previous reports[197]. The presence of arginine in H2A.1 probably stabilises the H2A monomer more as compared to lysine in H2A.2 which thermodynamically makes the H2A.1-H2B dimer less stable. Further, the ability of arginine to form higher number of H-bonds compared to lysine was also reflected in our FRAP assay and MDS studies.

As discussed earlier, the altered stability of H2A-H2B dimer may have its implication in the nucleosome stability. Previously reported MDS studies focussed on histone octamer-DNA interactions revealed that the H2A-H2B dimer is the least stable part of the nucleosome and could make a significant contribution to the histone-DNA interaction dynamics[204]. We found that H2A.1 isoform gives rise to more stable nucleosomes although the H2A.1-H2B dimers

were less stable. This is consistent thermodynamically as a less stable dimer would favour a more stable nucleosome. Similar to our observations, for H2A.Z variant it was reported that H2A.Z-H2B dimer was unstable as compared to the canonical H2A-H2B[102], however, the nucleosome was found to be more stable[205]. A more stable nucleosome is expected to cause hindrance to chromatin mediated processes like transcription, replication and repair. Previously, the HAR domain of H2A, which comprises of the residue 16 to 20 of the N-terminal tail, has been implicated in transcriptional repression owing to its ability to govern nucleosome dynamics by interacting with the minor groove of DNA[194]. The HAR domain, although, was initially identified in yeast, it was later shown to be important in humans as well[4]. In addition, S16A substitution at the HAR domain was found to disrupt its repressive ability[4]. Our data shows that S16T substitution does not significantly alter nucleosome dynamics.

## 6.5. Conclusion

In conclusion, H2A.1-containing nucleosomes are more stable owing to M51L and K99R substitutions that have the most prominent effect on cell proliferation, suggesting, that the nucleosome stability is intimately linked with the physiological effects observed. Possibly, the increased nucleosome stability resulting from H2A.1 incorporation, contributes to the contextual alteration in global gene expression pattern, which collectively promotes the attainment of different physiological states.

# VII. H2AUB119 IN HEPATOCELLULAR CARCINOMA

## Chapter 7

*Preliminary work done to identify the factors on which the context dependent role of H2A.1 depends*

**This work was published in the journal "Experimental Biology and Medicine"**

## 7.1. Introduction

Histones, which organize chromatin, often undergo a variety of post-translational modifications (PTMs)[206]. These modifications function either by disrupting chromatin contacts or by affecting the recruitment of various proteins to the chromatin and thereby, regulating transcription[206]. The context dependent effect of H2A.1 observed by us prompted us to investigate the possible role of histone modifications in this regard.

Histone H2A mono-ubiquitination is one such functionally important modification. Histone H2A is monoubiquitylated on lysine 119 in its C-terminal tail[207]. Histone H2A is the first protein identified to be modified by ubiquitin in cells[208] and is one of the most abundant ubiquitinated proteins in the nucleus[209,210]. Interestingly, H2A ubiquitin ligases were found in transcription repressor complexes, such as the PRC1, BCoR, E2F6.com-1, and 2A-HUB[211,212,213]. H2Aub by PRC1 (Polycomb Repressive Complex 1) type complexes creates a binding site for Jarid2–Aebp2–containing PRC2 and promotes H3K27 trimethylation on H2Aub containing nucleosomes[214]. H2Aub is thus an integral component of a positive feedback loop establishing H3K27me3 mark and maintaining repressive state of the chromatin[215]. Further, it was identified that RING1B mediated H2Aub is required for polycomb targeted gene silencing[216]. In addition, the fact that H2A deubiquitinases (DUBs) like Usp16, Usp21, Usp22 and 2A-DUB are usually required for genes activation provides a second line of evidence for the gene silencing function of H2Aub[217,218,219]. Notably, monoubiquitinated histone H2A and H2B were found to be dramatically down-regulated in prostate and breast tumors, respectively[220,221]. Despite its abundance, apart from the study in which mono-ubiquitinated histone H2A was markedly down-regulated in prostate cancer[220], this functionally important mark has not been investigated in other cancers. In addition, the drop in H2Aub is a pre-requisite for increase in H3S10 phosphorylation (H3S10p) during S to G2/M

transition *in vitro*[222]. Interestingly, H3S10p itself is a mark that is indispensable for cellular transformation[223]. Whether, an antagonistic effect of these modifications on each other may exist *in vivo*, has not been reported.

We began this part of the work by investigating the association of H2A isoform expression with the level of H2Aub119 in hepatocellular carcinoma model system in rat. We also investigated the association of H2Aub119 with the histone marks H3S10p and H4ac.

## 7.2. Methods

### *7.2.1 Animal experimentation*

Same as described in the chapter V (page number 124).

### *7.2.2 Histone isolation*

Same as described in the chapter V (page number 129).

### *7.2.3 AUT-PAGE analysis*

Same as described in the chapter V (page number 132).

### *7.2.4 Western blot analysis*

Same as described in the chapter V (page number 131). For western, histones were resolved on SDS–PAGE, transferred, and probed with respective antibodies from Millipore or Abcam or CST (see *Annexure I*) according to the manufacturer's instructions.

### *7.2.5 MALDI-TOF/TOF mass spectrometry*

Core histones spots on were subjected to in-gel digestion for identification by MALDI-MS fingerprint mapping as described earlier[224,225]. Core histones separated on SDS-AUT two-dimensional gel were visualized by SDS-silver staining method. Spots that appeared on the gel

were numbered and subjected to MALDI-TOF/TOF mass spectrometry in a two-step process. First the proteins were digested using trypsin protease. Then, the tryptic peptides were subjected to MALDI-MS and the MALDI-MS spectrum was compared to database for identification of histone protein. See *Annexure IV* in *Appendix* for details of chemicals required.

a) *In gel digestion:* Core histones spots on gel were subjected to in-gel digestion for identification by MALDI-MS fingerprint mapping as described. Core histone spot of interest was excised and cut into small pieces (~1$mm^3$) using a scalpel and placed into microcentrifuge tube. Freshly prepared reducing solution was added ten volumes of the gel pieces and vortexed until brownish stain disappeared. Gel pieces were rinsed few times with water to stop the reaction. Next, the gel pieces were washed with 25mM ammonium bicarbonate for 15min. Gel pieces were then dehydrated with 25mM ammonium bicarbonate in 50% v/v acetonitrile. Washing/rehydration step was repeated and gel pieces were dried in vacuum centrifuge.

Gel pieces were then covered with 10mM DTT solution and reduced for 1hr at 56$^0$C. Gel pieces were cooled to room temperature and DTT solution was replaced with 55mM iodoacetamide solution. Gel pieces were incubated at room temperature for 45min in dark with occasional vortexing. Gel pieces were washed, dehydrated and dried as mentioned in the previous paragraph. Gel pieces were rehydrated in 1 volume of 20µg/ml Trypsin in 25mM ammonium bicarbonate by vortexing for 5min. The gel pieces were covered in minimum volume of 25mM ammonium bicarbonate and incubated overnight at 37$^0$C.

Peptides were recovered as follows. Two volumes of water were added to the gel pieces, vortexed for 5 min. and sonicated for 5min. The peptide solution was removed and transferred to a fresh microcentrifuge tube. Two additional extractions were performed with 5% v/v TFA in 50% v/v acetonitrile. Recovered peptides were dried completely in vacuum centrifuge and

reconstituted in 10µl of 5% v/v TFA in 50% v/v acetonitrile. Recovered peptides were stored at -20°C until required for sample preparation for MALDI-MS analysis.

b) *Sample preparation and analysis:* One microliter of recovered peptides and 1µl of peptide matrix solution (20mg/ml HCCA in 0.1% v/v TFA in 50% v/v acetonitrile) were pipetted onto the sample target (plate), mixed and allowed to dry. External calibration was prepared by mixing peptide standard mixture and peptide matrix solution similarly. Sample target (plate) was inserted into the MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Ultraflex II) and analysed as recommended by manufacturer.

In brief, mass spectra were acquired on reflector ion positive mode. Database searching for protein masses was carried out using MASCOT search engine (version 2.2.03) by comparing peptide masses with those in NCBInr protein database *(database version: NCBInr_20080812.fasta)* in rattus spp. The searches were carried out with trypsin digestion, one missed cleavage, fixed carbamidomethylation of cysteine residues and optional oxidation of methionine with 100ppm mass tolerance for monoisotopic peptide masses.

### *7.2.6 Cell line maintenance*

Experiments were performed with the neoplastic CL38 and and pre-neoplastic CL44 cells of rat liver origin. The methodology for their maintenance is same as described in chapter V (page number 126).

### 7.2.7 *HDACi treatment*

Histone deacetylase inhibitor (HDACi), Trichostatin A was dissolved in absolute ethanol to prepare stock solution of 1mM. The cells were treated with a working concentration of 200nM of TSA for 48hrs. 1000 cells were seeded in 96 well plates. Rest of steps were performed as per MTT assay described in chapter V (page number 128).

### *7.2.8 cDNA synthesis, RT-PCR and Real-time PCR*

Described in chapter III (page number 81). See *Annexure III* in *Appendix* for primer sequences.

### *7.2.9 MTT assay*

Same as described in the chapter V (page number 128).

## 7.3. Results

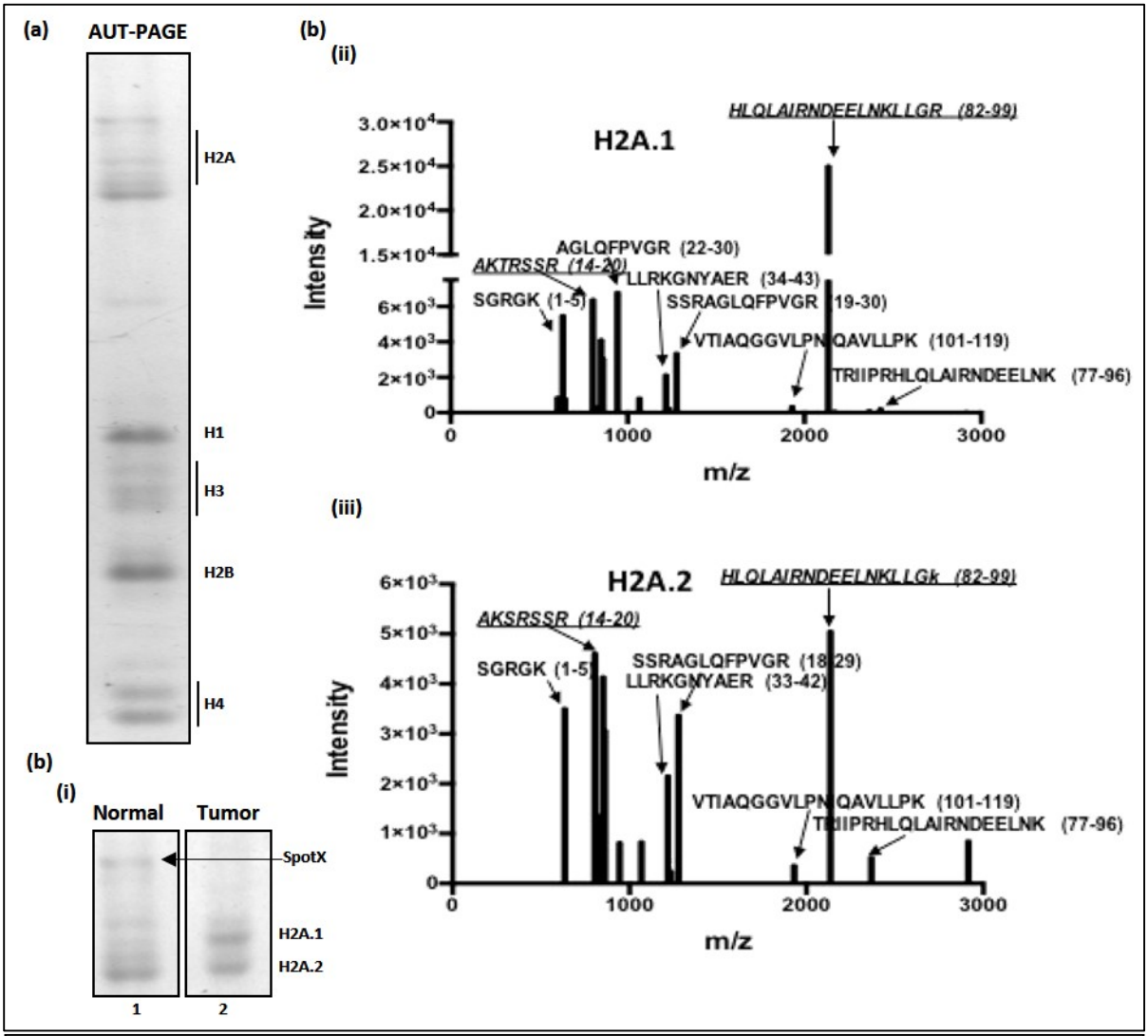



**Figure 7.1: H2Aub levels in normal and tumour tissues.** (a) Single dimensional AUT-PAGE profile of purified histones from normal tissue. Region marked spans the mobility of respective histones. (b) (i) Comparative AUT profile of normal and tumour tissues revealed differential levels of Spot X in enlarged H2A region. Mass spectrometry analysis of spot X excised from AUT-PAGE identified it to be b (ii) H2A.1 and (ii) H2A.2. The peptides specific for H2A.1 or H2A.2 are underlined. AUT-PAGE- Acetic Acid-Urea-Triton Polyacrylamide Gel Electrophoresis.

### 7.3.1 H2A ubiquitination markedly decreases in hepatocellular carcinoma

Histones were isolated from normal and tumor tissues and were resolved on AUT-PAGE [**Figure 7.1a**]. A prominent decrease in Spot X was observed in tumour sample [**Figure 7.1b (i)**]. Increase in H2A.1 and decrease in H2A.2 were also observed in tumour which we have previously reported [**Figure 7.1b (i)**][3]. MS analysis of Spot X suggested that the band corresponds to H2A [**Figure 7.1b (ii)(iii)**]. Both H2A.1 and H2A.2 were picked up. Upshift of the band of H2A suggested the occurrence of possible mono-ubiquitination. This was confirmed by subjecting the region of the AUT-PAGE to western blotting with anti-H2Aub [**Figure 7.2 (a)(b)**]. We included liver cell lines derived from the same model system for our studies, CL44 (pre-neoplastic), and CL38 (neoplastic). Similar changes in H2Aub levels were observed on comparison of histones isolated from both the cell lines [**Figure 7.2 (c)(d)**]. Our results suggest that overall there is decrease in the level of H2Aub in NDEA-induced liver cancer model.

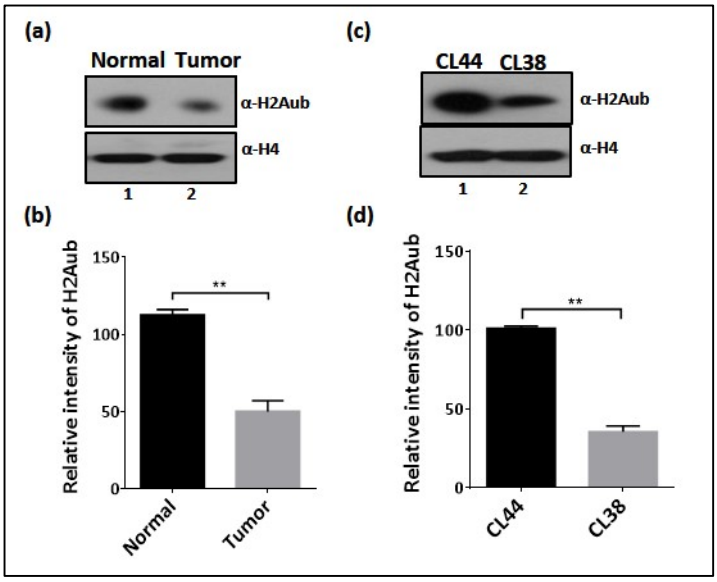



**Figure 7.2: H2Aub levels in normal and transformed states.** Western blotting with anti-H2Aub119 antibody of isolated histones revealed (a) decreased H2Aub119 levels in tumour compared to normal, and also (c) reduced levels of H2Aub119 in CL38 cells (neoplastic) compared to CL44 (pre-neoplastic) cells. H4 was used as a loading control. (b) & (d) Graphs showing quantification of H2Aub119 levels obtained from densitometry analysis of three independent blots similar to the ones depicted in (a) and (c). Each histogram represents mean densitometry + SEM (**p, 0.01).

### 7.3.2 Decrease in H2Aub is probably due to Usp21 and associates with upregulation of lipocalin 2

We looked for the levels of Usp21 as it was demonstrated to bring about H2A119 deubiquitination during liver regeneration[226]. Both in tumour tissue and neoplastic cell line CL38 a marked increase in the expression level of Usp21 was observed by quantitative [**Figure 7.3a (i)**] and semi-quantitative PCR [**Figure 7.3a (ii)**]. We did not find any change in the expression level of the major H2A ubiquitinase RING1 [**Figure 7.3a (iii)**], suggesting that the deubiquitination observed might be primarily owing to upregulation of the deubiquitinase.

H2Aub has been previously shown to localize to the promoter of genes like lipocalin 2 and repress their expression[226]. Hence, we checked the expression level of lipocalin 2 gene LCN2. A marked increase in the expression of LCN2 was observed in the CL38 cells and tumour tissue compared to their normal counterparts by quantitative [**Figure 7.3b (i)**] and semi-quantitative PCR [**Figure 7.3b (ii)(iii)**].

### 7.3.3 Increased cell proliferation may contribute to loss of H2Aub in tumor

Ubiquitination of H2A is dynamic, as suggested by the observations that global levels of

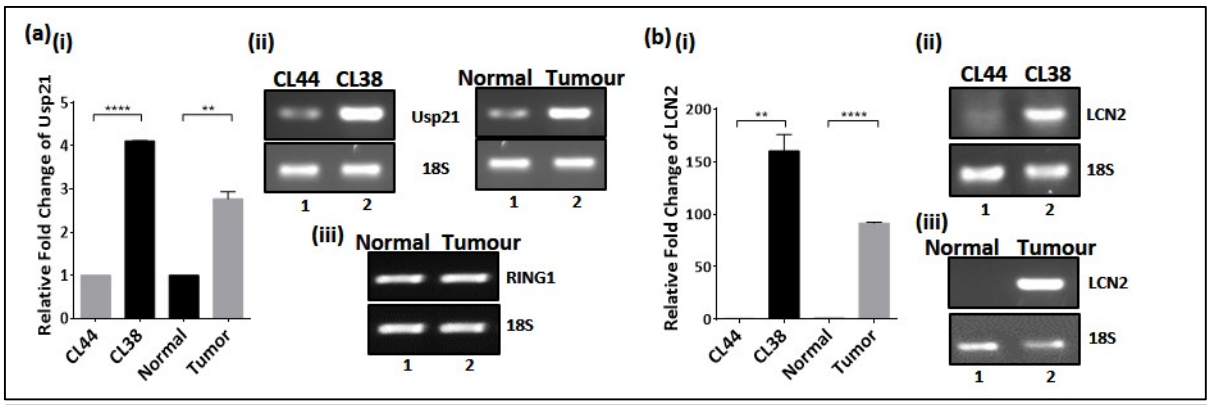



**Figure 7.3: Expression of Usp21 and LCN2 in normal vs transformed samples.** (a) Levels of Usp21, a H2A deubiquitinase enzyme in rat liver cell lines (CL44 and CL38) and tissues (Normal and Tumour). (i) Real time PCR and (ii) RT-PCR. (iii) RT-PCR of H2A ubiquitinase, RING1 levels in tissues. 18S rRNA was used as a loading control. (b) Level of lipocalin 2 (LCN2) in rat liver cell lines and tissues (i) Real time PCR and (ii) RT-PCR.

H2Aub vary during the cell cycle[222]. Drop in the level of H2Aub has been shown to be a prerequisite for increase in the mitotic mark H3S10p during S to G2/M transition of cells[222]. Along with H3S10p increase, H3-H4 hypoacetylation is also seen during mitosis[227]. Indeed, when we looked into levels of H3S10p and H4ac in both tissues and cell lines [**Figure 7.4a (i)(ii)**], we observed a marked increase of the H3S10p as opposed to decrease in H2Aub and H4ac. We further investigated the possible relation between the levels of H2Aub and cell proliferation. In CL38, increased percentage of cells in S and G2/M population was observed as compared to CL44 [**Figure 7.4a (iii)(iv)**]. Consistent with this, H2Aub was low in CL38 cells and in tumour [**Figure 7.4a (i)(ii)**]. Next, we enriched CL38 cells in G1-phase of the cell cycle by serum starvation [**Figure 7.4b (ii)**]. An increase is H2Aub was found in G1-enriched

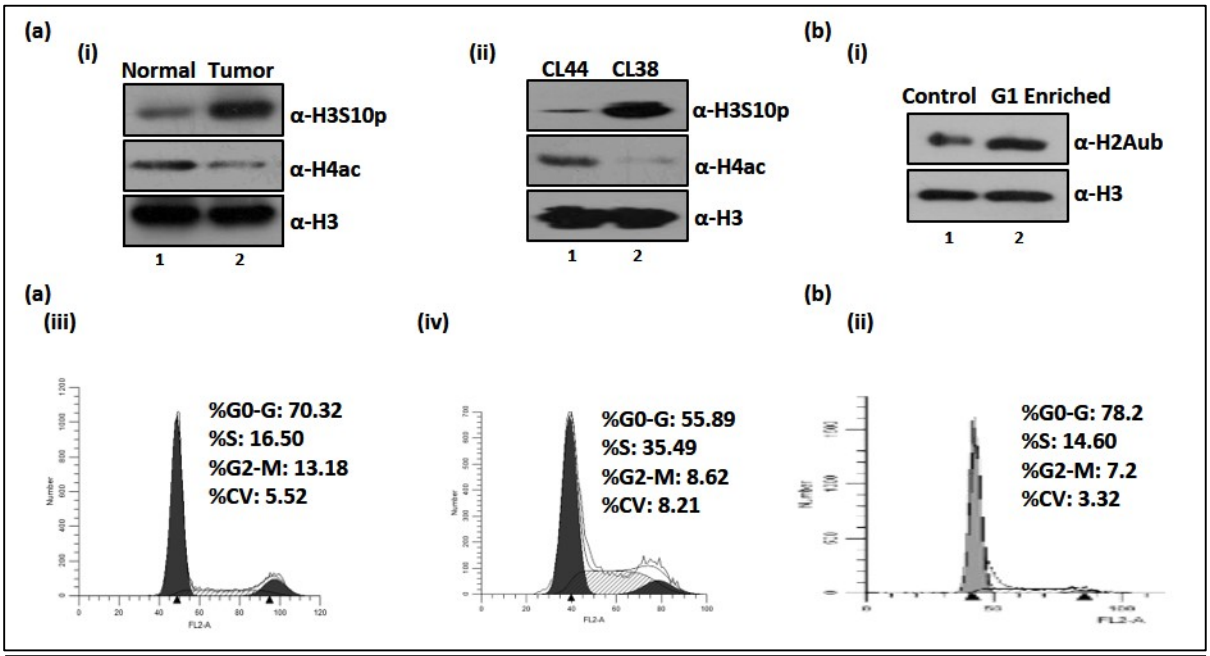



**Figure 7.4: Effect of cell cycle on H2Aub.** (a) Western blotting with anti-H3S10p and anti-H4ac antibody of isolated histones from (i) normal and tumour tissues and (ii) CL44 (pre-neoplastic) and CL38 cells (neoplastic). H3 was used as a loading control. (iii) Cell cycle profile of (iii) CL44 and (iv) CL38 cells. (b) (i) Western blotting with anti-H2Aub antibody of isolated histones from control and G1 enriched CL38 cells. (ii) Cell cycle profile of serum starved CL38 cells showing G1 enrichment.

cells [**Figure 7.4b (i)**] strengthening the inverse relationship between H2Aub and cell proliferation.

As we observed decrease in H4ac in tumour, we wanted to investigate the effect of increase in H4ac on H3S10p and H2Aub. As histone deacetylase inhibitors (HDACi) are known to decrease cell proliferation[228], we treated CL44 and CL38 cells with HDACi Trichostatin A. Histones were isolated 48hrs post drug-treatment and the levels of histone modifications were observed by immuno-blotting [**Figure 7.5a, b**]. The activity of the drug was confirmed

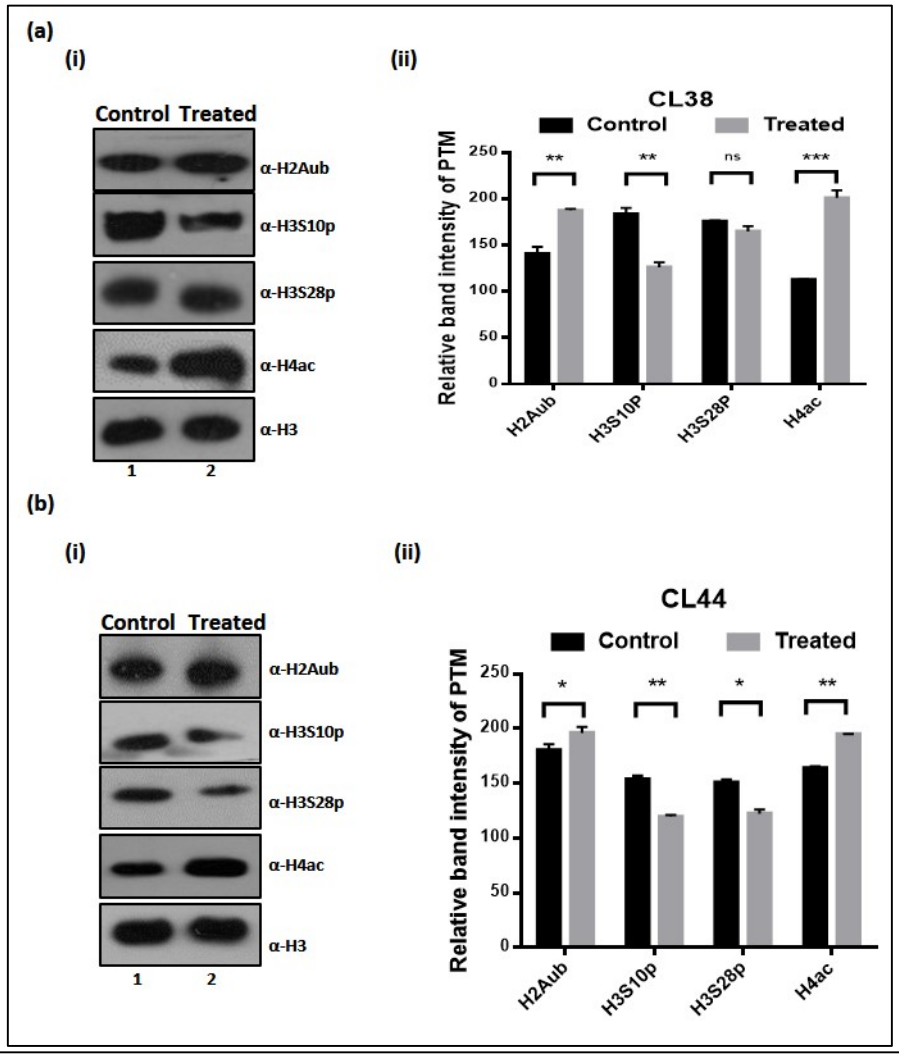



**Figure 7.5: Effect of HDACi treatment on H2Aub.** Histone PTM profile with indicated antibodies on 48hr treatment of (a) CL38 cells and (b) CL44 cells with HDACi TSA. (ii) Graphs showing quantification of histone PTM levels obtained from densitometry analysis of three independent blots similar to the ones depicted in (i). HDACi- Histone deacetylase inhibitor, TSA- Trichostatin A. The error bars depicted in the figure are of SEM (standard error of mean). ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$).

by increase in the histone H4ac levels [**Figure 7.5a, b**]. An increase in H2Aub was observed with a decrease in the level of H3S10p in both CL38 and CL44 cells [**Figure 7.5a, b**]. A

decreasing trend was observed for H3S28p (another mitotic mark) as well in both the cell lines [**Figure 7.5a, b**].

## 7.4. Discussion

Reduction in H2Aub has been reported in human pancreatic cancer[220] but otherwise this mark is poorly studied in cancers. In this study, we found marked decrease in H2Aub and increase in expression of gene lipocalin 2 in NDEA-HCC model. Interestingly, lipocalin 2 has been proposed to increase migration and metastasis of breast cancer cells[229]. We found the deubiquitinase Usp21 to be profoundly upregulated in HCC and no alteration in the level of ubiquitinase RING1. However, role of other ubiquitinases and deubiquitinases cannot be ignored. Along with H2Aub decrease, we have also observed a decrease in H4ac and increase in H3S10p. Our data thus suggests for the first time an inverse correlation between the global levels of cellular transformation mark H3S10p and H2Aub *in vivo*.

Upon treatment with HDACi, we found an increase in the level of H2Aub, H4ac and drop in H3S10p. Similar pattern was observed in both CL38 and CL44 cells, however, more pronounced effect was observed in CL38 cells. This might probably be due to the higher HDAC activity in transformed cell line. Previous studies have shown a decrease in H2Aub upon treatment with HDACi, valproic acid[230,231]. These differences in observation can be attributed to the use of different types of HDACi in the study. TSA is a class I and II specific inhibitor, without any specificity towards either of the enzymes. On the other hand, VPA used in previous report has been shown to have selectivity towards HDAC2 amongst other HDACs[232]. Further, whether this increase in H2Aub seen upon TSA treatment is solely because of change in cell cycle profile of cells and H3S10p or whether there is involvement of cross-talk between the histone acetylation mark and H2Aub would be an interesting prospect to investigate.

## 7.5. Conclusion

In summary, we here show that NDEA administration leads to stable transformation of liver to HCC. This is associated with changes in shape and size of the nucleus along with increase in cell proliferation markers. At the molecular level, there is a marked decrease in the level of H2Aub in HCC. The loss of H2Aub also appears to be partly due to increase cellular proliferation and treatment of cells with HDACi increases H2Aub in cells. Usp21 like in case of liver regeneration probably mediates this loss. This leads to upregulation of lipocalin 2, a growth factor whose role has been implicated in cancer progression. Future work should focus on understanding the functional implication of histone marks and whether the loss of H2Aub is a common hallmark of cancers.

