Biochemical Characterisation of KDM2A

Jin Chuan Zhou

周锦川

Corpus Christi College

Doctor of Philosophy in Biochemistry

Department of Biochemistry

University of Oxford

October 2012

Declaration

I declare that this thesis was fully composed by myself and the research presented here is my own except where otherwise stated.

Jin Chuan Zhou

1st October 2012

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Rob for taking me as his first PhD student and for his invaluable mentoring and supervision throughout the four years of my PhD. I would like to thank all members of the Klose lab for their daily guidance and support. In particular I would like to extend special thanks to the people who read the first draft of this thesis for their helpful comments and suggestions. I also would like to thank the Medical Research Council for financially supporting my PhD. And I would like to say a very special thank you to Ying for making my last year at Oxford such a special time.

最后，我想感谢我的父母，感谢你们对我的培养和教导，感谢你们支持我的学业，让我为牛津的留学生活画上一个完美的句号。

Table of contents

TABLE OF CONTENTS **4**

FIGURE LIST **9**

TABLE LIST **11**

ABSTRACT **12**

ABBREVIATIONS **14**

1. CHAPTER ONE – INTRODUCTION **17**

1.1. CHROMATIN 17

The “bead on a string” model 17

The elusive higher order chromatin 19

1.2. NUCLEOSOME 23

The nucleosome core particle 23

Histone-DNA interaction and nucleosome positioning 25

Genome-wide nucleosome distribution 26

A genomic code for nucleosome organisation 29

Chromatin remodelling and nucleosome positioning 31

Nucleosome positioning in gene regulation 32

1.3. HISTONE MODIFICATIONS 37

The histone code 37

Histone acetylation in transcription regulation 41

Histone methylation in transcription regulation 46

1.4. DNA METHYLATION 64

DNA methylation in higher eukaryotes 64

CpG islands 70

1.5. INTERPRETATION OF DNA METHYLATION 77

Methyl-CpG binding proteins 77

Non-methyl-CpG binding proteins 80

1.6. KDM2 82

1.7. AIMS OF THIS THESIS 90

2. CHAPTER TWO – MATERIALS AND METHODS 92

2.1. COMMON SOLUTIONS 92

2.2. BACTERIAL METHODS 92

Media 92

Antibiotics 93

Culture on plates 94

Liquid culture 94

Preparing chemically competent strains 95

Transformation 96

2.3. ISOLATION OF EXOGENOUS DNA FROM BACTERIA 96

Small scale plasmid DNA preparation 96

Large scale plasmid DNA preparation 97

Large scale preparation of Bacmid DNA 97

2.4. MANIPULATION OF DNA AND CLONING 98

Restriction endonuclease digestion of DNA 98

Methylation of DNA fragments at CpGs 99

Polymerase chain reaction (PCR) 100

DNA Sequencing 100

Agarose gel extraction of DNA fragments 101

Ligation independent cloning (LIC) 101

Mutagenesis 102

Radioactive labelling of DNA fragments 103

2.5. INSECT CELL AND BACULOVIRUS VECTOR EXPRESSION SYSTEM 103

Insect cell culture 103

Generation of recombinant baculovirus genome 104

Transfection of Sf9 cells and isolation of P1 virus 105

Amplification of viral stock 106

Infection of Sf9 cells for protein production 106

2.6. EXPRESSION OF PROTEINS IN *E. COLI* CELLS 107

2.7. PROTEIN METHODS 108

Measuring protein concentration 108

Dialysis 109
SDS-PAGE and Coomassie blue staining 109
Semi-dry transfer to nitrocellulose membrane and Western blot 110
2.8. PROTEIN PURIFICATION 112
Chromatography media and solutions 112
Ion exchange chromatography 112
Gel filtration chromatography 113
Purification of 6XHis and Strep II tagged proteins from *E. coli* cells 113
Purification of Flag-tagged proteins from Sf9 cells 115
Purification of CBP tagged proteins 116
Purification of Taq DNA polymerase by heat precipitation and biphasic extraction 116
Purification of core histones 117
2.9. CHROMATIN RECONSTITUTION METHODS 118
Generation of 601 DNAs 118
Installation of methyl lysine analogues (MLAs) into core histones 119
Refolding of histone octamer 120
Reconstitution of nucleosome core particle 121
Purification of recombinant nucleosomes by sucrose gradient sedimentation 122
2.10. DNA BINDING EXPERIMENTS 122
Electrophoretic mobility shift assay (EMSA) 122
Surface plasmon resonance (SPR) 123
2.11. ENZYME ACTIVITY ASSAY 125
Histone demethylase assay by KDM2A 125
Histone methyltransferase assay by Set2 126
3. CHAPTER THREE – ENZYMATIC ACTIVITY AND DNA BINDING PROPERTY OF KDM2A 128
3.1. KDM2A IS A H3K36 MONO- AND DI-DEMETHYLASE 128
Expression and purification of recombinant KDM2A 129
KDM2A6 removes H3K36 mono- and di-methylation from histones *in vitro* 139
KDM2A removes H3K36 methylation from nucleosomes *in vitro* 145
3.2. KDM2A SPECIFICALLY RECOGNISES NON-METHYL CPGS THROUGH THE ZINC FINGER CXXC DOMAIN 151

KDM2A binds preferentially to non-methylated CpG DNA *in vitro* 151
Quantification of KDM2A CXXC domain binding affinity by surface plasmon resonance 160
3.3. DISCUSSION 166
KDM2A displays mono- and di-methylated H3K36 demethylase activity against both free histones and chromatin 166
KDM2A binds preferentially to CpG DNA via the CXXC domain 170
KDM2A is a monovalent DNA-binding protein 173
4. CHAPTER FOUR – KDM2A CHROMATIN BINDING SPECIFICITY 176
4.1. ASSEMBLY OF RECOMBINANT CHROMATIN BY THE SALT DILUTION METHOD .. 176
Purification of recombinant Taq DNA polymerase 177
Generation of the 601 nucleosome positioning sequence DNA 182
Purification of recombinant M.SssI and CpG methylation of the 601 DNA 184
Expression and purification of recombinant core histones 186
Reconstitution and purification of histone octamer 189
Reconstitution and purification of nucleosomes 190
4.2. KDM2A BINDING TO CHROMATIN 194
KDM2A preferentially binds to non-methylated CpG dinucleotides in the linker region of a 216 bp mononucleosome 194
KDM2A binds preferentially to non-methylated CpG dinucleotides in the linker region between two nucleosomes. 199
4.3. DISCUSSION 203
KDM2A preferentially interacts with linker DNA *in vitro* 203
Why does KDM2A only bind to linker DNA? 205
How does KDM2A interact with *in vivo* chromatin? 206
5. CHAPTER FIVE – ROLE OF THE PHD DOMAIN IN KDM2A 210
5.1. KDM2A PHD DOMAIN IS NOT REQUIRED FOR BINDING TO CHROMATIN 210
Expression and purification of KDM2A PHD domain 210
KDM2A PHD domain does not bind specifically to any known histone modification 214
Installation of methyl analogues on histone H3 lysine 4 216
Neither KDM2A PHD domain or H3K4 trimethylation is required for KDM2A binding to chromatin 217
5.2. DISCUSSION 221

The KDM2A PHD domain does not have histone binding function 221

Possible functions of the KDM2A PHD domain..... 223

6. CHAPTER SIX – GENERAL DISCUSSION AND FUTURE DIRECTIONS.....225

6.1. Chromatin accessibility is important for KDM2A localisation 225

6.2. KDM2A contributes to a transcriptional permissive chromatin environment at CGIs 229

REFERENCES 236

APPENDIX.....280

PUBLICATIONS.....284

Figure list

Figure 1.1 – Interphase chromatin under the electron microscope 19
Figure 1.2 – Model for a fractal organisation of chromatin 22
Figure 1.3 – Crystal structure of the nucleosome core particle 24
Figure 1.4 – Nucleosome distribution surrounding the transcription start site (TSS) 27
Figure 1.5 – Favourable and repressive sequences cooperate to pinpoint nucleosome position 30
Figure 1.6 – Post-translational modifications of core histones 38
Figure 1.7 – Mechanisms of action for histone modifications in chromatin regulation 40
Figure 1.8 – The two families of histone lysine demethylase (HKDM) enzymes 50
Figure 1.9 – H3K4 and H3K36 methylations regulate gene transcription initiation and elongation respectively 60
Figure 1.10 – CpG islands 71
Figure 1.11 – CGI promoters during normal development 75
Figure 1.12 – Mechanism of methyl-CpG binding proteins mediated repression 79
Figure 1.13 – Crystal structure of the zinc finger CXXC domain in DNMT1 and CFP1 81
Figure 3.1 – Baculovirus expression vector system (BEVS) 131
Figure 3.2 – Expression and purification of full length KDM2A via BEVS. 134
Figure 3.3 – Purification of KDM2A6 138
Figure 3.4 – KDM2A6 removes H3K36me1 and H3K36me2 from calf thymus and recombinant histones 140
Figure 3.5 – Formaldehyde dehydrogenase (FDH)-coupled demethylase assay 143
Figure 3.6 – Histone H3K36 methyltransferase assay by recombinant yeast Set2 147
Figure 3.7 – KDM2A removes H3K36 methylation from nucleosomes *in vitro* 149
Figure 3.8 – KDM2A preferentially binds to non-methylated CpG DNA 155
Figure 3.9 – Efficient KDM2A demethylase activity is dependent on a functional CXXC domain 159
Figure 3.10 – Surface plasmon resonance 161
Figure 3.11 – Sensorgram of KDM2A binding to the DNA probe 162
Figure 3.12 – K_D measurement of KDM2A determined by SPR 165
Figure 4.1 – Purification of Taq DNA polymerase 180
Figure 4.2 – Purification of the 601 DNA sequence synthesised by PCR 183
Figure 4.3 – Purification of M.SssI and CpG-DNA methylation 186
Figure 4.4 – Purification of recombinant core histones 188
Figure 4.5 – Reconstitution of histone octamer 190
Figure 4.6 – Reconstitution and purification of nucleosome core particle 193
Figure 4.7 – KDM2A binding to mononucleosome 196
Figure 4.8 – KDM2A binding to dinucleosome 202

Figure 5.1 – Purification of KDM2A PHD domain 213
Figure 5.2 – Installation of H3K4 trimethyl-lysine analogue 217
Figure 5.3 – KDM2A PHD domain does not bind to H3K4me3 nucleosomes 220
Figure 6.1 – zinc finger CXXC proteins KDM2A and CFP1 contribute to an open and transcriptionally permissive environment at CGIs 231

Table list

Table 1 281

Jin Chuan Zhou
Corpus Christi College
Trinity 2012

Abstract

Mammalian genomes are characterised by unique regions of non-methylated DNA known as CpG islands (CGIs). These genomic elements are characterised by a high density of CpGs and an elevated GC content compared to the surrounding, bulk of the genome. CGIs are prevalently associated with the 5' end of genes and represent key nucleation sites where specific transcription factors and chromatin modifiers are recruited to impact on gene function.

This thesis is focused at understanding the biochemical properties of the recently discovered H3K36-specific histone demethylase, KDM2A. This enzyme is specifically recruited to CGIs but how it interfaces with local chromatin *in vivo* remains unknown. Using defined chromatin templates *in vitro*, this study demonstrates that KDM2A binding to DNA relies on a zinc finger CXXC domain that preferentially recognizes non-methylated CpGs. In particular, nucleosomes represent a major barrier to KDM2A binding and chromatin substrates are interpreted by the CXXC domain through specific interaction with CpGs within linker DNAs. Moreover, the adjacent PHD domain does not contribute to KDM2A binding to chromatin.

Together these observations suggest that sequence, methylation status and accessibility of DNA define how CGI chromatin is interpreted by CXXC domain proteins. In particular, the precise targeting of KDM2A to CGIs contributes to the

creation of a unique chromatin architecture that highlights gene regulatory regions within large and complex mammalian genomes.

Abbreviations

A	ampere
aa	amino acids
APS	ammonium persulphate
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
°C	degree centigrade
C-	carboxyl
CGI	CpG island
Ci	curie
cpm	counts per minute
Da	dalton
dNTP	deoxyribonucleoside triphosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ES	embryonic stem
g	relative centrifugal force
h	hour

Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HAT	histone acetyltransferase
HDAC	histone deacetylase
HKDM	histone lysine demethylase
HKMT	histone lysine methyltransferase
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilobase
kDa	kilodalton
kcpm	kilo counts per minute
L	litre
M	molar
MBD	methyl CpG binding domain
mg	milligram
min	minute
mol	mole
N-	amino
ng	nanogram
nm	nanometre
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pH	$-\log_{10}[H^+]$
PHD	plant homeo domain
PMSF	phenylmethylsulfonyl fluoride

RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
S	second
SAM	S-adenosyl-L-methionine
SDS	sodium dodecyl sulphate
SELEX	systematic evolution of ligands by exponential enrichment
SPR	surface plasmon resonance
TAE	Tris-acetate-EDTA
TB	Tris-borate
TCA	trichloroacetic acid
Tris	tris(hydroxymethyl)aminomethane
TEMED	tetramethylethylenediamine
µg	microgram
µl	microliter
µM	micromolar
µm	micrometer
UV	ultra violet
V	volt

1. Chapter One – Introduction

1.1. Chromatin

The "bead on a string" model

Genetic information is stored in the form of DNA. The human genome contains approximately 3.2 x 10^9 nucleotides (Alberts, 2002). If these were joined together, the genome would be approximately 2 m long. Since the discovery of the double helical structure of the DNA molecule (Watson and Crick, 1953), effort has been focused at understanding how such a long, negatively charged, polymer can be efficiently enclosed within a 6 µm wide nucleus while allowing important processes like DNA replication, gene transcription, DNA repair and chromosome segregation to occur.

Early insights came from the analysis of interphase chromosomes isolated from a comprehensive set of mammalian tissues. It revealed that approximately 90% of the total chromosome mass consists of DNA-histone complex, with the remaining fraction characterised by non-histone proteins (Mirsky and Ris, 1947, 1951). This complex of DNA and nuclear proteins was defined as chromatin. The later identification and biochemical isolation of the basic repeating units in chromatin, known as nucleosomes, was achieved by treating chromatin with DNA endonucleases such as micrococcal nuclease (MNase) derived from *Staphylococcus*

aureus bacteria (Hewish and Burgoyne, 1973; Noll and Kornberg, 1977). It led to the suggestion that the first layer of DNA compaction in the nucleus starts with the wrapping of the long DNA filament around an array of nucleosome core units, forming a 10 nm wide chromatin fibre. Within this fibre, the nucleosomes were described as “beads” in what was known as the “beads on a string” model (Kornberg, 1974). Crosslinking of extracted chromatin subunits by dimethyl suberimidate allowed further characterisation of the exact stoichiometry of the histone components and it was proposed that an octameric histone complex forms the core of a nucleosome (Thomas and Kornberg, 1975). Electron micrographs of chromatin fibres derived from interphase nuclei (Figure 1.1) (Olins and Olins, 1974) and small angle neutron scattering (SANS) analysis of nucleosomes released from MNase-digested interphase chromatin (Hjelm et al., 1977) supported this view of chromatin organisation. However, it was only until more recently that X-ray diffraction data of *in vitro* reconstituted nucleosome provided a high-resolution map of the nucleosome core allowing the precise interactions between histones and the DNA double helix to be deciphered (Luger et al., 1997a).

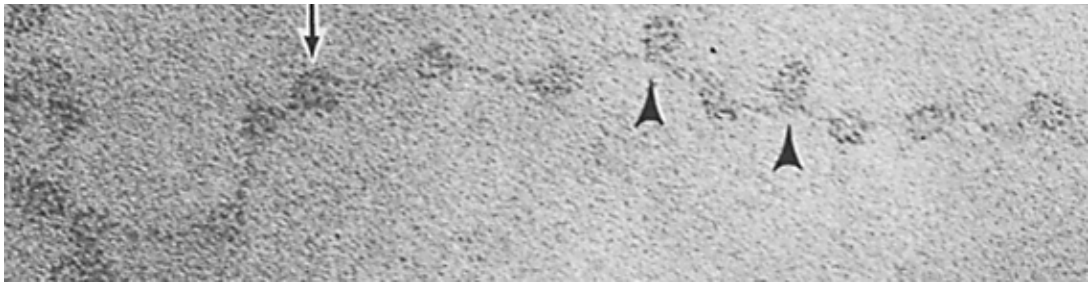


Figure 1.1 – Interphase chromatin under the electron microscope

Isolated interphase chromatin spread derived from chicken erythrocyte nuclei is stained with uranyl acetate and visualised under EM. Nucleosomes (indicated by fat and thin arrows) appear as "beads on a string". EM magnification is 357,000 X. Figure adapted from (Olins et al., 1976). Originally published in J Cell Biol *68*, 787-793.

The elusive higher order chromatin

The packaging of DNA into a 10 nm chromatin fibre allows for an order of 6-fold compaction (Fussner et al., 2011), which is difficult to reconcile with the highly condensed structure we can observe in mitotic chromosomes. Even within interphase nuclei, two forms of chromatin are known to be present: a highly condensed, transcriptionally inactive "heterochromatin" and a less condensed, transcriptionally active "euchromatin" (Alberts, 2002). These two chromatin domains vary significantly in terms of their sensitivity to MNase digestion as a result of the different accessibility of their underlying DNA (Levy and Noll, 1981). Heterochromatin domains are less accessible than euchromatin regions due to molecular crowding and volume exclusion, the effect of which has been evidenced by fluorescent microscopy analysis and measurement of the diffusion rate within heterochromatic regions (Bancaud et al., 2009). It is clear indeed that regardless of mitotic or interphase state, additional levels of chromatin folding must occur in the cell. To date, the exact nature of this second layer of chromatin organisation (the first being the nucleosome) remains elusive. It has long been noted however that in the presence of Mg^{2+} and histone H1, a linker DNA-specific histone protein, the

10 nm chromatin filament can adopt a 30 nm solenoidal structure (Finch and Klug, 1976). Reconstitution of regularly spaced nucleosome arrays has demonstrated that a 30 nm chromatin filament can be formed *in vitro* in the presence of the linker histone H1 (Robinson, 2006; Routh et al., 2008). The crystal structure of an *in vitro* obtained tetra-nucleosome complex, assembled in the absence of the linker histone H1 but in the presence of high levels of Mg^{2+}, suggested that a 30 nm fibre may be generated by stacking oligonucleosomes in a zigzag ribbon fashion (Schalch et al., 2005). The *in vivo* biological relevance of these *in vitro* structures, which rely on high concentrations of divalent cations, has however recently been questioned (Fussner et al., 2011).