# VIII. CONSOLIDATED DISCUSSION AND CONCLUSION

Chapter 8

## 8.1. Discussion

Wealth of information has emerged over the past two decades demonstrating the functionally important role of histones apart from their fundamental role in packaging the genome. Remarkably, new discoveries are continued to be made unravelling the functions of these highly conserved proteins which were previously thought to be "passive". Not only do histone PTMs and variants regulate gene expression patterns for proper development of an organism but also in cancer, which is often imagined as "development gone wrong", a variety of histone PTM and variant mediated mechanisms are perturbed leading to silencing of tumor suppressor genes and activation of oncogenes. The work described in this thesis not only re-emphasizes the functionally important role of histones but also, demonstrates that substitutions with even very similar amino acids, like that seen in histone isoforms, can have marked consequences on the chromatin dynamics that may contribute to attainment of specific patho-physiological sates.

Liver is the metabolic hub of the body and is involved in detoxification of substances. Prolonged insult, with a carcinogen for example, may lead to altered gene expression pattern in hepatocytes. Hence, through xenograft transplantation experiments it was important to demonstrate that the observed change in the expression profile of H2A.1 was indeed a stable change that is acquired during the process of transformation of liver cells and is not merely a response to insult by NDEA.

On resolving histones on AUT-PAGE we observed that along with the increase in intensity of H2A.1 band there was a marked decrease in the intensity of band corresponding to H2Aub119 (H2Aub) in tumor tissue. This raised the possibility that H2A.1 specifically undergoes mono-ubiquitination in normal liver cells. Probably, deubiquitination of H2A.1 in tumor tissue leads to the perception of increased abundance of the unmodified H2A.1. However, mass

spectrometry analysis suggests that both H2A.1 and H2A.2 undergo mono-ubiquitination and it is not an H2A.1 specific phenomenon. In addition, we also do find an increased transcript level of H2A.1 further suggesting that it is indeed upregulated in HCC. Interestingly, mono-ubiquitinated histone H2A was markedly down-regulated in human prostate cancer[220]. Also, this mark has been reported to significantly reduce during the process of liver regeneration[233]. Reduction of H2Aub at the promoter of LCN2 leads to derepression of this gene[233]. Consistent with these reports, we also found a marked up regulation of LCN2 in tumour tissue and transformed cell line with the decrease in H2Aub. Interestingly, lipocalin 2 has been proposed to increase migration and metastasis of breast cancer cells[229]. It would be interesting to investigate the status of H2Aub119 in breast cancer.

Ectopic overexpression of H2A.1 leads to increase in cell proliferation but the association appears to be context dependent. Remarkably, studies done with human samples and cell lines also point towards the context dependent role of histone isoforms. For example, the human H2A1C isoform, which was initially reported to be downregulated in CLL[2]. However, in a later study with higher number of samples H2A1C levels were found to be higher in CLL patients compared to samples from healthy individuals[179]. Further, H2A1C was found to promote proliferation in MCF7 cells[4], however, it showed anti-proliferative effect in gall bladder cancer cell line[2].

The first question is that why the pro-proliferative effect of H2A.1 was only observed in CL38 cells but not in CL44 cell line? Notably, in CL44 cells H2Aub levels are relatively higher compared to CL38. H2Aub plays a very important role in governing proliferation of cells. As discussed above H2Aub governs the expression of growth factor genes like lipocalin 2. Further, we and others have shown that its levels inversely correlate with the proliferation of cells[222]. Although, the decrease in H2Aub does not occur from a specific isoform but probably its

dominance over H2A.1 in governing cell proliferation makes its drop a prerequisite for the pro-proliferative effect of H2A.1 to be imparted. Another, instance which suggests that upregulation of H2A.1 might be context dependent is during liver regeneration. Liver regeneration involves rapid cell proliferation and is associated with a marked decrease in the level of H2Aub[233], however, increase in H2A.1 expression is not observed. Probably, drop in H2Aub is sufficient to promote cell proliferation. However, during cancer progression incorporation of H2A.1 alongside drop in H2Aub promotes persistent increase in cell proliferation required for tumor growth. The long half-life of H2A.1[234] makes it a suitable candidate to perform such a role. More studies are needed to ascertain these speculations. Notably, decrease in H2Aub level in human pancreatic cancer have been reported[220]. In addition, higher expression of H2A1C has been seen in chemo-resistance in pancreatic cancer cell lines[192]. It would be very interesting to investigate whether H2A1C is in general upregulated in pancreatic cancer. Further, Lipocalin 2 and H2A1C expression is upregulated in breast cancer. Whether an associated decrease in H2Aub also occurs remains to be investigated.

Overexpression of H2A.1 promoted proliferation of CL38 cells and this pro-proliferative effect was markedly reduced on carrying out the R99K and M51L substitutions. Based on these remarkable results we started to investigate the effect of these substitutions on the biochemical and biophysical properties of the nucleosome and its subcomplexes. As the substitutions are expected to have subtle effects on the stability/structure of histone complexes we opted to perform *in vitro* experiments with recombinant histones. On attempting recombinant expression, we found that H2A.2 and H4 did not express in the *E. coli* host. Post carrying out a systematic investigation of the parameters governing heterologous expression we conclude that destabilisation is the root cause behind lack of recombinant protein expression in bacteria.

Further, we propose that the different expression parameters like rare codon content, mRNA secondary structure and GC content of the transcripts etc. modulate translation kinetics and thus, folding of proteins which ultimately governs expression. Of course, a far greater number of proteins need to be characterised in different experimental settings to validate the hypothesis, but for now, we are able to put forth a unifying hypothesis that not only explains our results but also, the intermittent success (and failures) of the prevalent hypotheses.

Our *in vitro*, *in silico* and *in cellulo* results show that H2A.1 forms a less stable dimer with H2B as compared to H2A.2 and this contributes to the increased stability of H2A.1 containing nucleosomes. The alterations in the stability of complexes are significant yet the differences are small as would be expected to result from substitutions with very similar amino acids. This is consistent with previous reports where L to M replacements altered the protein stability only by 0.4 to 1.9 kcal/mol[196]. However, H2A.1 is a major H2A isoform and consistent with this we found that this protein is present globally across the chromatin. This suggests that on its increased abundance in certain patho-physiological states, even subtle differences brought about by it may have significant consequences on gene regulation which may ultimately affect cell physiology. Indeed, our cell based assays corroborate the same. Most remarkably, M51L and K99R alterations significantly impacted nucleosome stability suggesting that nucleosome dynamics is intimately linked with the non-redundant functionality of H2A.1.

## 8.2. Conclusion

NDEA administration leads to stable transformation of liver to HCC. This is associated with changes in shape and size of the nucleus of cancer cells [**Figure 8**] along with increase in cell proliferation markers cyclin D1 and PCNA. At the molecular level, there is a marked increase in the expression of histone H2A isoform H2A.1. This allows it to outcompete other isoforms

like H2A.2 in pairing with histone H2B. H2A.1-H2B dimer is less stable than H2A.2-H2B dimer. However, on incorporation into the nucleosome this thermodynamically favours the formation of a more stable nucleosome with H2A.1. Further, there is an associated decrease in

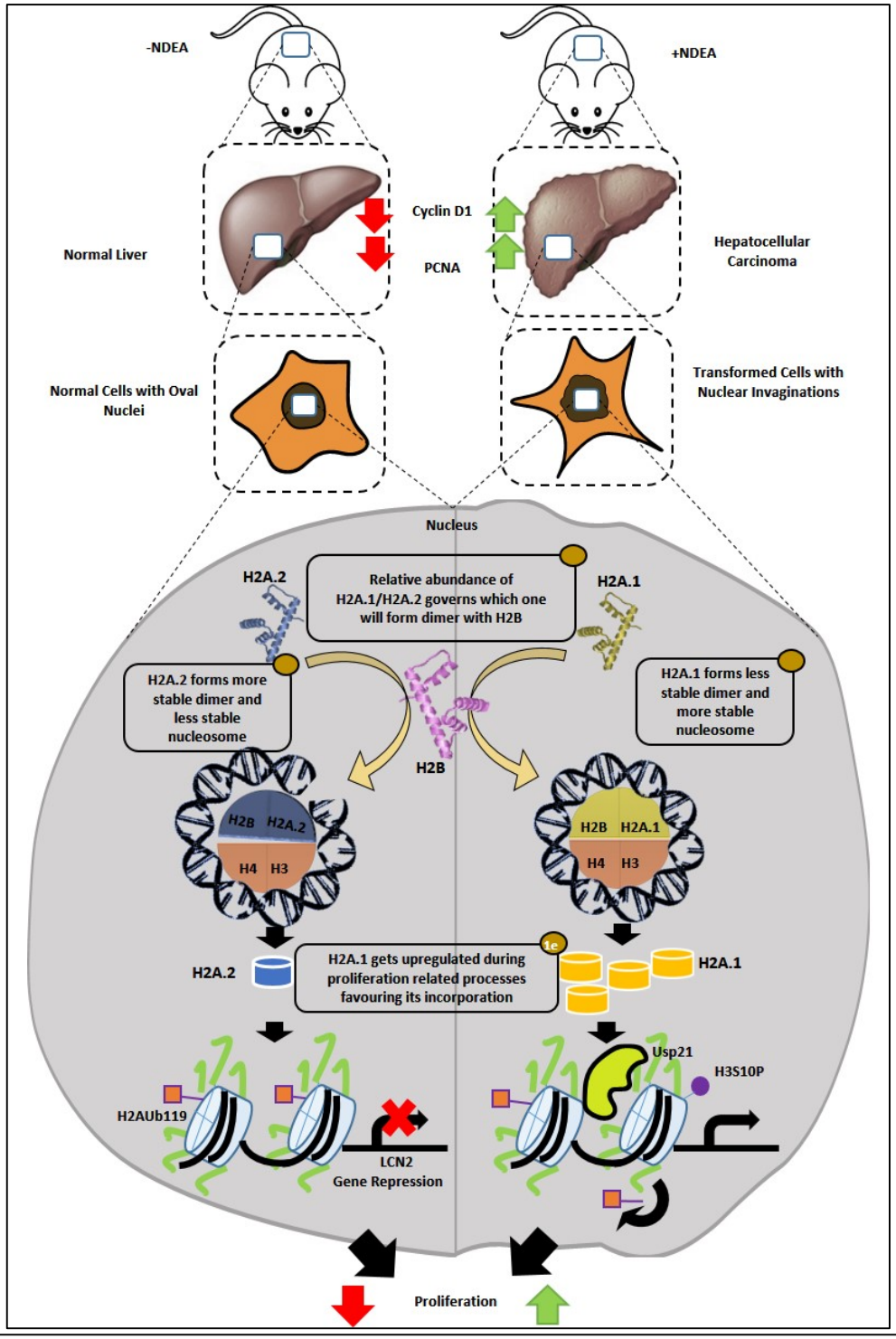



**Figure 8: Probable scenario in hepatocellular carcinoma.**

the level of H2Aub during the progress of HCC and increase in marks like H3S10p that has been shown to be indispensable for neoplastic transformation. The decrease in H2Aub is mediated by deubiquitinases like Usp21. These global changes in chromatin possibly leads to local changes at the promoter of genes like lipocalin 2 favouring its expression. Lipocalin 2, being a growth factor, in turn promotes the proliferation of cells.

In summary, H2A.1-containing nucleosomes are more stable owing to M51L and K99R substitutions that have the most prominent effect on cell proliferation, suggesting, that the nucleosome stability is intimately linked with the physiological effects observed. Possibly, the increased nucleosome stability resulting from H2A.1 incorporation, contributes to the contextual alteration in global gene expression pattern, which collectively promotes the attainment of different physiological states. This possibility of the non-redundant function, when extended to the plethora of the histone isoforms encoded by multiple genes (H2A, 16 genes; H2B, 22 genes; H3, 14 genes; H4, 14 genes; and H1, 6 genes), truly increases the complexity of the epigenome by many folds. Undoubtedly, such complexity is the necessity for multicellular organisms as the diversity in the epigenome plays central role in cell type specific gene expression. This in turn leads to specialized functions in thousands of cell types with the same genome.

# SALIENT FINDINGS AND FUTURE DIRECTIONS

## Salient Findings

Histone isoforms, like the variants, have been reported to be differentially expressed in a variety of patho-physiological states. Whether the histone isoforms actively contribute to the attainment of such states is not clear. Furthermore, the molecular basis of the non-redundancy remains elusive. The work done in the thesis provides insights into these aspects of histone biology and also provides solutions to some of the challenges faced during *in vitro* expression and purification of histone and its complexes.

- We found histones like H2A.2 and H4 do not express robustly in bacteria. The prevalent hypothesis to explain lack of recombinant protein expression like codon bias hypothesis and speculations regarding mRNA secondary structure and GC content of the transcripts do not provide a general explanation nor solution to the problem. Degradation mediated by host proteases probably is the root cause behind lack of recombinant protein expression. Further, our results provide compelling evidence that the codon bias usage, mRNA secondary structure and GC content of the transcript modulate the kinetics of translation, that in turn, influences protein folding and hence, degradation. Expression could be obtained by incorporation of the commonly used purification tags, like GST or MBP, of appropriate size and position leading to accumulation.
- A protocol was developed to quickly obtain large quantities of highly pure histones using bacterial expression system for GST pull-down and reconstitution experiments. An optimised method is presented to quickly reconstitute and purify properly folded H2A-H2B dimers, H3-H4 tetramers and histone octamers for *in vitro* experiments.

- The stability of H2A.1-H2B dimer is less than that of H2A.2-H2B. The M51L alteration at the H2A-H2B dimer interface results in lesser number of hydrophobic interactions leading to decreased stability of H2A.1-H2B dimer. The K99R substitution, although did not contribute significantly towards dimer stability, it seemed to have a destabilising effect along with M51L on H2A.1-H2B dimer. H2A.1 containing nucleosomes are more stable than H2A.2 nucleosomes. The increased stability of H2A.1 nucleosomes results from M51L and K99R substitutions that results in increased number of hydrogen bonds and hydrophobic interactions. The altered nucleosome dynamics on H2A.1 incorporation may lead to modulation of transcription or other chromatin mediated processes.
- H2A.1 exhibits drastically altered expression pattern in different normal tissues and human cancer cell lines (H2A1C in humans). H2A.1 and H2A.2 are present globally across the chromatin and overexpression of the particular isoform leads to its enrichment in chromatin. H2A.1 promotes cell proliferation in a context dependent manner but does not affect cell migration. Substitution R99K has the most drastic effect on the non-redundant functionality of H2A.1 followed by L51M substitution. The two substitution has synergistic effect when mutated together. The same substitutions have the most prominent effect on the stability of nucleosome complex and sub-complex suggesting that the nucleosome stability is intimately linked with the physiological effects observed.
- H2Aub markedly decreases in hepatocellular carcinoma (HCC). Usp21, a H2A deubiquitinase, is probably responsible for H2Aub. H2Aub levels shows an inverse correlation with H3S10 phosphorylation (H3S10p) and the proliferative state of the cells. Treatment of cells with histone deacetylase inhibitor trichostatin A results in

increase of H2Aub and decrease in H3S10p. Downregulation of H2Aub is also associated with increased expression of lipocalin 2. Lipocalin 2, being a growth factor, promotes proliferation of cells.

## Future Directions

The work described in this thesis elucidates the non-redundant functionality of H2A.1 isoform and its basis. However, much remains to be understood regarding this emerging regulator of epigenetic landscape. Based on our own results we speculate that the non-redundant functionality of H2A.1 is context dependent and studies from other groups also suggest the same. This raises the question what determines the context? It is reasonable to speculate that other factors that govern the epigenetic landscape of cells might contribute or pronounce the non-redundant functional effect of H2A.1. The lack of these factors, which may include specific binding partners and occurrence of post-translational modifications (or lack of it) etc., in CL44 cell line might explain the observed result. Indeed, our results suggest that H2Aub119 might be one factor however this correlation needs to be examined further to arrive at any conclusions.

We speculate that the increased nucleosome stability resulting from H2A.1 incorporation results in contextual alteration in gene expression pattern. It would be interesting to understand what subset of genes are regulated by H2A.1 in different contexts. Further, the markedly variable tissue specific expression pattern of H2A.1 makes one wonder what are the possible genes the regulation of which are possibly affected by H2A.1 incorporation.

The expression level analysis in different organs also shows that particularly the expression of H2A.1 (or H2A1C) varies in different tissues. Similarly, in humans particularly the expression level of H2A1C alters markedly in different contexts. Notably, sequence analysis shows the

presence of CRAS (Coding Region Activator Sequence) element in H2A.2 that is absent in H2A.1. CRAS element has been previously demonstrated to enhance the expression of H2A.2 by four fold. Possibly, the presence of CRAS element in H2A.2 leads to its more stable expression pattern as compared to the varied expression level seen for H2A.1. However, what *cis* or *trans* acting factors regulate the expression of H2A.1 would be a very important aspect to elucidate.

Further, the varied expression level of H2A.1 in different organs raises questions like what happens to H2A1C expression, for example, in brain cancer as in brain its expression level is already very high. This becomes even more important to understand as H2A1C expression pattern varies in different cancers. The expression analysis in different organs was carried out in rat. The scenario in human tissues and many human cancers remains unknown and should be investigated.

# BIBLIOGRAPHY

1. Luger K, Mäder a W, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389(6648):251–260.

2. Singh R, Mortazavi A, Telu KH, Nagarajan P, Lucas DM, Thomas-Ahner JM, Clinton SK, Byrd JC, Freitas MA, Parthun MR. Increasing the complexity of chromatin: functionally distinct roles for replication-dependent histone H2A isoforms in cell proliferation and carcinogenesis. Nucleic acids research. 2013;41(20):9284–95.

3. Khare SP, Sharma A, Deodhar KK, Gupta S. Overexpression of histone variant H2A.1 and cellular transformation are related in N-nitrosodiethylamine-induced sequential hepatocarcinogenesis. Experimental biology and medicine (Maywood, N.J.). 2011;236(1):30–5.

4. Su CH, Tzeng TY, Cheng C, Hsu MT. An H2A histone isotype regulates estrogen receptor target genes by mediating enhancer-promoter-3'-UTR interactions in breast cancer cells. Nucleic Acids Research. 2014;42(5):3073–3088.

5. Tyagi M, Khade B, Khan SA, Ingle A, Gupta S. Expression of histone variant, H2A.1 is associated with the undifferentiated state of hepatocyte. Experimental biology and medicine (Maywood, N.J.). 2014:1535370214531869.

6. Waddington CH. The Epigenotype. International Journal of Epidemiology. 2012;41(1):10–13.

7. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315–322.

8. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, et al. Human DNA methylomes at base resolution show widespread epigenomic differences . Nature. 2009;462(7271):315–322.

9. Baylin SB. DNA methylation and gene silencing in cancer. Nat Clin Prac Oncol.

10. Parker J. Nd Es Sc. 2005;1(March):352–355.

11. Kulis M, Esteller M. 2 - DNA Methylation and Cancer. In: Genetics ZH and TUBT-A in, editor. Epigenetics and Cancer, Part A. Vol. Volume 70. Academic Press; 2010. p. 27–56.

12. Feinberg a P, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89–92.

13. Cooper GM. Chromosomes and Chromatin. 2000.

14. Kornberg RD. Chromatin Structure: A Repeating Unit of Histones and DNA. Science. 1974;184(4139):868–871.

15. Robinson PJJ, Rhodes D. Structure of the "30 nm" chromatin fibre: A key role for the linker histone. Current Opinion in Structural Biology. 2006;16(3):336–343.

16. Wong H, Victor J-M, Mozziconacci J. An All-Atom Model of the Chromatin Fiber Containing Linker Histones Reveals a Versatile Structure Tuned by the Nucleosomal Repeat Length Chen P, editor. PLoS ONE. 2007;2(9):e877.

17. Robinson PJJ, Fairall L, Huynh VAT, Rhodes D. EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(17):6506–6511.

18. ConsortiumInternational HGS. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931–945.

19. Elgin SCR. Heterochromatin and gene regulation in Drosophila. Current Opinion in Genetics & Development. 1996;6(2):193–202.

20. Burgess RJ, Zhang Z. Histone chaperones in nucleosome assembly and human disease. Nat Struct Mol Biol. 2013;20(1):14–22.

21. Radman-Livaja M, Rando OJ. Nucleosome positioning: how is it established, and why does it matter? Developmental biology. 2010;339(2):258–266.

22. Nelson DM, Ye X, Hall C, Santos H, Ma T, Kao GD, Yen TJ, Harper JW, Adams PD. Coupling of DNA Synthesis and Histone Synthesis in S Phase Independent of Cyclin/cdk2 Activity. Molecular and Cellular Biology . 2002;22 (21 ):7459–7472.

23. Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. The Human and Mouse Replication-Dependent Histone Genes. Genomics. 2002;80(5):487–498.

24. Baxevanis AD, Landsman D. Histone Sequence Database: A Compilation of Highly-Conserved Nucleoprotein Sequences. Nucleic Acids Research . 1996;24 (1 ):245–247.

25. Van Holde KE. Chromatin. Springer Science & Business Media; 2012.

26. Luger K, Rechsteiner TJ, Richmond TJ. Expression and purification of recombinant histones and nucleosome reconstitution. In: Becker P, editor. Methods in molecular biology (Clifton, N.J.). Vol. 119. Humana Press; 1999. p. 1–16. (Methods in Molecular Biology™).

27. Arents G, Moudrianakis EN. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proceedings of the National Academy of Sciences . 1995;92 (24 ):11170–11174.

28. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. Journal of molecular biology. 2002;319(5):1097–1113.

29. Ramakrishnan V. Histone structure and the organization of the nucleosome. Annu. Rev. Biophys. Biomol. Struct. 1997;26:83–112.

30. Klug A, Rhodes D, Smith J, Finch JT, Thomas JO. A low resolution structure for the histone core of the nucleosome. Nature. 1980;287(5782):509–516.

31. Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A. Structure of the nucleosome core particle at 7 A resolution. Nature. 1983;311(5986):532–537.

32. Russo E, Giancotti V, Crane-Robinson C, Geraci G. Histone H1 and chromatin higher order structure does histone H1 exhibit specific self-association? International Journal of Biochemistry. 1983;15(4):487–493.

33. Yun M, Wu J, Workman JL, Li B. Readers of histone modifications. Cell Research. 2011;21(4):564–578.

34. Legube G, Trouche D. Regulating histone acetyltransferases and deacetylases. EMBO Reports. 2003;4(10):944–947.

35. Dokmanovic M, Clarke C, Marks PA. Histone Deacetylase Inhibitors: Overview and Perspectives. Molecular Cancer Research . 2007;5 (10 ):981–989.

36. Wood A, Shilatifard A. Posttranslational Modifications of Histones by Methylation. In: Chemistry BT-A in P, editor. Proteins in Eukaryotic Transcription. Vol. Volume 67. Academic Press; 2004. p. 201–222.

37. Feng W, Yonezawa M, Ye J, Jenuwein T, Grummt I. PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. Nat Struct Mol Biol. 2010;17(4):445–450.

38. Kallestad L, Christensen K, Woods E, Milavetz B. Transcriptional repression is epigenetically marked by H3K9 methylation during SV40 replication. Clinical Epigenetics. 2014;6(1):21.

39. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y. Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain. Current Biology. 2002;12(12):1052–1058.

40. Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y, Struhl K. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes & Development . 2002;16 (12 ):1518–1527.

41. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nature structural & molecular biology. 2007;14(11):1025–1040.

42. Rossetto D, Avvakumov N, Côté J. Histone phosphorylation: A chromatin modification involved in diverse nuclear events. Epigenetics. 2012;7(10):1098–1108.

43. Paull TT, Rogakou EP, Yamazaki V, Kirchgessner CU, Gellert M, Bonner WM. A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Current biology : CB. 2000;10(15):886–895.

44. Abraham RT. Cell cycle checkpoint signaling through the ATM and ATR kinases. Genes & Development . 2001;15 (17 ):2177–2196.

45. Mah L-J, El-Osta A, Karagiannis TC. γH2AX: a sensitive molecular marker of DNA damage and repair. Leukemia. 2010;24(4):679–686.

46. LEE HTAE, KIM SEKYE, CHOI MIRAN, PARK JIH, JUNG KHWA, CHAI YGYU. Effects of the activated mitogen-activated protein kinase pathway via the c-ros receptor tyrosine kinase on the T47D breast cancer cell line following alcohol exposure. Oncology Reports. 2013;29(3):868–874.

47. Lau ATY, Lee S-Y, Xu Y-M, Zheng D, Cho Y-Y, Zhu F, Kim H-G, Li S-Q, Zhang Z, Bode AM, et al. Phosphorylation of Histone H2B Serine 32 Is Linked to Cell Transformation. The Journal of Biological Chemistry. 2011;286(30):26628–26637.

48. Choi HS, Choi BY, Cho Y-Y, Mizuno H, Kang BS, Bode AM, Dong Z. Phosphorylation of Histone H3 at Serine 10 is Indispensable for Neoplastic Cell Transformation. Cancer research. 2005;65(13):5818–5827.

49. Sun J-M, Chen HY, Espino PS, Davie JR. Phosphorylated serine 28 of histone H3 is associated with destabilized nucleosomes in transcribed chromatin. Nucleic Acids Research. 2007;35(19):6640–6647.

50. Yang W, Xia Y, Hawke D, Li X, Liang J, Xing D, Aldape K, Hunter T, Yung WKA, Lu Z. PKM2 Phosphorylates Histone H3 and Promotes Gene Transcription and Tumorigenesis. Cell. 2012;150(4):685–696.

51. Hake SB, Allis CD. Histone H3 variants and their potential role in indexing mammalian genomes: The “H3 barcode hypothesis.” Proceedings of the National Academy of Sciences of the United States of America. 2006;103(17):6428–6435.

52. Jin C, Felsenfeld G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes and Development. 2007;21(12):1519–1529.

53. Hake SB, Garcia BA, Kauer M, Baker SP, Shabanowitz J, Hunt DF, Allis CD. Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(18):6344–6349.

54. Waterborg JH. Sequence analysis of acetylation and methylation in two histone H3 variants of alfalfa. Journal of Biological Chemistry . 1990;265 (28 ):17157–17161.

55. McKittrick E, Gafken PR, Ahmad K, Henikoff S. Histone H3. 3 is enriched in covalent modifications associated with active chromatin. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(6):1525–1530.

56. Loyola A, Bonaldi T, Roche D, Imhof A, Almouzni G. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Molecular cell. 2006;24(2):309–316.

57. Goldberg AD, Banaszynski LA, Noh K-M, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X. Distinct factors control histone variant H3. 3 localization at specific genomic regions. Cell. 2010;140(5):678–691.

58. Chow C, Georgiou A, Szutorisz H, e Silva AM, Pombo A, Barahona I, Dargelos E, Canzonetta C, Dillon N. Variant histone H3. 3 marks promoters of transcriptionally active genes during mammalian cell division. EMBO reports. 2005;6(4):354–360.

59. Mito Y, Henikoff JG, Henikoff S. Genome-scale profiling of histone H3. 3 replacement patterns. Nature genetics. 2005;37(10):1090–1097.

60. Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y. Histone H3. 1 and H3. 3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell. 2004;116(1):51–61.

61. Ray-Gallet D, Woolfe A, Vassias I, Pellentz C, Lacoste N, Puri A, Schultz DC, Pchelintsev NA, Adams PD, Jansen LET. Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3. 3 to maintain chromatin integrity. Molecular cell. 2011;44(6):928–941.

62. Drané P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3. 3. Genes & development. 2010;24(12):1253–1265.

63. Goldberg AD, Banaszynski LA, Noh K-M, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, et al. Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. Cell. 2010;140(5):678–691.

64. Faast R, Thonglairoam V, Schulz TC, Beall J, Wells JRE, Taylor H, Matthaei K, Rathjen PD, Tremethick DJ, Lyons I. Histone variant H2A. Z is required for early mammalian development. Current Biology. 2001;11(15):1183–1187.

65. Thatcher TH, Gorovsky MA. Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic acids research. 1994;22(2):174–9.

66. Jackson JD, Gorovsky MA. Histone H2A. Z has a conserved function that is distinct from that of the major H2A sequence variants. Nucleic Acids Research. 2000;28(19):3811–3816.

67. Guillemette B, Bataille AR, Gévry N, Adam M, Blanchette M, Robert F, Gaudreau L. Variant histone H2A. Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biol. 2005;3(12):e384.

68. Xu Y, Ayrapetov MK, Xu C, Gursoy-Yuzugullu O, Hu Y, Price BD. Histone H2A. Z controls a critical chromatin remodeling step required for DNA double-strand break repair. Molecular cell.

2012;48(5):723–733.

69. Swaminathan J, Baxter EM, Corces VG. The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin. Genes & development. 2005;19(1):65–76.

70. Rangasamy D, Greaves I, Tremethick DJ. RNA interference demonstrates a novel role for H2A. Z in chromosome segregation. Nature structural & molecular biology. 2004;11(7):650–655.

71. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823–837.

72. Kelly TK, Miranda TB, Liang G, Berman BP, Lin JC, Tanay A, Jones PA. H2A. Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. Molecular cell. 2010;39(6):901–911.

73. Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G. H3. 3/H2A. Z double variant–containing nucleosomes mark'nucleosome-free regions' of active promoters and other regulatory regions. Nature genetics. 2009;41(8):941–945.

74. Yang X, Noushmehr H, Han H, Andreu-Vieyra C, Liang G, Jones PA. Gene reactivation by 5-aza-2′-deoxycytidine–induced demethylation requires SRCAP–mediated H2A. Z insertion to establish nucleosome depleted regions. PLoS Genet. 2012;8(3):e1002604.

75. Zhang H, Roberts DN, Cairns BR. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. Cell. 2005;123(2):219–231.

76. Meneghini MD, Wu M, Madhani HD. Conserved histone variant H2A. Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell. 2003;112(5):725–736.

77. Fried VA. MacroH2A, a core histone containing a large nonhistone region. Science. 1992;257(5075):1398–1400.

78. Costanzi C, Pehrson JR. MACROH2A2, a new member of the MACROH2A core histone family. Journal of Biological Chemistry. 2001;276(24):21776–21784.