Early *in vivo* observation of a 30 nm-like chromatin structure was reported from small angle X ray scattering (SAXS) measurements of chromatin derived from a diverse range of mitotic and interphase cell types (Langmore and Paulson, 1983; Widom and Klug, 1985; Woodcock et al., 1984). Later electron microscopy (EM) and electron spectroscopic imaging analysis (ESI) of starfish sperm chromatin (Bazett-Jones, 1992; Horowitz et al., 1994) have also provided evidence of a 30 nm chromatin organisation. However, this view was disputed when cryo-EM analysis of mitotic chromosomes in HeLa cells failed to detect any higher order 30 nm structure (Eltsov et al., 2008; Maeshima et al., 2010; McDowall et al., 1986). Unlike conventional EM, cryo-EM does not require chemical fixation and alcohol dehydration during sample processing, and is therefore favoured because less prone to artefact formation. Recently, a comprehensive study of HeLa mitotic

chromosomes that combined cryo-EM, SAXS and ultra-SAXS may have brought an end to the long-standing controversy. It has demonstrated that the 30 nm diffraction peak observed in previous SAXS studies of mitotic chromosomes did not originate from the chromatin itself, for which only a 11 nm peak could be detected, but from chromatin-associated ribosome aggregates (Nishino et al., 2012). Therefore, instead of bearing a rigid structure composed of periodic units, it is likely that chromosomes consist simply of an irregular meshwork of chromatin fibres. Based on the new SAXS results, it was suggested that a flexible and dynamic way of chromatin folding is likely to be also present in most interphase nuclei (Maeshima et al., 2010).

A recent Hi-C (a variant of the chromosome conformation capture, 3C technique) mapping study of large scale chromatin domain interactions in a human lymphoblastoid cell line has proposed a novel model where chromatin has been described to adopt a fractal organisation (Figure 1.2) (Lieberman-Aiden et al., 2009). According to this model, chromatin fibres can fold into globular structures through local chromatin looping, a prominent feature of nuclear organisation (Misteli, 2007). Each globule can be seen as a monomer. Multiple monomers cluster together to form a bigger globule, the next level monomer. This hierarchy is repeated, creating larger and larger chromatin domains. The ultimate level of chromatin organisation is likely to be represented by whole chromosome territories. These have been recently visualised through 3D fluorescent *in situ* hybridisation (FISH) painting of human fibroblast nuclei (Bolzer et al., 2005).

Moreover, regions of open, active and closed, inactive chromatin are kept distinct from each other, clearly demarcated through the action of insulator proteins (Vogelmann et al., 2011). Interestingly, the fractal nature of such proposed chromatin organisation had already been predicted to exist following previous diffusion rate studies in differentially packed chromatin states (Bancaud et al., 2009). Importantly, the proposed system allows for transition between open and closed chromatin conformations without altering the underlying 10 nm structure. This is necessary to allow for dynamic changes in global chromatin organisation such as those observed by ESI during early embryogenesis (Imhof et al., 2010). Recently, a follow-up Hi-C study comparing global chromatin region association in human and mouse has indicated that such way of spatial organisation may represent a conserved feature in metazoan genomes (Dixon et al., 2012).

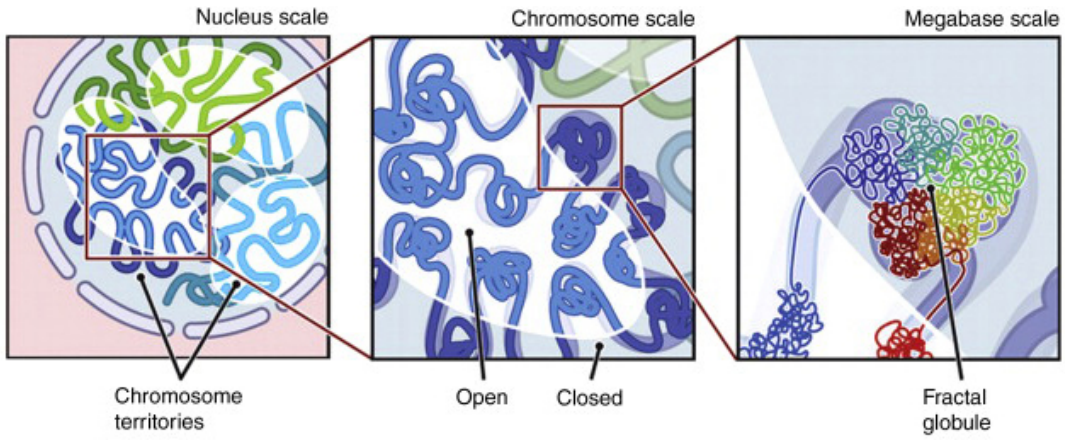


Figure 1.2 – Model for a fractal organisation of chromatin

Recent Hi-C data suggests that high order chromatin may be organised with disordered, fractal nature starting from whole chromosome territories down to local chromatin domains. Reprinted from Trends in Biochemical Sciences *36*, Fussner, E., Ching, R.W., and Bazett-Jones, D.P. Living without 30nm chromatin fibers, 1-6, Copyright (2011), with permission from Elsevier.

1.2. Nucleosome

The nucleosome core particle

Since a precise knowledge of the higher order chromatin structure in eukaryotes is currently lacking, the nucleosome, the fundamental unit of chromatin, remains to date the best studied form of chromatin organisation.

In a nucleosome, two molecules of each of the four types of canonical core histones H2A, H2B, H3 and H4 assemble together in a disc-shaped octameric core, around which 147 bp of DNA double helix is tightly wrapped 1.65 turns in a left-handed coil (Figure 1.3) (Luger et al., 1997a). This is known as the canonical nucleosome core particle. In the “beads on a string” model, each nucleosome core particle is separated by a variable length of linker DNA. A nucleosome is technically defined by the nucleosome core particle plus the connected linker DNA on one side although the term can be used more generally to refer to just the nucleosome core particle.

The core histones are among the most evolutionarily conserved proteins from yeast to humans (DeLange et al., 1969) and share a common domain architecture.

This comprises of a C-terminal structural motif, known as the histone fold, and an extended N-terminal unstructured tail (Luger et al., 1997a). A histone fold is characterised by three α helices joined by two loops. The specific folding of this domain is critical for the union of the four types of core histones into the octamer. This process can be thought to occur in a sequential fashion according to the strength of interaction between each histone molecule. Firstly, two heterodimers are formed between the H3 and H4 histones and between the H2A and H2B histones. Secondly, two H3-H4 dimers combine together to form an H3-H4 tetramer. Thirdly, two H2A-H2B dimers join one H3-H4 tetramer to form a complete octamer.

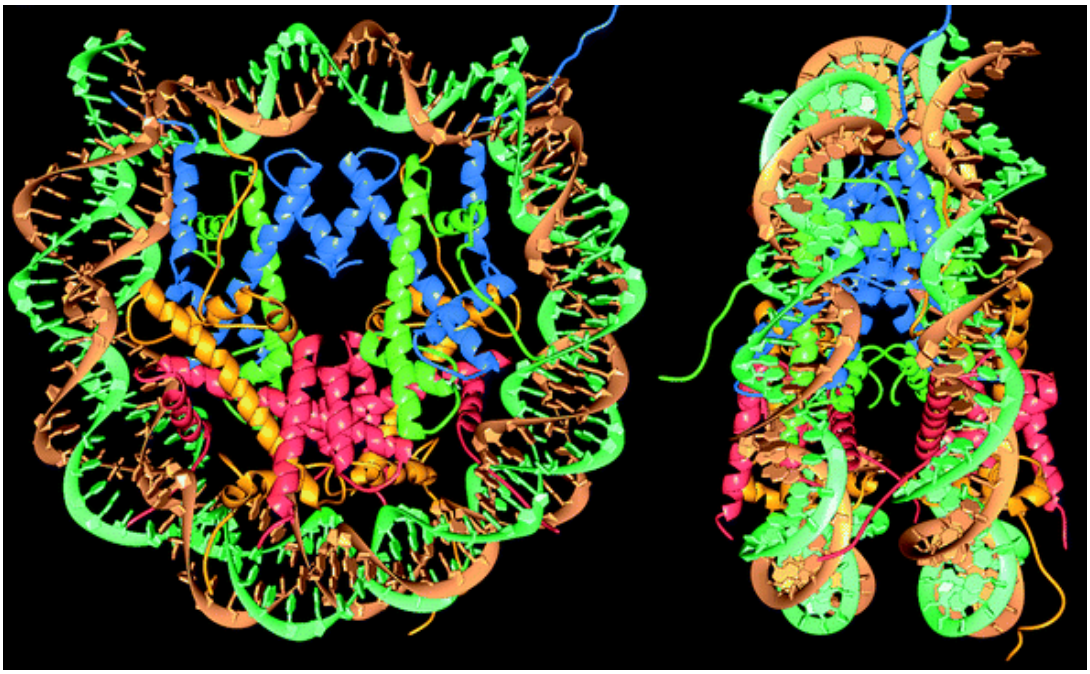


Figure 1.3 – Crystal structure of the nucleosome core particle

The four core histones H2A (yellow), H2B (pink), H3 (blue), H4 (green) form an octameric core surrounded by 146 bp of DNA. Views are looking down the DNA superhelical axis (left) and perpendicular to it (right). Figure adapted and reprinted by permission from Macmillan Publishers Ltd: Nature, (Luger et al., 1997a), copyright (1997).

Histone-DNA interaction and nucleosome positioning

Histones are small basic proteins rich in lysines and arginines. These residues contribute to an overall positive charge that serves to neutralise the negatively charged backbone of the DNA. Importantly, in a nucleosome numerous hydrogen bonds exist between the amino acid backbone of the histones and the phosphodiester backbone of the DNA (Luger et al., 1997a). The non-specific nature of these interactions allows virtually any DNA sequence to be packaged into a nucleosome, in fitting with the need to accommodate the vast diversity of sequences represented in all eukaryotic genomes. However, as the DNA winds around the histone octamer, sharp bending of the double helix occurs at both major and minor grooves with a 5 bp periodicity (Widom, 2001). Because some nucleotide sequences are more easily bent than others (Satchwell et al., 1986), nucleosomes do display DNA sequence preference and tend to occupy positions in the DNA that are most thermodynamically favourable (Widom, 2001). *In vitro* SELEX (systematic evolution of ligands by exponential enrichment) studies have demonstrated that such difference in DNA-nucleosome binding affinity can be of several orders of magnitude (Thastrom et al., 1999). Nonetheless, synthetic nucleosome positioning sequences derived from the SELEX enrichment process are

likely to represent extreme cases of DNA-histone interaction. *In vivo*, nucleosome affinity for a specific DNA sequence does not need to be as strong as that observed in an artificial setting. It is rather intuitive to think that in a cell, sequence-extrinsic factors, like transcription activators or chromatin remodelling enzymes, must dominate over sequence-intrinsic determinants in order to regulate nucleosome positions as required and depending on the on-going DNA function. Indeed, accumulating research evidence suggests that interplay between DNA-encoded sequence preference and active nucleosome sliding and/or eviction is responsible for the genome wide nucleosome organisation (Sadeh and Allis, 2011; Valouev et al., 2011).

Genome-wide nucleosome distribution

With the advent of massively parallel sequencing technology, high-resolution maps of nucleosome locations in the genome of a number of organisms from yeast to human have become recently available (Barski et al., 2007; Lee et al., 2007b; Mavrich et al., 2008b; Schones et al., 2008; Valouev et al., 2008). A common theme of nucleosome organisation started to emerge at the majority of gene-coding regions in these model organisms (Figure 1.4). The transcription start site (TSS) is usually preceded by a putative nucleosome free region (NFR) flanked by two highly phased (i.e. preferentially positioned) nucleosomes, denoted as -1 and +1 with respect to the TSS (Jiang and Pugh, 2009). Such phasing is progressively lost at

farther distances from the TSS. A statistical positioning model has been proposed to account for this phenomenon (Fedor et al., 1988; Mavrich et al., 2008a). It suggests that an array of nucleosomes is initially packed against a physical "barrier" that prevents nucleosome formation, presumably located in the NFR, which is known to be occupied by the transcription machinery. The degree of phasing of nucleosomes emanating from this barrier is then dictated by the statistical probability of each nucleosome to adopt a given position as they randomly fluctuate along the DNA template.

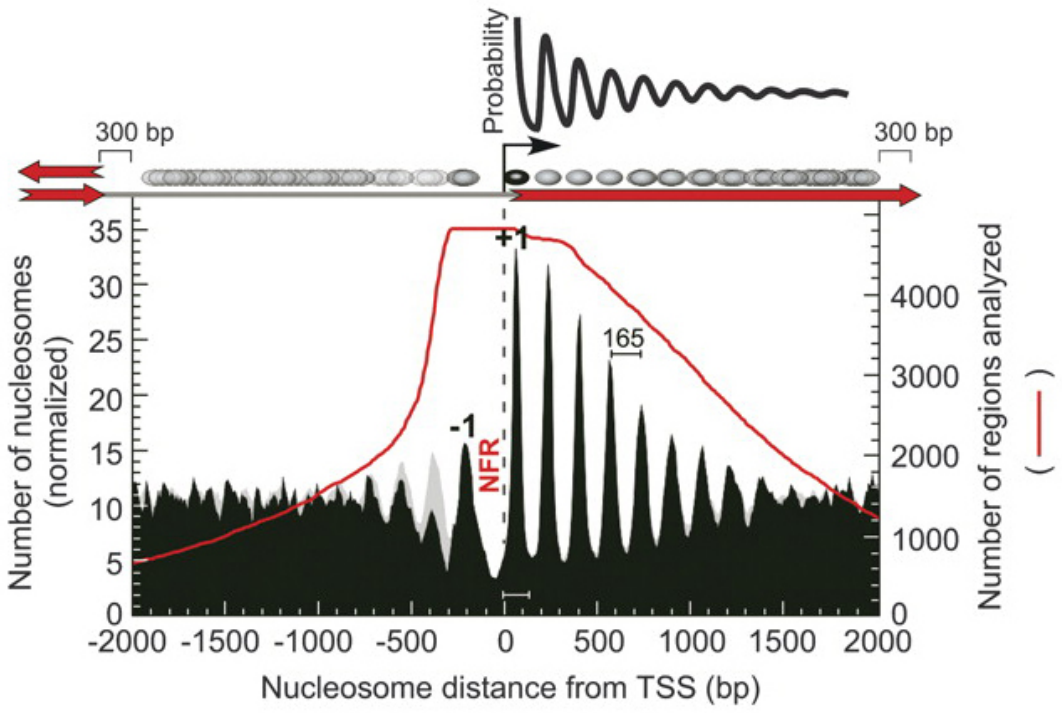


Figure 1.4 – Nucleosome distribution surrounding the transcription start site (TSS)

Shown here is a composite view of nucleosome positions around 4799 TSS in *S. cerevisiae*. -1 and +1 nucleosomes are among the most highly phased and flank the

nucleosome free region (NFR) with the degree of phasing fading further away from the TSS. This pattern is in accordance with the proposed statistical model of nucleosome positioning (top). Figure from (Mavrich et al., 2008a). Originally published in Genome Res *18*, 1073-1083.

This stereotypical view of nucleosome localisation around TSS applies well to the major class of promoters known (Cairns, 2009; Lenhard et al., 2012). These are associated with housekeeping genes that are constitutively and ubiquitously expressed in all cell types (Carninci et al., 2006). However, as discussed below, a different nucleosome landscape can be observed at other promoter classes, such as those linked to genes that are specific to terminally differentiated tissues and whose expression is consequently subject to tight regulation (Bai and Morozov, 2010).

It needs to be noted that the conservation of nucleosome organisation genome wide across species is not perfect and subtle discrepancies can be observed, even in relation to the highly phased nucleosomes. For example, the *Drosophila* -1 nucleosome is less well positioned with respect to the *Saccharomyces* counterpart while the +1 nucleosome is generally shifted an average 75 bp further downstream of the TSS, giving rise to a slightly larger NFR in flies (Mavrich et al., 2008b). These differences are thought to have an impact on how the TSS is engaged by RNA polymerase II (Pol II) and may reflect species-specific mechanisms of transcription regulation.

A genomic code for nucleosome organisation

In an attempt to account for the non-random pattern of nucleosome distribution observed *in vivo*, computational modelling studies using *in vitro* positioned nucleosome sequences have led to the proposition of an evolutionarily conserved genomic "code" that has the ability to dictate nucleosome positioning in a cell (Kaplan et al., 2008; Segal et al., 2006). It was suggested that rigid DNA sequences such as poly dAdT tracts (Iyer and Struhl, 1995) disfavour nucleosome assembly and are therefore enriched at specific genomic loci like the NFRs as a consequence of evolutionary selection. Importantly, this sequence bias is most pronounced at gene regulatory regions including promoters and enhancers (He et al., 2010; Sekinger et al., 2005), suggesting a functional relevance may exist.

The universality of such nucleosome positioning code however is questionable, as evolutionary sequence divergence around TSS is known to exist between lower and higher eukaryotes. As discussed below, in contrast to yeast and worm where AT-rich sequences clearly dominate at promoter regions (Johnson et al., 2006; Lee et al., 2007b), most mammalian promoters are enriched in G/C base pairs (Saxonov, 2006), which positively correlate with nucleosome occupancy *in vitro* but to much lesser extent *in vivo* (Tillo and Hughes, 2009; Valouev et al., 2011). Interestingly, a recent study using human primary T cells has shown that a nucleotide combination of G/C-rich core region flanked by A/T-rich sequences may be embedded at mammalian promoters. These short stretches of DNA have been suggested to act

as nucleosome “container” sites and have the capacity to accurately pinpoint the position of the most highly phased nucleosomes (Figure 1.5) (Valouev et al., 2011), which can in turn function as barriers to statistically distribute neighbouring nucleosomes.

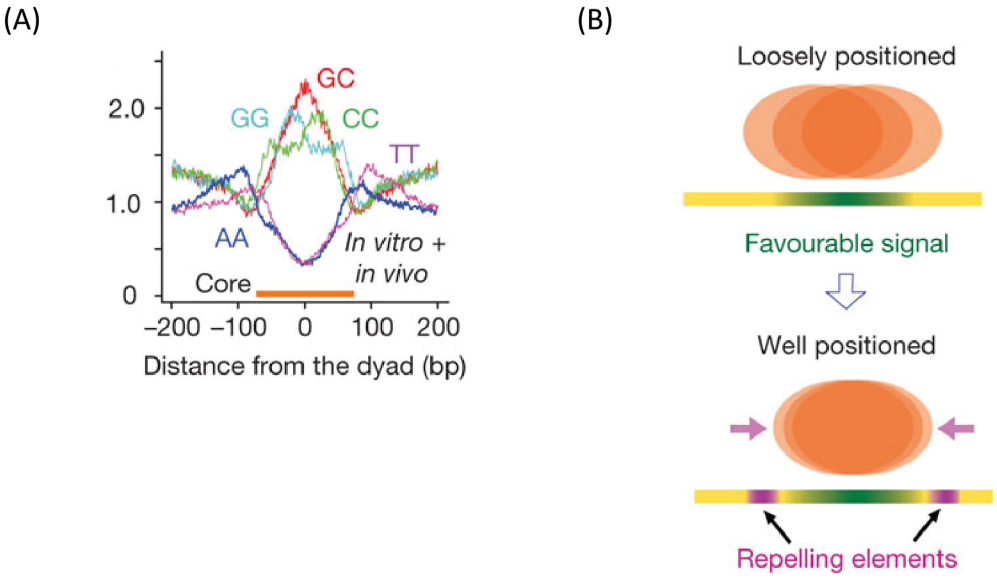


Figure 1.5 – Favourable and repressive sequences cooperate to pinpoint nucleosome position

(A) Combination of sequence signals observed for both *in vitro* and *in vivo* nucleosome positions indicate a prevalence of GC-rich core flanked by AA/TT-rich side sequence. (B) Schematic illustrating how nucleosome favouring and repelling signals can act in concert to dictate precise nucleosome position. Figure adapted and reprinted by permission from Macmillan Publishers Ltd: Nature, (Valouev et al., 2011), copyright (2011).