79. Talbert PB, Ahmad K, Almouzni G, Ausió J, Berger F, Bhalla PL, Bonner WM, Cande WZ, Chadwick BP, Chan SWL. A unified phylogeny-based nomenclature for histone variants. Epigenetics & chromatin. 2012;5(1):1.

80. Chakravarthy S, Gundimella SKY, Caron C, Perche P-Y, Pehrson JR, Khochbin S, Luger K. Structural characterization of the histone variant macroH2A. Molecular and cellular biology. 2005;25(17):7616–7624.

81. Buschbeck M, Di Croce L. Approaching the molecular and physiological function of macroH2A variants. Epigenetics. 2010;5(2):118–123.

82. Chadwick BP, Willard HF. Cell cycle–dependent localization of macroH2A in chromatin of the inactive X chromosome. The Journal of cell biology. 2002;157(7):1113–1123.

83. Angelov D, Molla A, Perche P-Y, Hans F, Côté J, Khochbin S, Bouvet P, Dimitrov S. The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Molecular cell. 2003;11(4):1033–1041.

84. Chadwick BP, Willard HF. A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. The Journal of cell biology. 2001;152(2):375–384.

85. Zhou J, Fan JY, Rangasamy D, Tremethick DJ. The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. Nature structural & molecular biology.

2007;14(11):1070–1076.

86. Gautier T, Abbott DW, Molla A, Verdel A, Ausio J, Dimitrov S. Histone variant H2ABbd confers lower stability to the nucleosome. EMBO reports. 2004;5(7):715–720.

87. Bao Y, Konesky K, Park Y, Rosu S, Dyer PN, Rangasamy D, Tremethick DJ, Laybourn PJ, Luger K. Nucleosomes containing the histone variant H2A. Bbd organize only 118 base pairs of DNA. The EMBO journal. 2004;23(16):3314–3324.

88. Gautier T, Abbott DW, Molla A, Verdel A, Ausio J, Dimitrov S. Histone variant H2ABbd confers lower stability to the nucleosome. EMBO reports. 2004;5(7):715–20.

89. Bao Y, Konesky K, Park Y-J, Rosu S, Dyer PN, Rangasamy D, Tremethick DJ, Laybourn PJ, Luger K. Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. The EMBO journal. 2004;23(16):3314–3324.

90. Soboleva TA, Nekrasov M, Pahwa A, Williams R, Huttley GA, Tremethick DJ. A unique H2A histone variant occupies the transcriptional start site of active genes. Nature structural & molecular biology. 2012;19(1):25–30.

91. Angelov D, Verdel A, An W, Bondarenko V, Hans F, Doyen C, Studitsky VM, Hamiche A, Roeder RG, Bouvet P. SWI/SNF remodeling and p300-dependent transcription of histone variant H2ABbd nucleosomal arrays. The EMBO journal. 2004;23(19):3815–3824.

92. Tolstorukov MY, Goldman JA, Gilbert C, Ogryzko V, Kingston RE, Park PJ. Histone variant H2A. Bbd is associated with active transcription and mRNA processing in human cells. Molecular cell. 2012;47(4):596–607.

93. Tyler-Smith C, Floridia G. Many paths to the top of the mountain: diverse evolutionary solutions to centromere structure. Cell. 2000;102(1):5–8.

94. Sullivan KF, Hechenberger M, Masri K. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. The Journal of cell biology. 1994;127(3):581–592.

95. Chen Y, Baker RE, Keith KC, Harris K, Stoler S, Fitzgerald-Hayes M. The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. Molecular and cellular biology. 2000;20(18):7037–7048.

96. Yoda K, Ando S, Morishita S, Houmura K, Hashimoto K, Takeyasu K, Okazaki T. Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proceedings of the National Academy of Sciences. 2000;97(13):7266–7271.

97. Black BE, Foltz DR, Chakravarthy S, Luger K, Woods VL, Cleveland DW. Structural determinants for generating centromeric chromatin. Nature. 2004;430(6999):578–582.

98. Malik HS, Vermaak D, Henikoff S. Recurrent evolution of DNA-binding motifs in the Drosophila centromeric histone. Proceedings of the National Academy of Sciences. 2002;99(3):1449–1454.

99. Leuba SH, Bustamante C, Van Holde K, Zlatanova J. Linker histone tails and N-tails of histone H3 are redundant: scanning force microscopy studies of reconstituted fibers. Biophysical journal. 1998;74(6):2830–2839.

100. Marion C, Roux B, Coulet PR. Role of histones H1 and H3 in the maintenance of chromatin in a compact conformation. FEBS letters. 1983;157(2):317–321.

101. Gilbert N, Allan J. Distinctive higher-order chromatin structure at mammalian centromeres. Proceedings of the National Academy of Sciences. 2001;98(21):11949–11954.

102. Placek BJ, Harrison LN, Villers BM, Gloss LM. The H2A.Z/H2B dimer is unstable compared to the dimer containing the major H2A isoform. Protein science : a publication of the Protein Society. 2005;14(2):514–522.

103. Suto RK, Clarkson MJ, Tremethick DJ, Luger K. Crystal structure of a nucleosome core particle containing the variant histone H2A. Z. Nature Structural & Molecular Biology. 2000;7(12):1121–1124.

104. Abbott DW, Ivanova VS, Wang X, Bonner WM, Ausió J. Characterization of the Stability and Folding of H2A. Z Chromatin Particles IMPLICATIONS FOR TRANSCRIPTIONAL ACTIVATION. Journal of Biological Chemistry. 2001;276(45):41945–41949.

105. Flaus A, Rencurel C, Ferreira H, Wiechens N, Owen-Hughes T. Sin mutations alter inherent nucleosome mobility. The EMBO journal. 2004;23(2):343–353.

106. Park Y-J, Dyer PN, Tremethick DJ, Luger K. A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. Journal of Biological Chemistry. 2004;279(23):24274–24282.

107. Abbott DW, Chadwick BP, Thambirajah AA, Ausió J. Beyond the Xi MacroH2A chromatin distribution and post-translational modifications in an avian system. Journal of Biological Chemistry. 2005;280(16):16437–16445.

108. Nusinow DA, Sharp JA, Morris A, Salas S, Plath K, Panning B. The histone domain of macroH2A1 contains several dispersed elements that are each sufficient to direct enrichment on the inactive X chromosome. Journal of molecular biology. 2007;371(1):11–18.

109. Okuwaki M, Kato K, Shimahara H, Tate S, Nagata K. Assembly and disassembly of nucleosome core particles containing histone variants by human nucleosome assembly protein I. Molecular and cellular biology. 2005;25(23):10639–10651.

110. Schwartzentruber J, Korshunov A, Liu X-Y, Jones DTW, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang D-AK, Tönjes M. Driver mutations in histone H3. 3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012;482(7384):226–231.

111. Conerly ML, Teves SS, Diolaiti D, Ulrich M, Eisenman RN, Henikoff S. Changes in H2A. Z occupancy and DNA methylation during B-cell lymphomagenesis. Genome research. 2010;20(10):1383–1390.

112. Hua S, Kallen CB, Dhar R, Baquero MT, Mason CE, Russell BA, Shah PK, Liu J, Khramtsov A, Tretiakova MS. Genomic analysis of estrogen cascade reveals histone variant H2A. Z associated with breast cancer progression. Molecular systems biology. 2008;4(1):188.

113. Svotelis A, Gévry N, Grondin G, Gaudreau L. H2A. Z overexpression promotes cellular proliferation of breast cancer cells. Cell Cycle. 2010;9(2):364–370.

114. Kim K, Punj V, Choi J, Heo K, Kim J-M, Laird PW, An W. Gene dysregulation by histone variant H2A.Z in bladder cancer. Epigenetics & Chromatin. 2013;6(1):34.

115. Valdés-Mora F, Song JZ, Statham AL, Strbenac D, Robinson MD, Nair SS, Patterson KI, Tremethick DJ, Stirzaker C, Clark SJ. Acetylation of H2A. Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. Genome research. 2012;22(2):307–321.

116. Dryhurst D, McMullen B, Fazli L, Rennie PS, Ausió J. Histone H2A. Z prepares the prostate specific antigen (PSA) gene for androgen receptor-mediated transcription and is upregulated in a model of prostate cancer progression. Cancer letters. 2012;315(1):38–47.

117. Lacroix M. Significance, detection and markers of disseminated breast cancer cells. Endocrine-related cancer. 2006;13(4):1033–1067.

118. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008;132(6):958–970.

119. Gévry N, Hardy S, Jacques P-É, Laflamme L, Svotelis A, Robert F, Gaudreau L. Histone H2A. Z is essential for estrogen receptor signaling. Genes & development. 2009;23(13):1522–1533.

120. Rangasamy D. Histone variant H2A.Z can serve as a new target for breast cancer therapy. Current medicinal chemistry. 2010;17:3155–3161.

121. Sporn JC, Kustatscher G, Hothorn T, Collado M, Serrano M, Muley T, Schnabel P, Ladurner AG. Histone macroH2A isoforms predict the risk of lung cancer recurrence. Oncogene. 2009;28(38):3423–3428.

122. Kapoor A, Goldberg MS, Cumberland LK, Ratnakumar K, Segura MF, Emanuel PO, Menendez S, Vardabasso C, LeRoy G, Vidal CI, et al. The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. Nature. 2010;468(7327):1105–1109.

123. Li X, Kuang J, Shen Y, Majer MM, Nelson CC, Parsawar K, Heichman KA, Kuwada SK. The atypical histone macroH2A1. 2 interacts with HER-2 protein in cancer cells. Journal of Biological Chemistry. 2012;287(27):23171–23183.

124. Williams AS, Marzluff WF. The sequence of the stem and flanking sequences at the 3′ end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. Nucleic Acids Research . 1995;23 (4 ):654–662.

125. Scharl EC, Steitz JA. The site of 3'end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. The EMBO journal. 1994;13(10):2432.

126. Sànchez R, Marzluff WF. The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. Molecular and cellular biology. 2002;22(20):7093–7104.

127. Dominski Z, Zheng L-X, Sanchez R, Marzluff WF. Stem-loop binding protein facilitates 3′-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. Molecular and cellular biology. 1999;19(5):3561–3570.

128. Arnold DR, Fran??on P, Zhang J, Martin K, Clarke HJ. Stem-loop binding protein expressed in growing oocytes is required for accumulation of mRNAs encoding histones H3 and H4 and for early embryonic development in the mouse. Developmental Biology. 2008;313(1):347–358.

129. Piña B, Suau P. Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons. Developmental biology. 1987;123(1):51–8.

130. Vassilev a P, Rasmussen HH, Christensen EI, Nielsen S, Celis JE. The levels of ubiquitinated histone H2A are highly upregulated in transformed human cells: partial colocalization of uH2A clusters and PCNA/cyclin foci in a fraction of cells in S-phase. Journal of cell science. 1995;108 ( Pt 3:1205–1215.

131. Pirngruber J, Johnsen SA. Induced G1 cell-cycle arrest controls replication-dependent histone mRNA 3' end processing through p21, NPAT and CDK9. Oncogene. 2010;29(19):2853–2863.

132. Rando OJ, Chang HY. Genome-Wide Views of Chromatin Structure. Annual Review of Biochemistry. 2009;78(1):245–271.

133. Singh MK, Scott TF, LaFramboise WA, Hu FZ, Post JC, Ehrlich GD. Gene expression changes in peripheral blood mononuclear cells from multiple sclerosis patients undergoing β-interferon

therapy. Journal of the Neurological Sciences. 2007;258(1–2):52–59.

134. Su X, Lucas DM, Zhang L, Xu H, Zabrouskov V, Davis ME, Knapp AR, Young DC, Payne PRO, Parthun MR, et al. Validation of an LC-MS based approach for profiling histones in chronic lymphocytic leukemia. Proteomics. 2009;9(5):1197–206.

135. Simon RH, Felsenfeld G. A new procedure for purifying histone pairs H2A+ H2B and H3+ H4 from chromatin using hydroxylapatite. Nucleic Acids Research. 1979;6(2):689–696.

136. Helliger W, Lindner H, Grübl-Knosp O, Puschendorf B. Alteration in proportions of histone H1 variants during the differentiation of murine erythroleukaemic cells. Biochemical Journal. 1992;288(3):747–751.

137. Ridsdale JA, Davie JR. Chicken erythrocyte polynucleosomes which are soluble at physiological ionic strength and contain linker histones are highly enriched hi β-globin gene sequences. Nucleic acids research. 1987;15(3):1081–1096.

138. Tanaka Y, Tawaramoto-Sasanuma M, Kawaguchi S, Ohta T, Yoda K, Kurumizaka H, Yokoyama S. Expression and purification of recombinant human histones. Methods. 2004;33(1):3–11.

139. Kubo M, Imanaka T. mRNA secondary structure in an open reading frame reduces translation efficiency in Bacillus subtilis. Journal of bacteriology. 1989;171(7):4080–2.

140. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Current Opinion in Biotechnology. 1995 [accessed 2015 Nov 11];6(5):494–500. http://www.sciencedirect.com/science/article/pii/0958166995800824

141. Zhang W, Xiao W, Wei H, Zhang J, Tian Z. mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in Escherichia coli. Biochemical and biophysical research communications. 2006 [accessed 2015 Dec 4];349(1):69–78. http://www.sciencedirect.com/science/article/pii/S0006291X06017189

142. Keller T, Mis D, Jia K, Wilke C. Reduced mRNA secondary-structure stability near the start codon indicates functional genes in prokaryotes. Genome biology and evolution. 2012;4(2):80–88.

143. Behura SK, Severson DW. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. Biological Reviews. 2013;88(1):49–61.

144. Alexeyev MF, Winkler HH. Gene synthesis, bacterial expression and purification of the Rickettsia prowazekii ATP/ADP translocase. Biochimica et Biophysica Acta (BBA) - Biomembranes. 1999 [accessed 2015 Dec 4];1419(2):299–306. http://www.sciencedirect.com/science/article/pii/S0005273699000784

145. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Current opinion in biotechnology. 1995;6(5):494–500.

146. Chen GF, Inouye M. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the Escherichia coli genes. Nucleic acids research. 1990;18(6):1465–1473.

147. Rosenberg a. H, Goldman E, Dunn JJ, Studier FW, Zubay G. Effects of consecutive AGG codons on translation in Escherichia coli, demonstrated with a versatile codon test system. Journal of Bacteriology. 1993;175(3):716–722.

148. Calderone TL, Stevens RD, Oas TG. High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in Escherichia coli. Journal of molecular biology. 1996 [accessed 2015 Dec 4];262(4):407–12. http://www.sciencedirect.com/science/article/pii/S0022283696905240

149. Gerchman SE, Graziano V, Ramakrishnan V. Expression of chicken linker histones in E. coli: sources of problems and methods for overcoming some of the difficulties. Protein expression and purification. 1994 [accessed 2015 Dec 4];5(3):242–51. http://www.sciencedirect.com/science/article/pii/S1046592884710370

150. Goodman DB, Church GM, Kosuri S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. Science . 2013;342 (6157 ):475–479.

151. Takyar S, Hickerson RP, Noller HF. mRNA helicase activity of the ribosome. Cell. 2005 [accessed 2015 Dec 4];120(1):49–58. http://www.sciencedirect.com/science/article/pii/S0092867404011468

152. Meyer IM, Miklós I. SimulFold: simultaneously inferring RNA structures including pseudoknots, alignments, and trees using a Bayesian MCMC framework. PLoS computational biology. 2007;3(8):e149.

153. Welch M, Villalobos A, Gustafsson C, Minshull J. You're one in a googol: optimizing genes for protein expression. Journal of The Royal Society Interface. 2009;6(Suppl_4):S467–S476.

154. Engebrecht J. Minipreps of Plasmid DNA. Current Protocols in Molecular Biology. 1990;1(1991):1.6.1–1.6.10.

155. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning. Cold spring harbor laboratory press New York; 1989.

156. Inoue H, Nojima H, Okayama H. High efficiency transformation of Escherichia coli with plasmids. Gene. 1990;96(1):23–28.

157. Guan K, Dixon JE. Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. Analytical biochemistry. 1991;192(2):262–267.

158. de Smit MH, van Duin J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proceedings of the National Academy of Sciences of the United States of America. 1990;87(19):7668–7672.

159. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell. 2010 [accessed 2015 Aug 16];141(2):344–54. http://www.sciencedirect.com/science/article/pii/S0092867410003193

160. Sharp PM, Li W-H. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research . 1987;15 (3 ):1281–1295.

161. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA Websuite. Nucleic Acids Research. 2008;36(Web Server):W70–W74.

162. Vergani L, Canneva F, Ghisellini P, Nicolini C. Expression, purification, and structural characterization of human histone H4. Protein expression and purification. 2002 [accessed 2015 Dec 4];24(3):420–8. http://www.sciencedirect.com/science/article/pii/S1046592801916018

163. Gottesman S. Proteases anf their targets in *Escherichia Coli*. Annual Review of Genetics. 1996;30(1):465–506.

164. Rosano GL, Ceccarelli E a. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. Microbial cell factories. 2009;8:41.

165. Cannarozzi G, Cannarrozzi G, Schraudolph NN, Faty M, von Rohr P, Friberg MT, Roth AC, Gonnet P, Gonnet G, Barral Y. A role for codon order in translation dynamics. Cell. 2010 [accessed

2015 Nov 25];141(2):355–67. http://www.sciencedirect.com/science/article/pii/S0092867410001893

166. Morse DE, Mosteller RD, Yanofsky C. Dynamics of Synthesis, Translation, and Degradation of trp Operon Messenger RNA in E. coli. Cold Spring Harbor Symposia on Quantitative Biology . 1969;34 :725–740.

167. Saunders R, Deane CM. Synonymous codon usage influences the local protein structure observed. Nucleic Acids Research. 2010;38(19):6719–6728.

168. Marin M. Folding at the rhythm of the rare codon beat. Biotechnology Journal. 2008;3(8):1047–1057.

169. Hoskins JR, Yanagihara K, Mizuuchi K, Wickner S. ClpAP and ClpXP degrade proteins with tags located in the interior of the primary sequence. Proceedings of the National Academy of Sciences. 2002;99(17):11037–11042.

170. Vergani L, Canneva F, Ghisellini P, Nicolini C. Expression, purification, and structural characterization of human histone H4. Protein expression and purification. 2002;24(3):420–8.

171. Martinez a, Knappskog PM, Olafsdottir S, Døskeland a P, Eiken HG, Svebak RM, Bozzini M, Apold J, Flatmark T. Expression of recombinant human phenylalanine hydroxylase as fusion protein in Escherichia coli circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme. The Biochemical journal. 1995;306 ( Pt 2:589–597.

172. Alvarez-Castelao B, Ruiz-Rivas C, Castaño JG. A critical appraisal of quantitative studies of protein degradation in the framework of cellular proteostasis. Biochemistry Research International. 2012;2012.

173. Sinclair G, Choy FYM. Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, Pichia pastoris. Protein Expression and Purification. 2002 [accessed 2015 Dec 4];26(1):96–105. http://www.sciencedirect.com/science/article/pii/S1046592802005260

174. Gu W, Zhou T, Wilke CO. A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. PLoS Computational Biology. 2010;6(2):e1000664.

175. Luger K, Rechsteiner TJ, Flaus a J, Waye MM, Richmond TJ. Characterization of nucleosome core particles containing histone proteins made in bacteria. Journal of molecular biology. 1997;272(3):301–311.

176. Tanaka Y, Tawaramoto-Sasanuma M, Kawaguchi S, Ohta T, Yoda K, Kurumizaka H, Yokoyama S. Expression and purification of recombinant human histones. Methods (San Diego, Calif.). 2004 [accessed 2015 Dec 4];33(1):3–11. http://www.sciencedirect.com/science/article/pii/S1046202303002883

177. Gloss LM, Placek BJ. The effect of salts on the stability of the H2A-H2B histone dimer. Biochemistry. 2002;41(50):14951–14959.

178. Shim Y, Duan MR, Chen X, Smerdon MJ, Min JH. Polycistronic coexpression and nondenaturing purification of histone octamers. Analytical Biochemistry. 2012;427(2):190–192.

179. Singh R, Harshman SW, Ruppert AS, Mortazavi A, Lucas DM, Thomas-Ahner JM, Clinton SK, Byrd JC, Freitas MA, Parthun MR. Proteomic profiling identifies specific histone species associated with leukemic and cancer cells. Clinical Proteomics. 2015;12(1):22.

180. Zhang H-H, Zhang Z-Y, Che C-L, Mei Y-F, Shi Y-Z. Array analysis for potential biomarker of gemcitabine identification in non-small cell lung cancer cell lines. International journal of clinical and experimental pathology. 2013;6(9):1734–46.

181. Olivares I, Ballester A, Lombardia L, Dominguez O, López-Galíndez C. Human immunodeficiency virus type 1 chronic infection is associated with different gene expression in MT-4, H9 and U937 cell lines. Virus Research. 2009;139(1):22–31.

182. Mitchell C, Willenbring H. A reproducible and well-tolerated method for 2/3 partial hepatectomy in mice. Nature protocols. 2008;3(7):1167–1170.

183. Glauert AM, Lewis PR. Biological specimen preparation for transmission electron microscopy. Princeton University Press; 2014.

184. Hawley TS, Hawley RG. Flow cytometry protocols. Springer; 2004.

185. van Meerloo J, Kaspers GJL, Cloos J. Cell sensitivity assays: the MTT assay. Cancer cell culture: methods and protocols. 2011:237–245.

186. Franken NAP, Rodermond HM, Stap J, Haveman J, Van Bree C. Clonogenic assay of cells in vitro. Nature protocols. 2006;1(5):2315–2319.

187. Zaret K. Micrococcal nuclease analysis of chromatin structure. Current protocols in molecular biology. 2005.

188. Shechter D, Dormann HL, Allis CD, Hake SB. Extraction, purification and analysis of histones. Nature protocols. 2007;2(6):1445–1457.

189. Kruger NJ. The Bradford method for protein quantitation. Basic protein and peptide protocols. 1994:9–15.

190. Ryan CA, Annunziato AT. Separation of histone variants and post-translationally modified isoforms by triton/acetic acid/urea polyacrylamide gel electrophoresis. Current Protocols in Molecular Biology. 2001:21–22.

191. Pramod KS, Bharat K, Sanjay G. Mass spectrometry-compatible silver staining of histones resolved on acetic acid-urea-Triton PAGE. Proteomics. 2009;9(9):2589–2592.

192. Yi Zhang, Xin Shi, Department of General Surgery, Zhongda Hospital of Southeast University, Nanjing 210009, Jiangsu Province C. Expression of HIST1H2AC and HIST1H2BC genes in pancreatic cancer cell line Puta8988 treated with pemetrexed. World Chinese Journal of Digestology. 2009;14.

193. Zhang K, Tang H. Analysis of core histones by liquid chromatography-mass spectrometry and peptide mapping. Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences. 783(1):173–179.

194. Parra M a, Wyrick JJ. Regulation of gene transcription by the histone H2A N-terminal domain. Molecular and cellular biology. 2007;27(21):7641–8.

195. Jin C, Felsenfeld G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & Development. 2007;21(12):1519–1529.

196. Gassner NC, Baase W a, Matthews BW. A test of the “jigsaw puzzle” model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(22):12155–12158.

197. Lipscomb L a, Gassner NC, Snow SD, Eldridge a M, Baase W a, Drew DL, Matthews BW. Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. Protein science : a publication of the Protein Society. 1998;7(3):765–773.

198. Sokalingam S, Raghunathan G, Soundrarajan N, Lee S-G. A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein.

PLoS ONE. 2012;7(7):e40410.

199. Mann CJ, Matthews CR. Structure and stability of an early folding intermediate of Escherichia coli trp aporepressor measured by far-UV stopped-flow circular dichroism and 8-anilino-1-naphthalene sulfonate binding. Biochemistry. 1993;32(20):5282–5290.

200. Pace CN, Scholtz JM. Measuring the conformational stability of a protein. Protein structure: A practical approach. 1997;2:299–321.

201. Wiedemann SM, Mildner SN, Boenisch C, Israel L, Maiser A, Matheisl S, Straub T, Merkl R, Leonhardt H, Kremmer E, et al. Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. The Journal of cell biology. 2010;190(5):777–791.

202. Haile JM. Molecular dynamics simulation. Wiley, New York; 1992.

203. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of chemical theory and computation. 2008;4(3):435–447.

204. Ettig R, Kepper N, Stehr R, Wedemann G, Rippe K. Dissecting DNA-Histone Interactions in the Nucleosome by Molecular Dynamics Simulations of DNA Unwrapping. Biophysical Journal. 2011;101(8):1999–2008.

205. Hoch D a., Stratton JJ, Gloss LM. Protein-Protein Förster Resonance Energy Transfer Analysis of Nucleosome Core Particles Containing H2A and H2A.Z. Journal of Molecular Biology. 2007;371(4):971–988.

206. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Research. 2011;21(3):381–395.

207. Böhm L, Crane-Robinson C, Sautière P. Proteolytic Digestion Studies of Chromatin Core-Histone Structure. European Journal of Biochemistry. 1980;106(2):525–530.

208. Goldknopf IRAL, Taylor CW, Baum RM, Yeoman LC, Olson J, Prestayko AW, Busch H, Sci HL. Isolation and Characterization of Protein Chromosomal Protein or during liver regeneration. 1975;(18):7182–7187.

209. Seon B, Sandberg AA. Structural Chromatin. 1979;76(12):6386–6390.

210. Atkinson JP, Top C, Immunol M, Robzyk K, Recht J, Osley MA. Rad6-dependent ubiquitination of histone H2B in yeast. Science (New York, N.Y.). 2000;287(5452):501–504.

211. Wang H, Wang L, Erdjument-bromage H. Role of histone H2A ubiquitination in Polycomb silencing. 2004;431(October):873–878.

212. Gearhart MD, Corcoran CM, Wamstad J a, Bardwell VJ. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Molecular and cellular biology. 2006;26(18):6880–6889.

213. Zhou W, Zhu P, Wang J, Pascual G, Ohgi KA, Lozach J, Glass CK, Rosenfeld MG. Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Molecular cell. 2008;29(1):69–80.

214. Kalb R, Latwiel S, Baymaz HI, Jansen PWTC, Müller CW, Vermeulen M, Müller J. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nature structural & molecular biology. 2014;21(6):569–71.

215. Bonner WM, West MH, Stedman JD. Two-dimensional gel analysis of histones in acid extracts of nuclei, cells, and tissues. European journal of biochemistry / FEBS. 1980;109(1):17–23.

216. Cao R, Tsukada Y-I, Zhang Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Molecular cell. 2005;20(6):845–54.

217. Zhang X-Y, Varthi M, Sykes SM, Phillips C, Warzecha C, Zhu W, Wyce A, Thorne AW, Berger SL, McMahon SB. The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. Molecular cell. 2008;29(1):102–11.

218. Zhao Y, Lang G, Ito S, Bonnet J, Metzger E, Sawatsubashi S, Suzuki E, Le Guezennec X, Stunnenberg HG, Krasnov A, et al. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Molecular cell. 2008;29(1):92–101.

219. Zhu P, Zhou W, Wang J, Puc J, Ohgi K a., Erdjument-Bromage H, Tempst P, Glass CK, Rosenfeld MG. A Histone H2A Deubiquitinase Complex Coordinating Histone Acetylation and H1 Dissociation in Transcriptional Regulation. Molecular Cell. 2007;27(4):609–621.

220. Glinskii AB, Yang Z, Li X, Hoffman RM. Essential Role for Activation of the Polycomb Group (PcG) Protein Chromatin Silencing Pathway in Metastatic Prostate Cancer. Cell cycle (Georgetown, Tex.). 2014;5(16):1886–1901.

221. Prenzel T, Begus-nahrmann Y, Kramer F, Hennion M, Hsu C, Gorsler T, Hintermair C, Eick D, Kremmer E, Simons M, et al. Estrogen-Dependent Gene Transcription in Human Breast Cancer Cells Relies upon Proteasome-Dependent Monoubiquitination of Histone H2B. 2011;48(6):5739–5754.

222. Joo H-Y, Zhai L, Yang C, Nie S, Erdjument-Bromage H, Tempst P, Chang C, Wang H. Regulation of cell cycle progression and gene expression by H2A deubiquitination. Nature. 2007;449(7165):1068–1072.

223. Choi HS, Choi BY, Cho YY, Mizuno H, Kang BS, Bode AM, Dong Z. Phosphorylation of histone H3 at serine 10 is indispensable for neoplastic cell transformation. Cancer Research. 2005;65(13):5818–5827.

224. Imai BS, Mische SM. Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity. Electrophoresis. 1999;20:601–605.

225. Jiménez CR, Huang L, Qiu Y, Burlingame AL. In-Gel Digestion of Proteins for MALDI-MS Fingerprint Mapping. Current protocols in protein science. 2001:14–16.

226. Nakagawa T, Kajitani T, Togo S, Masuko N, Ohdan H, Hishikawa Y, Koji T, Matsuyama T, Ikura T, Muramatsu M, et al. Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes and Development. 2008;22(1):37–49.

227. Li Y, Kao GD, Garcia BA, Shabanowitz J, Hunt DF, Qin J, Phelan C, Lazar MA. A novel histone deacetylase pathway regulates mitosis by modulating Aurora B kinase activity. Genes & development. 2006;20(18):2566–79.

228. Sakajiri S, Kumagai T, Kawamata N, Saitoh T, Said JW, Koeffler HP. Histone deacetylase inhibitors profoundly decrease proliferation of human lymphoid cancer cell lines. Exp Hematol. 2005;33(1):53–61.

229. Yang J, Bielenberg DR, Rodig SJ, Doiron R, Clifton MC, Kung AL, Strong RK, Zurakowski D, Moses MA. Lipocalin 2 promotes breast cancer progression. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(10):3913–8.

230. Krämer OH, Zhu P, Ostendorff HP, Golebiewski M, Tiefenbach J, Peters MA, Brill B, Groner B,

Bach I, Heinzel T, et al. The histone deacetylase inhibitor valproic acid selectively induces proteasomal degradation of HDAC2. The EMBO journal. 2003;22(13):3411–20.

231. Bommi P V, Dimri M, Sahasrabuddhe A a, Khandekar J, Dimri GP. The polycomb group protein BMI1 is a transcriptional target of HDAC inhibitors. Cell cycle (Georgetown, Tex.). 2010;9(13):2663–73.

232. Göttlicher M, Minucci S, Zhu P, Krämer OH, Schimpf A, Giavara S, Sleeman JP, Lo Coco F, Nervi C, Pelicci PG, et al. Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. EMBO Journal. 2002;20(24):6969–6978.

233. Nakagawa T, Kajitani T, Masuko N, Ohdan H, Hishikawa Y, Koji T, Matsuyama T, Ikura T, Muramatsu M, Ito T, et al. Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes and Development. 2008;22(1):37–49.

234. Grove GW, Zweidler a. Regulation of nucleosomal core histone variant levels in differentiating murine erythroleukemia cells. Biochemistry. 1984;23(19):4436–43.

# APPENDIX

# *Annexure I – Chemicals and Antibodies*

## *List of Chemicals*

| Serial Number | Chemical Name | Company |
|---|---|---|
| 1 | Acrylamide | Sigma |
| 2 | Agarose | USB |
| 3 | Ammonium persulfate | USB |
| 4 | Acetic Acid | Qualigens |
| 5 | Aprotinin | Sigma |
| 6 | Amido Black | Sigma |
| 7 | Ammoniacal Silver nitrate | Qualigens, SD fine |
| 8 | Ammonium hydroxide | Qualigens |
| 9 | Bovine Serum Albumin fraction V | Roche-Sigma |
| 10 | Bradford reagent | Biorad |
| 11 | BME | Sigma |
| 12 | Coomassie brilliant blue R-250 | USB |
| 13 | Calcium chloride | SD fine |
| 14 | Crystal Violet | Sigma |
| 15 | Disodium Hydrogen Phosphate | Qualigens |
| 16 | Dihydrogen Sodium phosphate | SD fine |
| 17 | Dimethyl Sulphoxide | Qualigens |
| 18 | EDTA | Sigma |
| 19 | EtBr | Sigma |
| 20 | ECL detection reagent | Millipore |
| 21 | Formaldehyde | Qualigenes |
| 22 | Femto | Pierce |
| 23 | Glycerol | Merck |
| 24 | Glycine | Sigma |
| 25 | Guanidine Chloride | Sigma |
| 26 | Hydrochloric Acid | Qualigens |
| 27 | IPTG | Sigma |
| 28 | Ketmin-50 | Themis Medicare |
| 29 | Leupeptin | Sigma |
| 30 | Magnesium Chloride | Qualigens |
| 31 | Methanol | Merck |
| 32 | MNase | USB |
| 33 | MTT | Sigma |
| 34 | N-N Methylene Bis-Acrylamide | Sigma, Amresco |
| 35 | Nonidet P-40 | Sigma |
| 36 | Potassium Acetate | SD fine |
| 37 | PMSF | Sigma |

| | | |
|---|---|---|
| 38 | Potassium Chloride | Glaxo |
| 39 | Pepstatin A | Sigma |
| 40 | Protamine Sulfate | Sigma |
| 41 | Potassium Acetate | SD fine |
| 42 | PVDF membrane | Millipore |
| 43 | Paraformaldehyde | Sigma |
| 44 | Riboflavin 5 phosphate | Sigma |
| 45 | RNASE A | Amresco |
| 46 | Sodium Chloride | Qualigens |
| 47 | Sodium Orthovandate | Sigma |
| 48 | Sodium Citrate | Merck |
| 49 | Sodium Deoxycholate | BDH |
| 50 | Sodium Bicarbonate | SD fine |
| 51 | Sodium Hydroxide | SD fine |
| 52 | Sulphuric Acid | Qualigens |
| 53 | Spermine | USB |
| 54 | Spermidine | USB |
| 55 | SDS | Sigma |
| 56 | Sucrose | Sigma |
| 57 | Tween-20 | Sigma |
| 58 | Triton X-100 | Sigma |
| 59 | Tris | Sigma |
| 60 | Trypsin | Sigma |
| 61 | TEMED | Sigma |
| 62 | TMAO | Sigma |
| 63 | TSA | Sigma |
| 64 | Urea | Sigma |
| 65 | VECTASTAIN | Vector Lab |

## *List of Antibodies*

| Serial Number | Antibody Name | Company | Catalogue Number |
|---|---|---|---|
| 1 | His | Abcam | Ab18184 |
| 2 | H4 | Millipore | 07-108 |
| 3 | Cyclin D1 | Santacruz | SC-450 |
| 4 | PCNA | Santacruz | SC-96 |
| 5 | FLAG | Sigma | F1804 |
| 6 | GFP | Roche | 11814460001 |
| 7 | GAPDH | Ambion | AM4300 |
| 8 | H2Aub119 | Millipore | D27C4 |
| 9 | H3S10P | Abcam | Ab5176 |
| 10 | H4ac | Abcam | Ab61257 |
| 11 | H3 | Upstate | 05-499 |

| | | | |
|---|---|---|---|
| 12 | H3S28P | Abcam | Ab5169 |
| 13 | Anti-rabbit HRP | Cell Signalling | 7074 |
| 14 | Anti-mouse HRP | Sigma | A-4416 |

# *Annexure II – Cell Lines/Strains*

| Serial Number | Cell Line/Strain | Origin | Medium |
|---|---|---|---|
| 1 | BL21 (DE3) pLysS | Bacteria | LB/2xTY |
| 2 | Rosetta (DE3) pLysS | Bacteria | LB/2xTY |
| 3 | CL38 | Rat Liver | MEM |
| 4 | CL44 | Rat Liver | MEM |
| 5 | HHL5 | Human Liver | DMEM |
| 6 | HEPG2 | Human Liver | RPMI |
| 7 | HACAT | Human Colon | DMEM |
| 8 | A431 | Human Colon | DMEM |
| 9 | HFE145 | Human Stomach | RPMI |
| 10 | KATOIII | Human Stomach | RPMI |
| 11 | AGS | Human Stomach | F12/K |

# *Annexure III – Primers*

## *List of Cloning Primers*

| Gene | Vector | Primer | Sequence |
|---|---|---|---|
| H2A.1 | pET3A/pET28a | F | GTCCATATGTCTGGACGCGGCAAACAA |
| | | R | GTCGGATCCTTATTTGCCCTTGGCCTTGTG |
| H2A.2 | pET3A/pET28a | F | GTCCATATGTCCGGCCGTGGCAAGCAA |
| | | R | GTCGGATCCTCACTTGCCCTTCGCCTTATGG |
| H2B.1 63.1 | pET3A/pET28a | F | GTTGTTCCATATGCCTGAGCCCGCCAAGTCC |
| | | R | GTCGGATCCTCACTTGGAGCTGGTGTACTTGG |
| H2B.1 68.1 | pET3A/pET28a | F | GTTGTTCCATATGCCTGAGCCCGCCAAGTCC |
| | | R | GTCGGATCCTCACTTGGAGCTGGTGTACTTGG |
| H2B.1 NP | pET3A/pET28a | F | GTTGTTCCATATGCCTGAGCCCGCCAAGTCC |
| | | R | GTCGGATCCTCACTTGGAGCTGGTGTACTTGG |
| H3.2 | pET3A/pET28a | F | GTTGTTCCATATGGCCCGTACTAAGCAGAC |
| | | R | GTCGGATCCTTAGGCCCGCTCCCCGCGGAT |
| H3.3 | pET3A/pET28a | F | GTTGTTCCATATGGCCCGAACCAAGC |
| | | R | GTCGGATCCTTAAGCTCTCTCTCCCCGTATC |
| H4 Rat | pET3A/pET28a | F | GTCGTCCCATATGTCTGGCAGAGGAAAGGGTG |
| | | R | GTCGGATCCCTAGCCTCCGAAGCCGTACA |
| H2A.1 | pGEx2T | F | GTCGGATCCATGTCTGGACGCGGCAAACAA |

| | | | |
|---|---|---|---|
| | | R | GTCGAATTCCTATTATTTGCCCTTGGCCTTGTG |
| H2A.2 | pGEx2T | F | GTCGGATCCATGTCCGGCCGTGGCAAGCAA |
| | | R | GTCGAATTCTTATCACTTGCCCTTCGCCTTATG |
| GST | N-ter pET3A | F | GTCTCTAGAAATAATTTTGTTTAACTTTAAGAAGG AGATATAGATATGTCCCCTATACTAGGTTA |
| | | R | GTCCATATGTTCGATATCCGATTTTGGAGG |
| GST | Empty | F | GTCCATATGTCCCCTATACTAGGTTATTGG |
| | | R | GTCGGATCCACCGGTCTCGAGCCCGGGTACCATG GTTCGATATCCGATTTTGGAGGATGGTCGCC |
| GST | C-ter pET3A | F | GTCGAATTCCTCGAGATGTCCCCTATACTA |
| | | R | GTCGGATCCTTATTTTGGAGGATGGTCGC |
| N-ter Domain GST | pET3A/pET28a | F | GTCCATATGTCCCCTATACTAGGTTATTGG |
| | | R | GACGGATCCTCATGGACAACCACCCAACAT |
| H2A.1 | peYFPn1 | F | GTCGAATTCATGTCTGGACGCGGCAAACAA |
| | | R | GTCGGATCCTTTGCCCTTGGCCTTGTGGT |
| H2A.2 | peYFPn1 | F | GTCGAATTCATGTCCGGCCGTGGCAAG |
| | | R | GTCGGATCCCTTGCCCTTCGCCTTATGGTG |
| H2A.1 | pcDNA | F | GTTGGATCGCCATGTCTGGACGCGGC |
| | | R | GCGGAATTCTTATTTGCCCTTGGCCTTGTGG |
| H2A.2 | pcDNA | F | GTTGGATCCGCCATGTCCGGCCGTGGC |
| | | R | GCGGAATTCTCACTTGCCCTTCGCCTTATGGTG |
| H2A.1 | SDM 16th | F | CAAGGCCAAGTCCCGCTCCTCCCGG |
| | | R | GGAGGAGCGGGACTTGGCCTTGGCG |
| H2A.1 | SDM 51st | F | TCCGGTGTACATGGCGGCTGTGCTG |
| | | R | CAGCACAGCCGCCATGTACACCGGA |
| H2A.1 | SDM 99th | F | CAAGCTGCTGGGCAAGGTGACCATCGCCC |
| | | R | CGATGGTCACCTTGCCCAGCAGCTTGTTG |
| H2A.2 | SDM 51st | F | GCCGGTATACCTGGCGGCGGTGCTGGAGTAC |
| | | R | GTACTCCAGCACCGCCGCCAGGTATACCGGC |

## *List of PCR Primers Used*

| Gene | Application | Primer | Sequence |
|---|---|---|---|
| Usp21 | RT/Real Time-PCR | F | ACTTCTCTCCGGCGTCTT |
| | | R | TGTGTGGTGAGCCATCTT |
| LCN2 | RT/Real Time-PCR | F | GGACCGAACGGTTCCAGG |
| | | R | CCCTGACGAGGATGGAAG |
| RING1 | RT PCR | F | CCAAGCGGTCCCTACGGCC |
| | | R | CCTCGATACTGGAGCTCA |
| GAPDH | RT/Real Time-PCR | F | GGATTTGGTCGTATTGGGCG |

| | | | |
|---|---|---|---|
| | | R | ATCGCCCCACTTGATTTTGG |
| *E. coli* GAPDH | RT/Real Time-PCR | F | CTCCGATGTTCGTTAAAG |
| | | R | CGTTGATAACTTTAGCCA |
| PCNA | RT-PCR | F | TCACAAAAGCCACTCCACTG |
| | | R | CATCTCAGAAGCGATCGTCA |
| Cyclin D1 | RT-PCR | F | CGTGGCCACCTGGATGCTAGAG |
| | | R | TGCAGCAACTCCTCGGGGCGGAT |
| H2A.1 | Rat RT/Real Time-PCR | F | CTGTGCTGGAGTACCTGACG |
| | | R | TGTGGTGGCTCTCAGTCTTC |
| H2A.2 | Rat RT/Real Time-PCR | F | GAAGACGGAGAGCCACCATA |
| | | R | GGAAGAGTAGGGCACACGAC |
| H2A1C | Human RT/Real Time-PCR | F | GGTGATTTTTGTCTGATTG |
| | | R | CTGCGTAGTTGCCTTTA |
| H2AA3 | Human RT/Real Time-PCR | F | CGGTGACTACTATCGCTGT |
| | | R | AGCCATGTAGACGGGC |
| H2A.1 | RT/Real Time-PCR | F | GTCCATATGTCTGGACGCGGCAAACAA |
| | | R | GTCGGATCCTTATTTGCCCTTGGCCTTGTG |
| H2A.2 | RT/Real Time-PCR | F | GTCCATATGTCCGGCCGTGGCAAGCAA |
| | | R | GTCGGATCCTCACTTGCCCTTCGCCTTATGG |
| H2B.1 | RT/Real Time-PCR | F | GTTGTTCCATATGCCTGAGCCCGCCAAGTCC |
| | | R | GTCGGATCCTCACTTGGAGCTGGTGTACTTGG |
| H4 | RT/Real Time-PCR | F | GTCGTCCCATATGTCTGGCAGAGGAAAGGGTG |
| | | R | GTCGGATCCCTAGCCTCCGAAGCCGTACA |
| T7 | RT | F | TAATACGACTCACTATAGGG |
| | | R | GCTAGTTATTGCTCAGCGGT |
| Ki67 | RT PCR | F | AGACGTGACTGGTTCCCAAC |
| | | R | ACTGCTTCCCGAGAACTGAA |
| 18S | RT PCR | F | CGCGGTTCTATTTTGTTGGT |
| | | R | AGTCGGCATCGTTTATGGTC |

# *Annexure IV - Media and Buffer Composition*

## a. Phosphate Buffered Saline

1. Dissolve the following in 800ml distilled $H_2O$:
    - 8g of NaCl
    - 0.2g of KCl
    - 1.44g of $Na_2HPO_4$
    - 0.24g of $KH_2PO_4$
2. Adjust pH to 7.4 with HCl.
3. Adjust volume to 1L with additional distilled $H_2O$.
4. Sterilize by autoclaving.

The final concentration of the components are as follows:

a. 137mM NaCl
b. 2.7mM KCl
c. 10mM $Na_2HPO_4$
d. 2mM $KH_2PO_4$

## b. 2x BBS

1. Add the following:
    - 266.56mg BES
    - 409.5mg NaCl
    - 6.65mg $Na_2HPO_4$
2. Adjust the pH to 6.95
3. Adjust volume to 25ml with distilled $H_2O$.
4. Sterile filter inside laminar hood

## c. Electrophoresis Buffer

1. Add the following:
    - 6.05g Tris
    - 28.5g Glycine
    - 1g SDS
2. Adjust volume to 1000ml with distilled $H_2O$.

## d. Phosphate Buffer

500ml of 1M potassium phosphate buffer pH 7.4 = 401ml of 1M $K_2HPO_4$ + 99ml of 1M $KH_2PO_4$

- 1M $K_2HPO_4$ = 87.09gm in 500ml
- 1M $KH_2PO_4$ = 13.6gm in 100ml

## e. TEV Buffer

| Reagent | Stock | Working | For 10ml |
|---|---|---|---|
| **Tris pH 8.0** | 1M | 50mM | 500µl |

| | | | |
|---|---|---|---|
| **EDTA** | 0.5M | 0.5mM | 10µl |
| **DTT** | 1M | 1mM | 10µl |
| **$H_2O$** | | | 9.48ml |

## f. FactorXa Buffer

| Reagent | Stock | Working | For 10ml |
|---|---|---|---|
| **Tris pH 8.0** | 1M | 20mM | 200µl |
| **NaCl** | 5M | 100mM | 200µl |
| **$CaCl_2$** | 1M | 2mM | 20µl |
| **$H_2O$** | | | 9.58ml |

## g. Chemicals for In-gel Digestion

| Reagent | Stock | Working | For 1ml |
|---|---|---|---|
| **$NH_4HCO_3$** | 1M (79g/mol) | 25mM | 79mg |
| **DTT** | 1M (154g/mol) | 10mM | 154mg |
| **Iodoacetamide** | 1M (185g/mol) | 55mM | 185mg |
| **Trypsin** | | 20ng/µl | |
| **$K_3[Fe(CN)_6]$** | 1M (329g/mol) | 30mM | 329mg |
| **Sodium thiosulphate** | 1M(158g/mol) | 100mM | 158mg |

## h. Transfer Buffer

| Reagent | Stock | Working | For 1L |
|---|---|---|---|
| **Glycine** | | 0.19M | 14gm |
| **Tris** | | 25mM | 3.03gm |
| **SDS** | 10% | 0.01% | 1ml |
| **Methanol** | 100% | 20% | 200ml |
| **$H_2O$** | | | Upto 1L |

## i. Chemicals for SDS-PAGE

**Acrylamide Main Stock:**

292g acrylamide and 8g bis-acrylamide are mixed and dissolved in 500ml of $H_2O$. Heating at $37^0$C may be required to dissolve acrylamide. Volume is made up to 1000ml.

**Resolving Gel (18%):**

| Reagent | Stock | Working | For 10ml |
|---|---|---|---|
| **Acrylamide – Bis-acrylamide (29:1)** | 30% | 18% | 6ml |
| **Tris pH 8.8** | 1.5M | 375mM | 2.5ml |

| | | | |
|---|---|---|---|
| **SDS** | 10% | 0.1% | 100µl |
| **APS** | 10% | 0.05% | 50µl |
| **TEMED** | | 0.05% | 5µl |
| **$H_2O$** | | | 1.345ml |

**Resolving Gel (10%):**

| Reagent | Stock | Working | For 10ml |
|---|---|---|---|
| **Acrylamide – Bis-acrylamide (29:1)** | 30% | 10% | 4.076ml |
| **Tris pH 8.8** | 1.5M | 375mM | 2.5ml |
| **SDS** | 10% | 0.1% | 100µl |
| **APS** | 10% | 0.05% | 50µl |
| **TEMED** | | 0.05% | 5µl |
| **$H_2O$** | | | 3.269ml |

**Stacking Gel (4%):**

| Reagent | Stock | Working | For 10ml |
|---|---|---|---|
| **Acrylamide – Bisacrylamide (29:1)** | 30% | 4% | 1.3ml |
| **Tris pH 6.8** | 1M | 125mM | 1.25ml |
| **SDS** | 10% | 0.1% | 100µl |
| **APS** | 10% | 0.05% | 50µl |
| **TEMED** | | 0.1% | 10µl |
| **$H_2O$** | | | 7.29ml |

## j. 2x Sample Buffer for Histones

| Reagent | Stock | Working | For 10ml |
|---|---|---|---|
| **Tris pH 6.8** | 1M | 125mM | 1.25ml |
| **Glycerol** | 100% | 20% | 2ml |
| **SDS** | 10% | 4% | 4ml |
| **β-Mercaptoethanol** | 14.3M | 300mM | 200µl |
| **Bromophenol blue** | 0.5% | 0.05% | 5mg |
| **$H_2O$** | | | 2.55ml |

## k. Chemicals for AUT-PAGE

**Resolving Gel:**

| Reagent | Stock | Working | For 30ml |
|---|---|---|---|
| **Acrylamide** | 60% | 25mM | 7.5ml |
| **Bis-acrylamide** | 2.5% | 10mM | 1.2ml |
| **Acetic acid** | | | 1.8ml |
| **Ammonium hydroxide** | | | 98.4µl |

| | | | |
|---|---|---|---|
| **Urea** | | 8M | 14.4gm |
| **$H_2O$** | | | Upto 27.31ml |
| **Vacuum to degas** | | | |
| **Triton-X** | 25% | | 675µl |
| **Riboflavin** | | | 2ml |
| **TEMED** | | | 150µl |

**Stacking gel:**

| Reagent | Stock | Working | For 10ml |
|---|---|---|---|
| **Acrylamide** | 60% | 25mM | 670µl |
| **Bis-acrylamide** | 2.5% | 10mM | 640µl |
| **Acetic acid** | | | 570µl |
| **Ammonium hydroxide** | | | 35µl |
| **Urea** | | 8M | 4.8gm |
| **$H_2O$** | | | Upto 9.3ml |
| **Vacuum to degas** | | | |
| **Riboflavin** | | | 650µl |
| **TEMED** | | | 50µl |

**Sample Preparation for AUT-PAGE:**

2ml of Urea-DTT solution (0.864gm urea, 14mg of DTT) was prepared. AUT-sample buffer was prepared by taking 1.8ml of Urea-DTT solution and adding 100µl ammonia and 100µl of phenolphthalein. 50µl of the sample buffer was added on 20µg of vacuum dried histones. It was mixed and incubated for 30mins at $37^0$C. Then 2.5µl acetic acid and 2.5µl concentrated dye (25µl acetic acid and a pinch of methylene blue in 500µl of Urea-DTT solution) was added.

## l. TELT Buffer

| Reagent | Stock | Working | For 40ml |
|---|---|---|---|
| **Tris-Cl pH 7.5** | 1M | 50mM | 2ml |
| **EDTA pH 8.0** | 0.5M | 62.5mM | 5ml |
| **Triton X-100** | 100% | 0.4% | 160µl |
| **LiCl** | | 2.5M | 4.239g |
| **$H_2O$** | | | Upto 40ml |

## m. SOB Media

| Reagent | Stock | Working | For 100ml |
|---|---|---|---|
| **Bactotryptone** | | 2% | 2gm |
| **Yeast Extract** | | 0.5% | 0.5g |
| **NaCl** | 5M | 10mM | 200µl |

| | | | |
|---|---|---|---|
| **KCl** | 3M | 2.5mM | 83.33µl |
| **$MgCl_2$** | 1M | 10mM | 1ml |
| **$MgSO_4$** | | 10mM | 0.246g |
| **$H_2O$** | | | Upto 100ml |

## n. Transforming Buffer

1. Add the following:
    - 1.0884g $MnCl_2.4H_2O$
    - 0.22053g $CaCl_2.2H_2O$
    - 1.8637g KCl
    - 0.30237g PIPES
2. Adjust pH to 6.7 with KOH
3. Adjust volume to 100ml with distilled $H_2O$.
4. Sterile filter

The final concentration of the components are as follows:

- 55mM $MnCl_2.4H_2O$
- 15mM $CaCl_2.2H_2O$
- 250mM KCl
- 10mM PIPES

## o. DMEM

1. Add the following:
    - DMEM
    - 81.4mg Non Essential Amino Acid
    - 0.293gm L-Glutamine (working 2mM)
    - 3.7gm Sodium Bicarbonate (working 44mM)
2. Adjust volume to 1000ml with distilled $H_2O$.
3. Sterile filter inside laminar hood

## p. MEM

1. Add the following:
    - 9.6gm MEM powder (Sigma, #M6043)
    - 81.4mg Non Essential Amino Acid
    - 0.293gm L-Glutamine (working 2mM)
    - 2.2gm Sodium Bicarbonate
    - 3.574gm HEPES
2. Adjust volume to 1000ml with distilled $H_2O$.
3. Sterile filter inside laminar hood

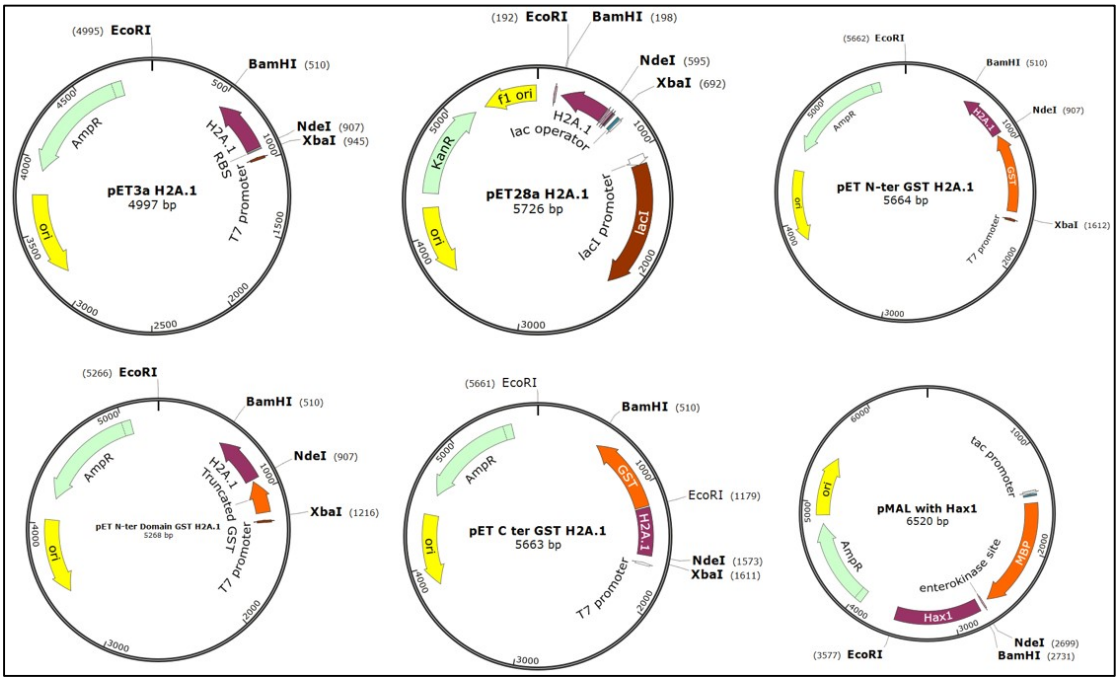
## q. RPMI

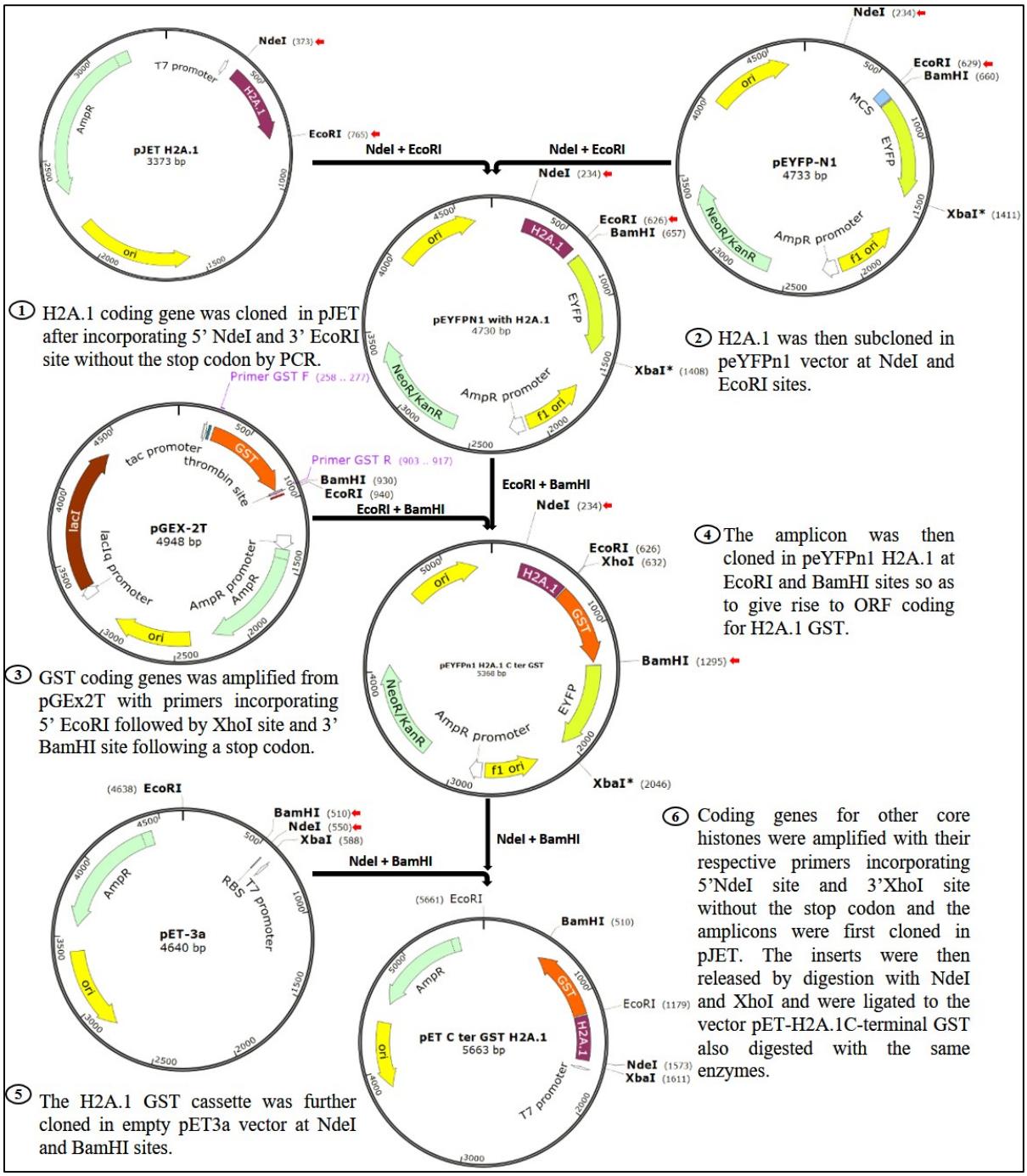
1. Add the following:
    - RPMI
    - 81.4mg Non Essential Amino Acid
    - 0.293gm L-Glutamine (working 2mM)
    - 2gm Sodium Bicarbonate
2. Adjust volume to 1000ml with distilled $H_2O$.
3. Sterile filter inside laminar hood

### r. Trypsin Composition

4. Add the following:
    - 2.5gm Trypsin (0.25%)
    - 0.2gm EDTA
    - 0.5gm Glucose
5. Adjust volume to 1000ml with distilled $H_2O$.
6. Sterile filter inside laminar hood

## *Annexure V - Construct Preparation Maps*




pJET H2A.1
3373 bp
T7 promoter
H2A.1
AmpR
ori
NdeI (373)
EcoRI (765)
NdeI + EcoRI
NdeI + EcoRI
pEYFP-N1
4733 bp
MCS
EYFP
NeoR/KanR
AmpR promoter
f1 ori
ori
NdeI (234)
EcoRI (629)
BamHI (660)
XbaI* (1411)
pEYFPN1 with H2A.1
4730 bp
EcoRI (626)
BamHI (657)
XbaI* (1408)
1 H2A.1 coding gene was cloned in pJET after incorporating 5' NdeI and 3' EcoRI site without the stop codon by PCR.
2 H2A.1 was then subcloned in peYFPn1 vector at NdeI and EcoRI sites.
Primer GST F (258 .. 277)
pGEX-2T
4948 bp
tac promoter
thrombin site
GST
lacI
lacIq promoter
AmpR promoter
AmpR
ori
Primer GST R (903 .. 917)
BamHI (930)
EcoRI (940)
EcoRI + BamHI
EcoRI + BamHI
3 GST coding genes was amplified from pGEx2T with primers incorporating 5' EcoRI followed by XhoI site and 3' BamHI site following a stop codon.
pEYFPn1 H2A.1 C ter GST
5368 bp
NdeI (234)
EcoRI (626)
XhoI (632)
BamHI (1295)
XbaI* (2046)
4 The amplicon was then cloned in peYFPn1 H2A.1 at EcoRI and BamHI sites so as to give rise to ORF coding for H2A.1 GST.
pET-3a
4640 bp
(4638) EcoRI
BamHI (510)
NdeI (550)
XbaI (588)
RBS
T7 promoter
AmpR
ori
Ndel + BamHI
Ndel + BamHI
5 The H2A.1 GST cassette was further cloned in empty pET3a vector at NdeI and BamHI sites.
pET C ter GST H2A.1
5663 bp
(5661) EcoRI
BamHI (510)
EcoRI (1179)
NdeI (1573)
XbaI (1611)
6 Coding genes for other core histones were amplified with their respective primers incorporating 5'NdeI site and 3'XhoI site without the stop codon and the amplicons were first cloned in pJET. The inserts were then released by digestion with NdeI and XhoI and were ligated to the vector pET-H2A.1C-terminal GST also digested with the same enzymes.