Chromatin remodelling and nucleosome positioning

To date, there is on-going controversy over the predictive power of these sequence-intrinsic rules in establishing nucleosome localisation *in vivo* (Zhang et al., 2009). It has been argued that a genomic code for nucleosome positioning is unlikely to be the major determinant for nucleosome organisation in a cell because of the presence of *trans*-acting factors like chromatin remodelling proteins. Indeed, a vast repertoire of ATP-dependent chromatin remodelling complexes is known to exist that has the capacity to reposition or evict nucleosomes over any particular DNA region (Saha et al., 2006). Recent triple knockout studies in *S. cerevisiae* have shown that in the absence of the three remodelling enzymes Isw1, Isw2 and Chd1, the existing organised pattern of nucleosome distribution in the genome fails to be maintained (Gkikopoulos et al., 2011). This effect is most striking over coding regions and less significant around TSS, where additional factors are likely to be involved in preserving the NFR. Interestingly, yeast mutants that carry null alleles for these three chromatin remodelers do not show mitotic defects under normal growth conditions but are more susceptible to various environmental stresses (Tsukiyama et al., 1999). This result suggests that these enzymes may have a more prominent role in regulating inducible promoters that do not usually display NFRs and whose associated genes are tightly controlled according to environmental cues. A complementary study in *S. cerevisiae* has indicated that formation of NFRs is likely due to DNA binding by transcription factors like Reb1 and Abf1, followed by

active sliding of the flanking nucleosomes (Hartley and Madhani, 2009). Importantly, this process seems to be dependent on the RSC remodelling complex. In contrast to Isw1, Isw2 and Chd1, RSC is essential for mitotic growth in yeast and estimated to be at least 10 times more abundant, which highlights the importance of keeping an open chromatin state at NFRs (Cairns et al., 1996; Laurent et al., 1992). Moreover, comparison studies of *in vitro* assembled chromatin in the presence or absence of ATP-dependent chromatin remodelers have indicated that chromatin remodelling complexes do have the ability to overcome intrinsic DNA-nucleosome preferences and relocate nucleosomes to less thermodynamically favourable positions (Perales et al., 2011; Whitehouse and Tsukiyama, 2006; Zhang et al., 2009). Indeed, the pattern of nucleosome distribution obtained in the presence of ATP-dependent chromatin remodelling is nearly indistinguishable from that of chromatin isolated from cells. Together these series of studies highlight the key role played by chromatin remodelers in actively shaping global nucleosome distribution *in vivo*.

Nucleosome positioning in gene regulation

Whether dictated by DNA sequence or chromatin remodelling enzymes, the striking conservation of nucleosome phasing around TSS found across species is strongly suggestive of nucleosomes acting as regulators of gene transcription. Two key properties of nucleosomes are thought to facilitate this role.

Firstly, nucleosomes have the capacity to limit the accessibility of their underlying DNA sequence by occluding protein binding sites including those recognised by transcription factors (Morse, 2007). As discussed below, this is valid for both the family of general transcription factors that mediate RNA Pol II recruitment and the gene-specific ones.

Secondly, nucleosomes are inherently dynamic in nature. Reports have indicated that nucleosomes are locally mobile and can transiently and spontaneously expose their underlying DNA (Li et al., 2005). These thermal fluctuations allow DNA sequences that are buried in the nucleosome to be briefly accessible to nearby proteins. This property is of particular relevance for transcription factors and chromatin remodelers as they can potentially trap the nucleosomal DNA during the unwrapping state, and initiate progressive sliding or eviction of the nucleosome core particle. The equilibrium constant measured for such nucleosome dynamics decreases towards the nucleosome dyad (i.e. the midpoint) suggesting that in a cell DNA unwrapping is likely to proceed from the nucleosome boundaries (Anderson and Widom, 2000; Tims et al., 2011). This is in accordance with a genome wide mapping study showing preferential localisation of promoter elements including TATA boxes, transcription factor recognition sequences and transcription initiation sites near the entry/exit point of nucleosomes (Albert et al., 2007).

The above mentioned features allow nucleosomes to dynamically modulate transcriptional output and contribute to the plasticity of gene promoters. Arguably such plasticity is most important for inducible promoters that have evolved to regulate the local concentration of transcription factors in response to changes in environmental stimuli (Tirosh and Barkai, 2008). The local concentration of transcription factors can be viewed as a reflection of the number of binding sites that are accessible at a given time, which is in turn dictated by nucleosome occupancy. Historically, this relationship has been first described using the *S. cerevisiae PHO5* gene promoter as a model (Reinke and Horz, 2004). During conditions of inorganic phosphate depletion, the activator Pho4 is recruited to a single constitutively accessible, low affinity binding site (UASp1) within a short region of the *PHO5* promoter. Following initial binding, chromatin remodeler-mediated removal of -1 to -4 nucleosomes then allows for transcriptional activation. It has been originally proposed that activation is dependent on the exposure of a second, high affinity binding site for Pho4 (UASp2) occluded by the -2 nucleosome (Venter et al., 1994). Such argument has been disputed by recent evidence of Pho4 interaction with UASp2 in the presence of the nucleosome (Adkins et al., 2004). Nonetheless, it is clear that nucleosome remodelling is an essential requirement for transcription to occur and that the events also involve local histone acetylation (Reinke and Hörz, 2003) through pathways described in more detail below. Importantly, over recent years, it has become clear that chromatin remodelling represents a general mode of action that is utilised by

many other inducible genes and is conserved from yeast to mammal (Bai et al., 2010; Barbaric et al., 1992; Bryant et al., 2008; Goriely et al., 2003; Moreira and Holmberg, 1998; Schones et al., 2008; Takahata et al., 2009; Zhang and Boothby, 2006). In the majority of cases, it is a prerequisite that some binding sites for the activator are already exposed prior to chromatin remodelling. Only a few exceptions have been reported where transcription factors are able to engage directly with the nucleosomal DNA sequence (Angelov et al., 2004; Perlmann, 1992; Perlmann and Wrange, 1988). Moreover, the process described above is bidirectional and gene silencing can occur via the re-deposition of nucleosomes over promoter (Adkins et al., 2004). In contrast to these inducible genes, the larger class of constitutively active genes are believed to display more stable chromatin architecture with transcription factor binding sites highly localised within the accessible NFR and therefore less dependent on chromatin remodelling activity.

In vitro gene transcription studies using reconstituted chromatin templates have shown that nucleosomes constitute a significant physical barrier during both initiation and elongation stages of gene transcription. Early studies with adenovirus major late promoter engineered in the presence or absence of nucleosomes have indicated that RNA Pol II cannot initiate transcription when a nucleosome is associated with the promoter (Lorch et al., 1987). Crystal structures of the yeast and human TATA-binding protein (TBP), a key member of the general transcription factors, have revealed that upon association of TBP with the TATA sequence, the DNA double helix is subject to significant distortions and undergoes partial

unwinding, an effect that is likely to be incompatible with nucleosome assembly (Kim et al., 1993a; Kim et al., 1993b; Nikolov et al., 1996). This implicates that RNA Pol II recognition sites need to be either constitutively available by residing inside NFRs as in the case of most housekeeping gene promoters or actively exposed upon signalling as for the class of inducible promoters. In all instances, the presence of RNA Pol II strongly anti-correlates with nucleosome occupancy independently of the expression state of the gene (Schones et al., 2008).

Upon TBP binding to the promoter sequence, a number of proteins, including RNA Pol II, are known to be recruited to form a preinitiation complex (PIC) immediately upstream of the TSS. Due to the position of the +1 nucleosome relative to the TSS in *Drosophila* and mammals, once fired, the elongating RNA Pol II in these organisms need to overcome the +1 nucleosome as it attempts to translocate along the DNA template. Early *in vitro* transcription assays using RNA Pol II purified from HeLa cell extracts have shown that the +1 nucleosome represents a formidable kinetic barrier to transcription elongation (Izban and Luse, 1992) and it is believed to induce promoter-proximal pausing (Hodges et al., 2009; Mavrich et al., 2008b; Schones et al., 2008), a widespread phenomenon that has been observed in both *Drosophila* and human genomes (Core et al., 2008; Guenther et al., 2007; Muse et al., 2007; Nechaev et al., 2010; Zeitlinger et al., 2007). Clearly, for transcription elongation to proceed, DNA-histone contacts must be inevitably disrupted with concomitant transfer or removal of the nucleosome.

Recent biophysical studies monitoring RNA Pol II dynamics during translocation along a nucleosomal template reconstituted *in vitro* favour the model which suggests a transient DNA loop may be formed to physically circumvent the elongating polymerase and mediate transfer of the blocking nucleosome upstream to the polymerase elongating position (Hodges et al., 2009; Studitsky et al., 1994; Studitsky et al., 1997). A wealth of evidence indicates that this process requires the intervention of additional proteins including the transcription elongation factor TFIIS (Kireeva et al., 2005), the histone chaperone FACT (Reinberg and Sims, 2006), the transcription factor c-MYC (Rahl et al., 2010) and the chromatin remodelling complex RSC (Carey et al., 2006).

Together the above studies demonstrate how the physical properties of the nucleosome alone can profoundly affect different stages of gene transcription.

1.3. Histone modifications

The histone code

In addition to the genetic code embedded in the DNA, over the last ten years a hypothesis has been advanced that an epigenetic code may exist, “written” on the tails of the core histones that form the nucleosome (Strahl and Allis, 2000; Turner, 2000). Indeed, the long flexible N-terminal tails of the histones provide an ideal

platform where an impressive array of chemical modifications is known to be harboured, together constituting an important layer of biological information (Figure 1.6) (Kouzarides, 2007). Because such information is found in close proximity to the DNA, it can be utilised to complement the aforementioned physical properties of the nucleosome in directly modulating DNA based processes.

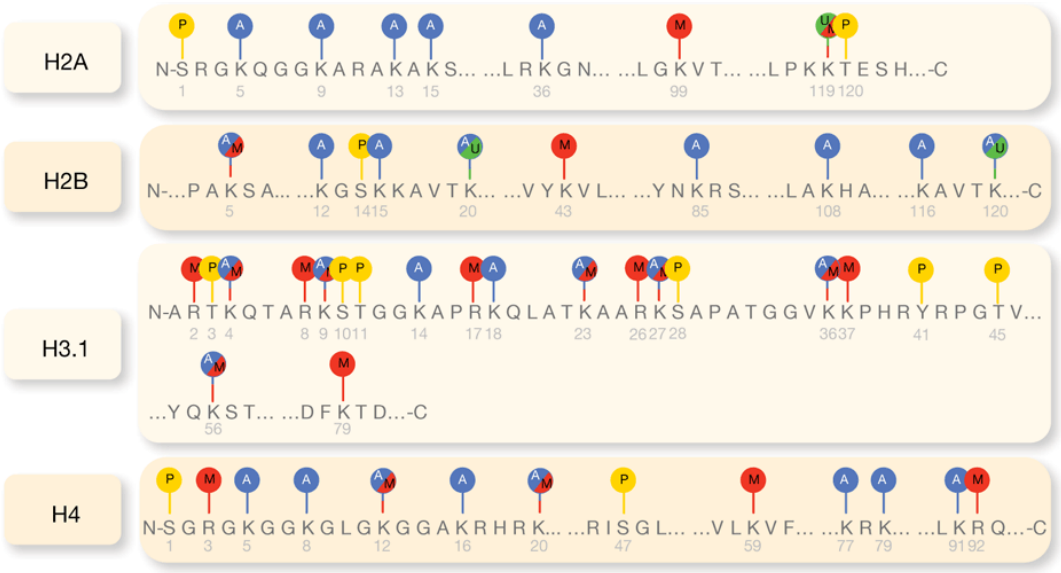


Figure 1.6 – Post-translational modifications of core histones

Diagram depicting the most common modifications present on the four core histones. It is not an exhaustive representation of all known histone modifications to date. P, phosphorylation. A, acetylation. M, methylation. U, ubiquitylation. Figure adapted and reprinted by permission from Macmillan Publishers Ltd: Nature Biotechnology, (Portela and Esteller, 2010), copyright (2010).

As discussed below, there are generally two ways the histone posttranslational modifications (HPTMs) are thought to affect genome function (Figure 1.7) (Allis et

al., 2006). Firstly, they can impact on the overall structure of chromatin by interfering with the histone-DNA contacts within the nucleosome or facilitating changes in the level of chromatin compaction. This can be achieved through charge disruption or steric occlusion. Secondly, they can act as nucleation sites and attract or repel classes of enzymes that have evolved to recognise specific HPTMs or combinations of HPTMs. These effector proteins have the capacity to interpret and translate the information encoded in the HPTMs into biologically relevant downstream functional outputs (Kouzarides, 2007). Interestingly, accumulating research evidence suggests that a single HPTM can be associated with different binding effector proteins that specialise in distinct cellular processes. For example, trimethylation on histone H3 lysine 36 (H3K36me3) in *D. melanogaster* is recognised by the male-specific lethal (MSL) complex that is required for transcriptional up-regulation of the fly male X chromosome (Larschan et al., 2007). In comparison, the human MRG15 protein, which is evolutionarily related to the MSL3 component of the *Drosophila* MSL complex (Marin and Baker, 2000), also binds to H3K36me3 but plays a role in the regulation of alternative pre-mRNA splicing (Luco et al., 2010). It is consequently clear that H3K36 methylation has acquired species-specific roles across evolution. This finding argues against the existence of a true epigenetic code where a given modification is associated to a unique predictable functional outcome. Instead, the function is likely to be dependent on the biological context where the modification is found. Recent research has evidenced functional analogies between the chromatin modifications

and the receptor tyrosine kinase modifications leading to the proposal of a signalling network model as an alternative view to the original histone code hypothesis (Schreiber and Bernstein, 2002).

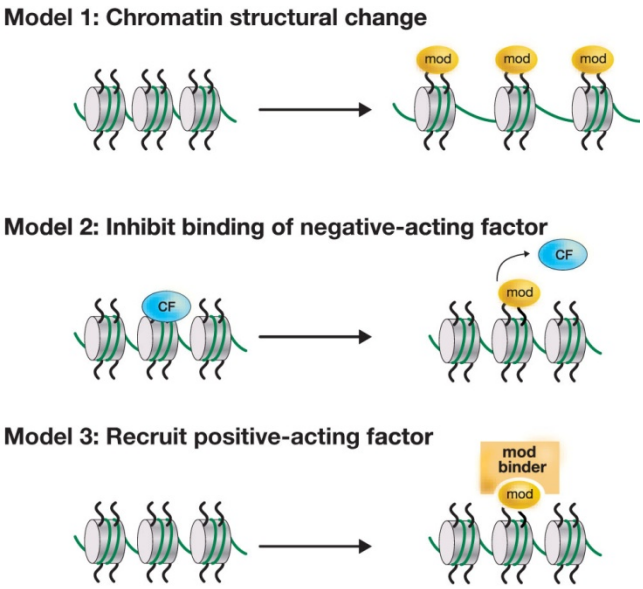


Figure 1.7 – Mechanisms of action for histone modifications in chromatin regulation

Histone modifications (mod) can exert their influence on chromatin state in several ways. Firstly, as in the case of acetylation, they can directly induce change in local chromatin structure (top). Secondly, they can occlude the binding site of specific chromatin-associated factors (CF) (middle). Thirdly, they can favour the recruitment of other chromatin-binders (mod binder) by acting as nucleation sites (bottom). Figure from Allis, C.D., Jenuwein, T., and Reinberg, D. (2006). Epigenetics (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press), reprinted with permission from Cold Spring Harbor Laboratory Press, copyright (2007).

Regardless of whether a strict histone code can be defined, a crucial advantage of having a HPTM system in support of the DNA based genetic code lies in its dynamic nature. Indeed, almost all of the HPTMs known to date are reversible and a

plethora of enzymes that catalyse the addition and removal of these marks has been recently identified. Importantly, patterns of HPTMs are known to be responsive to environmental triggers and can be maintained accordingly to mediate both short-term and long-term, possibly cross-generational, cellular memory. Arguably, the potential to dynamically establish, maintain and erase epigenetic programmes in response to environmental influences over time is one of the greatest strengths of the HPTM system.

The number of HPTMs known has been ever increasing over recent years. They can be broadly separated into large chemical modifications that include ubiquitylation and sumoylation for example, and smaller chemical modifications like acetylation, methylation and phosphorylation.

The functional relevance of histone acetylation and methylation in relation to gene transcription will be discussed further here.

Histone acetylation in transcription regulation

Histone acetylation is known to occur on lysine residues on the N-terminal tails of all four core histones. The presence of an acetyl group has the capacity to counteract the intrinsic positive charge of the lysine side chain and therefore reduce the affinity between histones and the negatively charged DNA molecule (Hong et al., 1993). The charge neutralisation is believed to induce two main

effects. First, *in vitro* binding assays using hyperacetylated nucleosomes have shown that acetylation can cause partial unfolding of the nucleosome structure exposing DNA binding sites previously buried within the nucleosomal core and consequently facilitating interaction by proteins like transcription factors (Lee et al., 1993; Vettese-Dadey et al., 1996). Secondly, chromatin assembly experiments where hyperacetylated histones have been incorporated into nucleosome arrays have demonstrated that histone acetylation antagonises compaction of the chromatin fibre into more condensed, higher order structures (Shogren-Knaak et al., 2006). In both cases, histone acetylation has a clear role in favouring open, transcription-permissive chromatin architecture with increased accessibility of associated DNA elements as evidenced by the high correlation between acetylation and DNase I hypersensitive sites (Hebbes et al., 1994). In accordance with this observation, recent genome wide mapping studies of histone acetylation have revealed that this modification is preferentially enriched at the 5' end of active genes including promoters and TSS in both yeast and higher eukaryotes with a few residue-specific histone acetylation occurring also over the transcribed regions of active genes (Bernstein et al., 2005; Pokholok et al., 2005; Roh et al., 2004; Roh et al., 2005; Wang et al., 2008). In particular, the level of acetylation seems to correlate proportionally with the transcriptional status of the gene with the most active genes being also the most heavily acetylated (Pokholok et al., 2005; Schubeler et al., 2004; Wang et al., 2008).

Another major mechanism by which histone acetylation mediates its transcriptional activation effect is the recruitment of ATP-dependent chromatin remodelling enzymes (Hassan et al., 2002; Horn and Peterson, 2001). Indeed, a number of remodelling complexes are known to contain a specific histone lysine-binding domain known as the bromodomain. Crystal structures of this domain obtained for several histone-associated proteins indicate that it can specifically recognise acetylated lysine residues on the histone tails and such binding occurs mainly through hydrophobic interactions (Dhalluin et al., 1999; Hudson et al., 2000; Jacobson et al., 2000; Owen et al., 2000). Studies using the well characterised yeast *PHO5* promoter have demonstrated that following acetylation of histones by the HAT Gcn5 specifically at promoter region, these acetyl marks can function as targeting sites to attract chromatin remodelers like the Swi/Snf complex. At the human *IFN-β* gene promoter, SWI/SNF recruitment by GCN5 acetylation occurs via the bromodomain of BRG1 subunit. SWI/SNF mediated nucleosome translocation in turn allows for loading onto chromatin of general transcription factors like TFIID and consequent gene activation (Agalioti et al., 2002; Gregory et al., 1998; Reinke and Hörz, 2003; Syntichaki et al., 2000).

Alternatively, studies of another yeast HAT, NuA4, have shown that acetylation at gene promoter can also lead to direct recruitment of TFIID with subsequent PIC assembly (Durant and Pugh, 2007). Importantly, such recruitment is again dependent on the recognition of the acetyl group by a bromodomain-containing subunit known as Bdf1. Interestingly, in addition to TFIID, Bdf1 has also been

shown to interact with the SWR-C complex, a Snf2 family related chromatin remodeler. It therefore reveals a potential pathway that brings together both direct and indirect effects of histone acetylation on the recruitment of general transcription factors like TFIID (Krogan et al., 2003b).

The first histone acetyltransferase (HAT) enzyme that has been shown to mediate the deposition of the acetyl group on histone tails was isolated from the macronucleus of the ciliate *Tetrahymena termophila* (Brownell et al., 1996). To date, several main families of histone acetyltransferases have been identified in eukaryotes including GNAT (GCN5-related N-acetyltransferase), MYST and CBP/p300. All of them function by catalysing the transfer of an acetyl group from acetyl-coenzyme A (acetyl-CoA) to the ε-amino group of lysine side chains but display some degree of overlapping specificity in terms of the specific substrate residue targeted (Sterner and Berger, 2000). Importantly, most HATs like the yeast Gcn5 are able to acetylate nucleosomal substrates only when found in multi-molecular complexes such as the SAGA or ADA complex evidencing the requirement for auxiliary chromatin-modulating factors in aiding their HAT activity and substrate specificity *in vivo* (Grant et al., 1997; Ruiz-Garcia et al., 1997). Moreover, certain HATs are known to be in complex with either the initiating form of RNA Pol II as in the case of the human CBP/p300 or with the elongating polymerase as exemplified by PCAF (Cho et al., 1998). These associations are likely to influence the genomic region targeted by individual HATs and may account for the different distribution of specific acetylated residues over gene promoter versus

coding region. The importance of correct targeting of histone acetylation is evidenced by the fact that aberrant acetylation can lead to erroneous transcription initiation from cryptic promoters within coding regions of genes (Carrozza et al., 2005).

Remarkable sequence identity has been observed in certain HATs like GCN5 between yeast and human homologues in support of an evolutionarily conserved function of histone acetylation in gene regulation across eukaryotes (Xu et al., 1998). Interestingly, most HATs contain functional bromodomains, suggesting that such domain may be used by HATs to interpret and further propagate or reinforce already established sites of acetylated chromatin. Association of the Gcn5 bromodomain with H4K16ac, for example, has been demonstrated to be important for the retention of the SAGA complex on acetylated nucleosomes after the initial recruiting factor has been removed (Hassan et al., 2002).

In parallel to HATs, a class of evolutionarily conserved histone deacetylases (HDACs) are known to counteract histone acetylation. Based on sequence analogy, HDACs are generally divided into three distinct groups (Yang and Seto, 2003). Enzymes of the first group display high homology with the yeast HDAC Rpd3 while members of the second groups are more related to the yeast Hda1 protein and show Zn^{2+} dependent HDAC activity. The third group includes the Sir-2 like proteins, which display NAD^{+} dependent catalytic activity. Similar to HATs, many HDACs function within multimeric enzyme complexes. These associating factors serve to locally

modulate the broad and overlapping substrate specificity displayed by HDACs genome-wide (Shahbazian and Grunstein, 2007).

Histone methylation in transcription regulation

Methylation of histones generally occurs on lysine and arginine residues on all core histone tails (Byvoet et al., 1972; Murray, 1964; Young et al., 2010). Unlike acetylation, up to three methyl groups can be added to one single residue side chain. Together with the large number of methylation sites on histone tails identified so far, the combinatorial coding potential represented by histone methylation is rather impressive and constitutes an ideal system for regulating complex dynamic cellular processes like gene transcription (Jenuwein and Allis, 2001; Lee et al., 2005a; Zhang and Reinberg, 2001). Because the installation of methyl groups does not alter the charge of the modified residue, it is believed that the main function of histone methylation is to provide nucleation sites for the recruitment of chromatin effector molecules that are able to interpret the methylation signal and alter the local chromatin state (Taverna et al., 2007). Indeed, a large number of protein domains have been found to specifically interact with methylated histone lysine or arginine residues including PHD fingers, chromodomains, Tudor domains, PWWP domains, WD40 repeats, CW domains, ankyrin repeats and MBT repeats (Bannister et al., 2001; Collins et al., 2008; Hoppmann et al., 2011; Kim et al., 2006; Lachner et al., 2001; Margueron et al.,

2009; Shi et al., 2006a; Wang et al., 2009; Wysocka et al., 2006; Yang et al., 2010). The diversity of these histone methyl-binding motifs is a reflection of the complexity of the histone methylation code. This is in agreement with the general view that histone methylation on a specific residue with a specific methylation state is interpreted by a particular chromatin reader that is associated with a defined nuclear function in a context dependent manner.

Early biochemical studies of chromatin modifiers in human and mice have identified an evolutionarily well conserved protein domain known as the SET domain that is capable of catalysing the transfer of methyl groups from S-adenosylmethionine (SAM) to the ε-amino group of histone lysine side chains (Rea et al., 2000). The domain was part of the SUV39H1 protein, the first ever characterised mammalian histone lysine methyltransferase (HKMT). Following this initial discovery, an extended family of HKMTs has been found over recent years that share the same catalytic domain. To date, all known HKMTs with the exception of Dot1, are SET domain dependent enzymes (Feng et al., 2002). Another family of protein methyltransferases characterised by the PRMT catalytic subunit is responsible for the deposition of methyl groups on histone arginine residues in eukaryotes (Bedford and Clarke, 2009).

In contrast to HATs, *in vitro* methylation studies have shown that HKMTs display much higher specificity both in terms of the substrate lysine modified and the degree of methylation introduced. Such specific activity is mainly attributable to

the presence of key residues contained within the active site of the enzyme. Depending on the physical property of the residues, these can sterically exclude lysine side chains carrying a certain number of methyl groups. For example, mutational analysis of the HKMT DIM-5 in *Neurospora crassa* has shown that a single Phe to Tyr point mutation is sufficient to alter the product specificity of the enzyme from tri- to mono-methylation (Zhang et al., 2003). Interestingly, unlike the previously mentioned HATs, several HKMTs including the yeast Set2, Dot1 and the human PR-Set7 display enzymatic activity that is selectively targeted to nucleosomal substrates in the absence of any associating factor (Nishioka et al., 2002b; Strahl et al., 2002; van Leeuwen et al., 2002). Because no structure of any SET domain dependent HKMTs in complex with nucleosome has been so far solved, it remains to be determined whether the nucleosome substrate specificity is attributable to the physical architecture of these proteins.

Although global histone methylation level is subject to a low turnover rate, local sites of histone methylation are known to be dynamically regulated (Byvoet et al., 1972; Duerre and Lee, 1974; Zee et al., 2012). Alongside the HKMTs, an ever growing number of histone demethylases that are responsible for the removal of methyl groups from histone tails has been recently the focus of intense research (Klose et al., 2006). So far, two main families of histone lysine demethylases (HKDMs) have been identified (Figure 1.8). These can be distinguished based on their conserved catalytic domains: the KDM1 family characterised by the FAD-dependent amine oxidase domain and the Jumonji C (JmjC) family which features a

Fe(II) and α-ketoglutarate (α-KG) dependent oxygenase domain. Sequence homology comparison has allowed further classification of the JmjC family members into subgroups that include KDM2, PHF, KDM3, KDM4, KDM5 and KDM6 (Klose et al., 2006; Pedersen and Helin, 2010). A combination of *in vitro* activity assay with recombinant histone peptides or nucleosomes and *in vivo* immunofluorescence (IF) imaging coupled with RNA knockdown approaches has allowed the substrate specificity of many of these HKDMs to be identified. Similar to the HKMTS, HKDMs demonstrate high degree of specificity with respect to both the lysine residue targeted and the number of methyl groups that can be removed.

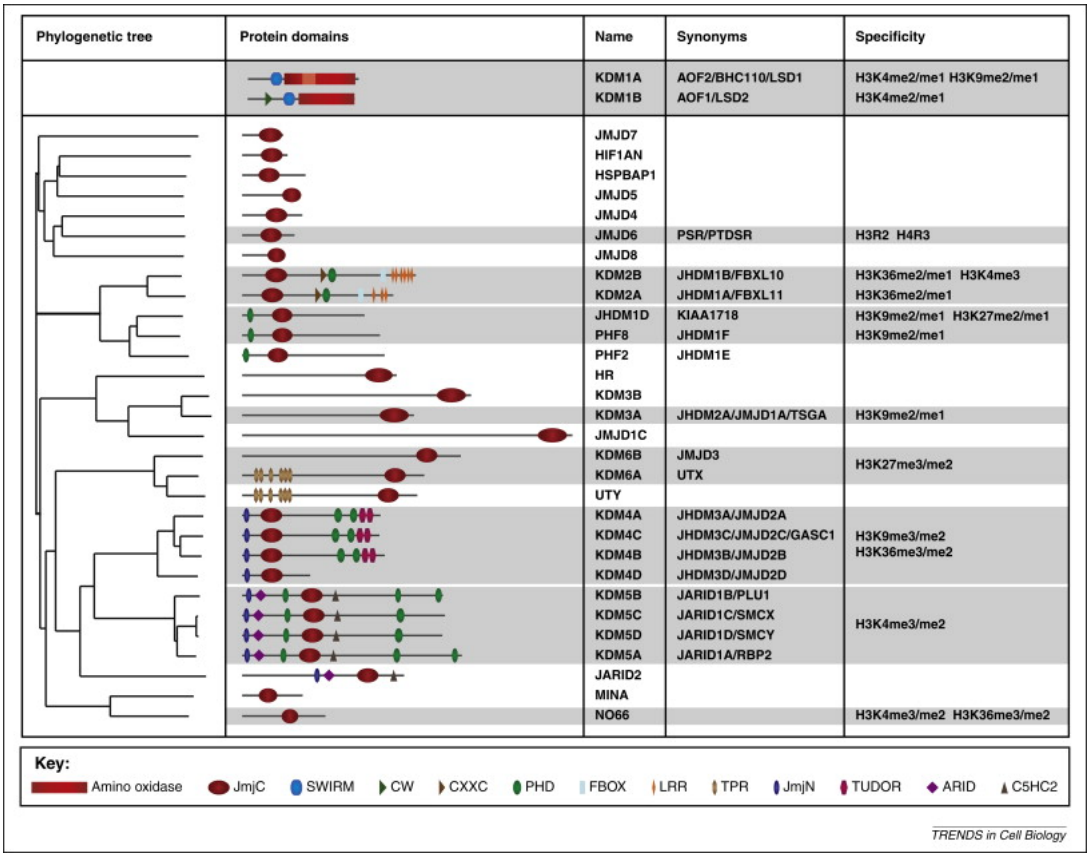


Figure 1.8 – The two families of histone lysine demethylase (HKDM) enzymes

Listed here are the currently known members of the two families of histone lysine demethylases. All conserved domains are shown. Note the Tower domain (light red box) splits the amino oxidase domain (dark red box) of KDM1A in two parts. Reprinted from Trends in Cell Biology *20*, Pedersen, M.T., and Helin, K., Histone demethylases in development and disease, 662-671, Copyright (2010), with permission from Elsevier.

Crystallographic data has shown that several mechanisms exist that ensure the unique substrate specificity of these histone demethylases. For example, biochemical studies have identified in KDM1, which catalyses the removal of H3K4me1 and H3K4me2, a particular protein domain known as the Tower domain. This region of the protein provides a docking site for KDM1 association with a regulator named Co-REST (co-repressor of RE1-silencing transcription factor) (Yang et al., 2006). Such interaction is necessary to confer KDM1 efficient H3K4 demethylase activity on nucleosome substrate *in vivo* and also to provide a means of recruiting KDM1 to Co-REST target genes where removal of H3K4 methylation is required (Lee et al., 2005b; Shi et al., 2004; Shi et al., 2005). Interestingly, chromatin immunoprecipitation (ChIP) experiments indicate that KDM1 can also associate with the nuclear androgen receptor (AR). Following AR binding, KDM1 is readily converted to a H3K9 demethylase that localise to AR responsive genes (Metzger et al., 2005). These results illustrate how the catalytic domain of KDM1 has evolved to accommodate different substrate specificity and the ultimate target of demethylation is strictly dependent on the nature of its associated factor. In contrast to KDM1, the exact conformation of the catalytic core and the amino acid

sequence surrounding the target residue on the histone tail contribute to a less promiscuous substrate specificity of the JmjC family of demethylases. Crystal structures of KDM4A, for example, have shown that due to the shape of the binding pocket surrounding the catalytic site, bending of the preferred H3K9 substrate peptide is necessary to facilitate insertion and contact with the co-factor α-KG. Such requirement implies that specific residues like glycine or proline need to be present at the tip of the bend within the peptide (Chen et al., 2006; Ng et al., 2007).

To date, the most well characterised histone methylation marks implicated in gene transcription regulation are H3K4, H3K36, H3K9, H3K27, H4K20 (Barski et al., 2007). Each of these modifications will be discussed in detail below.

H3K4 methylation

Early ChIP analysis of both chicken β-globin and yeast mating-type loci indicated H3K4 methylation as a hallmark of euchromatin (Litt et al., 2001). Indeed, genome wide H3K4 methylation distribution has been observed to correlate highly with regions of "open" acetylated chromatin domains (Wang et al., 2008). In particular, H3K4me3 has been shown in *S. cerevisiae* to be deposited preferentially at the 5' end of active genes by the HKMT Set1 (Pokholok et al., 2005; Santos-Rosa et al., 2002). ChIP coupled with DNA microarray analysis in human embryonic stem cells

have confirmed that H3K4me3 association with the TSS of actively transcribed genes is conserved in mammals and occurs genome wide (Guenther et al., 2007). Several mechanisms have been uncovered which allow for the precise targeting of Set1 and therefore H3K4me3 at the 5' end of genes. Firstly, co-immunoprecipitation analysis have shown that Set1 physically associates with both the Ser-5 phosphorylated form of the early elongating RNA Pol II and the PAF1 component of the transcription elongation complex (Figure 1.9) (Krogan et al., 2003a; Ng et al., 2003). Interestingly, Set1 does not interact with the Ser-2 phosphorylated form of RNA Pol II, which corresponds to the late phase of transcription elongation (Phatnani and Greenleaf, 2006). Such binding specificity may be responsible for the preferential enrichment of H3K4me3 around gene promoters as opposed to the body of genes. Secondly, ChIP experiments have shown that H2BK123ub1 is required for the proper establishment of H3K4 methylation likely by creating a local chromatin environment that is permissive to Set1 activity (Lee et al., 2007a; Nakanishi et al., 2009; Sun and Allis, 2002). Importantly, H2BK123ub1 is known to be placed by the ubiquitin-conjugating enzyme Rad6, which also interacts with the elongating RNA Pol II via PAF1 (Robzyk et al., 2000; Wood et al., 2003). Thirdly, as discussed in more detail below, recent studies have shown that the mammalian CFP1 component of the SET1 complex can specifically recognise regions of non-methylated CpG DNA whose genome wide distribution highly correlates with the 5' end of genes (Thomson et al., 2010).

The best studied function of H3K4me3 is to favour gene transcription initiation through the coordinated recruitment of specific activators and the concomitant exclusion of repressor elements. Indeed, evidence suggests that in human H3K4me3 is recognised by both the PHD domain of the BPTF component of the NURF chromatin remodelling complex that facilitates nucleosome translocation or eviction at gene promoters and also by the SGF29 protein, part of the SAGA HAT complex (Vermeulen et al., 2010; Wysocka et al., 2006). In addition, reports suggest methylation at H3K4 can preclude docking of the HDAC repressor complex NuRD on the H3 tail at gene promoters (Nishioka et al., 2002a). Finally, structural data has revealed that methylation at lysine 4 residue prevents interaction of the mammalian DNMT3L, a *de novo* DNA methyltransferase-associated factor, with the histone H3 tail (Ooi et al., 2007). It has been therefore suggested that H3K4 methylation may act to inhibit the *de novo* deposition of DNA methylation at promoter regions, a mechanism that, as discussed later, is critical to long-term gene silencing in cells.

Recent genome-wide mapping indicates that H3K4me2 distribution closely resembles that of tri-methylation with peaks observed at the 5' end of genes. In contrast, H3K4me1 preferentially localises at enhancer elements (Heintzman et al., 2007). The genomic function of these two H3K4 methylation states is still poorly characterised.

H3K36 methylation

H3K36 methylation is a chromatin mark that has been associated with different aspects of transcription regulation (Wagner and Carpenter, 2012). It is ubiquitously present in both lower and higher eukaryotic organisms where it has acquired species-specific functions. In *S. cerevisiae*, all three methylation states are deposited by a single H3K36-specific HKMT known as Set2 (Strahl et al., 2002). While H3K36me1 remains poorly characterised, the genomic distributions of di- and trimethylation states are known. H3K36me3 is prevalently enriched over the transcribed region of genes with peaks rising towards the 3' end and levels positively correlating with the transcriptional activity of the gene (Pokholok et al., 2005). The specific distribution of this modification state is attributable to Set2 interaction with the elongating RNA Pol II. Similarly to Set1, Set2 directly associates with the phosphorylated C-terminal domain (CTD) and the transcription elongation factor PAF1 (Krogan et al., 2003c; Xiao et al., 2003). However, unlike Set1, the SRI domain of Set2 preferentially interacts with RNA Pol II phosphorylated at Ser-2, a form that generally accumulates over the body and the 3' end of genes (Figure 1.9) (Bannister et al., 2005; Bell et al., 2007). In contrast, H3K36me2 targeting is independent of the SRI domain and its localisation strictly correlates with regions of active transcription but not with the gene transcriptional rate (Rao et al., 2005; Youdell et al., 2008). These observations have led to speculation that H3K36me2 and me3 may act within distinct biological contexts. Further investigation is required to elucidate the exact nature of these.