# PUBLICATIONS

Biotechnology Reports 11 (2016) 62–69

Contents lists available at ScienceDirect

Biotechnology Reports

journal homepage: www.elsevier.com/locate/btre

# Incorporation of a tag helps to overcome expression variability in a recombinant host

Saikat Bhattacharya[a], Divya Reddy[a], Raja Reddy[b], Asmita Sharda[a], Kakoli Bose[b], Sanjay Gupta[a,*]

[a] Epigenetics and Chromatin Biology Group, Gupta Lab, India
[b] Integrated Biophysics and Structural Biology Lab, Cancer Research Institute, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Kharghar, Navi Mumbai 410210, MH, India

ARTICLE INFO

*Article history:*
Received 8 March 2016
Received in revised form 8 June 2016
Accepted 27 June 2016
Available online 18 July 2016

*Keywords:*
Rare codons
Histones
Truncated GST
Misfolding
Translation

ABSTRACT

Epigenetics have witnessed a renewed interest over the past decade and assays with recombinant histones has become an important tool for uncovering various aspects of histone biology. However, at times absence of recombinant histone accumulation in bacteria is encountered which is also commonly observed for many eukaryotic proteins in general. In this study, we have investigated the effect of multiple parameters on heterologous expression of proteins. We show that there is marked variability in the accumulation of H2A.2, H2B.1, H3.2 and H4 in the recombinant host, possibly owing to translational variability and degradation by the host proteases. We found that the variability could be overcome by incorporation of the commonly used purification tags, like GST or MBP, of appropriate size and position. Our results provide compelling evidence that transcript parameters like rare codon and GC content, mRNA secondary structure etc. together modulate translation kinetics and govern recombinant protein accumulation.


## 1. Introduction

Histones are highly basic proteins with the primary function of packaging the DNA of an organism in a very organised fashion inside the nucleus. Histones undergo various post-translational modifications (PTMs) and have sequence divergent forms known as histone variants. The functional importance of PTMs and variants has led to massive interest in uncovering the outcome of the biochemical changes in the composition of nucleosomes and how these changes are brought about. For conducting *in vitro* experiments recombinant expression has become a very powerful tool for obtaining abundant amounts of highly purified histones with specific PTMs or amino acid composition.

*E. coli* has been popularly used as heterologous host for robust recombinant production of histones like for most other proteins. However, at times, drastic differences in expression levels of histones is observed in bacteria with complete absence of recombinant histone expression in the worst scenarios [16,28]. This has been attributed to the presence of rare codons in the coding sequence of histones. As per the codon usage bias hypothesis, rare codons may play a role in determining levels of protein expressed in a heterologous host [3]. Apart from the absolute number of such codons, parameters like rare codon clusters [12], their number in the 5′-end of a transcript [7,29] and first six codons [6] and the ones coding for arginine [4,22] are considered important determinants. Likewise, high frequency of arginine residues is suspected to negatively affect expression of histone H1 and H5 [8]. Although, codon optimization proves to be useful in some cases, however, at times, the lack of expression persists [1,28] suggesting that other factors might be involved [reviewed by Ref. [31]].

*Abbreviations:* HAX-1, human protein HCLS-1 associated protein X-1; NAP1, nucleosome assemble protein 1; DUSP1, dual specificity phosphatase 1; PP1, protein phosphatase 1; IPTG, Isopropyl β-D-1-thiogalactopyranoside; RT-PCR, reverse transcriptase polymerase chain reaction; CAI, codon adaptation indexes; RBS, ribosome-binding site; TMAO, trimethylamine oxide; GST, glutathione-*S*-transferase; MALDI, matrix-assisted laser desorption/ionization; MBP, maltose binding protein; GAPDH, glyceraldehyde phosphate dehydrogenase.

* Corresponding author at: Epigenetics and Chromatin Biology Group, Gupta Lab, Cancer Research Institute, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Kharghar, Navi Mumbai 410210, MH, India.

*E-mail addresses:* sbhattacharya@actrec.gov.in (S. Bhattacharya), dvelga@actrec.gov.in (D. Reddy), rkuppili@actrec.gov.in (R. Reddy), asharda@actrec.gov.in (A. Sharda), kbose@actrec.gov.in (K. Bose), sgupta@actrec.gov.in (S. Gupta).

http://dx.doi.org/10.1016/j.btre.2016.06.002

We also encountered lack of accumulation of few of the histone isoforms on attempting heterologous expression in *E. coli* but did not find a stringent correlation with rare codon content. Interestingly, the transcripts for these histone proteins were being formed in the host. Our results provide compelling evidence that the codon bias usage, mRNA secondary structure and GC content of the histone transcript together modulate the post-transcriptional steps and govern recombinant protein accumulation. Notably, we found that the variability in accumulation could be overcome by incorporation of the purification tags, like GST or MBP, of appropriate size and position. Further, we extended these correlations to six non-histone proteins and show that this probably, can be used as a general strategy to obtain heterologous expression.

## 2. Materials and methods

### 2.1. Construct preparation

The coding sequences of the genes were amplified from the cDNA synthesized from RNA (treated with DNaseI) isolated from cell lines. The amplicons were cloned into pTZRT57 vector (Thermo scientific). The cloned fragments were sequenced. For subsequent cloning into different expression vectors, the coding sequences were amplified with primers incorporating the appropriate restriction sites and were subcloned, maintaining the correct reading frame. More details pertaining to cloning are available on request. Please see Supplementary Fig. S8 and S9 for a few of the construct maps.

### 2.2. Growth and IPTG induction of transformed bacterial expression hosts

A single colony was inoculated from the plates of transformed bacteria in 5 mL or 20 mL LB media and incubated at 37 °C until the $OD_{600}$ reached between 0.3 and 0.6. Induction of recombinant protein expression were carried out with 0.2 mM IPTG. When lack of expression was encountered, titration of IPTG concentration was carried out ranging 0.2 mM–2 mM. The cultures were induced for 3 h at 37 °C or overnight at 18 °C. To investigate the effect of trimethylamine oxide (TMAO), it was added at a working concentration of 60 mM, 1.5 h post IPTG induction and the cells were harvested 1.5 h post-addition. The soluble and the insoluble fractions of proteins were separated by resuspending the cells in buffer containing 50 mM Tris-Cl pH 8.0, 0.5% Triton X-100 and 100 μg/ml lysozyme followed by three rounds of sonication, each for 30 s at 30% amplitude. The lysate was then centrifuged at 27000*g* for 30 min at 4 °C. The supernatant and pellet, thus obtained contains the soluble proteins and the insoluble proteins respectively. The proteins were resolved by 18% SDS-PAGE followed by Coomassie staining (Brilliant Blue R250).

### 2.3. RT-PCR and real-time PCR

Total RNA was extracted and treated with DNaseI from previously collected bacterial cells as per the manufacturer's (Macherey-Nagel) instructions. RNA (2 μg) was subjected to reverse transcription using M-MLV Reverse Transcriptase and random hexamer primers according to the manufacturer's instructions. cDNAs were then amplified with the corresponding gene-specific primer sets, designed to amplify the total coding sequence. The PCR products were analyzed on a 1% agarose gels. The cDNA synthesized was further used for Real-time PCR experiments with Syber green dye. The expression levels were plotted as relative fold change with respect to GAPDH (glyceraldehyde 3-phoshphate dehydrogenase). Similar results were obtained when fold change was plotted with respect to Ampicillin expression which is expressed from the vector backbone (data not shown).

### 2.4. Rare codon, CAI and RNA secondary structure prediction

Rare codons were predicted using the Caltor Prediction tool (http://people.mbi.ucla.edu/sumchan/caltor.html). CAI (Codon Adaptation Index) for the coding regions was calculated using CAIcal server (http://genomes.urv.es/CAIcal/) and RNA structures were analyzed by Vienna RNA Websuite program on the web at http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi.

### 2.5. Accession numbers

H2A.1 (GenBank: JX661508.1), H2A.2 (GenBank: JX661509.1), H2B.1 63.1 (XM_002725263.1), H2B.1 68.1(XM_002725268.1 updated to XM_002725268.2), H4 (NM_001123469.1), H3.3 (GenBank: BC006497.2), HAX-1 (GenBank: AK290626.1), DUSP1 (NM_004417.3), PP1α (NM_002708.3), HLA (GenBank: CAA59215.1), β2M (NM_004048.2). H2B.1 XM_002725263.1 and H2B.1 XM_002725268.1 are referred to as H2B.1 63.1 and H2B.1 68.1 respectively in this article.

## 3. Results

### 3.1. Lack of recombinant histone expression is a commonly encountered phenomenon

During purification of rat histone proteins using recombinant methods, we encountered lack of expression of few of the transcripts. There was lack of accumulation of proteins H2A.2 (lane 6), H2B.1 68.1 (lane 10) and H4 (lane 16) even though we used the *E. coli* BL21 (DE3) pLysS strain [Fig. 1a]. No accumulation was observed even at lower growth temperatures post-induction and at varied IPTG concentrations (data not shown). RT-PCR [Fig. 1b (i)] and real-time PCR data demonstrates that not only were the transcripts produced, but also, the relative transcript levels of H2A.2, H2B.1 68.1and H4 were comparable to H2A.1 [Fig. 1b (ii)]. The expression of histones was next attempted in Rosetta (DE3) pLysS, a codon-optimized strain, as the presence of rare codons is speculated to interfere with translation. However, the lack of accumulation of H2A.2 (lane 6) and H4 (lane 16) [Fig. 1a lower image] persisted.

### 3.2. Recombinant expression does not stringently correlate with rare codon parameters

Analysis of the rare codon parameters of the histone transcripts was carried out to look for possible correlation between accumulation level of proteins and the proposed parameters. Several inconsistences with the codon bias hypothesis were noted. For example, the H3.3 transcript, in addition to 12 rare codons for arginine, contains the highest percentage (16.91%) and number of rare codons in a cluster (four) [Table 1]. Robust accumulation of H3.3 in *E. coli* and lack of accumulation of H2A.2 with the least percentage (6.15%) and rare codons in a cluster (zero), is contradictory to codon bias hypothesis [Table 1]. Further, no correlation between the presence of rare codons at the 5′-end of a transcript with expression levels was observed ["First 6 codons", Table 1]. These contradictions were also reflected in codon adaptation indexes (CAI), which is a proposed important parameter [25] [Table 1]. For instance, H2A.2 having the highest CAI of 0.706 amongst the histone transcripts didn't express in either of the strains [Fig. 1a,c].

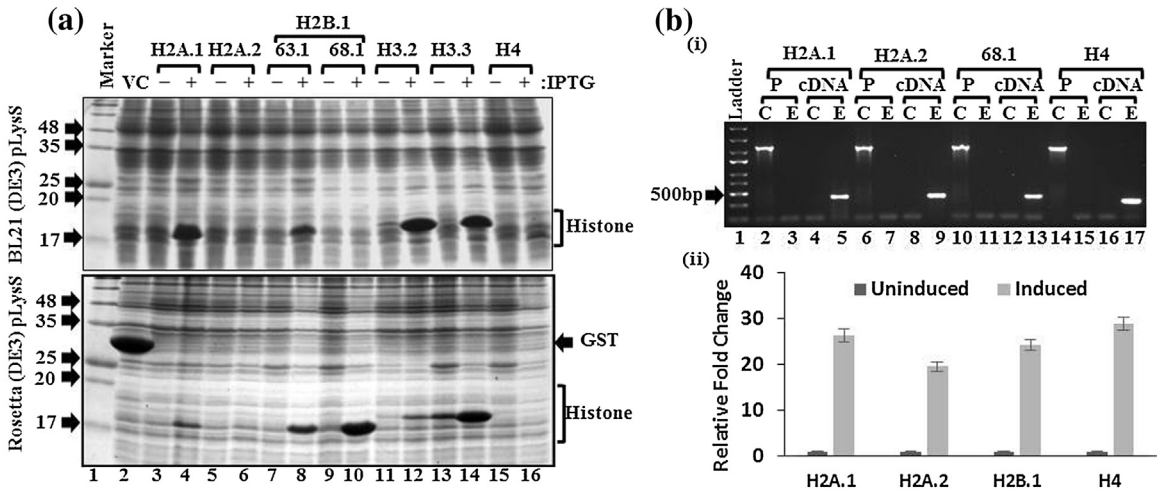



**Fig 1.** Expression status of recombinant proteins in *E. coli*. (a) SDS-PAGE analyses (18%, coomassie stained) of whole cell lysates of BL21 (DE3) pLysS (upper image) and Rosetta (DE3) pLysS (lower image) harbouring the pET3a histone constructs, induced with 0.2 mM IPTG at 37 °C for 2 h. pET3a empty vector was used as vector control in BL21 (DE3) pLysS (marked as VC). pGEX2 T empty vector expressing GST was used as an induction control in Rosetta (DE3) pLysS (marked as VC) . The region marked as Histone spans the molecular weight range of the histones being expressed. (b) For checking expression status at the transcript level (i) semi-quantitative RT-PCR was done. 'cDNA' denotes cDNA synthesized by RT-PCR using random hexamer primers followed by PCR amplification with respective gene specific primers. For control (C) cDNA from bacteria having empty pET3a vector was amplified using corresponding gene specific primers and for the experimental sets (E) cDNA from bacteria harbouring pET3a histone constructs was amplified using respective histone primers. H2A.1 was used as a positive control. 'P' denotes experiment to validate DNA free preparation of RNA, in which C is control for amplification where respective plasmids were used as a template. E is the experimental set in which DNA removal by DNaseI from RNA preparation was validated. (ii) Real time PCR to denote the relative fold change upon induction for respective constructs.

### *3.3. mRNA secondary structure and GC content of transcripts does not always correlate with protein expression levels*

As an alternative explanation for lack of recombinant protein accumulation the importance of factors like strong mRNA secondary structure and the 'GC' content of the transcript have been emphasized in relation to transcription and translation [8,14,15]. Therefore, we investigated the possible correlation of mRNA secondary structure and GC content of the histone transcript on protein expression. Vienna RNA web suite [10] was used to maintain consistency with the previously published literature [32] for predicting a possible long-range secondary structure spanning the ribosome-binding site (RBS), translation initiator AUG and the first seven codons. Again, a general pattern did not emerge from our analysis, considering the complexity of the overall secondary structure, the structure near the ATG translation initiation codon, free energy of the structure (see Supplementary Fig. S1) and the% GC of the full-length mRNA transcripts as well as the first six codons (see Supplementary Fig. S1). Most strikingly, the mRNA secondary structure, the free-energy values (−14.20 kcal/mol) and GC content were identical in the case of the three transcripts of H2B.1 (see Supplementary Fig. S1). Thus, their differential expression pattern strongly suggests that the mRNA secondary structure and/or GC content may not be the primary cause behind the results obtained. Although, the secondary structures are predicted, nonetheless, the differential expression pattern observed for the same transcript in different strains cannot be explained based on mRNA secondary structure differences.

### *3.4. Incorporation of N-terminus tag affects recombinant protein accumulation*

Tags are frequently used in purification of proteins. We wanted to test the effect of tags on recombinant protein expression. Interestingly, with 6xHis tag accumulation of H2A.2 protein (lane 6) was seen even without using a codon optimized strain [Fig. 2a]. This further highlights the lack of direct correlation between rare

**Table 1**
*In silico* rare codon analysis of histone transcripts.

| Amino Acid | Rare Codon | H2A.1 | H2A.2 | H2B.1 63.1 | H2B.1 68.1 | H3.2 | H3.3 | H4 |
|---|---|---|---|---|---|---|---|---|
| Arg | CGA | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| | CGG | 2 | 2 | 1 | 1 | 5 | 3 | 2 |
| | AGG | 2 | 0 | 0 | 0 | 1 | 5 | 0 |
| | AGA | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| Gly | GGA | 1 | 1 | 3 | 1 | 0 | 1 | 2 |
| | GGG | 0 | 1 | 1 | 1 | 3 | 3 | 1 |
| Iso | AUA | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Leu | CUA | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Pro | CCC | 3 | 2 | 3 | 3 | 3 | 5 | 1 |
| Thr | ACG | 1 | 2 | 2 | 2 | 4 | 1 | 0 |
| | Consecutive | 0 | 0 | 2 | 2 | 2 | 4 | 2 |
| | First 6 Codons | 1 | 0 | 1 | 1 | 0 | 1 | 2 |
| | Frequency | 7.69% | 6.15% | 7.93% | 6.34% | 11.76% | 16.91% | 8.73% |
| | CAI | 0.673 | 0.706 | 0.625 | 0.654 | 0.666 | 0.644 | 0.617 |
| | Total | 10 | 8 | 10 | 8 | 16 | 23 | 9 |
| | Expression | √ | X | √ | X | √ | √ | X |

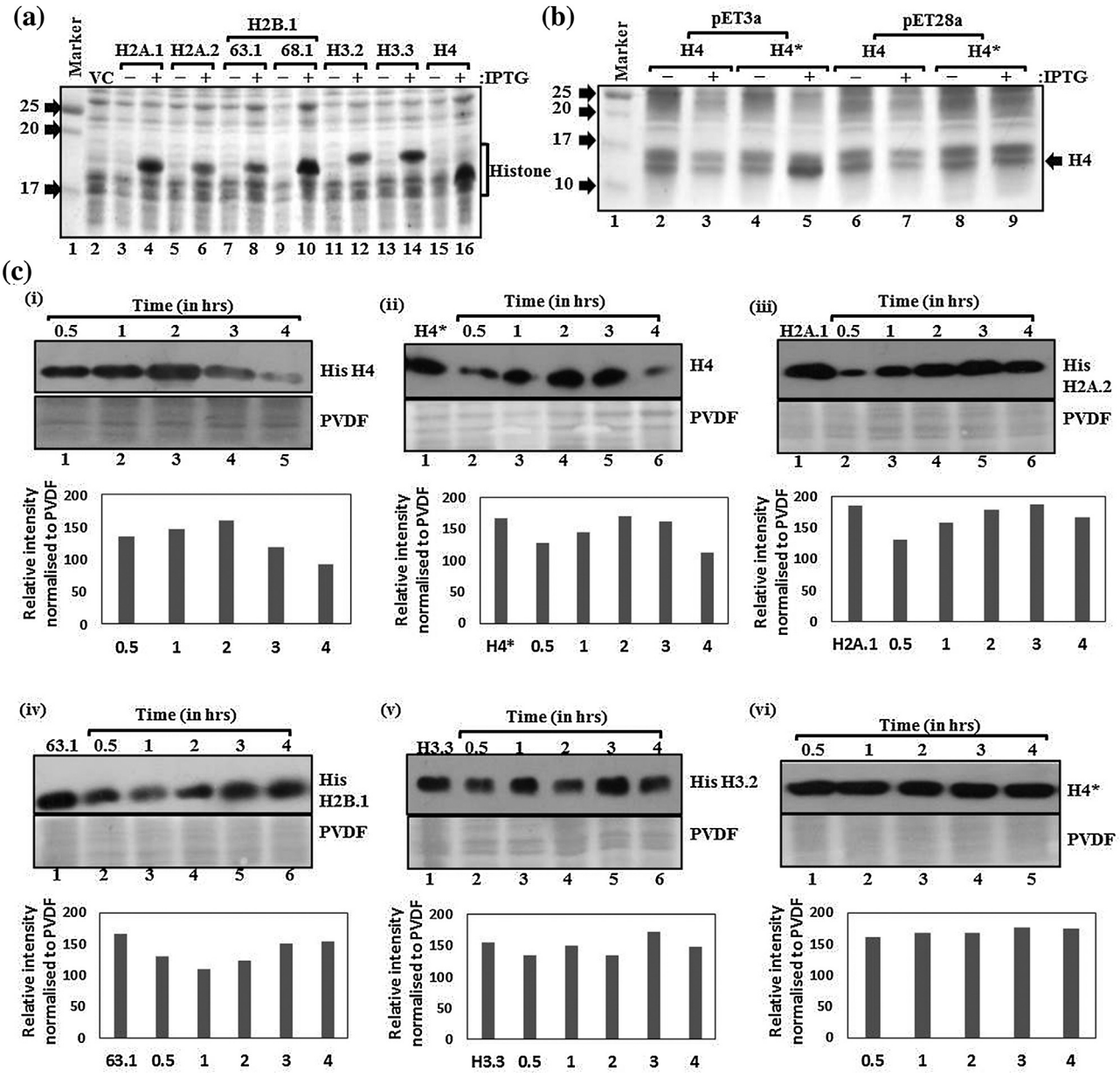



**Fig. 2.** Effect of incorporation of N-terminal 6X His tag: (a) SDS-PAGE analysis of total cell lysates of BL21 (DE3) pLysS harbouring 6X his tag histone constructs, induced with 0.2 mM IPTG at 37 °C for 2 h (b) 18% coomassie stained gel showing the expression status of histone H4 with and without 6x His tag, induction done at 37 °C for 2 h with 0.2 mM IPTG. H4* denotes codon optimized H4. (c) Western blotting analysis of total cell lysates of BL21 (DE3) pLysS harbouring various histone constructs at different time points post-induction with 0.2 mM IPTG, with anti-His antibody for (i)(iii)–(v) and H4 antibody for (ii)(vi). H4* was selected as a negative control for the experiment, as it does not undergo degradation. The intensity quantification of the bands in the western blots is also depicted below the blots.

codons and protein accumulation. Expression was also seen in case of pET28a H2B.1 68.1 (lane 10) and H4 (lane 16) but the results were "inconsistent". Importantly, this clearly suggested translational variability or the possibility of involvement of protein degradation or both behind the varied accumulation pattern observed for different proteins with changing parameters.

We wanted to investigate this aspect further. We included a codon optimized H4 transcript, in our studies ($H4^*$) that expresses robustly in bacteria. Notably, expression was not obtained in pET28a vector [Fig. 2b]. A time-course experiment was carried out post-induction to monitor H4 accumulation with non-codon optimized H4 in pET28a [Fig. 2c (i)] and pET3a [Fig. 2c (ii)] vector backbone. A decrease in the intensity of band in the immunoblot analysis was observed. Similar decrease was observed in previous report [30] although the kinetics observed was different. Comparable observations were made for H2A.2, H2B.1 68.1 and H3.2 proteins as well [Fig. 2c (iii)–(v)], again, with differing kinetics. Changing growth medium or incorporation of protein stabilising osmolytes like trimethylamine oxide (TMAO) and glycerol did not significantly improve recombinant protein accumulation (data not shown).

### *3.5. Incorporation of GST tag overcomes variability in recombinant protein accumulation*

Strains like BL21 (DE3) pLysS and Rosetta (DE3) pLysS, that are deficient in *OmpT* spare GST fusion proteins from degradation [9]. Hence, GST was cloned upstream of the histones already cloned in

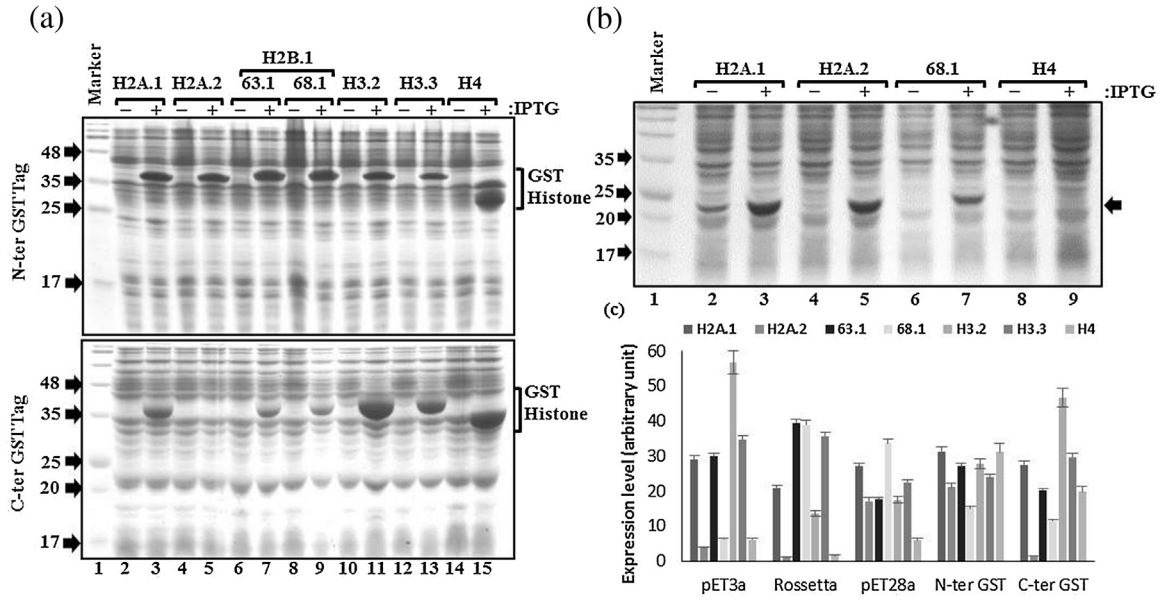



**Fig. 3.** Size and position of the tag are critical: (a) SDS-PAGE (18% coomassie stained) analysis of cell lysates of BL21 (DE3) pLysS harbouring N-terminal GST tagged (upper image) and with C-terminal GST (lower image) histone constructs induced with 0.2 mM IPTG 37 °C for 2 h. (b) Expression status of histones tagged with N-terminal domain of GST (truncated GST), protein production induced by adding 0.2 mM IPTG at 37 °C for 2 h. (c) Graph showing the relative levels of various histones cloned in different vectors and expressed in different strains.

pET3a (*construct details available on request*). This rules out contribution of difference in vector backbone and copy number towards expression level. After incorporation of GST, indeed, consistent levels of accumulation were seen for all the constructs, including H2B.1 68.1 and H4, as judged by the presence of an intense band around ~35 kD in the induced samples in coomassie stained gels [Fig. 3a].

To test whether the location of the tag is also a determinant, the GST tag was cloned at the C-terminus of histones. Apart from H2A.2 (lane 5), all constructs expressed robustly, although, there was a considerable difference in expression levels of different proteins [Fig. 3a lower image]. Importantly, this rules out the possibility that the expression achieved was due to altered stability of the 5′ end mRNA or 5′ end amino acid residues or enhanced translation initiation resulting from tag incorporation. Quantification depicted in Fig. 3c indicates that with N-ter GST tag most consistent level of recombinant protein accumulation is achieved.

Next, we asked whether the size of the tag influences accumulation levels. To address this, the N-terminal domain of GST comprising of 86 amino acid residues was cloned upstream of the histones in pET3a vector and their expression was attempted. The results were similar to that obtained with the 6xHis tag with stable expression of H2A.2 and H2B.1 and the H4 protein being expressed inconsistently [Fig. 3b]. This clearly suggested that the size of the tag with respect to that of protein of interest is an important determinant.

### *3.6. Incorporation of a suitable tag can possibly be used as a general procedure to obtain recombinant protein expression*

We included a few non-histone proteins, HAX-1, DUSP1, DUSP1 Mutant, HLA, β2 M and NAP1 for which correlation between protein accumulation and the different proposed parameters was not observed (see Supplementary Fig. S2). The proteins HAX-1, DUSP1, DUSP1 Mutant and NAP1 were tagged with 6xHis, GST and MBP and their expression was checked in BL21 (DE3) pLysS. Quantitative depiction of the three independent experiments is depicted in Fig. 4 and representative gel pictures can be seen in Supplementary Fig. S3. Indeed, the expression of HAX1 and DUSP1 with GST tag was achieved, however, the level was very low. For DUSP1 mutant no expression was observed in BL21 (DE3) pLysS.

Earlier, the result with N-terminal truncated GST tag upstream of histones suggested that the ratio of the size of the tag to that of the protein might be important. Thus, the expression of the non-histone proteins was attempted with a bigger tag, MBP. Comparison of lane 7 and lane 9 of Supplementary Fig. S3 a(i) clearly suggests that yield was significantly enhanced when MBP tag was used instead of GST. Notably, from the quantification plots it is apparent that once a tag of appropriate size has been incorporated then the expression in Rosetta (DE3) pLysS is higher than BL21 (DE3) pLysS [Fig. 4], probably, owing to the more robust translation resulting from increased tRNA supplies for the rare codons.