To date, many lines of evidence suggest that both H3K36 methylation states contribute to suppression of transcription initiation over the coding region of genes. Indeed, many gene bodies contain cryptic transcription start sites that can be erroneously activated by the wave of acetylation that accompanies the elongating RNA Pol II. Studies in *S. cerevisiae* have shown that H3K36me2 functions as nucleation site that is recognised by the Eaf3 subunit of the Rpd3 HDAC complex which in turn mediates removal of the acetyl groups from nucleosomes in the wake of transiting RNA Pol II (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005; Li et al., 2009a; Reid et al., 2004; Youdell et al., 2008). Yeast proteome-wide peptide array analysis has recently identified Nto1, which is part of the NuA3 HAT complex, as a specific binder of H3K36me3 (Shi et al., 2006b). Whether such recognition also occurs *in vivo* warrants further investigation but this suggests that H3K36me3 may also contribute to transcription activation via histone acetylation hints to the potential complexity of the H3K36 methylation system and, as exemplified below, the need to interpret this modification based on the biological context.

While the yeast Set2 is responsible for the deposition of all three methyl groups on H3K36, in mammals a number of H3K36 methyltransferases have been identified which display preference for a specific methylation state (Wagner and Carpenter, 2012). Numerous *in vitro* assays using recombinant nucleosome or octamer substrates and *in vivo* RNA knockdown studies indicate that the nuclear receptor SET domain containing 1 (NSD1) and 2 (NSD2) proteins can catalyse the addition of

up to two methyl groups on H3K36 while the human orthologue of Set2, known as SETD2, is responsible for tri-methylation (Hu et al., 2010; Kuo et al., 2011; Li et al., 2009b; Lucio-Eterovic et al., 2010; Nimura et al., 2009; Qiao et al., 2010; Yuan et al., 2009). Depletion of NSD3 by RNAi has been shown to lead to a decrease in H3K3me3 (Rahman et al., 2011) but whether this enzyme is specific for mono, di or tri-methylation has not been examined *in vitro* yet. In a similar fashion to the yeast Set2, SETD2 also interacts with the Ser-2 phosphorylated form of RNA Pol II and such binding occurs within a multi-protein complex that also incorporates ISW1 and Spt6 (Yoh et al., 2008). Recent ChIP sequencing data indicates conservation of H3K36me3 distribution exists between *S. cerevisiae* and human, where this modification also correlates with actively transcribed genes and peaks towards their 3' ends (Barski et al., 2007). Interestingly, a more careful inspection of mammalian H3K36me3 distribution has revealed its specific enrichment over exons, in particular at their 5' ends (Hon et al., 2009; Spies et al., 2009). Even more intriguing was the observation that H3K36me3 levels reflect exon usage as this mark is noticeably depleted in the spliced out exons compared to the spliced in ones. Therefore, H3K36me3 appears to play a role in the alternative splicing of RNA transcripts, a phenomenon that is most prevalent in higher eukaryotes. Indeed, a cross talk is thought to be present between splicing activity and H3K36me3 deposition. SETD2 recruitment is directly influenced by the level of splicing (de Almeida et al., 2011; Kim et al., 2011) while MRG15, a known member of the retinoblastoma binding protein 2 (RBP2) H3K4 demethylase complex

(Hayakawa et al., 2007), specifically recognises H3K36me3 via its chromodomain (Zhang et al., 2006), an interaction which in turn leads to the recruitment of the splicing regulator polypirimidine tract-binding protein (PTB) (Luco et al., 2010). Recently, the p52 isoform of PC4 and SF2-interacting protein 1 (Psip1) has also been shown to bind H3K36me3 via a PWWP domain and this is required for the recruitment of splicing factors like SRSF1 (Pradeepa et al., 2012).

The function of H3K36me3 in exon definition can also be observed in worms (Kolasinska-Zwierz et al., 2009). Here, MES-4 (maternal effect sterile 4) and MET-1 (histone methyltransferase like 1), the orthologues of mammalian NSD1 and SETD2 respectively, are known to coordinate H3K36 di- and trimethylation (Andersen and Horvitz, 2007; Bender et al., 2006). Although the role of MET-1 in alternative splicing is poorly characterised, it has been suggested to perform a role similar to the yeast Set2 in preventing inappropriate transcription initiation, specifically in the context of germ line development (Andersen and Horvitz, 2007). Reports indicate that MES-4 is also involved in regulation of germ line-expressed genes although in a trans-generational fashion and independently of RNA Pol II activity (Furuhashi et al., 2010; Rechtsteiner et al., 2010). It has been suggested that MES-4 contributes to the maintenance of a defined expression program in the progeny germ line partially by counteracting Polycomb group-mediated repression (Gaydos et al., 2012).

In *Drosophila*, H3K36 methylation is known to specialise in yet another unique function: dosage compensation. Here, to compensate for the presence of only one X chromosome in males, transcription of the genes on the male X chromosome is up-regulated two-fold with respect to the genes on the autosomes. The fly orthologues of human NSD1 and SETD2 enzymes are known as MES-2 and HYBP respectively (Bell et al., 2007). According to *in vivo* protein-specific siRNA knockdown, both proteins display conserved di- and trimethylation state specificity. Like in yeast, these two methylation states are found in correspondence of transcribed regions although H3K36me2 tends to peak closer to the 5' end of genes as opposed to H3K36me3, which is more abundant near the 3' region. Such divergence is also reflected in their opposing influence with regard to H4K16 acetylation (Bell et al., 2007). Indeed, RNAi experiments indicate that while reduction of H3K36me3 alone leads to an increase in acetylation genome wide, reduction of both H3K36me2 and me3 produces the opposite effect. Interestingly, unlike in *S. cerevisiae*, where alteration of H3K36me2 and me3 is responsible for global changes in histone acetylation, in flies the effect is limited to the H4K16 residue hinting to the possibility that different mechanisms are at play in the two organisms. Recent studies in dosage compensation have helped to clarify the link between H3K36me3 and H4K16 acetylation in *Drosophila*. As mentioned above, the male specific lethal (MSL) complex is recruited to specific binding sites on the fly male X chromosome through a combination of H3K36me3 interaction and recognition of GA-rich sequence motifs (Alekseyenko et al., 2008; Bell et al., 2008).

Once established, MSL mediates H4K16 acetylation through its HAT subunit MOF (Male absent on the first) (Gelbart et al., 2009; Gu et al., 1998). Further spreading of the MSL complex over the X chromosome occurs through binding of H3K36me3 by the chromodomain of the MSL3 subunit, a homologue of the yeast Eaf3 (Bell et al., 2008). The increased H4K16 acetylation in turn favours RNA Pol II elongation through X-linked gene bodies likely by facilitating mobility over a de-compacted chromatin template and therefore leading to transcription up-regulation (Larschan et al., 2011).

Yeast cells with defective Set2 are viable but display abnormal transcription elongation as evidenced by the different sensitivity to the drug 6-azauracil (Kizer et al., 2005). In human, SETD2 is an essential protein, and its mutations have been implicated in sporadic renal cell carcinoma (Duns et al., 2010). Equally, defects in NSD1 and NSD2 have been associated with the Wolf-Hirschhorn syndrome and various forms of cancer (Nimura et al., 2009; Wagner and Carpenter, 2012). Together these disorders highlight the importance of an efficient H3K36 methylation system.

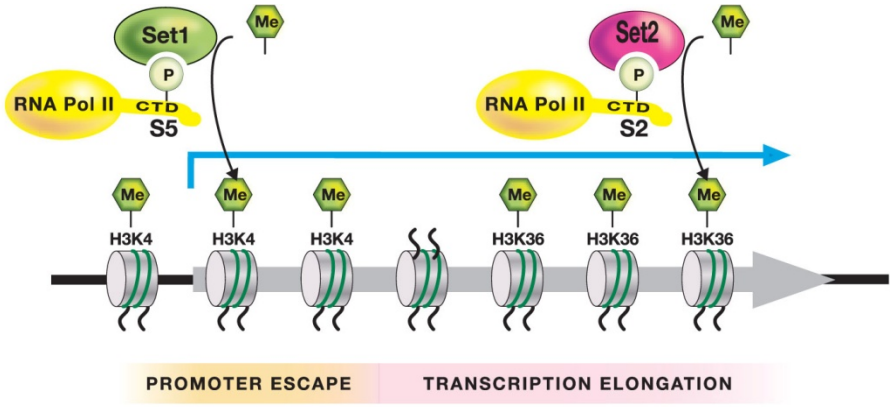


Figure 1.9 – H3K4 and H3K36 methylations regulate gene transcription initiation and elongation respectively

The early elongating Ser-5 phosphorylation on the C-terminal domain (CTD) of RNA Pol II recruits Set1 which mediates H3K4 methylation at gene promoter while the late elongating Ser-2 modified polymerase attract Set2 leading to deposition of H3K36 methylation over the transcribed region of the gene. Figure from Allis, C.D., Jenuwein, T., and Reinberg, D. (2006). Epigenetics (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press), reprinted with permission from Cold Spring Harbor Laboratory Press, copyright (2007).

H3K9 methylation

The first ever characterised HKMT is the mammalian SUV39H1 enzyme, which displays specific catalytic activity towards H3K9 (Rea et al., 2000). IF studies in mouse embryonic fibroblasts (MEFs) have shown that both SUV39H1 and SUV39H2 contribute to H3K9me3 while another HMT, known as G9a, is responsible for mono and di-methylation on the same residue (Rice et al., 2003). In opposite fashion to H3K4me3, which is a characteristic feature of euchromatin domains, H3K9me3 is an important attribute of pericentromeric heterochromatin with a genome-wide distribution that highly correlates with regions of hypoacetylated histones.

The principal mechanism that leads to the establishment of H3K9 methylation and therefore heterochromatin formation is through a RNA interference (RNAi) dependent pathway. Studies focused at characterising the structure of centromeres in fission yeast have shown that these genomic regions contain large amounts of repetitive elements, which are subject to bidirectional transcription (Steiner et al., 1993; Takahashi et al., 1991). These transcripts are processed into short interfering RNAs (siRNAs) by the dicer component of the RNAi machinery and lead to the localisation of the RNAi silencing complex (RISC) to centromeric domains (Volpe et al., 2002). Co-immunoprecipitation experiments demonstrate that RIK1, a WD-domain containing subunit of the Clr4 complex, which is the equivalent of SUV39H1 in *S. pombe*, can physically interact with RISC resulting into the specific targeting of H3K9 methylation at centromeric regions (Zhang et al., 2008). The H3K9 methyl group is in turn interpreted by the chromo domain of heterochromatin protein 1 (HP1), which, as the name suggests, is a heterochromatin-specific protein known to be responsible for the position-effect variegation (PEV) observed in the *Drosophila* genome (Bannister et al., 2001; Eissenberg et al., 1990; Lachner et al., 2001). Once bound, HP1 can attract further H3K9 methylation activity initiating a self-perpetuating mechanism of heterochromatin spreading to neighbouring regions (Hall et al., 2002; Nakayama et al., 2000; Nakayama et al., 2001). Accumulating evidence suggests that the HP1/SUV39H1 complex provides a platform for the downstream recruitment of additional chromatin modifiers including HDACs and chromatin remodelers that

cooperatively induce changes in local chromatin structure and ultimately gene silencing (Sugiyama et al., 2007).

H3K27 methylation

Methylation at H3K27 has been historically associated with gene silencing specifically at a handful of well-studied developmental genes like the HOX genes in *Drosophila*. These genes are known to carry Polycomb group (PcG) responsive elements (PRE) which, as the name indicates, serve to recruit various PcG repressive complexes (Bantignies and Cavalli, 2011). The first two of these complexes to be isolated are the Polycomb repressive complex 1 (PRC1) and 2 (PRC2). They display biochemically distinct components but act in concert to mediate an evolutionarily conserved mechanism of long term gene silencing and critically contribute to the maintenance of established cell lineages during embryogenesis. Not surprisingly, the HMT responsible for the deposition of H3K27 methyl group, known as EZH2, is a core member of the PRC2 complex (Cao et al., 2002). Similarly to H3K9 methylation, H3K27me3 functions as a binding site to attract the chromo domain of the Polycomb (PC) component of PRC1 complex. EM analysis has shown that the presence of PRC1 can in turn alter local chromatin state possibly through direct interaction with nucleosomes (Francis et al., 2004).

In addition to PcG-mediated gene repression, H3K27 methylation has also been implicated in X chromosome inactivation and tissue-specific imprinting (Plath et al., 2003; Yamasaki-Ishizaki et al., 2007).

H4K20 methylation

Extensive research has been carried out in recent years in understanding the contribution of H4K20 methylation to transcription regulation. This modification is catalysed by PR-Set7, the main histone methyltransferase responsible for the deposition of H4K20me1 and SUV4-20H1 and SUV4-20H2, which mediate conversion to di and tri-methylation states from PR-Set7 modified histone H4 tail (Nishioka et al., 2002b; Yang et al., 2008).

Early studies have demonstrated the exclusive localisation of H4K20me1 at transcriptionally silent regions in *Drosophila* polytene chromosomes in support of a function of this histone modification in gene repression possibly by influencing chromatin compaction (Lu et al., 2008; Nishioka et al., 2002b). Evidence suggests that the human transcription repressor L3MBTL1 can induce chromatin compaction in a H4K20 mono and di-methylation-dependent manner (Trojer et al., 2007). In support of its association with gene repression, H4K20me1 is also enriched on the inactive X chromosome in mice (Kohlmaier et al., 2004). However, recent genome-wide mapping of H4K20me1 in human T cells has revealed a high

correlation of this mark with sites of active transcription although a mechanistic link between H4K20 methylation and gene activity is lacking (Barski et al., 2007).

These apparent conflicting results may be partly attributable to the multifunctional role played by PR-Set7. Indeed, levels of this enzyme and consequently H4K20me1 have been observed to dynamically shift across phases of the cell cycle and such behaviour is believed to be important not only for gene expression but a number of other key cellular events including cell cycle progression, DNA replication and DNA damage response (Beck et al., 2012; Oda et al., 2010). All these activities can lead to indirect effects on specific local gene expression state. It is possible that H4K20 methylation may have a general global transcription facilitator role while still retaining defined local repressive function.

1.4. DNA methylation

DNA methylation in higher eukaryotes

Similar to histones, DNA can be enzymatically modified in the form of methylation, specifically on the adenine and cytosine bases. Firstly identified in *Mycobacterium tuberculosis* (Johnson and Coghill, 1925), DNA methylation is a widespread genomic feature that can be found in all three domains of life: bacteria, archea and eukaryotes (Iyer et al., 2011). However, a few organisms exist where DNA

methylation is absent or detectable at very low levels (Antequera et al., 1984; Gowher et al., 2000; Lyko et al., 2000; Proffitt et al., 1984; Simpson et al., 1986). In those species whose genomes are subject to DNA methylation, the relative abundance and functional contribution of this modification to genome regulation varies. In this thesis, the biological importance of DNA methylation in mammalian genomes will be discussed.

In contrast to invertebrates which generally display mosaic pattern of DNA methylation, genomes of vertebrate organisms are subject to extensive methylation (Tweedie et al., 1997), with methylated bases representing about 1% of the total DNA content, the highest level observed in the animal kingdom (Ehrlich et al., 1982). Early comparison studies of animal DNA from different taxonomic groups have revealed that this covalent modification occurs almost exclusively on the cytosine base in mammals (Vanyushin et al., 1970) and predominantly in the context of the dinucleotide sequence CpG on both strands of the DNA helix. This has been evidenced by recent single-base resolution mapping of DNA methylation by sodium bisulphite sequencing (BS-Seq) in both human embryonic stem cells and foetal lung fibroblasts (Lister et al., 2009).

A class of evolutionarily conserved DNA methyltransferase enzymes specialise at depositing CpG methylation genome wide. Early investigation using retroviral genomes injected into mouse pre-implantation embryos has shown that methylation is established during the early stages of embryogenesis (Jahner et al.,

1982) by the *de novo* DNA methyltransferases DNMT3a and DNMT3b (Okano et al., 1998). Indeed, mouse embryonic stem (ES) cells lacking both DNMT3a and DNMT3b are unable to methylate the transfected proviral DNA and a number of genomic elements including imprinted genes such as *Igf2*, the promoter of the X-inactive specific transcript (*Xist*) gene and centromeric satellite repeats (Okano et al., 1999). Inactivation of DNMT3a in mice leads to postnatal death while loss of DNMT3b or both is early embryonic lethal suggesting the two enzymes are likely to have distinct genomic targets and functions during embryonic development. Moreover, another member of the DNMT3 family known as DNMT3L (Aapola et al., 2001), which lacks the conserved methyltransferase motifs, has been found to be required for enhancing *de novo* methyltransferase activity through direct physical association with DNMT3a and DNMT3b in human 293T cells (Chedin et al., 2002; Suetake et al., 2004). Crystallographic analysis indicates that DNMT3L can modulate DNMT3a activity by forming a tetrameric complex consisting of two active sites separated about one DNA helical turn apart. *In vitro* DNA methylation assay has revealed that such fixed distance between the two catalytic sites may be responsible for the 8-10 bp periodic distribution of methylated CpGs observed across the genome (Ferguson-Smith and Greally, 2007; Jia et al., 2007; Jurkowska et al., 2008). Due to the extent of DNA contacts, the oligomeric enzyme can only methylate its substrate in a non-processive fashion (Gowher and Jeltsch, 2001; Jia et al., 2007). Genetic analysis in mice has shown that deletion of the DNMT3L gene severely affects male germ cell development and leads to azoospermia. The

presence of DNMT3L is also required for the correct establishment of DNA methylation over maternally imprinted loci and retrotransposon elements (Bourc'his and Bestor, 2004; Bourc'his et al., 2001).

Sucrose gradient sedimentation assays have shown that the majority of DNMT3a and DNMT3b proteins are found tightly associated with nucleosomes within regions of methylated CpG rich sequences and repetitive elements *in vivo* (Jeong et al., 2009). In fitting with this observation, recombinant DNMT3a can efficiently methylate DNA in a nucleosomal context (Gowher et al., 2005). Additional biochemical studies have demonstrated that recognition of nucleosomes by DNMT3a and DNMT3b is likely to be mediated by both PWWP and PHF-like ADD domains, contained within the N-terminal portion of the protein, interacting with H3K36me3 and H3K4me0 respectively (Dhayalan et al., 2010; Zhang et al., 2010). Recent crystal studies have also evidenced the capacity of the DNMT3L subunit to specifically associate with the non-methylated form of H3K4 residue (Ooi et al., 2007).