### *3.7. Solubility of proteins doesn't necessarily correlate with accumulation*

Solubility of proteins is proposed to be inversely correlated with accumulation due to increased accessibility to host proteases [13,20,23,27]. The solubility profile of histones expressed using pET28a vector in BL21 (DE3) pLysS strain at 37 °C is depicted in Fig. 5a. All the histones were mostly insoluble. Further, we tested the solubility of H2A.1 with different tags. The experiments were carried out at 18 °C, as lower temperature may lead to lower metabolic rate allowing better folding and solubility. The codon optimized Rosetta (DE3) pLysS strain was also used as codon optimisation have been shown to alter protein solubility [21]. We did find altered solubility for 6x His-tagged H2A.1 at 18 °C in Rosetta (DE3) pLysS strain. Major alteration in solubility profile was observed with GST and MBP tag [Fig. 5b]. Therefore, we checked the solubility of the non-histone proteins with GST and MBP tag at 37 °C and 18 °C in both the strains [Fig. 5c]. However, we didn't find any correlation between protein solubility and expression, as irrespective of solubility, we could achieve robust accumulation.

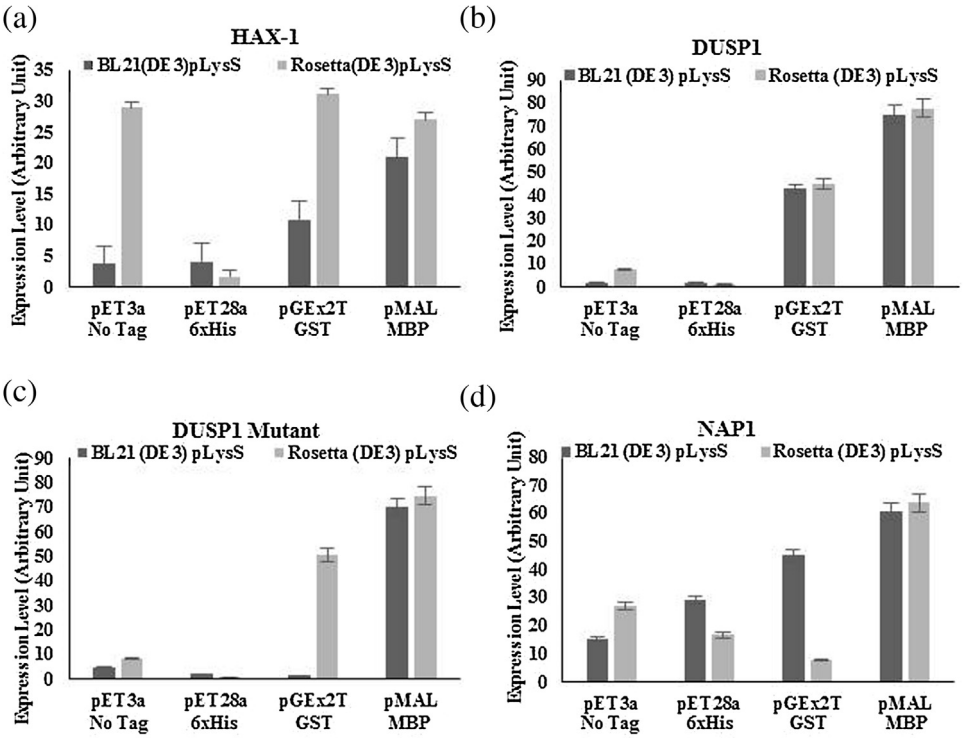



**Fig. 4.** Testing the hypothesis on Non-Histone proteins. Quantitative data of the constructs cloned in various vectors and expressed either in BL21 (DE3) pLysS or Rosetta (DE3) pLysS strain. (a) is for Hax1, (b) DUSP1, (c) DUSP1 mutant, (d) NAP1. See Supplementary Fig. S3 for representative gel pictures.

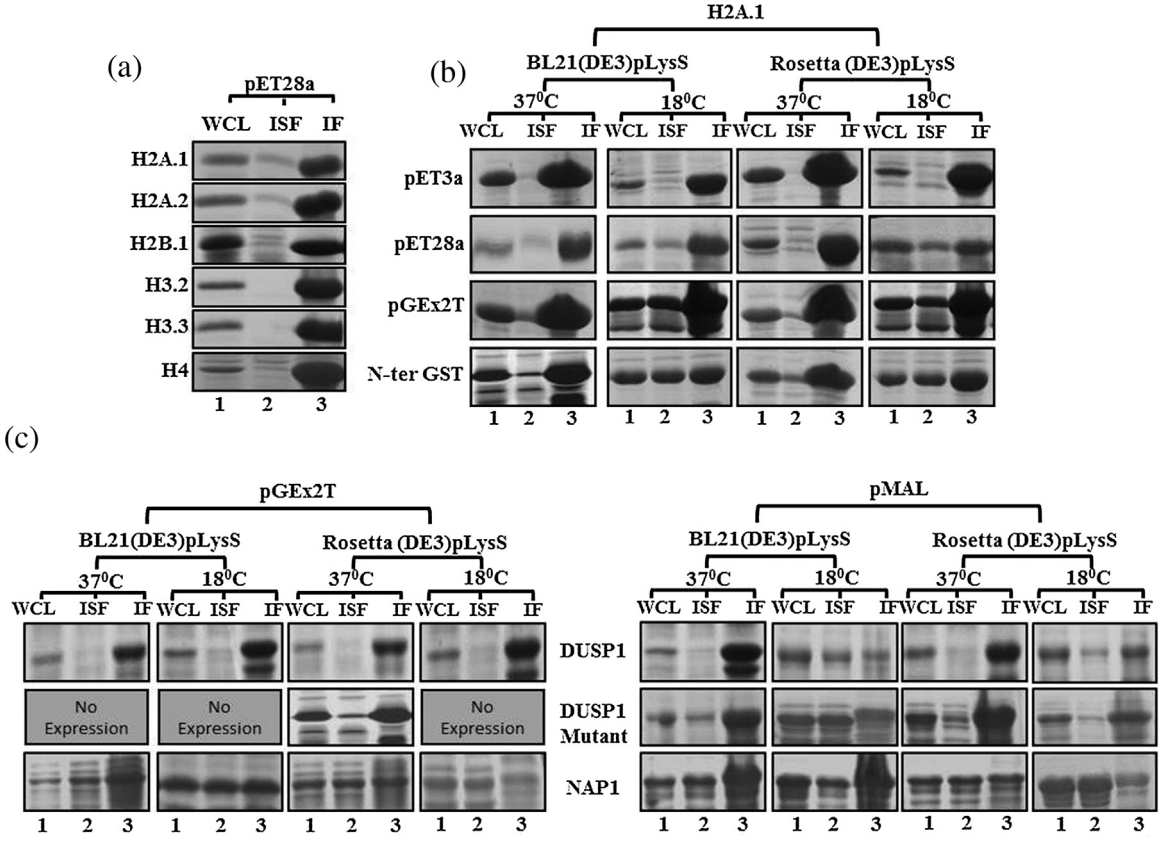



**Fig. 5.** Solubility analysis of various proteins. Various fractions, WCL—Whole cell lysate, ISF—Induced soluble fraction and IF—Insoluble fraction, post induction with IPTG were collected and separated on SDS gel. (a) Solubility profile of histones cloned in pET28a vector and expressed in BL21 (DE3) pLysS at 37 °C, induction by 0.2 mM IPTG for 2 h. (b) SDS gel with comparative analysis of H2A.1 when cloned and expressed in various vectors, in BL21 (DE3) pLysS and Rosetta (DE3) pLysS at 37 °C and 18 °C. (c) Comparative solubility profile of Non-Histone proteins DUSP1, DUSP1 mutant and NAP1 when expressed as GST and MBP fusions (cloned in pMAL) in both the bacterial strains at 37 °C and 18 °C.

## 4. Discussion

Heterologous expression is a very powerful tool for producing recombinant proteins. Many reports provide guidelines to optimise recombinant protein expression, however, not many studies address why many a times the proposed hypotheses fail to explain the lack of recombinant protein expression. Factors like rare codons need not be detrimental to protein accumulation and codon-optimisation, at times, may be unfavourable for protein expression. Possibly, rather than a barrier to the translation machinery, the various transcript parameters like rare codons or mRNA secondary structure (to which GC content also contributes), are modulators of translation kinetics. The overall translation speed can be modulated through tRNA recycling at the ribosome [5] and attenuation of *trp* operon presents a classic example of this [19].

In such scenario, the question arises that why sometimes there is complete lack of expression? Our immunoblot data provides interesting insights that helps us to understand the underlying reason. The intensity of bands did not show a unidirectional pattern suggesting translational variability or degradation or both during the course of induction of recombinant protein expression. In fact, translational variability may itself lead to degradation as translation and protein folding are coupled in bacteria. Presence of rare codons can affect the rate of translation and would influence the folding, thus local protein structure [24]. The local protein structure probably in turn may define the response of targeting and degradation by the host protease. Therefore, altering the "translation kinetics code" inherent in the sequence of the transcript may lead to complete lack of accumulation or may result in improved accumulation. This also emphasizes that the position of a rare codon should be more important than its absolute number. Subsequently, coding sequences having more rare-codons, even though would have slower translation kinetics, may lead to better accumulation if present at positions which promote "better" folding. This hypothesis is consistent with the proposition that rare codons may provide the genetic instruction for the regulation of rate of protein synthesis to allow proper folding of the nascent polypeptide [17].

The challenge remains to establish a predictable correlation between any of the proposed parameters and expression, which can help us to engineer means of circumventing expression variability. Unfortunately, it appears it is not going to be very straight forward. An earlier study showed that for the same transcript replacing the rare codons with the major ones or changing the GC content can independently bring about improved expression [26]. Clearly, not only are these parameters independently important, but correlations amongst secondary structure, the GC content and the codon usage of the transcript are significant [11]. As an alternative, we found that incorporation of a tag leads to much more consistent and predictable outcomes of protein expression and the size of the tag with respect to that of proteins is also important. The above hypothesis also gives us the reason as to why some proteins, which do not efficiently express without a tag, are sometimes robustly expressed as fusion with known tags like GST and MBP. Importantly, an N-terminus tag appears to be better suited. Probably, at the N-terminus of a fusion, the folding of GST/MBP precedes that of the downstream unit. However, when GST is present downstream of a protein, its folding might be influenced by its fusion partner leading to the attainment of a misfolded form.

## 5. Conclusions

To summarise, translational variability arising due to the combined effect of rare codon and GC content, mRNA secondary structure etc. leads to inefficient accumulation of heterologous proteins, possibly by inducing degradation of the protein. The expression variability may be averted by incorporating an N-terminal tag of suitable size. Finally, the steps depicted in Supplementary Fig. S4 should allow a researcher to troubleshoot expression problems in minimum time with maximum probability of success.

## Competing interests

No competing interests.

## Author's contributions

SB prepared constructs, performed expression analysis, purified proteins, and contributed to manuscript preparation. DR prepared constructs, performed expression analysis, carried out cell line maintenance and transfections. RR prepared constructs and performed expression analysis. AS prepared constructs. KB and SG contributed to experimental designing and manuscript preparation. All authors read and approved the final manuscript.

## Acknowledgments

The authors are grateful to Dr. Karolin Luger, Colorado State University for providing codon optimized construct of H4. We are also thankful to Dr. Francoise Carlotti, LUMC-Netherlands for providing DUSP-1 wild type and mutant constructs in mammalian expression vector from which they were subcloned. SB and DR are supported by CSIR India fellowship, AS by UGC India fellowship and RR by ACTREC fellowship.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.btre.2016.06.002.

## References

[1] M.F. Alexeyev, H.H. Winkler, Gene synthesis, bacterial expression and purification of the *Rickettsia prowazekii* ATP/ADP translocase, Biochim. Biophys. Acta Biomembr. 1419 (1999) 299–306, doi:http://dx.doi.org/10.1016/S0005-2736(99)00078-4.
[3] S.K. Behura, D.W. Severson, Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes, Biol. Rev. 88 (2013) 49–61, doi:http://dx.doi.org/10.1111/j.1469-185X.2012.00242.x.
[4] T.L. Calderone, R.D. Stevens, T.G. Oas, High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*, J. Mol. Biol. 262 (1996) 407–412, doi:http://dx.doi.org/10.1006/jmbi.1996.0524.
[5] G. Cannarozzi, G. Cannarrozzi, N.N. Schraudolph, M. Faty, P. von Rohr, M.T. Friberg, A.C. Roth, P. Gonnet, G. Gonnet, Y. Barral, A role for codon order in translation dynamics, Cell 141 (2010) 355–367, doi:http://dx.doi.org/10.1016/j.cell.2010.02.036.
[6] G.F. Chen, M. Inouye, Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes, Nucleic Acids Res. 18 (1990) 1465–1473, doi:http://dx.doi.org/10.1093/nar/18.6.1465.
[7] M.H. de Smit, J. van Duin, Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 7668–7672, doi:http://dx.doi.org/10.1073/pnas.87.19.7668.
[8] S.E. Gerchman, V. Graziano, V. Ramakrishnan, Expression of chicken linker histones in *E. coli*: sources of problems and methods for overcoming some of the difficulties, Protein Expr. Purif. 5 (1994) 242–251, doi:http://dx.doi.org/10.1006/prep.1994.1037.
[9] S. Gottesman, Proteases and their targets in *Escherichia coli*, Annu. Rev. Genet. 30 (1996) 465–506, doi:http://dx.doi.org/10.1146/annurev.genet.30.1.465.
[10] A.R. Gruber, R. Lorenz, S.H. Bernhart, R. Neubock, I.L. Hofacker, The Vienna RNA websuite, Nucleic Acids Res. 36 (2008) W70–W74, doi:http://dx.doi.org/10.1093/nar/gkn188.
[11] W. Gu, T. Zhou, C.O. Wilke, A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes, PLoS Comput. Biol. 6 (2010) e1000664, doi:http://dx.doi.org/10.1371/journal.pcbi.1000664.

[12] J.F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*, Curr. Opin. Biotechnol. 6 (1995) 494–500, doi:http://dx.doi.org/10.1016/0958-1669(95)80082-4.
[13] J.F. Kane, D.L. Hartley, Formation of recombinant protein inclusion bodies in *Escherichia coli*, Trends Biotechnol. 6 (1988) 95–101.
[14] T. Keller, D. Mis, K. Jia, C. Wilke, Reduced mRNA secondary-structure stability near the start codon indicates functional genes in prokaryotes, Genome Biol. Evol. 4 (2012) 80–88, doi:http://dx.doi.org/10.1093/gbe/evr129.
[15] M. Kubo, T. Imanaka, mRNA secondary structure in an open reading frame reduces translation efficiency in *Bacillus subtilis*, J. Bacteriol. 171 (1989) 4080–4082.
[16] K. Luger, T.J. Rechsteiner, T.J. Richmond, Expression and purification of recombinant histones and nucleosome reconstitution, in: P. Becker (Ed.), Methods in Molecular Biology (Clifton, N.J.), Methods in Molecular Biology™, Humana Press, 1999, pp. 1–16, doi:http://dx.doi.org/10.1385/1-59259-681-9:1.
[17] M. Marin, Folding at the rhythm of the rare codon beat, Biotechnol. J. 3 (2008) 1047–1057, doi:http://dx.doi.org/10.1002/biot.200800089.
[19] D.E. Morse, R.D. Mosteller, C. Yanofsky, Dynamics of synthesis translation, and degradation of trp operon messenger RNA in *E. coli*, Cold Spring Harb. Symp. Quant. Biol. 34 (1969) 725–740, doi:http://dx.doi.org/10.1101/SQB.1969.034.01.082.
[20] M. Murby, M. Uhlén, S. Ståhl, Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*, Protein Expr. Purif. 7 (1996) 129–136.
[21] G.L. Rosano, E. a Ceccarelli, Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain, Microb. Cell Fact. 8 (2009) 41, doi:http://dx.doi.org/10.1186/1475-2859-8-41.
[22] A.H. Rosenberg a, E. Goldman, J.J. Dunn, F.W. Studier, G. Zubay, Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system, J. Bacteriol. 175 (1993) 716–722.
[23] Rudolph, R., 1996. Successful protein folding on an industrial scale. Princ. Pract. protein Fold.
[24] R. Saunders, C.M. Deane, Synonymous codon usage influences the local protein structure observed, Nucleic Acids Res. 38 (2010) 6719–6728, doi:http://dx.doi.org/10.1093/nar/gkq495.
[25] P.M. Sharp, W.-H. Li, The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications, Nucleic Acids Res. 15 (1987) 1281–1295, doi:http://dx.doi.org/10.1093/nar/15.3.1281.
[26] G. Sinclair, F.Y.M. Choy, Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*, Protein Expr. Purif. 26 (2002) 96–105, doi:http://dx.doi.org/10.1016/S1046-5928(02)00526-0.
[27] P. Suthakaran, J. Balasubramanian, M. Ravichandran, V. Murugan, L.N. Ramya, K.K. Pulicherla, Studies on lipidification of streptokinase: a novel strategy to enhance the stability and activity, Am. J. Ther. 21 (2014) 343–351.
[28] Y. Tanaka, M. Tawaramoto-Sasanuma, S. Kawaguchi, T. Ohta, K. Yoda, H. Kurumizaka, S. Yokoyama, Expression and purification of recombinant human histones, Methods 33 (2004) 3–11, doi:http://dx.doi.org/10.1016/j.ymeth.2003.10.024.
[29] T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, Y. Pilpel, An evolutionarily conserved mechanism for controlling the efficiency of protein translation, Cell 141 (2010) 344–354, doi:http://dx.doi.org/10.1016/j.cell.2010.03.031.
[30] L. Vergani, F. Canneva, P. Ghisellini, C. Nicolini, Expression, purification, and structural characterization of human histone H4, Protein Expr. Purif. 24 (2002) 420–428, doi:http://dx.doi.org/10.1006/prep.2001.1601.
[31] M. Welch, A. Villalobos, C. Gustafsson, J. Minshull, You're one in a googol: optimizing genes for protein expression, J. R. Soc. Interface 6 (2009) S467–S476, doi:http://dx.doi.org/10.1098/rsif.2008.0520.focus.
[32] W. Zhang, W. Xiao, H. Wei, J. Zhang, Z. Tian, mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in *Escherichia coli*, Biochem. Biophys. Res. Commun. 349 (2006) 69–78, doi:http://dx.doi.org/10.1016/j.bbrc.2006.07.209.

**Advances in Bioscience and Biotechnology, 2016, 7, 133-141**
Published Online March 2016 in SciRes. http://www.scirp.org/journal/abb
http://dx.doi.org/10.4236/abb.2016.73012

# A Simple Method to Produce Sub-Nucleosome Complexes of High Purity *In Vitro*

**Saikat Bhattacharya, Sanjay Gupta***

Epigenetics and Chromatin Biology Group, Gupta Lab, Cancer Research Institute, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Kharghar, Navi Mumbai, India
Email: sbhattacharya@actrec.gov.in, *sgupta@actrec.gov.in

Received 25 January 2016; accepted 5 March 2016; published 8 March 2016



## Abstract

**With the identification of increasing number of chromatin modifiers, histone variants, histone post-translational modifications and their cross-talk, it is essential to validate these findings and interactions *in vitro* for which pure histone complexes are required. Although, the production of such complexes has been described earlier but still it remains a challenge for a non-specialist lab. Here we describe a protocol to quickly obtain large quantities of highly pure histones using bacterial expression system for GST pull-down and reconstitution experiments. In addition, we describe methods to quickly reconstitute and purify H2A/H2B dimers, H3/H4 tetramers and histone octamers for *in vitro* experiments. We demonstrate that these sub-complexes are properly folded and are hence, true representatives of the actual substrates *in vivo*. We also show that histones have a propensity to be non-specifically cleaved by proteases. Our results suggest that TEV protease is the most suitable protease while working with histones. The methodology described here should allow researchers to purify histone complexes in three days enabling functional and structural analyses of histone variants, mutants and post-translational modifications.**

## Keywords

**Histone, Nucleosome, H2A/H2B Dimer, Histidine-Tag, Recombinant**

## 1. Introduction

The eukaryotic genome is packaged inside the nucleus of only few microns in diameter [1] with the aid of highly

*Corresponding author.

**How to cite this paper:** Bhattacharya, S. and Gupta, S. (2016) A Simple Method to Produce Sub-Nucleosome Complexes of High Purity *In Vitro*. *Advances in Bioscience and Biotechnology*, **7**, 133-141. http://dx.doi.org/10.4236/abb.2016.73012

basic proteins known as histones. The core histones comprise of H2A, H2B, H3 and H4. H1 is the linker histone. H2A and H2B form heterodimers and two copies each of H3 and H4 form a tetramer. Two H2A/H2B dimer and one H3/H4 tetramer form the octamer core around which ~147 bp of DNA is wrapped to give rise to nucleosome, which is the fundamental repeating unit of the chromatin [2]. Apart from the well-established fact now that the canonical histones undergo a variety of functionally important post translational modifications (PTM's), sequence divergent forms of the histones also exist which are known as histone variants [3]. In addition, novel histone modifiers and binding partners like histone chaperones are being identified with redundant or supposedly non-redundant functions.

With the identification of increasing number of chromatin modifiers, histone variants, histone post-translational modifications and their cross-talk, it is essential to validate these findings and interactions *in vitro* for which pure histone complexes are required. Large quantity of histones can be purified by isolation from cells [4], however, they are highly heterogeneous in terms of variants and PTM's [5] [6]. This does not allow the investigation of, for example, the possible contribution of a specific variant on structural organization of chromatin or the effect of a particular modification in recruiting chromatin modifiers or affecting nearby modifications. In such scenario use of recombinant histones becomes the method of choice to study the effect of histone variants or modified histones on the structure and stability of chromatin. Further, recombinant histones can be used to reconstitute nucleosome sub-complexes, H2A/H2B dimer, H3/H4 tetramers and H2A-H2B-H3-H4 octamers, which would fare as better representatives of *in vivo* substrates in *in vitro* assays.

Histones when expressed in bacteria form inclusion bodies and hence, their purification first requires solubilizing them with denaturing agents like urea or guanidinium chloride followed by ion exchange chromatography and gel filtration [7] [8]. Requirement of using multiple chromatography steps makes it technically and practically challenging for a non-specialist lab to employ these methods to purify recombinant histones. In addition, although, the production of nucleosome complexes with purified histones has been described earlier [9] it requires considerable expertise and resources to successfully execute the protocol. Further, a major challenge faced for the use of full length histone proteins for carrying out *in vitro* assays and reconstitution is their aggregation in commonly used buffers.

In the present study, we have evaluated various methods to purify histone proteins. We found that GST solubilizes histones enabling affinity purification and use in GST pull down assays. In addition, we came up with a simple 6xHis tag based purification system by modifying and combining existing protocols that enables quick reconstitution of H2A/H2B dimers, H3/H4 tetramers and histone octamers for *in vitro* experiments. Further, we biophysically characterize these complexes and demonstrate that these are properly folded.

## 2. Materials and Methods

### 2.1. Protein Expression and Fraction Preparation

Histone constructs in the desired vector were transformed in BL21 (DE3) pLysS. After overnight incubation, a single colony was inoculated from the plates of transformed BL21 (DE3) pLysS in 5 mL or 20 mL LB media and incubated at 37˚C until the OD600 reached between 0.3 and 0.6. Induction was carried out with 0.2 mM IPTG. The cultures were induced for 3 hrs at 37˚C or overnight at 18˚C. Post induction the bacteria were harvested and processed. The soluble and insoluble fractions of proteins were separated by re-suspending the cells in buffer containing 50 mM Tris-Cl pH 8.0, 0.5% Triton X-100 and 100 µg/ml lysozyme followed by three rounds of sonication, each for 30 seconds at 30% amplitude. The lysate was then centrifuged at 27000 g for 30 mins at 4˚C. The supernatant and pellet thus obtained contains the soluble proteins and the insoluble proteins respectively. The proteins were resolved by 18% SDS-PAGE followed by Coomassie staining (Brilliant Blue R250).

### 2.2. GST Affinity Purification

The fraction consisting of the soluble proteins was incubated with the equilibrated Glutathione conjugated beads for 2 hours at 4˚C. After incubation, the beads were washed with buffer containing 20 mM Tris-Cl pH 8.0, 200 mM NaCl, 1 mM EDTA pH 8.0, 0.5% Nonidet P-40, 2 µg/ml aprotinin, 1 µg/ml leupeptin, 0.7 µg/ml pepstatin and 25 µg/ml PMSF. The protein was eluted with the elution buffer containing 50 mM Tris-Cl pH 8.0 and 10 mM reduced glutathione.

### 2.3. Ni-NTA Affinity Purification

The insoluble pellets were dissolved in 50 mM Tris–HCl buffer (pH 8.0) containing 500 mM NaCl, 5% glycerol, and varying concentration of urea (see figure for more details), and were centrifuged at 27,000 g for 20 min at 4˚C. The insoluble pellets were discarded. The supernatants containing the 6xHis-tagged histones were mixed with 50% slurry of nickel–nitrilotriacetic acid (Ni–NTA) agarose beads (Qiagen), and the samples were rotated at 4˚C. After a 60 min rotation, the beads were washed with 50 mM Tris–HCl buffer (pH 8.0) containing 500 mM NaCl, 5% glycerol, 6 M urea, and 10 mM imidazole. The Ni–NTA agarose beads bound to the 6xHis-tagged proteins were washed further with the same buffer. The 6xHis-tagged histones were eluted by a linear gradient of imidazole from 5 to 300 mM. The eluted fractions were analysed by loading on 18% SDS-PAGE.

### 2.4. Reconstitution of H2A/H2B Dimer and H3/H4 Tetramer

To reconstitute the H2A/H2B dimer, the His6-tagged histones H2A and H2B or H3 and H4 in 10 mM HCl were mixed in a 1:1 stoichiometry. Then, the mixture was rapidly diluted three folds by adding 50 mM Tris–HCl buffer (pH 8.0) containing 10 mM DTT and 2 mM EDTA with continuous stirring. The diluted mixture is then dialyzed against 1 liter of 50 mM Tris–HCl buffer (pH 8.0) containing 10 mM DTT, 2 mM EDTA and 2 M NaCl for 1 h at room temperature, and the dialysis was continued overnight at 4˚C. After an overnight dialysis, the impurities and excess histones were precipitated. The dialysate was then concentrated and loaded onto equilibrated gel filtration column and the peak fraction was collected. The salt concentration can be reduced by stepwise dialysis.

### 2.5. Reconstitution of Histone Octamer

The reconstituted H2A/H2B dimer and H3/H4 tetramer were mixed in 1:1 stoichiometry and dialysed overnight against 2 liter of 50 mM Tris–HCl buffer (pH 8.0) containing 10 mM DTT, 2 mM EDTA and 2 M NaCl at 4˚C. The dialysate was then loaded onto equilibrated gel filtration column and the peak fraction was collected.

### 2.6. Mass Spectrometric Analysis

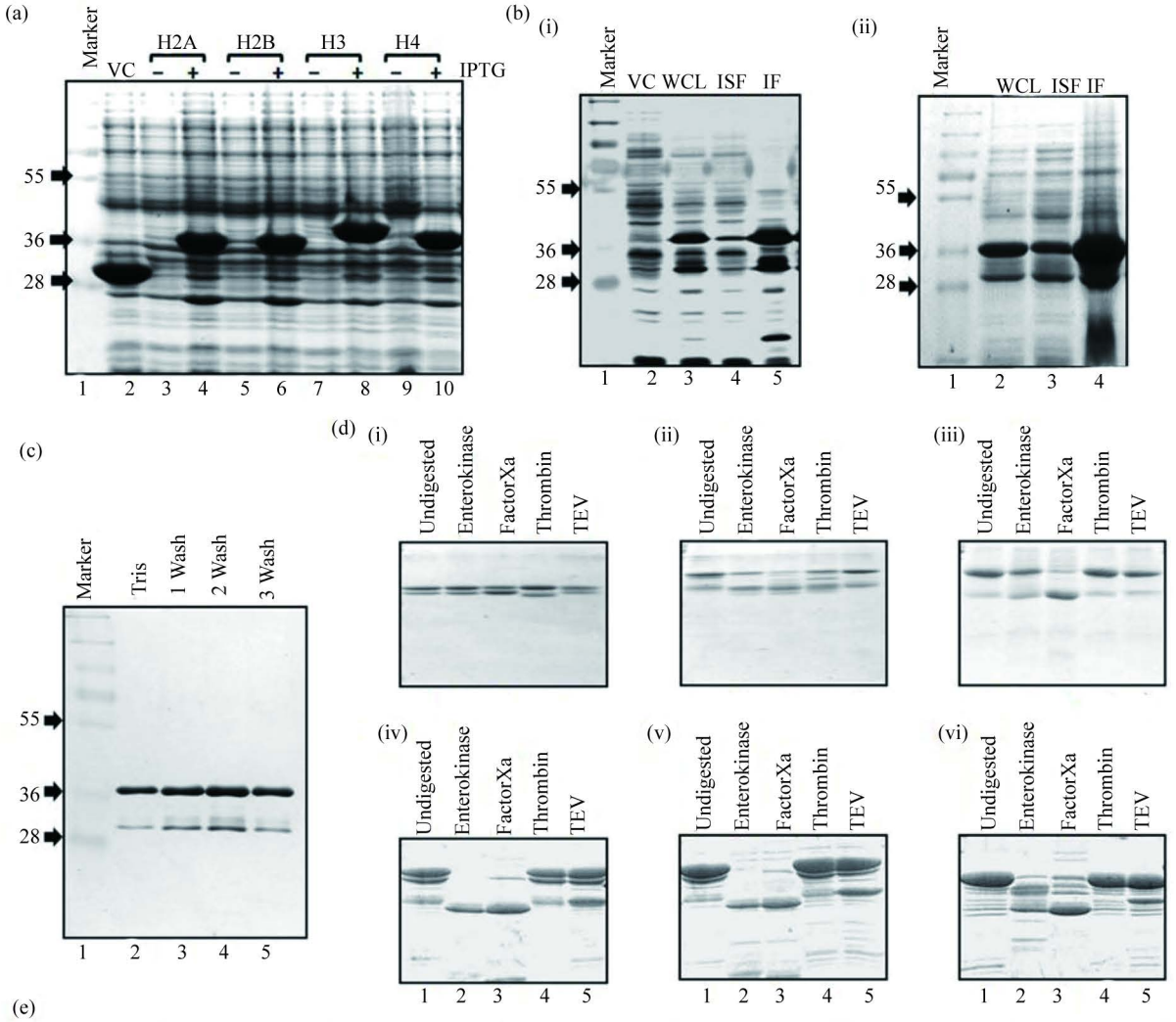
Protein spots of interest were manually excised and cut into fine pieces. These were destained, chemically reduced, alkylated and subjected to in gel digestion using 20 ng/µl Trypsin (Sigma) at 37˚C for 16 - 18 hrs. Peptides thus obtained were extracted from gel pieces, vacuum dried and reconstituted in 10 µl of 0.1% TFA. Samples were then mixed with matrix solution ($\alpha$-Cyano-4-hydroxycinnamic acid), dried and mass spectra were acquired on a Bruker Daltonics Ultraflex II in reflector positive ion mode. Database searching for protein masses was carried out using Mascot search engine (version 2.2.03) by comparing peptide masses with those in NCBInr protein database.

## 3. Results

### 3.1. Purification of Recombinant Histones Using Affinity Purification

Histones are highly insoluble when expressed in bacteria. One way of circumventing this issue is by co-expressing histones, for example H2A and H2B together in tandem, in bacteria which at times makes them soluble [10]. However, this method does not ensure that all histones expressed will be soluble [10]. As an alternative approach, we tried to use the commonly used GST tag as it often solubilizes protein and also enables use of affinity purification. Affinity purification may yield ~90% pure proteins after only one round of purification and is technically less challenging than ion-exchange chromatography.

We found the histones expressed robustly with the GST tag (**Figure 1(a)**), however, the expressed proteins were largely insoluble when growth temperature post-induction of recombinant protein expression was maintained at 37˚C. The data of only H2A is shown but similar results were obtained for all the core histones (**Figure 1(b)(i)**). On attempting expression at 18˚C, approximately 10% of the protein, as judged by densitometric analysis, was recovered in the soluble fraction (**Figure 1(b)(ii)**). Proteins were purified by affinity purification. The GST-Histone fusion proteins can be used for carrying out GST-pull down assays for investigating direct physical interactions of putative binding partners of histones.




(e)

| | GST | | H2A | | H2B | | H3.2 | | H3.3 | | H4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E | O | E | O | E | O | E | O | E | O | E | O |
| Enterokinase | X | X | X | √ | X | √ | X | √ | X | √ | X | √ |
| Factor Xa | X | X | X | √ | X | √ | X | √ | X | √ | X | √ |
| Thrombin | X | X | √ | √ | X | X | X | X | X | X | √ | - |
| TEV | X | X | X | X | X | X | X | X | X | X | X | X |

**Figure 1.** Expression and purification of GST tagged Histones; (a) Coomassie stained SDS gel loaded with whole cell lysates of BL21 (DE3) pLysS expressing GST Histones. (b)(i) Solubility analysis of GST-Histones expressed at 37˚C. WCL-Whole cell lysate, ISF-Induced soluble fraction, IF-Insoluble fraction (ii) Solubility analysis of GST-Histones expressed at 18˚C overnight. (c) SDS PAGE showing purified GST Histones. (d)(i) Proteolysis of GST tag with all the used enzymes. (ii)-(vi) Enzymatic action of proteases on GST tagged H2A, H2B, H3 and H4 histones. (e) Table summarised to show the protease activity of all the enzymes used on all the purified GST-Histones.

## 3.2. Removal of GST Tag from Recombinant Proteins

GST being of 26 kDa size and a homodimer itself is expected to interfere with the hetero-dimerisation of histones. This makes cleavage of the GST tag essential for carrying out H2A/H2B dimerisation or H3/H4 tetramerisation. We tested the efficacy of some commonly used proteases for histone purification. Incubation of only

GST protein did not lead to any cleaved products (**Figure 1(d)(i)**) as expected. However, to our surprise, we found proteases like FactorXa and Enterokinase, non-specifically cleave the histones as can be seen with the disappearance of the band corresponding to GST fusion and appearance of band corresponding to the GST protein (**Figures 1(d)(ii)-(vi)**). By mass spectrometry we confirmed that the protein we purified were indeed GST-Histone fusion. On screening for different proteases, TEV protease was found not to cleave histones non-specifically. The expected and observed pattern of protease activity is tabulated in **Figure 1(e)**. Although, thrombin is expected to cleave H4 protein but the cleavage site is just after four amino acids. This makes the band shift difficult to be perceived on SDS-PAGE (**Figure 1(d)(vi)**). Post GST removal the monomeric histones were found to precipitate that is commonly observed with full length histones. This limits their use in downstream reconstitution experiments. However, for purification of smaller peptides of histones, which might be soluble, our results suggest that TEV protease should be used.

### 3.3. 6xHis Tag Is Better Suited for Affinity Purification of Histones

We opted to use N-terminal 6xHis tag for histone purification keeping in consideration that a small tag like 6xHis is not expected to interfere in the oligomerization of histones (N-terminal tails of histones protrude out of the nucleosomes). Unlike the case with GST tag, this would enable reconstitution without the need of removal of affinity tag.

We used pET28a vector to express histones in BL21 (DE3) pLysS. The recombinant protein expression was robust (**Figure 2(a)**). On analyzing different fractions, the histones were found to be completely part of insoluble fraction at both strategies were employed. Urea was the preferred denaturant to solubilize proteins from the insoluble fraction as GdmHCl doesn't allow solubilized 6xHis proteins to bind to Ni-NTA column. Upon further titration with different urea concentrations, 6 M urea was found as the optimal denaturant concentration to solubilize these proteins (**Figure 2(c)**). The solubilized proteins robustly bound to Ni-NTA beads (**Figure 2(d)**). Elution with buffers containing increasing concentration of imidazole suggested that histones could be very efficiently eluted with 150 mM imidazole concentration with almost no non-specific proteins (**Figure 2(e)**). Thus,

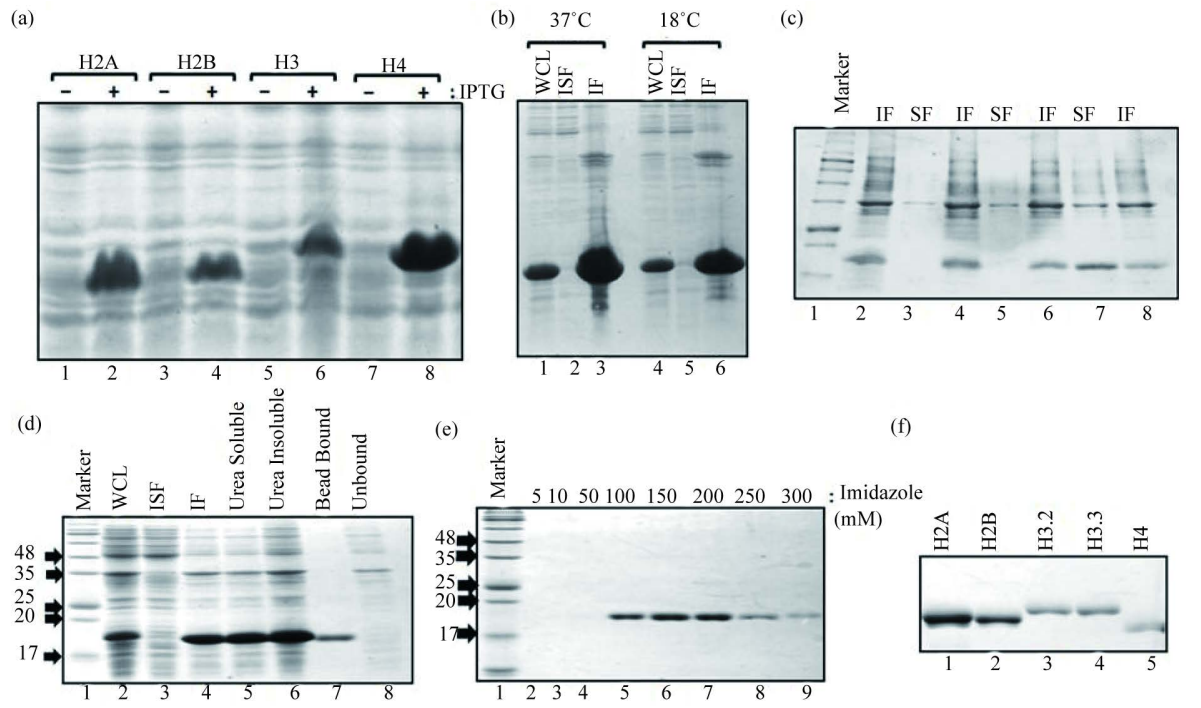



**Figure 2.** Expression and purification of 6X His tag histones; (a) 18% SDS PAGE showing the expression of 6xHis-Histones. (b) Solubility analysis of His tag Histones at 37˚C and 18˚C. (c) Solubilisation of histones with different urea concentrations. Lanes 2, 4, 6, 8 correspond to insoluble fractions and lanes 3, 5, 7 to fractions solubilized with 2 M, 4 M and 6 M urea. (d) Complete profile of purification strategy for 6x His tagged histones. (e) Elution profile of tagged histones at different concentrations of imidazole. (f) SDS PAGE showing all the purified histones.

all the core histones could be very efficiently and conveniently purified with this one step purification process.

### 3.4. Histone Monomers Are Not Properly Folded in Solution

To remove denaturant from the purified proteins, the eluted fraction was overnight dialyzed against 50 mM phosphate buffer. However, often it led to the precipitation of majority of histones in a concentration dependent manner. Histones being highly basic proteins dialyzing them against an acidic solution may enhance solubility. Hence, histones were dialyzed against 10 mM HCl to remove urea. By this method the precipitation could be circumvented irrespective of the concentration of histones. For carrying out subsequent analysis the concentrated histones could be diluted in phosphate or tris buffer without precipitation. Afterwards, we wanted to see whether the histone monomers attain their native confirmation on removal of the denaturant. Circular dichroism spectra showed a dip at around 200 nm for all the histone monomers suggesting that the monomers were not folded properly and were largely in random coil (**Figure 3(a)**). This might be attributed to the fact that the alpha 2 helix of histones is largely hydrophobic in nature.

### 3.5. Histone Oligomers Can Be Conveniently Reconstituted from Affinity Purified Histones

Next, we wanted to reconstitute H2A/H2B dimers and H3/H4 tetramers as 1) they would be better substrates for carrying out *in vitro* assays, 2) would be properly folded owing to the burial of hydrophobic residues and 3) would be easier to use for octamerisation (discussed later).

In this regard, we used the protocol of rapid dilution of histones for forming oligomers which was previously described [5]. We found that even H3/H4 tetramers could be *in vitro* reconstituted directly from 10 mM HCl (**Figure 3(c)**) without urea denaturation by this method which was not previously reported. Post rapid dilution the protocol involves binding of the mix to heparin column and subsequent elution [5]. We found dialyzing the rapid dilution mix against 2 M NaCl leads to precipitation of excess histone monomers leaving the properly folded dimers in soluble form thus eliminating the requirement of Heparin column. In ideal scenario, with usage of equimolar histones, there should be no or minimum precipitate post dialysis. The precipitate (if formed) could be removed by centrifuging the sample at 10,000 g for 20 mins. The dimerization could be confirmed by loading the soluble fraction onto a gel filtration column. The dimers thus obtained were purified by size-exclusion chromatography using HiLoad 16/60 Superdex-200 gel filtration column (GE). The dimer eluted at a volume of 82 - 84 mL (**Figure 3(b)(i)**) consistent with previous reports [2]. The molecular weight of H2A and H2B being very similar (H2A-14119.5Da, H2B-13906.1Da) (**Figure 3(b)(ii)**), the peak fractions from gel filtration were resolved onto AUT-PAGE, which separates proteins on the basis of hydrophobicity, to confirm the formation of heterodimers (**Figure 3(b)(iii)**).

### 3.6. H2A/H2B Dimers and H3/H4 Tetramers Are Better Substrates for Octamer Reconstitution than Histone Monomers

Once we were able to obtain properly folded H2A/H2B dimers and H3/H4 tetramers, we next wanted to test whether these complexes can be used to reconstitute histone octamers. Equimolar ratio of H2A/H2B dimer and H3/H4 tetramer were mixed and were dialysed against 2 M NaCl. As the H2A/H2B and H3/H4 complexes are independently stable under high NaCl concentrations, no precipitation was observed. This is unlike the use of histone monomers where errors in quantification can lead to large amount of protein precipitation (discussed later). Post dialysis the histone octameric core particle was formed which could be confirmed and purified by loading onto size exclusion chromatography column with elution at an expected volume ~65 ml (**Figure 3(d)**). The excess H2A/H2B dimer eluted at 84 mL.

### 3.7. Histone Oligomers Are Properly Folded

Post reconstitution and purification, secondary structure validation of the H2A/H2B dimer and H3/H4 tetramer were carried out by far-UV CD. The spectra dip at 222 and 208 nm, characteristic of helices confirmed that the histone oligomers, which has a high proportion (40%) of helical structure (**Figure 4(a)**), is properly folded (**Figure 4(b)(i)**) for H2A/H2B dimer, (c)(i) for H3/H4 tetramer]. The greater dip at 208 nm compared to 222 nm highlights the greater proportion (~60%) of randomly coiled structure of the dimer and tetramer. The tertiary structure of the H2A/H2B dimer and H3/H4 tetramer was validated by near UV CD and fluorescence spectros-

copy. H2A has three tyrosine residues at position 39, 50 and 57 and H2B have five tyrosine residues at position 37, 40, 42, 83 and 121. Histones do not contain tryptophan, hence, the fluorescence of histone proteins is majorly attributed to the tyrosine. The bell shaped curve with peak near 278 nm in the near-UV CD spectra is in agreement with signature tyrosine peak between 275 and 282 nm (**Figure 4(b)(iii)**). The dimers exhibited a fluorescence emission maximum at 305 nm (**Figure 4(b)(ii)**) which is expected for tyrosine. Similar pattern was observed for the H3-H4 tetramers (**Figure 4(c)(ii)**).

## 4. Discussion

We have described a very convenient and robust method to express, purify and reconstitute sub-nucleosomal complexes of histones for carrying out *in vitro* studies. The proteins were purified by one step affinity chromatography. GST tag could be used to solubilise and purify histones, however, due to the precipitation of monomeric histones post GST tag removal the strategy was not useful especially considering that only 10% of the

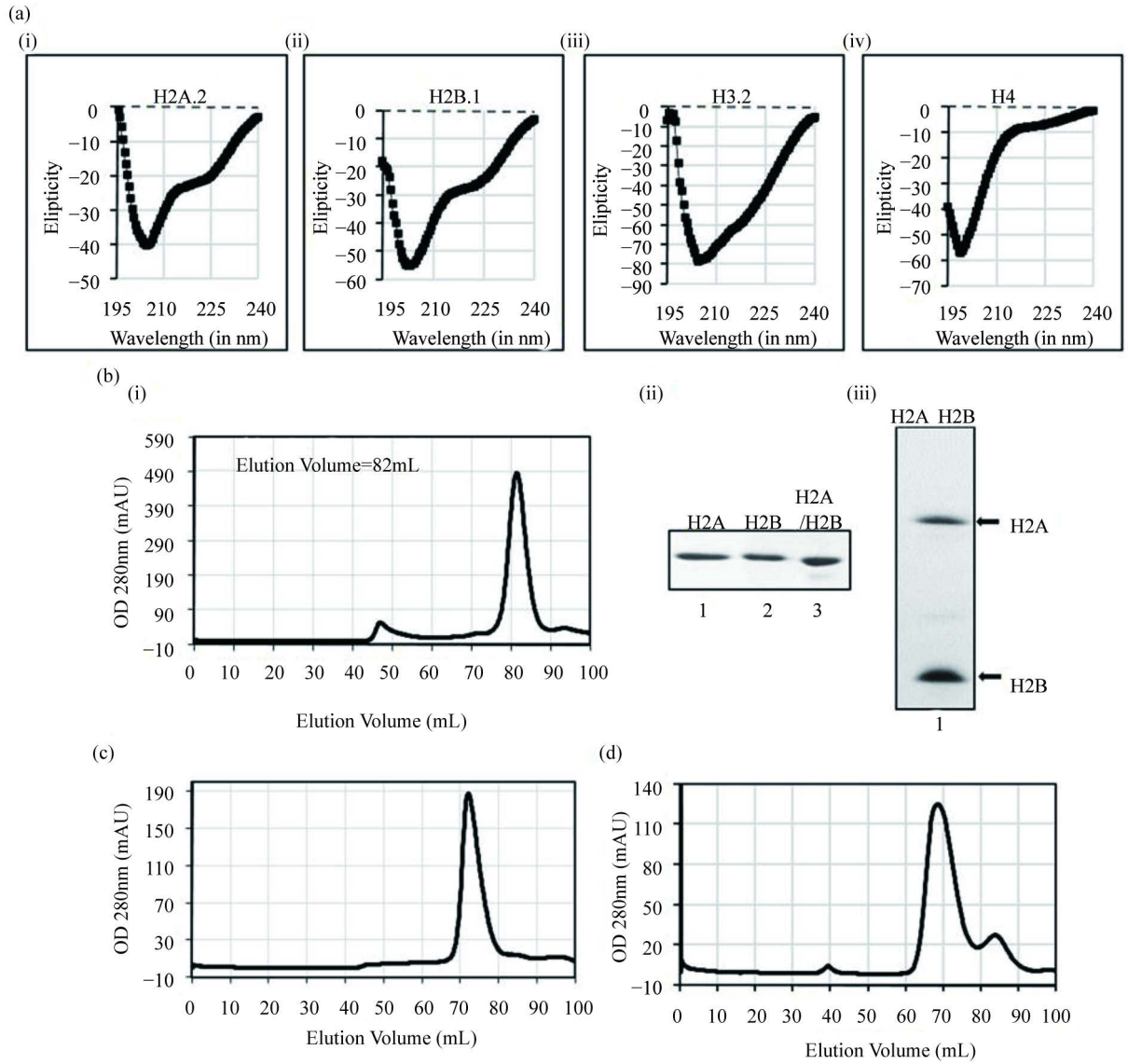



**Figure 3.** Reconstitution of histone dimer, tetramer and octamer: (a)(i)-(iv) Far-UV CD spectra of all the purified histone monomers showing their unfolded state. (b)(i) Graph corresponding to elution profile of H2A/H2B dimer, eluting at 82 ml. (ii) SDS PAGE of the above collected dimer. (iii) AUT PAGE of H2A/H2B dimer, showing distinctly resolved H2A and H2B. (c) Elution profile of reconstituted H3/H4 tetramer eluting at 72 ml. (d) Histone octamer elution peak at 65 ml.

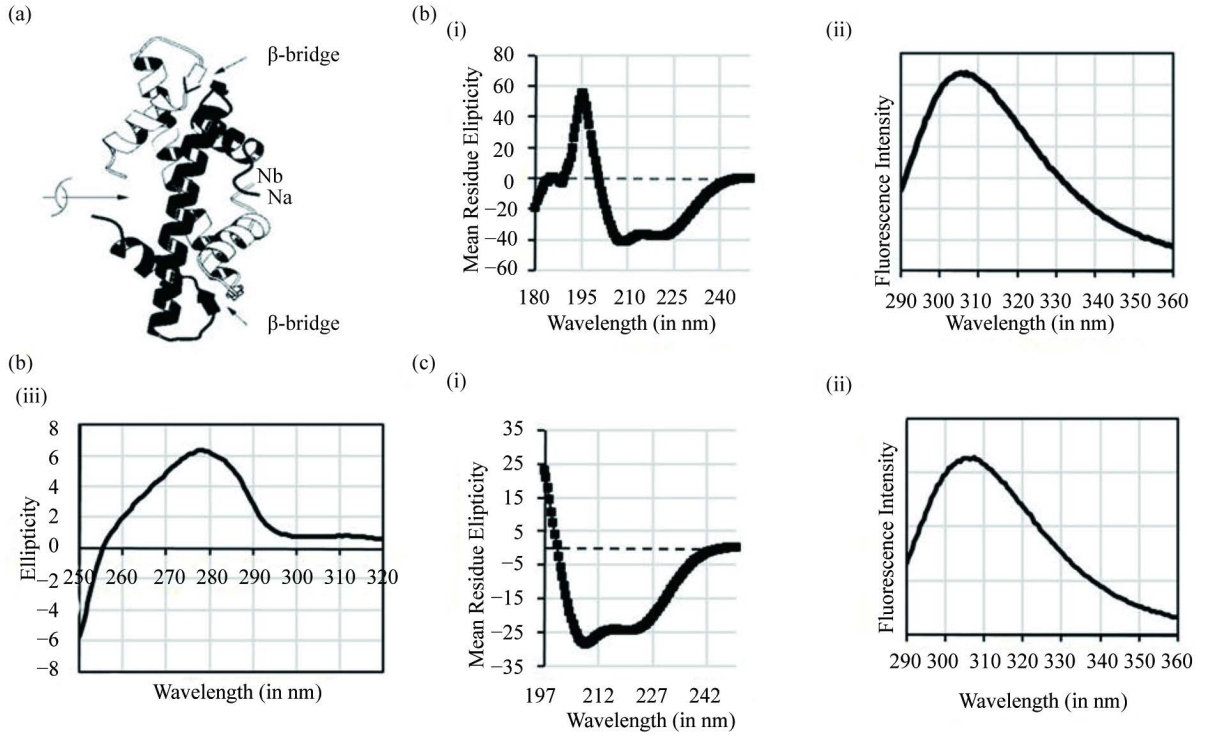



**Figure 4.** Secondary structure analysis of histone oligomers: a) Pictorial representation of structure of a histone H2A/H2B dimer. (b)(i) Far UV CD spectra of reconstituted dimer. (ii) Fluorescence spectral scan of reconstituted dimer. (iii) Near UV CD spectra for tertiary structure determination of dimers. (c)(i) Far UV CD spectra of reconstituted H3/H4 tetramer. (ii) Fluorescence spectral scan of tetramer.

protein is solubilised on GST tag incorporation. Nonetheless, the strategy can be useful to carry out *in vitro* GST pull down assays. Another striking thing which emerged from our data is the degradation of histones by proteases FactorXa and Enterokinase, which have propensity to non-specifically cleave at basic residues. Hence, care is to be taken while choosing the protease to remove tags from fusion proteins consisting of histones. TEV protease was found to be the best suited for this purpose. For obtaining recombinant histones in bulk amount, Ni-NTA affinity purification under denaturing conditions was very convenient and robust technique. In addition, 6xHis tag being small and as N-terminus of histones are not involved in formation of the complexes, the His-tagged proteins could be directly used in reconstitution experiments without removal of tag. Further, we documented that the reconstituted complexes are properly folded which makes them useful for carrying out *in vitro* studies pertaining to histone variants/mutants and modifications. During the formation of octamer equimolar ratio of the core histones assemble. Accurate quantification of histone monomers is crucial to prevent precipitation of proteins and successful reconstitution. However, histones lack tryptophan and have very less tyrosine and phenylalanine. This makes their accurate quantification difficult using colorimetric approach. In this regard use of reconstituted H2A/H2B dimers and H3/H4 tetramers for histone octamer reconstitution is more convenient rather than adding all the four histone monomers stoichiometrically. In our experience using the reconstituted dimers and tetramers stoichiometrically for octamerisation gives more consistent results than using the individual monomers. The histone octamer elutes at 65 mL. The H3/H4 tetramer elutes at 72 mL, whereas the H2A/H2B dimer elutes at 84 mL. Hence, it is advised to use a little excess of H2A/H2B dimer during octamer reconstitution as they are easier to distinguish from the octamer peak during gel filtration as compared to H3/H4 tetramer peak.

## 5. Conclusion

We have described a protocol in which histones can be purified and its oligomeric structures reconstituted in very few steps in three days. Histones can be purified in bulk by affinity purification and then via use of rapid dilution method followed by dialysis, H2A/H2B dimers and H3/H4 tetramers can easily be reconstituted. For the

octamerisation, the reconstituted and purified dimers and tetramers are mixed proportionately followed by dialysis. These reconstituted and purified octamers which have their native structure are better substrates for use in *in vitro* assays. We also showed that adding GST tag partially solubilises the otherwise insoluble histones which can be used for *in vitro* assays. Further, histones have a propensity to be non-specifically cleaved by proteases and TEV protease should be opted for while working with histones.

## Acknowledgements

The authors are grateful to all members of Gupta Lab, ACTREC for valuable discussion. S. B is supported by CSIR, India fellowship.

## Conflict of Interest

The authors hereby declare no conflicts of interest.

## References

[1] Rocha, W. and Verreault, A. (2008) Clothing up DNA for All Seasons: Histone Chaperones and Nucleosome Assembly Pathways. *FEBS Letters*, **582**, 1938-1949. http://dx.doi.org/10.1016/j.febslet.2008.03.006

[2] Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal Structure of the Nucleosome Core Particle at 2.8A Resolution. *Nature*, **389**, 251-260. http://dx.doi.org/10.1038/38444

[3] Kamakaka, R.T. and Biggins, S. (2005) Histone Variants: Deviants? *Genes & Development*, **19**, 295-316. http://dx.doi.org/10.1101/gad.1272805

[4] Simon, R.H. and Felsenfeld, G. (1979) A New Procedure for Purifying Histone Pairs H2A+H2B and H3+H4 from Chromatin Using Hydroxylapatite. *Nucleic Acids Research*, **6**, 689-696. http://dx.doi.org/10.1093/nar/6.2.689

[5] Gloss, L.M. and Placek, B.J. (2002) The Effect of Salts on the Stability of the H2A−H2B Histone Dimer. *Biochemistry*, **41**, 14951-14959. http://dx.doi.org/10.1021/bi026282s

[6] Ridsdale, J.A. and Davie, J.R. (1987) Chicken Erythrocyte Polynucleosomes Which Are Soluble at Physiological Ionic Strength and Contain Linker Histones Are Highly Enriched hi *β*-Globin Gene Sequences. *Nucleic Acids Research*, **15**, 1081-1096. http://dx.doi.org/10.1093/nar/15.3.1081

[7] Suto, R.K., Clarkson, M.J., Tremethick, D.J. and Luger, K. (2000) Crystal Structure of a Nucleosome Core Particle Containing the Variant Histone H2A.Z. *Nature Structural & Molecular Biology*, **7**, 1121-1124. http://dx.doi.org/10.1038/81971

[8] Shim, Y., Duan, M.-R., Chen, X., Smerdon, M.J. and Min, J.-H. (2012) Polycistronic Coexpression and Nondenaturing Purification of Histone Octamers. *Analytical Biochemistry*, **427**, 190-192. http://dx.doi.org/10.1016/j.ab.2012.05.006

[9] Luger, K., Rechsteiner, T.J., Flaus, A.J., Waye, M.M.Y. and Richmond, T.J. (1997) Characterization of Nucleosome Core Particles Containing Histone Proteins Made in Bacteria1. *Journal of Molecular Biology*, **272**, 301-311. http://dx.doi.org/10.1006/jmbi.1997.1235

[10] White, C.L., Suto, R.K. and Luger, K. (2001) Structure of the Yeast Nucleosome Core Particle Reveals Fundamental Changes in Internucleosome Interactions. *The EMBO Journal*, **20**, 5207-5218. http://dx.doi.org/10.1093/emboj/20.18.5207

Exp Biol Med (Maywood) OnlineFirst, published on May 6, 2016 as doi:10.1177/1535370216649262

*Brief Communication*

# Histone H2A mono-ubiquitination and cellular transformation are inversely related in N-nitrosodiethylamine-induced hepatocellular carcinoma

**Saikat Bhattacharya[1], Divya Reddy[1], Arvind Ingle[2], Bharat Khade[1] and Sanjay Gupta[1]**

[1]Epigenetics and Chromatin Biology Group, Gupta Lab, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Cancer Research Institute, Kharghar, Navi Mumbai, MH 410210, India; [2]Laboratory Animal Facility, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Cancer Research Institute, Kharghar, Navi Mumbai, MH 410210, India

Corresponding author: Sanjay Gupta. Email: sgupta@actrec.gov.in

## Abstract

Aberrant changes in histone post-translational modifications are encountered frequently in diseases like cancer. Although histone H3 post-translational modifications have been extensively studied in context of diseases, the functionally important histone H2A PTM H2A119ub (H2Aub) has not gained much attention. In this study, we report that H2Aub markedly decreases in hepatocellular carcinoma. Usp21, a H2A deubiquitinase, is probably responsible for decrease in H2Aub. In addition, the H2Aub levels showed an inverse correlation with H3S10 phosphorylation (H3S10p) and the proliferative state of the cells. Downregulation of H2Aub is also associated with increased expression of growth factor gene lipocalin 2. Interestingly, we show that treatment of cells with histone deacetylase inhibitor trichostatin A results in increase of H2Aub and decrease in H3S10p. Our work for the first time suggests the in vivo association of H3S10p, H4ac, and H2A119ub with cellular transformation.

**Keywords:** Histone post-translational modifications, deubiquitination, lipocalin 2, histone deacetylase inhibitor, cancer, epigenetics

***Experimental Biology and Medicine*** **2016; 0: 1–6. DOI: 10.1177/1535370216649262**

## Introduction

Histones, which organize chromatin, often undergo a variety of post-translational modifications (PTMs).[1] These modifications function either by disrupting chromatin contacts or by affecting the recruitment of various proteins to the chromatin and thereby, regulating transcription.[1] Histone H2A mono-ubiquitination at lysine 119 (H2Aub) is one such modification identified for polycomb-targeted gene silencing; H2Aub by PRC1 type complexes creates a binding site for Jarid2–Aebp2–containing PRC2 and promotes H3K27 trimethylation on H2Aub containing nucleosomes.[2] H2Aub is thus an integral component of a positive feedback loop establishing H3K27me3 mark and maintaining repressive state of the chromatin.[3,4]

Histone H2A was the first protein identified to be ubiquitinated and is one of the most abundant ubiquitinated protein in the nucleus.[4] Despite its abundance, apart from the study in which mono-ubiquitinated histone H2A was markedly down-regulated in prostate cancer,[5] this functionally important mark has not been investigated in other cancers. In addition, the drop in H2Aub is a pre-requisite for increase in H3S10p during S to G2/M transition in vitro.[6] Interestingly, H3S10p itself is a mark that is indispensable for cellular transformation.[7] Whether an antagonistic effect of these modifications on each other may exist in vivo has not been reported.