Investigation of global methylation patterns in mice embryos indicates that DNA methylation by the *de novo* methyltransferases is dynamically regulated during early development (Feng et al., 2010). Interestingly, immunofluorescence imaging of one cell embryo has shown that immediately after fertilization both paternal and maternal genomes first undergo active and passive demethylation respectively (Mayer et al., 2000). This wave of demethylation ensures the erasure of parental

epigenetic programs and contributes to the totipotent nature of the zygote. Such genome wide reduction in DNA methylation is allowed to proceed until the blastocyst stage, following which the action of *de novo* methyltransferases promotes the gradual restoration of methylation level to that observed at birth (Monk et al., 1987).

In contrast to germ cells, which undergo a second wave of global demethylation, active removal of DNA methylation in somatic cells is rare and strictly limited to specific cell types and gene loci. Indeed, the established pattern of DNA methylation is faithfully maintained in somatic cells following each DNA replication by the maintenance DNA methyltransferase DNMT1 (Li et al., 1992) in a semi-conservative fashion (Bestor and Ingram, 1983; Song et al., 2010). Biochemical studies indicate that DNMT1 can physically interact with the DNA polymerase clamp loader complex (PCNA) (Chuang et al., 1997). Such association allows DNMT1 to track alongside the DNA replication machinery and catalyse the addition of methyl groups onto the nascent chain of hemi-methylated DNA originating from the replication fork. In fitting with the maintenance function, *in vitro* binding assay has demonstrated a marked preference of DNMT1 for hemi-methylated DNA as opposed to the non-methylated substrate (Yoder et al., 1997). Recent structure data of DNMT1 suggests that the presence of a zinc finger CXXC domain, a DNA binding module that has been shown in other chromatin modifiers to specifically interact with non-methylated CpG dinucleotides (Voo et al., 2000), can sterically preclude insertion of DNA into the catalytic pocket when DNMT1 is in contact with

non-methylated CpGs (Song et al., 2010). The stability and heritability of DNA methylation through cell generations makes it an important epigenetic system of cellular memory.

The major function of DNA methylation is transcriptional repression (Bird and Wolffe, 1999) whether in the context of X-inactivation (Panning and Jaenisch, 1996; Wolf et al., 1984), genomic imprinting (Li et al., 1993), retrotransposon silencing (Bourc'his and Bestor, 2004), chromosome stability (Dodge, 2004; Gaudet, 2003) or tissue-specific gene expression (Bruniquel and Schwartz, 2003). There are two known mechanisms that account for the effect of DNA methylation on gene expression. Firstly, the presence of a methyl group in the major groove of the DNA helix can prevent binding of a number of transcription activators that recognise CpG containing sequences at promoter regions (Bell and Felsenfeld, 2000; Iguchi-Ariga and Schaffner, 1989; Mancini et al., 1999; Renda et al., 2007; Watt and Molloy, 1988; Zhu et al., 2003). Secondly, methylated DNA can recruit a class of methyl-CpG binding domain (MBD) proteins (Defossez and Stancheva, 2011) that can target repressor complexes to methylated gene promoters and mediate gene silencing (Jones et al., 1998; Nan et al., 1998).

Aberrant DNA methylation system is associated with a number of disorders (Amir et al., 1999; Ehrlich, 2003) and many forms of cancer (Kulis and Esteller, 2010), highlighting the importance of DNA methylation for normal genome function.

CpG islands

Although DNA methylation is widespread across the genome, its distribution is not uniform. Methyl-CpGs (meCpGs) prevalently accumulate over repetitive and transposon elements as well as exons, introns and 3' UTR of genes but are significantly depleted at most transcription start sites (TSS) and CpG islands (CGIs) (Figure 1.10) (Lister et al., 2009). It is estimated that around 70% of mammalian CpGs are methylated (Ehrlich et al., 1982). The remaining non-methylated fraction of CpGs are clustered into CGIs, segments of the genome historically isolated by cleavage using the methylation-sensitive HpaII restriction enzyme and therefore initially named as "HpaII tiny fragments" (Bird et al., 1985). These constitute around 1% of the full genome, have an average length of 1 kb and display amorphous DNA sequences, except for a fraction containing repeat elements (e.g. Alu repeats) (Lander et al., 2001).

According to a commonly accepted definition of CGIs (Gardiner-Garden and Frommer, 1987), these genomic regions are characterised by ≥ 50% GC content and ≥ 0.6 observed over expected (obs/exp) ratio of CpG frequency. In comparison, the average genome GC content is around 41% with an obs/exp CpG ratio of only 0.2 (Josse et al., 1961; Lander et al., 2001).

Such base composition discrepancy between CGIs and the bulk of the genome is believed to be a direct consequence of the innate resistance of CGIs to global DNA methylation (Cooper et al., 1983). Indeed 5-methylcytosine is known to mutate

spontaneously to thymine via deamination resulting into a C to T mutation that cannot be efficiently corrected despite the existence of a dedicated repair system (Hendrich et al., 1999; Millar, 2002). The mutability of 5-methylcytosine has been proposed as the origin of CpG deficiency in the non-CGI fraction of the genome over evolutionary time (Bird, 1980).

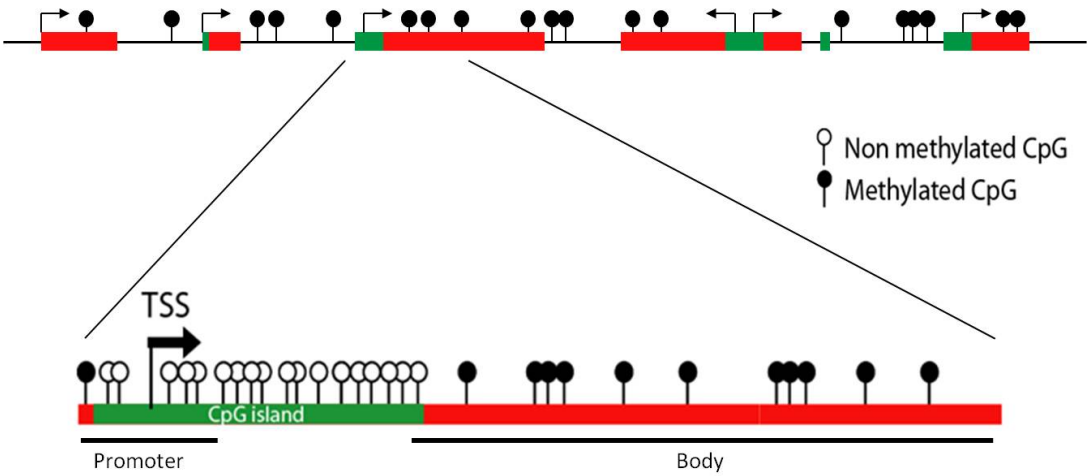


Figure 1.10 – CpG islands

CpG islands (CGIs) are clusters of non-methylated CpGs with higher than genome average CpG density and GC content. They are predominantly associated with the 5' end of most human genes.

A major limitation of the algorithms that have been developed to bioinformatically predict CGI distribution across different genomes is the inevitable dependence on the arbitrary nature of the threshold selected and, importantly, the fact that they do not take into consideration the methylation status of predicted CGI sequences.

As a consequence, initial discordance arose when CGI predictions were compared between the mouse and the human genomes. More recently a biochemical approach using affinity purification via CXXC protein domain was developed which takes advantage of the non-methylated status of CGIs (Illingworth et al., 2008). By applying this new method, a striking conservation of CGI patterning between humans and mice was revealed (Illingworth et al., 2010). Since the emergence of genome-wide DNA methylation is believed to coincide with the invertebrate to vertebrate transition in evolution (Tweedie et al., 1997), the resulting appearance of CGI regions that are refractory to DNA methylation is necessarily an event specific to vertebrate genomes. Moreover, the high degree of conservation of CGIs between different species suggests an important role in vertebrate genome function.

Although the functional significance of CGI in gene regulation is not well understood, their almost exclusive association with promoter elements, as revealed by recent mapping studies (Illingworth et al., 2010; Maunakea et al., 2010), strongly supports a role in transcription initiation. Indeed at least 70% of mammalian genes carry a CGI containing promoter (Saxonov, 2006), including almost all house-keeping genes as well as many tissue-specific genes (Larsen et al., 1992; Zhu et al., 2008). It is estimated that around 50% of CGIs mapped in the mouse and human genomes are associated with currently annotated TSSs with the remaining fraction, denoted as “orphan” CGIs, distributed across intragenic and intergenic regions (Illingworth et al., 2010). To date, RNA sequencing data indicates

that around 40% of these orphan CGIs are associated with novel transcription initiation sites although this proportion is expected to increase significantly as more sequencing results are collected. It has been suggested that most if not all orphan CGIs represent alternative promoters associated with nearby annotated genes or sites of non-coding RNA transcript initiation (Illingworth et al., 2010; Maunakea et al., 2010).

That all CGIs almost unequivocally correspond to some form of transcription initiation is in agreement with the observation that chromatin at these promoters is enriched in trimethylation of lysine 4 on histone H3 (H3K4me3), a mark of transcription initiation (Guenther et al., 2007). Indeed, presence of RNA Pol II at these sites as well as nascent RNA transcripts can be observed (Core et al., 2008). Conversely, recent focus on the function of histone H3 lysine 36 dimethylation (H3K36me2), a modification that contributes to transcription repression in yeast (Li et al., 2009a; Strahl et al., 2002; Youdell et al., 2008), reveals that this modification is specifically depleted at CGIs in keeping with the idea of CGI promoting a chromatin environment that is permissive to transcription initiation (Blackledge et al., 2010; Zhou et al., 2011).

Genome-wide profiling data suggests that CGIs present a unique chromatin architecture whereby the CG content and the length of CGI positively correlates with the degree of nucleosome depletion observed at NFRs (Fenouil et al., 2012). Such nucleosome deficiency could be partially due to an active exchange process in

which canonical nucleosomes are replaced with less stable variant forms (Jin et al., 2009). These features are accompanied by highly phased nucleosomes immediately adjacent to the NFR and overlapping DNase I hypersensitivity, a general indicator of open chromatin structure (Boyle et al., 2008; Fenouil et al., 2012). Importantly the above stated characteristics are independent of any productive transcription elongation events (Core et al., 2008; Guenther et al., 2007).

Such promoter architecture generally remains unperturbed throughout cellular differentiation. This is especially valid for those CGI promoters associated with ubiquitously expressed housekeeping genes (Meissner et al., 2008) (Figure 1.11). However, studies in mouse ES cells have evidenced the existence of a class of bivalent CGI genes where activating H3K4me3 and repressive H3K27me3 chromatin marks coexist in the same promoter (Bernstein et al., 2006). This type of promoters belongs to the unique group of highly regulated developmental genes (e.g. some of the HOX loci) that are transcriptionally poised in pluripotent cells but become either fully activated or repressed upon cellular differentiation. This is concomitant to the erasure of one of the two modifications (Bernstein et al., 2006). Moreover, in case of permanent silencing of the gene (i.e. in the absence of PcG-mediated repression), DNA hypermethylation of the CGI promoter can be observed (Meissner et al., 2008). Interestingly, at these genes, there seem to be an antagonistic relationship between the presence of H3K4me2 and the level of DNA methylation suggesting that this particular modification may have a role in

protecting CGIs from *de novo* methylation (Meissner et al., 2008; Mohn et al., 2008; Weber et al., 2007).

In addition to the above two classes of genes, a minority of CGI promoters is associated with tissue-specific genes that acquire DNA methylation during the transition from pluripotency to terminally differentiated cell types (e.g. neurons) as a result of normal developmental program (Mohn et al., 2008) (Figure 1.11). Additionally, a subset of CGI gene promoters are only expressed in germ lines (e.g. the testis-specific MAGE family of genes) while subject to methylation in somatic tissues (De Smet et al., 1999).

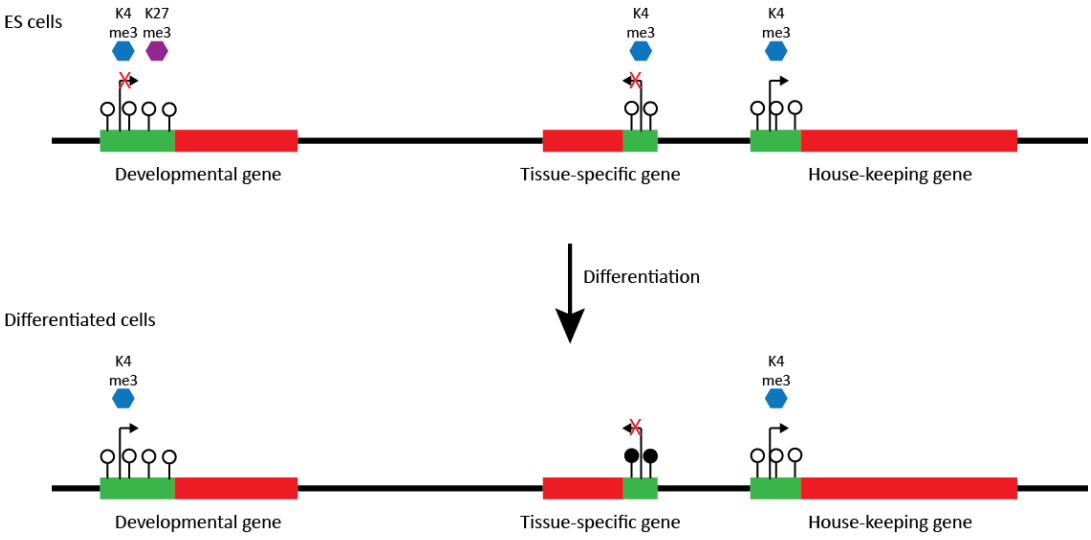


Figure 1.11 – CGI promoters during normal development

Diagram illustrating changes in histone and DNA methylations in three different types of CGI promoter-associated gene during normal development. Shown on the left, many highly regulated developmental CGI gene promoters are non-methylated and display co-localisation of both H3K4me3 and H3K27me3 chromatin

marks in ES cells (Bernstein et al., 2006). These genes are believed to be temporarily silenced through PcG-mediated repression. Upon differentiation, depending on the developmental lineage pathway, either H3K4me3 or H3K27me3 is removed on each promoter and the gene becomes active or repressed accordingly. If active, the gene promoter remains devoid of DNA methylation. If permanently repressed, *de novo* DNA methylation with concomitant loss of H3K27me3 occurs (Meissner et al., 2008). Shown in the middle, a minority of CGI promoters belong to tissue-specific genes. These promoters are non-methylated, also feature H3K4me3 but are predominantly silent in ES cells (Mohn et al., 2008). The presence of H3K4me3 strongly anti-correlates with DNA methylation and vice versa. Indeed, upon differentiation, most of these promoters lose H3K4me3, acquire DNA methylation and become permanently silenced. Only a handful instead become active in specialised tissue types (e.g. neurons). In such cases, an increase in H3K4me3 is observed (Meissner et al., 2008). Shown on the right, the vast majority of CGI promoters are associated with housekeeping genes. These constitutively display H3K4me3, absence of DNA methylation and are ubiquitously expressed in both pluripotent and differentiated cell types. Green and red boxes represent CGI promoter and gene body regions respectively. Open and closed lollipops represent non-methylated and methylated CpGs respectively. Blue and purple hexagonals represent H3K4me3 and H3K27me3 respectively. Vertical bar indicates the position of the TSS and the arrow the direction of transcription. Red X over the arrow indicates that transcription is silenced.

It is unclear how CGIs in general remain refractory to the wave of global DNA methylation during embryogenesis. Several hypotheses have been proposed. Given their strong correlation with transcription, it is possible that transcription factors or the transcription machinery itself may act to exclude DNA methylation from these regions (Straussman et al., 2009). It has also been suggested that the H3K4me3 in CGI chromatin could interfere with binding of DNMT3L, a regulatory component of the *de novo* DNA methyltransferase system, to its recognition sites (Ooi et al., 2007). Alternatively, CGIs could be depleted of DNA methylation through an active, targeted demethylation process although the molecular

mechanism involved is still undefined (Wu and Zhang, 2010). Recent efforts in understanding a novel DNA modification, have pointed at 5-hydroxymethylation by the TET1 enzyme (Tahiliani et al., 2009) as a new attractive route to remove DNA methylation (Guo et al., 2011) at CGIs, where TET1 is found to preferentially localise (Ficz et al., 2011; Pastor et al., 2011; Williams et al., 2011; Wu et al., 2011).

1.5. Interpretation of DNA methylation

Methyl-CpG binding proteins

Early work aimed at elucidating the molecular mechanism that gives rise to DNA methylation induced silencing effect has identified the methyl-CpG binding protein MeCP2, a previously known Rett syndrome associated factor, as a critical translator of the repressive DNA methylation signal (Boyes and Bird, 1991). Further sequence comparison analysis has allowed the discovery of a number of DNA binding proteins that are related to MeCP2 and share the same methyl-CpG-binding domain (MBD) (Clouaire and Stancheva, 2008). Indeed, with the exception of the Kaiso family, all currently known methyl-CpG-binding proteins are MBD domain-dependent. Despite conservation of the DNA binding module, evidence suggests that each MBD protein is likely to be targeted independently to distinct, non-overlapping genomic regions. For example, recent biochemical characterisation of

MeCP2 DNA binding specificity has shown a clear preference of its MBD domain for methyl-CpGs that are surrounded by AT-rich sequences (Klose et al., 2005).

Understanding of how recognition of methylated DNA by MBD proteins can result into transcription repression came from biochemical fractionation analysis of MeCP2 associated complex in *Xenopus* oocyte extract. It was found that MeCP2 co-purifies with the HDAC Sin3A and such interaction is important for MeCP2-mediated repression effect (Figure 1.12) (Nan et al., 1998). Indeed, the presence of HDACs and other repressive HKMTs can induce the formation of a permanently silenced promoter environment. Additional work has shown that other members of the family like MBD2 and MBD3 also associates with the NuRD chromatin remodelling and deacetylase complex (Wade et al., 1999; Zhang et al., 1999). Therefore, it seems that recruitment of HDAC corepressor elements represents a unified mechanism for MBD protein-dependent transcriptional repression.