Using a hepatocellular carcinoma (HCC) model of rat, we show that the level of H2Aub and H4ac markedly decreases in tumor and is inversely correlated with gain in H3S10p. The enzyme Usp21 probably brings about this loss in ubiquitination. Upon treatment with HDAC inhibitor, the scenario is reversed with increase in H2Aub, H4ac, and loss of H3S10p, again pointing towards a possible correlation amongst these marks. Our results indicate that the drop in H2Aub might be functionally important in cancer progression and emphasizes the need of studying this modification and the cross talk with other modifications in more number of cancers.

## Materials and methods

### Animal experimentation

All the experiments were performed on male Sprague-Dawley rats (spp. *Rattus norvegicus*) after approval from

the Institutional Animal Ethics Committee, ACTREC. Protocol used to induce HCC is as previously described.[8] For liver tissue transplantation, small piece of ~3 $mm^2$ size liver and tumor tissue collected from the donor mice were washed in RPMI medium. Small skin incision was made at the flank region of NOD-SCID mice, and liver/tumor tissue was implanted aseptically under the subcutis. For histology analysis, liver tissues were excised, washed with ice-cold saline and either fixed in formalin for hematoxylin and eosin (H&E) staining or snap-frozen in liquid $N_2$. Further, IHC for proliferating cell nuclear antigen (PCNA) (Santacruz SC-96) and CycinD1 (SC-450) was done using VECTASTAIN® ABC kit (Vector Lab, P6200).

### Transmission electron microscopy

Liver tissues were fixed with 3% glutaraldehyde and post-fixed with 1% osmium tetraoxide. Grids were contrasted by use of alcoholic uranyl acetate for 1 min and lead citrate for 30 s. The grids were then observed under a Carl Zeiss LIBRA120 EFTEM.

### Histone extraction, acetic acid urea triton (AUT)-PAGE and Western blotting

Histones were extracted and purified as earlier.[9] AUT PAGE was done as described previously.[8] Gels were documented as image files and proteins spots were quantitated by ImageJ software (v1.42q, National Institutes of Health). For western, histones were resolved on SDS-PAGE, transferred, and probed with respective antibodies from Millipore or Abcam or CST (H2Aub, D27C4; H3S10p, M-06-570; H3S28p, ab-5169; Pan-acetyl, ab-61257; H3, M-05-499; H4, 07-108) according to the manufacturer's instructions (Millipore/Abcam).

### Mass spectrometry

Histone spots of interest were subjected to matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) (Bruker Daltonics, Bremen, Germany; Ultraflex II) as mentioned else were.[8]

### Cell lines and HDAC inhibitors

Rat liver cell lines CL44 (preneoplastic) and CL38 (neoplastic) were used in the study. The cell lines were cultured in DMEM (invitrogen) media with 10% FBS and were maintained at 37°C with 5% $CO_2$ and 100 mg/ml streptomycin. Histone deacetylase inhibitor (HDACi), Trichostatin A (TSA: Sigma, T8552) was dissolved in absolute ethanol to prepare stock solution and cells were treated with different concentrations of TSA.

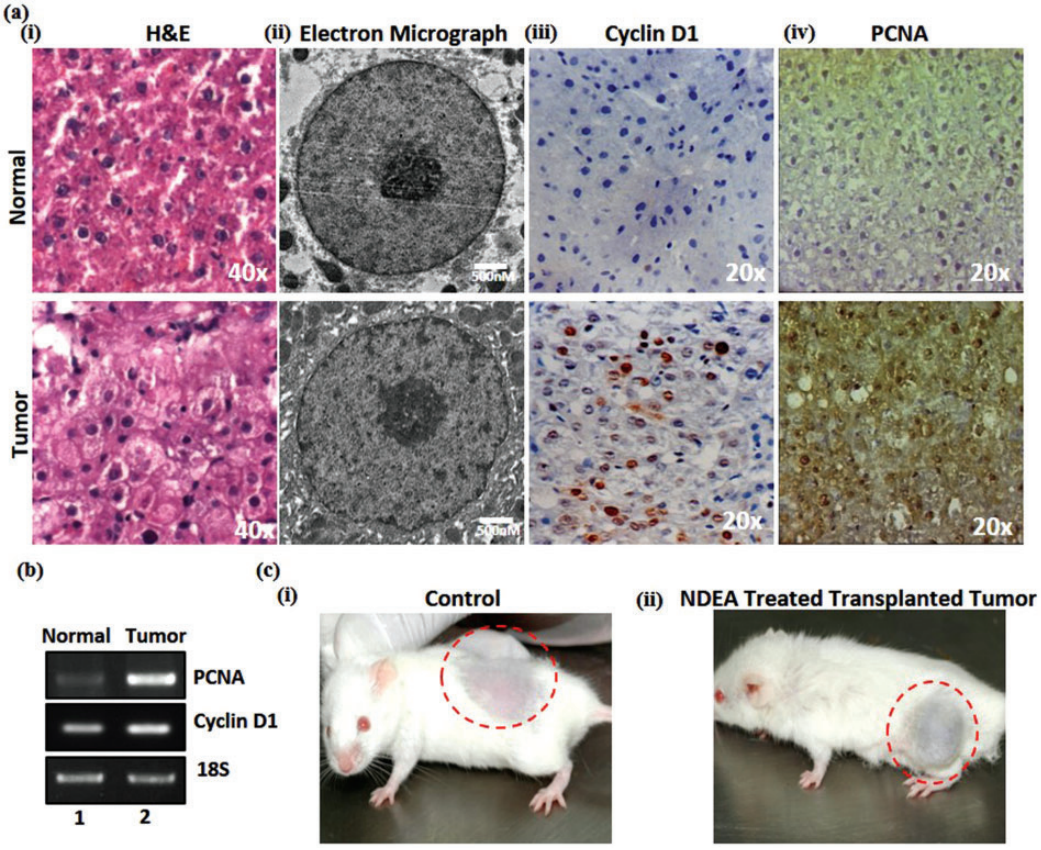



**Figure 1** NDEA-mediated liver cell transformation. (a) (i) Comparative H&E, (ii) Electron micrographs, (iii) IHC of Cyclin D1 and (iv) PCNA of normal and four-month NDEA-treated liver tissues suggesting development of HCC. (b) (i) Semi-quantitative RT-PCR of Cyclin D1 and PCNA of the same tissues. (c) Implanting liver of (i) normal and (ii) NDEA-treated to NOD-SCID mice confirms cellular transformation in four months NDEA-treated rats.
NDEA: N-nitrosodiethylamine; H and E: hematoxylin and eosin; HCC: hepatocellular carcinoma; IHC: immuno-histochemistry. (A color version of this figure is available in the online journal.)

### RT-PCR and real-time PCR

Total RNA (5 μg) prepared using TRIzol (invitrogen, 5596-026) was subjected to cDNA synthesis using random primers (Fermentas, K1632) as per the manufacturer's instructions. cDNA was subjected to PCR (NEB, M0271L) or real-time (SYBER green, Agilent Tech. 600882) analysis for the analysis of respective genes – **PCNA** (F: TCACAAAAG CCACTCCACTG, R: CATCTCAGAAGCGATCGTCA), **CyclinD1** (F:CGTGGCCACCTGGATGCTAGAG, R:TGCA GCAACTCCTCGGGGCGGAT), **Usp21** (F: ACTTCTCTCC GGCGTCTT, R:TGTGTGGTGAGCCATCTT), **RING1** (F: CCAAGCGGTCCCTACGGCC, R:CCTCGATACTGGAGC TCA), LCN2 (F:GGACCGAACGGTTCCAGG, R:CCCTGA CGAGGATGGAAG), **18SrRNA** (F:CGCGGTTCTATTTTGT TGGT, R:AGTCGGCATCGTTTATGGTC).

### Cell proliferation assay

Proliferation was quantified by the ability of viable cells to reduce tetrazolium salt 3-(4,5-dimethylthiazol-2-y)-2, 5-diphenyl tetrasodium bromide (MTT) to a colored formazan product. Cell proliferation was then expressed as the percentage of absorbance (of formazan crystals) obtained in control cultures.

## Results

### Nitrosodiethylamine is a potent inducer of HCC in Sprague-Dawley rats

Nitrosodiethylamine (NDEA) administration for four months led to the development of HCC as confirmed by H&E staining, with cancer cells exhibiting evident atypia and large nuclei (Figure 1(a, i)). In addition, electron microscopy revealed several indentations giving rise to a more irregular shaped nucleus (Figure 1(a, ii)). This was further corroborated by high Cyclin D1 (Figure 1(a, iii)) and PCNA (Figure 1(a, iv)) expression in tumor sections by immunohistochemistry analysis and also their marked upregulation at the transcript level (Figure 1(b)). Also, 3 $mm^2$ liver tissue was excised from normal and NDEA administered animals (with tumor) and implanted into NOD-SCID-mice. Two weeks post-implantation tumors developed only in the mice with tumor tissue (Figure 1(c)). These results demonstrate the stable transformation of liver cells by administration of NDEA.

### H2A ubiquitination markedly decreases in HCC

Histones were isolated from tissues and were resolved on AUT-PAGE (Figure 2(a)). A prominent decrease in Spot X

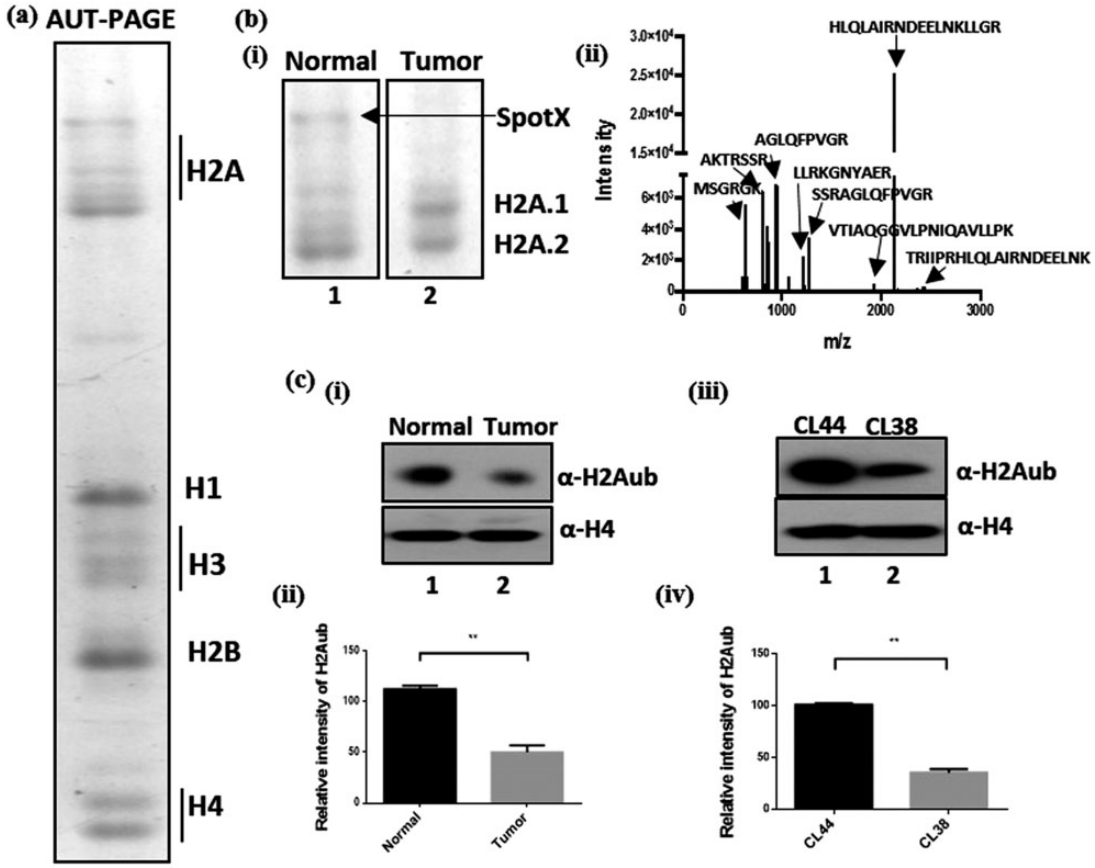



**Figure 2** H2Aub levels in normal and tumor tissues. (a) Single dimensional AUT-PAGE profile of purified histones from normal tissues. Region marked spans the mobility of respective histones. (b) (i) Comparative AUT profile of normal and tumor tissues revealed differential levels of Spot X in enlarged H2A region. (ii) Mass spectrometry confirmed the identity of spot X to be as H2A. (c) Western blotting with antiH2Aub antibody revealed (i) decrease levels in HCC compared to normal, and also (iii) reduce levels of H2Aub in CL38 (Neoplastic) compared to CL44 (preneoplastic) cell lines. H4 was used as a loading control. (ii) & (iv) Densitometry analysis indicated decrease H2Aub in tumor and in CL38, respectively. Each histogram represents mean densitometry + SEM (*$p = 0.05$, **$p = 0.01$).
AUT-PAGE: acetic acid urea triton polyacrylamide gel electrophoresis.

was observed in tumor sample (Figure 2(b, i)). Increase in H2A.1 and decrease in H2A.2 were also observed in tumor which we have previously reported (Figure 2(b, i)).[8] MS–MS analysis of Spot X suggested that the band corresponds to H2A (Figure 2(b, ii)). Upshift of the band of H2A suggested the occurrence of mono-ubiquitination. This was further validated by subjecting the region of the AUT-PAGE to Western blotting with anti-H2Aub (Figure 2(c, i,ii)). We included liver cell lines derived from the same model system for our studies, CL44 (pre-neoplastic), and CL38 (neoplastic). Similar changes in H2Aub levels were observed on comparison of histones isolated from both the cell lines (Figure 2(b, iii,iv)). Our results suggest that overall there is decrease in the level of H2Aub in NDEA-induced liver cancer model.

### Decrease in H2Aub is probably due to Usp21 and associates with upregulation of lipocalin 2

We looked for the levels of Usp21 as it was demonstrated to bring about H2A119 deubiquitination during liver regeneration.[10] Both in tumor tissue and neoplastic cell line CL38, a marked increase in the expression level of Usp21 was observed by quantitative (Figure 3(a, i)) and semi-quantitative PCR (Figure 3(a, ii)). We did not find any change in the expression level of the major H2A ubiquitinase RING1 (Figure 3(a, iii)), suggesting that the deubiquitination observed might be primarily owing to upregulation of the deubiquitinase.

H2Aub has been previously shown to localize to the promoter of genes like lipocalin 2 and repress their expression.[10] Hence, we checked the expression level of lipocalin 2 gene LCN2. A marked increase in the expression of LCN2 was observed in the CL38 cells and tumor tissue compared to their normal counterparts (by quantitative (Figure 3(b, i)) and semi-quantitative PCR (Figure 3(b, ii,iii)).

### Increased cell proliferation may contribute to loss of H2Aub in tumor

Ubiquitination of H2A is dynamic, as suggested by the observations that global levels of H2Aub vary during the cell cycle.[6] Drop in the level of H2Aub has been shown to be a prerequisite for increase in the mitotic mark H3S10p during S to G2/M transition of cells.[6] Along with H3S10p increase, H3/H4 hypoacetylation is also seen during mitosis.[11] Indeed, when we looked into levels of H3S10p and H4ac in both tissues and cell lines (Figure 3(c, i,ii)), we observed a marked increase of H3S10p as opposed to decrease in H2Aub and H4ac. We further investigated the

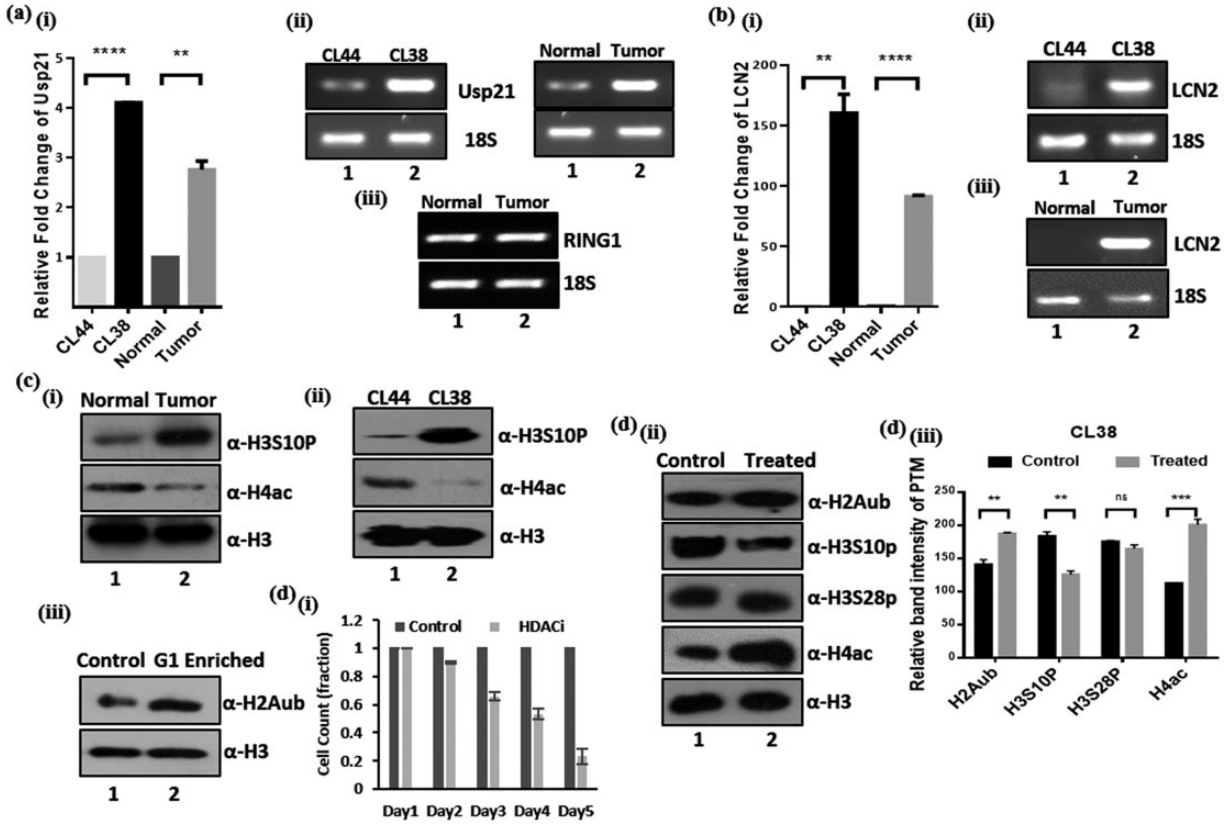



**Figure 3** Decreased H2Aub associates with upregulation of lipocalin 2 expression. (a) Levels of Usp21, a H2A deubiquitinase enzyme in both cell lines (CL44 and CL38) and tissues (normal and HCC). (i) Real-time PCR and (ii) semi-RT-PCR. (iii) Semi-RT-PCR of H2A ubiquitinase, RING1 levels in tissues. 18 S rRNA was used as a loading control. (b) Level of lipocalin 2 (LCN2) in cell lines and tissues (i) Real-time PCR, (ii) semi-RT-PCR in both cell lines, and (iii) semi-RT-PCR in normal and HCC. (c) Relative protein levels of H3S10p and H4ac in (i) normal and HCC, (ii) CL38 and CL44 cell lines, and (iii) H2Aub in control and G1 enriched cells. (d) HDACi, TSA treatment of CL38 cell line (i) time-dependent decrease in cell proliferation as observed by MTT assay, (ii) histone PTM profile with indicated antibodies, and (iii) densitometry analysis of the western data.
HDACi: histone deacetylase inhibitor; TSA: trichostatin A.

possible relation between the levels of H2Aub and cell proliferation. In CL38, increased percentage of cells in S and G2/M population was observed as compared to CL44 (Figure S1(a, i,ii)). Consistent with this, H2Aub was low in CL38 cells and in tumor (Figure 2(c, i,ii)). Next, we enriched CL38 cells in G1-phase of the cell cycle by serum starvation (Figure S1(b)). An increase is H2Aub was found in G1-enriched cells (Figure 3(c, iii)) strengthening the inverse relationship between H2Aub and cell proliferation.

As we have seen decrease in H4ac in tumor, we wanted to investigate the effect of increase in H4ac on H3S10p and H2Aub. As histone deacetylase inhibitors (HDACi) are known to decrease cell proliferation,[12] we treated CL44 and CL38 cells with HDACi Trichostatin A. We observed decrease in proliferation of both CL44 (data not shown) and CL38 cells (Figure 3(d, i)) by MTT assay. Histones were isolated 48 h post drug-treatment and the levels of histone modifications were observed by immuno-blotting (Figure 3(d, ii)). The activity of the drug was confirmed by increase in the histone H4ac levels (Figure 3(d, ii,iii)). An increase in H2Aub was observed with a decrease in the level of H3S10p and H3S28p (another mitotic mark).

## Discussion

Reduction in H2Aub has been reported in human pancreatic cancer[5] but otherwise this mark is poorly studied in cancers. In this study, we found marked decrease in H2Aub and increase in expression of gene lipocalin 2 in NDEA-HCC model. Interestingly, lipocalin 2 has been proposed to increase migration and metastasis of breast cancer cells.[13] We found the deubiquitinase Usp21 to be profoundly upregulated in HCC and no alteration in the level of ubiquitinase RING1. However, role of other ubiquitinases and deubiquitinases cannot be ignored. Along with H2Aub decrease, we have also observed a decrease in H4ac and increase in H3S10p. Our data thus suggest for the first time an inverse correlation between the global levels of cellular transformation mark H3S10p and H2Aub in vivo.

Upon treatment with HDACi, we found an increase in the level of H2Aub, H4ac, and drop in H3S10p. Previous studies have shown a decrease in H2Aub upon treatment with HDACi, valproic acid.[14,15] These differences in observation can be attributed to the use of different types of HDACi in the study. TSA is a class I and II specific inhibitor, without any specificity towards either of the enzymes.

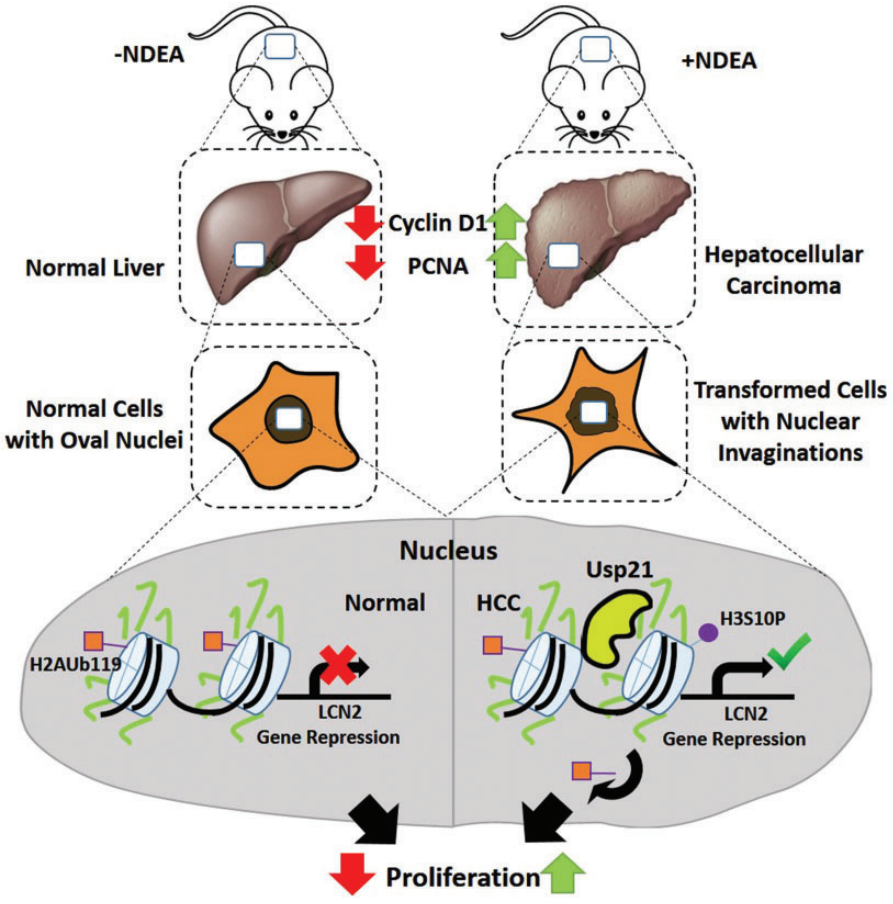



**Figure 4** Proposed model for decrease in H2Aub and its association with other histone marks in HCC. NDEA treatment for four months led to stable cellular transformation with gross morphological changes in the nuclear shape compared to normal liver and increase in expression of cell proliferation markers (PCNA). At the molecular level, global decrease in H2Aub and increase in H3S10p levels was seen with simultaneous increase in lipocalin 2 (LCN2) expression, thus favoring cellular transformation of normal liver to HCC. (A color version of this figure is available in the online journal.)

On the other hand, VPA used in previous report has been shown to have selectivity towards HDAC2 amongst other HDACs.[16] Further, whether this increase in H2Aub seen upon TSA treatment is solely because of change in cell cycle profile of cells and H3S10p or whether there is involvement of cross-talk between the histone acetylation mark and H2Aub would be an interesting prospect to investigate.

In summary, we here show that NDEA administration leads to stable transformation of liver to HCC. This is associated with changes in shape and size of the nucleus along with increase in cell proliferation markers. At the molecular level, there is a marked decrease in the level of H2Aub in HCC. The loss of H2Aub also appears to be partly due to increase cellular proliferation and treatment with HDACi increases H2Aub in cells. Usp21 like in case of liver regeneration probably mediates this loss. This leads to upregulation of lipocalin 2, a growth factor whose role has been implicated in cancer progression (Figure 4). Future work should focus on understanding the functional implication of histone marks and whether the loss of H2Aub is a common hallmark of cancers.

## Authors' contribution

SB majorly performed experiments. DR contributed in culture maintenance. AI carried out animal implantation. BK carried out histology and IHC experiments. SB and DR contributed in manuscript writing SG planned the experiments and wrote the manuscript.

### ACKNOWLEDGEMENTS

The authors thank Dr. H. M. Rabes (University of Munich, Germany) for providing the CL44 and CL38 cell lines, ACTREC support for Gupta Lab and are grateful to all members of Gupta Lab, ACTREC for valuable discussion. S. B and D. R are supported by CSIR fellowship.

### DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### REFERENCES

1. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res* 2011;**21**:381–95
2. Kalb R, Latwiel S, Baymaz HI, Jansen PWTC, Müller CW, Vermeulen M, Müller J. Histone H2A monoubiquitination promotes histone H3 methylation in polycomb repression. *Nat Struct Mol Biol* 2014;**21**:569–71
3. Bonner WM, West MH, Stedman JD. Two-dimensional gel analysis of histones in acid extracts of nuclei, cells, and tissues. *Eur J Biochem* 1980;**109**:17–23
4. Matsui S-I, Seon BK, Sandberg AA. Disappearance of a structural chromatin protein A24 in mitosis: implications for molecular basis of chromatin condensation. *Proceedings of the National Academy of Sciences* 1979;**76**:6386–6390
5. Glinskii AB, Yang Z, Li X, Hoffman RM. Essential role for activation of the polycomb group (PcG) protein chromatin silencing pathway in metastatic prostate cancer. *Cell Cycle* 2014;**5**:1886–1901
6. Joo H-Y, Zhai L, Yang C, Nie S, Erdjument-Bromage H, Tempst P, Chang C, Wang H. Regulation of cell cycle progression and gene expression by H2A deubiquitination. *Nature* 2007;**449**:1068–72
7. Choi HS, Choi BY, Cho YY, Mizuno H, Kang BS, Bode AM, Dong Z. Phosphorylation of histone H3 at serine 10 is indispensable for neoplastic cell transformation. *Cancer Res* 2005;**65**:5818–5827
8. Khare SP, Sharma A, Deodhar KK, Gupta S. Overexpression of histone variant H2A.1 and cellular transformation are related in N-nitrosodiethylamine-induced sequential hepatocarcinogenesis. *Exp Biol Med (Maywood)* 2011;**236**:30–5
9. Bonenfant D. Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. *Mol Cell Proteomics* 2005;**5**:541–52
10. Nakagawa T, Kajitani T, Togo S, Masuko N, Ohdan H, Hishikawa Y, Koji T, Matsuyama T, Ikura T, Muramatsu M, Ito T. Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev* 2008;**22**:37–49
11. Li Y, Kao GD, Garcia BA, Shabanowitz J, Hunt DF, Qin J, Phelan C, Lazar MA. A novel histone deacetylase pathway regulates mitosis by modulating aurora B kinase activity. *Genes Dev* 2006;**20**:2566–79
12. Sakajiri S, Kumagai T, Kawamata N, Saitoh T, Said JW, Koeffler HP. Histone deacetylase inhibitors profoundly decrease proliferation of human lymphoid cancer cell lines. *Exp Hematol* 2005;**33**:53–61
13. Yang J, Bielenberg DR, Rodig SJ, Doiron R, Clifton MC, Kung AL, Strong RK, Zurakowski D, Moses MA. Lipocalin 2 promotes breast cancer progression. *Proc Natl Acad Sci U S A* 2009;**106**:3913–8
14. Krämer OH, Zhu P, Ostendorff HP, Golebiewski M, Tiefenbach J, Peters MA, Brill B, Groner B, Bach I, Heinzel T, Göttlicher M. The histone deacetylase inhibitor valproic acid selectively induces proteasomal degradation of HDAC2. *EMBO J* 2003;**22**:3411–20
15. Bommi PV, Dimri M, Sahasrabuddhe AA, Khandekar J, Dimri GP. The polycomb group protein BMI1 is a transcriptional target of HDAC inhibitors. *Cell Cycle* 2010;**9**:2663–73
16. Göttlicher M, Minucci S, Zhu P, Krämer OH, Schimpf A, Giavara S, Sleeman JP, Lo Coco F, Nervi C, Pelicci PG, Heinzel T. Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO J* 2002;**20**:6969–78

*(Received February 3, 2016, Accepted April 19, 2016)*