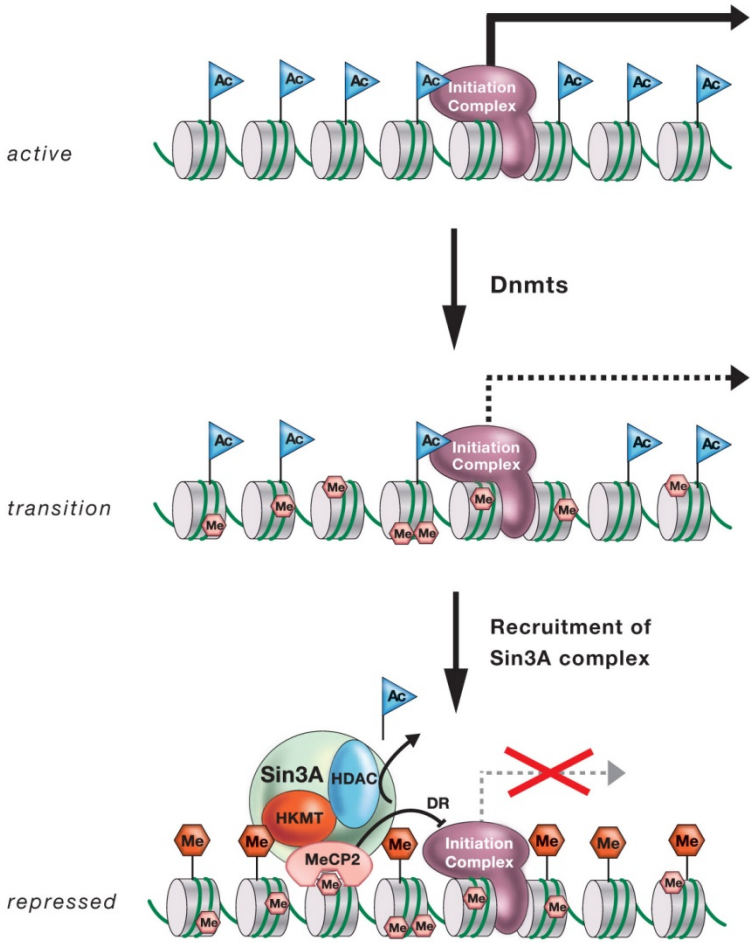


Figure 1.12 – Mechanism of methyl-CpG binding proteins mediated repression

An initially active gene promoter (top) is subject to *de novo* DNA methylation (middle) causing transcriptional silencing. The methyl-CpGs recruit the mediator MeCP2, which in turn promotes the local assembly of the Sin3A histone deacetylase complex (HDAC) and specific histone lysine methyltransferases (HKMTs). Removal of acetylation and installation of repressive methylation marks together contribute to a permanently silenced chromatin state. Figure from Allis, C.D., Jenuwein, T., and Reinberg, D. (2006). Epigenetics (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press), reprinted with permission from Cold Spring Harbor Laboratory Press, copyright (2007).

Non-methyl-CpG binding proteins

In similar fashion to methyl-CpG targeted gene silencing, lack of DNA methylation favours the recruitment of a different set of nuclear proteins that share a common DNA binding module known as the zinc finger CXXC domain.

The CXXC domain is a cysteine-rich protein domain found to be highly conserved through a numbered of DNA binding proteins such as DNMT1, MLL1, MBD1, CFP1, KDM2A/B and TET1 (Bestor et al., 1988; Blackledge et al., 2010; Cross et al., 1997; Frauer et al., 2011; Ma et al., 1993; Voo et al., 2000). Structural data indicate that each three of the six highly conserved cysteine residues within the CXXC domain constitute a particular sequence motif, CGXCXXC, which is able to coordinate a single zinc ion. Together the two motifs dictate the folding of the protein domain and contribute to its overall crescent-like structure. The domain was first shown in CFP1 to bind preferentially to non-methylated CpG dinucleotides *in vitro* (Voo et al., 2000). The molecular details of this specific interaction emerged from recent structural studies of the MLL1 CXXC domain in complex with DNA (Figure 1.13) (Allen et al., 2006; Cierpicki et al., 2009). It was clear that the CXXC domain carries a short extended loop that is able to insert into the major groove of the DNA helix and mediate CG base pair recognition. This is achieved through both protein side chain and backbone interactions, the latter playing a critical role in allowing the protein to discriminate between methylated and non-methylated CG states. In addition to base specific recognition, several additional residues are involved in

electrostatic and hydrophobic interactions with the minor groove of DNA contributing to the overall stability of the binding complex.

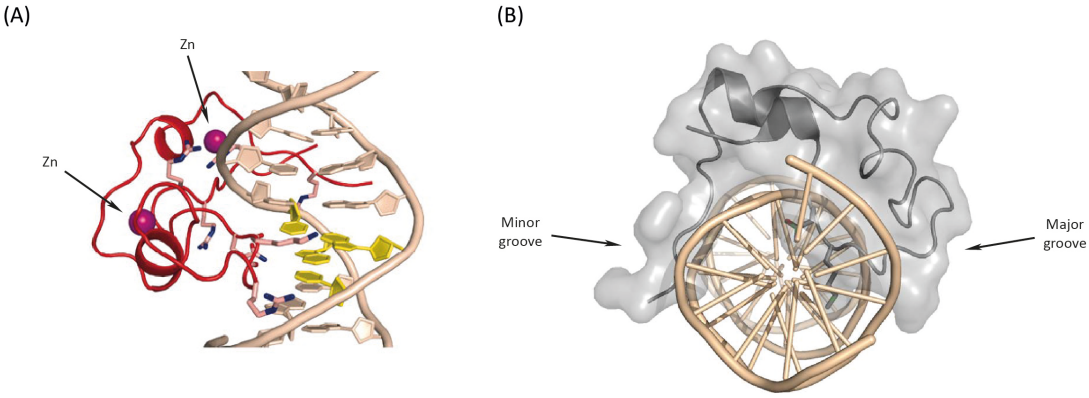


Figure 1.13 – Crystal structure of the zinc finger CXXC domain in DNMT1 and CFP1

(A) DNMT1 CXXC domain (red) in contact with a portion of DNA (pink). The view is perpendicular to the DNA helix. Arrows highlight the two zinc ions (purple) that are coordinated by the cysteines within the CXXC domain. Note the loop from the domain inserting into the major groove of the DNA helix and mediating recognition of the CG base pairs (yellow). (B) CFP1 CXXC domain in complex with DNA viewed looking down the DNA helix. Note both major and minor grooves of the DNA helix are in contact with the CXXC domain. Figure from (Song et al., 2010) Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2010). Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation. Science *331*, 1036-1040. Reprinted with permission from AAAS.

Given the specificity of the CXXC domain-DNA interaction characterised *in vitro*, it was hypothesised that the CXXC domain could mediate protein localisation to non-methylated CG-rich sequences of the genome, namely CGIs. Indeed, a functional study of CXXC3 in MBD1 first confirmed the ability of this domain to act as a targeting module to non-methylated CpGs *in vivo* (Jorgensen et al., 2004). Following this early observation, recent genome wide map of CFP1 localisation in

mice further supported the idea that the CXXC domain may represent a unified mechanism for the recruitment of several chromatin modifiers to CGIs (Thomson et al., 2010).

1.6. KDM2

It is worth mentioning that among the class of CXXC domain containing proteins, two distinct groups can be found, namely the MLL and the KDM2 families, which are known to be associated with H3K4 and H3K36 methylation systems respectively. As evidenced previously, both modifications play important roles in the regulation of gene transcription. While accumulating research over the years has contributed to the appreciation of the biological function of MLL1, which belongs to the historically well characterised group of Trithorax proteins (Schuettengruber et al., 2011), relatively little is known regarding the genomic role of the two KDM2 proteins (KDM2A and KDM2B), the founding members of the only recently discovered family of JmjC domain containing histone lysine demethylase enzymes (Tsukada et al., 2005).

KDM2A was originally identified from a human T cell cDNA library through yeast two-hybrid screen and bioinformatically from the public GenBank sequence database (Cenciarelli et al., 1999; Winston et al., 1999). Because the protein incorporates a C-terminal F-box domain and adjacent leucine-rich repeats, KDM2A

was initially named as FBXL11 (F-box and leucine-rich repeat protein 11) and characterised as a member of the F-box family of proteins that associate with the Skp1 component of the ubiquitin ligase complex SCF (Skp-Cullin-F-box).

The F-box motif was first discovered in the cell cycle regulator cyclin F as a 40 amino acid (aa) sequence that mediates Skp1-specific interaction (Bai et al., 1996). Historically, F-box-containing proteins are known to form integral components of the SCF complex and are responsible for the ubiquitylation of cellular proteins destined for degradation via the proteasome pathway (Ho et al., 2008). However, it was not until recently that FBXL11 was rediscovered as the pioneering member of a novel class of HKDMs (Tsukada et al., 2005). In contrast to the family of KDM1 enzymes, which are FAD-dependent amine oxidases, FBXL11 features a previously poorly characterised JmjC catalytic core domain. In view of its newly uncovered function, FBXL11 was renamed as KDM2A (Lysine demethylase 2A), together with its paralogue KDM2B, the other member of the KDM2 family.

The demethylase activity of KDM2A was originally detected in a single fraction of crude HeLa cell nuclear extract using an *in vitro* formaldehyde release assay (Tsukada et al., 2005). The assay was designed to search for enzymatic reactions that are chemically similar to that catalysed by the family of bacterial DNA demethylases AlkB (Figure 3.4A) (Trewick et al., 2002). Initial protein chromatography steps were able to assign the isolated activity to two protein bands visualised on silver-stained gel. Subsequent peptide mass fingerprint

analysis revealed the upper protein band to correspond to FBXL11. Further deletion experiments of this protein allowed mapping of the specific activity to its N-terminal JmjC domain, a previously bioinformatically identified metallo-binding domain which was believed to be related to the ancestral cupin superfamily of metalloenzymes (Clissold and Ponting, 2001). Because of their close homology, JmjC domain-containing proteins were predicted to carry metal ion-dependent enzymatic activity (Cikala et al., 2004; Trewick et al., 2005). Indeed, the activity of all JmjC family of demethylase enzymes known to date depend on the presence of Fe(II) as a cofactor in the active site.

Recent structural data has revealed that the JmjC domain of KDM2A carries the typical His-X-Asp/Glu-X_n-His motif, found also in other JmjC family of enzymes. In this motif, the charged side chains of these three highly conserved amino acids act together to coordinate one Fe(II) ion. This binding is further stabilised by the close presence of one α-KG molecule, which chelates Fe(II) through its C-1 carboxylate and C-2 ketone (Han et al., 2007; Hausinger, 2004).

The exact chemical reaction that allows for the JmjC-mediated removal of methyl groups from histone tail is essentially the same as that catalysed by the DNA demethylase AlkB, well known for its role in DNA repair of alkylation damage in *E. coli* (Trewick et al., 2002). The reaction involves initial hydroxylation of the methyl-lysine residue (Yu et al., 2006). Because the resulting hydroxyl intermediate is

inherently unstable, it is spontaneously released under the form of formaldehyde leaving the non-methylated lysine as the final product.

In the case of KDM2A, enzymatic assays in which recombinant protein was incubated with differentially methylated free histone, mononucleosome and oligonucleosome substrates have demonstrated that demethylase activity is specific for histone H3 lysine 36 mono- and di-methylation (Tsukada et al., 2005). Indeed, it was found that KDM2A is insensitive to H3K36me3, which is the substrate of another JmjC-containing demethylase, known as KDM4A (Lin et al., 2008).

Although no crystal structure of the JmjC domain in complex with its H3K36 substrate is currently available, analysis of the catalytic domain alone suggests that the predicted lysine binding pocket in the active site is likely to be formed by residues S145, Y199, F215, N298, with the two aromatic side chains trapping the methyl-lysine in a mechanism analogous to that utilised by the H3K4-specific chromodomain of the *Drosophila* CHD1, and commonly found in other histone lysine binding modules (Flanagan et al., 2005). However, the basis for KDM2A discrimination between di- and tri-methylation states remains to be established.

In evolutionary terms, the substrate specificity of KDM2A JmjC domain has been shown to be conserved in both *Drosophila* and *S. cerevisiae*. Indeed, the fly homologue of KDM2A, known as dKDM2, displays an almost identical protein architecture including JmjC, zinc finger CXXC, PHD and F-box domains. Purification

studies suggest dKDM2 is a core component of the *Drosophila* dRAF complex, which also includes several PcG proteins like PSC (Posterior sex combs) and dRING (Lagarou et al., 2008). The latter is an E3 ubiquitin ligase generally found in *Drosophila* PRC1 complex. Indeed, similarly to PRC1, dRAF can efficiently monoubiquitylate H2AK119, a mark of transcription repression (Wang et al., 2004) while its dKDM2 component catalyses specific removal of H3K36me2 (Lagarou et al., 2008). In the yeast KDM2A homologue, named Jhd1, only the JmjC and PHD domain are present with the latter positioned nearest to the N-terminal. Two independent studies have revealed the conserved substrate preference of Jhd1 JmjC domain for H3K36me2 (Fang et al., 2007; Kim and Buratowski, 2007). Gene deletion and overexpression experiments indicate that the demethylase activity of Jhd1 functions to antagonise H3K36 methylation by the HKMT Set2 and consequently contribute to the maintenance of an appropriate H3K36me2 distribution genome wide.

Similarly to KDM2A, KDM2B also has been shown to carry specific activity towards di-methylated lysine 36 on histone H3 (H3K36me2) (He et al., 2008). Recently, protein overexpression studies of KDM2B in HeLa cells and MEFs have suggested that this enzyme may also mediate removal of H3K4me3 mark *in vivo* (Frescas et al., 2007; Tzatsos et al., 2009). However, a number of experiments using both recombinant and immunoprecipitated native KDM2B enzymes have reported discording results in terms of its catalytic activity towards H3K4 trimethylated histones *in vitro* (He et al., 2008; Janzer et al., 2012; Tzatsos et al., 2009). Therefore,

the enzymatic specificity of KDM2B in relation to this particular histone mark remains a matter of contention.

Interestingly, despite the conservation of enzymatic activity and the high homology of protein domains between KDM2A and KDM2B, multiple lines of evidence suggest that the two demethylases may play distinct biological functions. While the mammalian KDM2A has never been reported in association with any stable protein complex, purification studies have shown that the paralogue KDM2B can interact with a number of nuclear proteins including members of the PcG family. Indeed, mammalian KDM2B has been observed within the corepressor BCOR complex alongside other PRC1 components and two distinct ubiquitin E3 ligases (Farcas et al., 2012; Gearhart et al., 2006; Sanchez et al., 2007). Genome wide analysis indicates that such complex is attracted to a group of CGIs that are associated with Polycomb response elements (PRE) via a mechanism dependent on the CXXC domain of KDM2B (Farcas et al., 2012).

In addition to PcG recruitment, a flurry of recent studies has led to the proposition that KDM2B may act as a repressor of the INK4a-ARF-INK4b locus, which encodes replicative senescence-promoting cell cycle inhibitors (Gil and Peters, 2006; He et al., 2008; Kottakis et al., 2011; Pfau et al., 2008; Tzatsos et al., 2011; Tzatsos et al., 2009). In MEFs, ectopically expressed KDM2B has been shown to antagonise the repressive action of miR-101, an EZH2-targeting micro RNA (miRNA), and therefore contribute to the up-regulation of EZH2. This in turn leads to the PRC2-mediated

repression of the INK4a-ARF-INK4b locus, a well-known PcG target. In the absence of the cyclin-dependent kinase inhibitors, namely p15 (INK4b) and p16 (INK4a), the cell is able to bypass the G1-S phase checkpoint, an important step towards immortalisation. However, the precise role of KDM2B in tumorigenesis is unclear as conflicting evidence demonstrate that both high and low levels of KDM2B protein can be observed in different forms of tumour.

In striking contrast with the cell proliferative function is the finding from a research group that KDM2B can be recruited to nucleoli through a putative nucleolar localisation signal (NoLS) (Frescas et al., 2007). It has been suggested that KDM2B functions as a repressor of ribosomal RNA (rRNA) synthesis by catalysing removal of H3K4me3 from ribosomal genes with consequent cell growth arrest.

KDM2B has also been shown to facilitate induced pluripotent stem cell (iPSC) reprogramming but disagreement exists over whether such function is dependent on the previously documented role in senescence (Liang et al., 2012; Wang et al., 2011).

Lastly, KDM2B has been implicated in regulating the expression of several genes including the transcription factors c-Jun and c-Fos, redox-regulators, chemokines and the non-coding RNA Xist by direct targeting to their promoter regions (Ge et al., 2011; Janzer et al., 2012; Koyama-Nasu et al., 2007). Given the apparent disparity of the documented functions of KDM2B, further work is required to assess the genuine nature of the above claims.

Like KDM2B, confusing reports have accumulated over recent years with regard to the biological function of its paralogue, KDM2A. The first published study in HeLa cells has shown KDM2A targeting to pericentromeric heterochromatin where it interacts with the heterochromatin protein HP1 and mediates transcriptional silencing of centromeric satellite repeats (Frescas et al., 2008). In agreement with this finding, recent SILAC (stable isotope labelling by amino acids in cell culture)-based proteomic analysis indicate that the presence of HP1 may facilitate KDM2A recognition of H3K9 methylated nucleosomes (Bartke et al., 2010). In similar fashion to KDM2B, KDM2A has been proposed as a suppressor of rRNA synthesis and a negative regulator of replicative senescence with both functions dependent on an active JmjC domain (Pfau et al., 2008; Tanaka et al., 2010). Interestingly, two distinct reports have evidenced that KDM2A can also demethylate substrates other than the canonical H3 histone. Indeed, KDM2A has been found to specifically remove methyl groups from both K218 and K221 residues on the p65 subunit of the transcription factor NF-κB (Lu et al., 2009a; Lu et al., 2009b). Such activity is inhibitory to NF-κB function.

It is important to stress that, in addition to the N-terminal JmjC domain, both KDM2A and KDM2B enzymes also incorporate a downstream CXXC domain, which may play an important role in selecting the genomic target of these demethylases through direct DNA sequence recognition. While *in vitro* electrophoretic mobility shift assay (EMSA) have demonstrated that the KDM2B CXXC domain is functional and is required for localisation of this protein to its target genes, no evidence so far

has been provided about the relevance of the same domain in KDM2A (Koyama-Nasu et al., 2007). Therefore, how KDM2A is recruited to the aforementioned target loci remains unknown.

1.7. Aims of this thesis

Since the discovery of the JmjC domain family of HKDMs, the urge to understand the biological function of these chromatin modifiers has been ever increasing. Following early studies aimed at characterising the enzymatic specificity of each demethylase, research focus is now shifting towards the appreciation of the functional outcome of such catalytic activity in the context of the genomic target and the associated protein complex. In particular, our current knowledge regarding the role of KDM2A is limited to sporadic studies with conflicting reports and a lack of mechanistic detail. In an attempt to achieve a better comprehension of the biological significance of KDM2A activity in relation to chromatin function, the work presented in this thesis uses an *in vitro* biochemical approach to address an important question: what is the major molecular mechanism that specifies KDM2A localisation in the genome?

Chapter 3 aims at investigating the enzymatic activity and the basic properties of the proposed major chromatin-targeting domain. Chapter 4 illustrates the setting up of a defined recombinant chromatin system with the purpose of clarifying the

mechanism that allows KDM2A to interpret its chromatin substrate. These series of experiments uncover an interesting chromatin binding specificity of KDM2A that may be relevant for other related chromatin readers. Finally, chapter 5 investigates the possibility of secondary chromatin targeting modules in aiding KDM2A binding specificity.

Together these works reveal the mechanistic details of KDM2A recognition of chromatin substrate and provide a molecular basis for the understanding of KDM2A function *in vivo*.

2. Chapter Two – Materials and Methods

2.1. Common solutions

Phosphate buffered saline (PBS) (1X) – 140 mM NaCl, 3 mM KCl, 2 mM KH_2PO_4, 10 mM Na_2HPO_4.

Tris-glycine SDS (TGS) (1X) – 25 mM Tris, 192 mM glycine, 0.1% SDS

Tris-acetate EDTA (TAE) (1X) – 40 mM Tris pH 8.0, 20 mM acetic acid, 1 mM EDTA

Tris-borate (TB) (1X) – 89 mM Tris pH 8.0, 89 mM boric acid.

2.2. Bacterial methods

Media

Luria-Bertani (LB) media – 10 g/L tryptone, 5 g/L yeast extract (Fisher Scientific), 10 g/L NaCl in Milli-Q H_2O. Adjusted to pH 7.0 and autoclaved for 20 min at 121°C.

LB agar – Recipe as per LB media with addition of 16 g/L of agar. Adjusted to pH 7.0 and stored at 4°C.

2XTY media – 16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl in distilled H_2O. Adjusted to pH 7.0 and autoclaved.

All media and agar plates were prepared by the media kitchen in the Department of Biochemistry, University of Oxford.

Antibiotics

Ampicillin – Ampicillin was diluted in Milli-Q H_2O to a stock concentration of 100 mg/ml and stored at -20°C. Ampicillin was thawed at RT and added to LB media or LB agar at a working concentration of 100 µg/ml.

Kanamycin – Kanamycin was diluted in Milli-Q H_2O to a stock concentration of 50 mg/ml and stored at -20°C. Kanamycin was thawed at RT and added to LB media or LB agar at a working concentration of 50 µg/ml.

Chloramphenicol – Chloramphenicol was diluted in ethanol to a stock concentration of 34 mg/ml and stored at -20°C. Chloramphenicol was thawed at RT and added to LB media or LB agar at a working concentration of 34 µg/ml.

Tetracycline – Tetracycline was diluted in ethanol to a stock concentration of 10 mg/ml and stored at -20°C. Tetracycline was thawed at RT and added to LB media or LB agar at a working concentration of 10 µg/ml.

Gentamycin - Gentamycin was diluted in Milli-Q H_2O to a stock concentration of 50 mg/ml and stored at -20°C. Gentamycin was thawed at RT and added to LB media or LB agar at a working concentration of 7 µg/ml.

Culture on plates

All LB agar plates with various combinations of the above listed antibiotics were prepared by the media kitchen in the Department of Biochemistry, University of Oxford, at the working concentrations mentioned above. Plates were stored at 4°C. Prior to use, plates were warmed in an incubator at 37°C for 1 hr. Bacteria were grown either by streaking a single colony on the plate, or by spreading 100 µl of liquid culture evenly over the surface of the plate using a sterile bent plastic rod. Plates were inverted and incubated overnight at 37°C.

Liquid culture

Various combinations of the above listed antibiotics were added to LB media at the working concentrations stated above and mixed thoroughly. For cultures destined for small scale Miniprep processing of plasmid DNA, a single bacterial colony was scraped from the plate using a Gilson pipette tip and dropped into 5 ml LB media.

For cultures destined for large scale Maxiprep processing of plasmid DNA, 150 ml LB media was used. Cultures were grown shaking at 250 rpm, 37°C overnight.

Preparing chemically competent strains

E. coli cells of the desired strain were streaked from liquid or glycerol stocks onto LB agar plate and incubated at 37°C overnight. A single colony was inoculated in 5 ml LB culture and grown shaking at 250 rpm, 37°C overnight. The 5 ml inoculate culture was added to 500 ml LB media and incubated shaking at 250 rpm, 37°C until OD at 600 nm reached 0.5. Cells were incubated on ice for 10 min and collected by centrifugation at 2,400 rpm, 4°C for 15 min. Using sterile technique, cells were resuspended in 165 ml of 0.22 µm filtered sterile buffer A (100 mM RbCl, 50 mM $MnCl_2$, 30 mM K-acetate, 10 mM $CaCl_2$, 15% glycerol, pH 5.8). Cells were incubated on ice for 45 min and recovered by centrifugation at 2,400 rpm, 4°C for 15 min. Cells were resuspended in 40 ml of 0.22 µm filtered sterile buffer B (10 mM MOPS, 10 mM $RbCl_2$, 75mM $CaCl_2$, 15% glycerol, pH 6.8) and incubated on ice for 15 min. Cells were then snap-frozen in liquid nitrogen in 200 µl aliquots and stored at -80°C.

Transformation

Competent cells were defrosted on ice for 10 min and 1 µl of plasmid DNA was added into 100 µl of competent cell solution while gently mixing with a pipette tip. Cells were incubated on ice for 20 min and then heat-shocked at 42°C for 1 min 30 s in a water bath. Cells were incubated on ice for 2 min. 800 µl of LB media was added to the transformation mixture and cells were incubated shaking at 250 rpm, 37°C for 1 hr. If the input plasmid DNA was a product of ligation reaction, cells were collected by centrifugation at 6,000 g for 1 min and resuspended in 100 µl LB media before plating on selective LB agar plate. If the input plasmid DNA was supercoiled, 30 µl of cells were plated without centrifugation on selective LB agar plate.

2.3. Isolation of exogenous DNA from bacteria

Small scale plasmid DNA preparation

A single colony of *E. coli* cells containing the plasmid DNA to be extracted was added to 5 ml LB media and grown shaking at 250 rpm, 37°C overnight. Cells were collected by centrifugation at 4,000 rpm, RT for 10 min. DNA was extracted from the cells using the QIAGEN Miniprep kit according to manufacturer recommendation.

Large scale plasmid DNA preparation

A single colony of *E. coli* cells containing the plasmid DNA to be extracted was added to 150 ml LB media and grown shaking at 250 rpm, 37°C overnight. Cells were collected by centrifugation at 4,000 rpm, RT for 10 min. DNA was extracted from the cells using the Sigma-Aldrich GenElute HP Maxiprep kit according to manufacturer recommendation with the exception that DNA was eluted with elution buffer that was pre-heated to 65°C. The eluted DNA was then concentrated following manufacturer recommendation and stored at -20°C.

Large scale preparation of Bacmid DNA

A single white colony from DH10Bac recombination plate was inoculated into 2 ml LB media containing 50 µg/ml kanamycin, 7 µg/ml gentamycin, 10 µg/ml tetracycline and grown shaking at 250 rpm, 37°C overnight. 1.5 ml of cells were collected by centrifugation at 14,000 g, RT for 1 min and resuspended in 0.3 ml of 0.22 µm filtered sterile solution A (15 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNase A). 0.3 ml of 0.22 µm filtered sterile solution B (0.2 M NaOH, 1% SDS) was added and gently mixed. The resulting lysate was incubated at RT for 5 min. 0.3 ml of 3 M sodium acetate pH 5.5 was then added and mixed gently. The lysate was incubated on ice for 10 min and precipitated by centrifugation at 14,000 g, 4°C for 10 min. 800 µl of isopropanol was added to the derived supernatant and mixed

gently. The solution was incubated on ice for 10 min and the Bacmid DNA was collected by centrifugation at 14,000 g, RT for 15 min. 500 µl of 70% ethanol was added to the precipitated DNA pellet. DNA was collected by centrifugation at 14,000 g, RT for 5 min. The supernatant was removed and the DNA pellet air-dried at RT until complete evaporation of ethanol. DNA was resuspended in 40 µl of EB buffer (QIAGEN) and incubated at 37°C for 20 min with occasional gentle mixing. DNA concentration was measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific) at 260 nm and the DNA was stored at -20°C.

2.4. Manipulation of DNA and cloning

Restriction endonuclease digestion of DNA

Restriction digests were performed in 20 µl volume for conventional analytical purposes (e.g. to examine the efficiency of CpG DNA methylation) or in 50 µl volume when used to linearize a cloning vector. Reactions were carried out at temperatures indicated by the manufacturer (New England Biolabs, Roche) and in the presence of 100 ng/µl of BSA where this was recommended. Linearized vector DNA was purified using the QIAquick Gel Extraction kit (QIAGEN) and eluted in 30 µl EB buffer. To analyse the extent of digestion, 100-200 ng of purified digestion product was resolved by agarose gel electrophoresis in 1X TAE and visualised by

ethidium bromide staining. The percentage of agarose varied depending on the size of the expected DNA fragment.

Methylation of DNA fragments at CpGs

Methylation of DNA at CpGs by pre-purified recombinant M.SssI was carried out according to the following ratio of reagents: 1 µg of template DNA was incubated in the presence of 2 µl of 10X NEBuffer 2 (New England Biolabs), 2 µl of 20X (3.2 mM) SAM (New England Biolabs) and 1 µg of pre-purified recombinant M.SssI in a 20 µl reaction volume. This ratio was maintained during scale-up methylation reactions. The reaction mixture was incubated at 37°C overnight. In the following morning, 2 µl of 20X (3.2 mM) SAM (New England Biolabs) and 1 µg of pre-purified recombinant M.SssI were added to the reaction which was then incubated at 37°C for 2 hr. The methylated DNA was purified via QIAGEN QIAquick Gel Extraction kit and eluted in 50 µl EB buffer. DNA concentration was measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific) at 260 nm and the DNA was stored at -20°C. To verify the extent of methylation, 100-200 ng of purified DNA was digested with the CpG methylation sensitive restriction enzyme HhaI, resolved by agarose gel electrophoresis in 1X TAE and visualised by ethidium bromide staining.

Polymerase chain reaction (PCR)

All PCR reactions were carried in 50 µl volume. 200 ng of pre-purified recombinant Taq DNA polymerase, 25 pmol of each forward and reverse primer, 10 nmol of dNTP mix (Fermentas) and 10 ng of plasmid template DNA were incubated together in the presence of PCR buffer (20 mM Tris HCl pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X 100). If the PCR products were destined for cloning, 1 unit of Phusion polymerase (Finnzymes) was used instead together with the commercial buffer provided with the polymerase. Cycling parameters varied depending on DNA template and primer properties. The PCR products were purified using the QIAquick PCR Purification kit (QIAGEN) according to manufacturer instructions. The DNA was eluted in 30 µl EB buffer.

DNA Sequencing

The plasmid DNA to be sequenced and the primers for DNA sequencing were diluted to concentrations indicated by Source BioScience LifeSciences and delivered to their sequencing facility at the University of Oxford.

Agarose gel extraction of DNA fragments

DNA was extracted from agarose gels via QIAquick Gel Extraction kit (QIAGEN) according to manufacturer instructions and the DNA eluted in 50 µl EB buffer.

Ligation independent cloning (LIC)

5 µg of the *sacB* gene-containing LIC-adapted vector plasmid was digested with the restriction endonuclease BsaI in a 50 µl reaction volume for 1 hr. The digestion products were separated by agarose gel electrophoresis in 1X TAE and visualised by ethidium bromide staining. The DNA band corresponding to the empty vector was excised out of the gel and purified via QIAquick Gel Extraction kit (QIAGEN). The DNA was eluted in 30 µl EB buffer. At the same time, the insert gene was PCR amplified with primers carrying the LIC cloning hybridisation sequence. The PCR product was purified via the QIAquick PCR Purification kit (QIAGEN) and eluted in 30 µl of EB buffer. The linearized empty vector and the amplified insert were each individually digested with 2 µl of 5 u/µl T4 DNA polymerase (Fermentas) in the presence of 2 µl of 2 mM dGTP and dCTP respectively, in a 50 µl reaction volume. 10 µl of 5X reaction buffer supplied with the enzyme (Fermentas) was also included into each reaction mixture. The reaction was first allowed to proceed for 30 min at 22°C, followed by heat inactivation at 75°C for 20 min. The DNA was extracted via QIAquick Gel Extraction kit and eluted in 30 µl EB buffer. The purified

empty vector and the insert gene DNA were mixed together in three separate Eppendorf tubes according to the following ratios: 0.2 µl, 1 µl and 2 µl of insert respectively with 1 µl of vector. The vector and the insert were allowed to hybridise for 30 min at RT. 1 µl of each hybridised plasmid was transformed into 100 µl of competent XL10 *E. coli* cells and plated on selective LB plate. The cells were grown at 37°C overnight. To screen for colonies with successful clones, the DNA was extracted via the QIAGEN Miniprep kit from 3 colonies for each plate and eluted in 30 µl EB buffer. 1 µl of each purified DNA was analysed with a suitable restriction endonuclease to ensure for the presence of the gene of interest. The full gene was then sequenced by the Source BioScience LifeSciences sequencing facility at the University of Oxford to ensure that no mutations were present in the cloned gene sequence.

Mutagenesis

All site-directed mutagenesis were performed using the QuickChange II XL site-directed mutagenesis kit (Stratagene) according to manufacturer instructions. All CpG mutations engineered in the 601 DNA sequence were introduced by PCR amplification using custom-synthesised mismatch primers (Invitrogen).

Radioactive labelling of DNA fragments

200 ng of template DNA was end labelled in a 30 µl reaction volume in the presence of 2 µl of 10 u/µl T4 polynucleotide kinase (New England Biolabs), 3 µl of 10X buffer supplied with the enzyme and 1.5 µl of 10 mCi/ml [γ-^{32}P] ATP (Perkin Elmer). The reaction was incubated for 1 hr at 37°C. If the end-labelled DNA was to be directly used for EMSA assay, it was purified via QIAquick PCR Purification kit (QIAGEN) and eluted in 75 µl EB buffer. If the end-labelled DNA was destined for nucleosome reconstitution, it was purified via MinElute Reaction Clean Up kit (QIAGEN) and eluted in 12 µl EB buffer. The radioactivity of 1 µl of the derived elution was measured in counts per minute (cpm) using a Bioscan QC-2000 β-counter. If the eluted DNA was to be used as binding probe for EMSA, the final volume of the elution was adjusted with EB buffer to reach 5,000 cpm/µl.

2.5. Insect cell and baculovirus vector expression system

Insect cell culture

Sf9 cells derived from Invitrogen were adapted to grow in suspension in serum free Sf900 II SFM media (Gibco) supplemented with 50 u/ml penicillin and 50 µg/ml streptomycin (Gibco) shaking at 100 rpm, 27°C. Cells were passaged by

centrifugation at 1,500 rpm, RT for 5 min to collect the cells, which were then resuspended in fresh Sf900 II SFM media at 0.5 x 10^6 cells per ml density. The growing cell density was maintained between 0.5 x 10^6 and 3 x 10^6 cells per ml. Sf9 cells were cryogenically stored by centrifugation at 1,500 rpm for 5 min followed by resuspension at 1.25 x 10^7 cells per ml in Sf900 II SFM media supplemented with 10% DMSO. Cells were stored in 2 ml aliquots at -80°C overnight before transfer to long-term storage in liquid nitrogen. To recover the cells from liquid nitrogen stocks, each 2 ml aliquot was defrosted at RT and diluted in 50 ml Sf900 II SFM media supplemented with antibiotics. Cells were incubated at 27°C overnight and passaged the following morning to remove dead cells.

Generation of recombinant baculovirus genome

Baculoviruses were generated using a previously LIC-adapted version of the pFastBac vector plasmid following manufacturer (Invitrogen) instructions except for the use of LB media during transformation of the pFastBac plasmid into the DH10Bac cells. Successfully recombined Bacmid DNA was extracted using the purification method described above. Diagnostic PCR using M13 forward and reverse primers was also performed according to manufacturer instructions.

Transfection of Sf9 cells and isolation of P1 virus

Sf9 cells were seeded in 6 well plate at 9 x 10^5 cells per 2 ml of Sf900 II SFM media supplemented with 50 u/ml penicillin and 50 µg/ml streptomycin and allowed to attach for 15 min at 27°C. For each well, 1 µg of Bacmid DNA and 8 µl of Cellfectin II (Invitrogen) were each diluted separately in 100 µl of Grace's un-supplemented media (Gibco). The diluted Bacmid DNA was gently mixed to the diluted Cellfectin II and incubated at RT for up to 15 min to allow formation of complex. The seeded Sf9 cells were washed once with 2 ml of Grace's un-supplemented media and incubated in 800 µl of the same media. The Bacmid-Cellfectin II mixture was then added drop-wise to the cells and mixed by gentle agitation of the well. Cells were incubated at 27°C for 5 hr, following which the transfection mixture was replaced with 2.5 ml of Sf900 II SFM media. Cells were incubated at 27°C for 72 hr and the supernatant then collected. This was cleared of dead cells via centrifugation at 1,500 rpm for 5 min and stored as P1 virus at 4°C protected from light sources. To analyse the efficiency of transfection, cells were collected by gentle scraping in 500 µl of ice-cold PBS and pelleted at 2,000 rpm for 1 min. The cell pellet was resuspended in 250 µl of lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% NP40, 1 mM EDTA) supplemented with complete protease inhibitor cocktail (Roche). The lysate was incubated while rotating for 15 min at 4°C and cleared via centrifugation at 13,000 rpm, 4°C for 10 min. The resulting supernatant was stored as whole cell

extract at -80°C. 15 µl of the extract was analysed by Western blotting using both protein (e.g. KDM2A) and affinity tag (e.g. Flag, Strep II) specific antibodies.

Amplification of viral stock

To amplify the P1 viral stock, 12 x 10^6 Sf9 cells were seeded on a 94 mm culture plate in 12 ml Sf900 II SFM media supplemented with penicillin and streptomycin. 300 µl of the P1 viral stock was added drop-wise to each 94 mm plate and gently mixed. The infection was allowed to proceed for 72 hr at 27°C. The supernatant was then processed as per P1 virus and stored as P2 viral stock at 4°C. Cells were also collected following the same procedure described above and the derived whole cell extract analysed by Western blotting. This cycle was repeated once to generate the P3 viral stock, which was then titrated to determine the optimal volume of virus inoculum for protein expression.

Infection of Sf9 cells for protein production

For large scale protein expression, Sf9 cells were passaged to 0.75 x 10^6 cells per ml density and allowed to grow shaking at 100 rpm, 27°C overnight to reach up to 1.5 x 10^6 cells per ml the next morning. 1 µl P3 virus was added to every 2 x 10^6 cells and the infection allowed to proceed for 48-72 hr. Cells were then collected by

centrifugation at 1,500 rpm for 5 min and the pellet was stored at -80°C until protein purification.

2.6. Expression of proteins in *E. coli* cells

All prokaryotically derived KDM2A proteins and the FDH protein were expressed from a LIC-adapted version of pNIC28 BsaI vector plasmid. The M.SssI enzyme was expressed from a pBAD24 plasmid. The Taq DNA polymerase was expressed from a pTAQ plasmid. The Set2 enzyme was expressed from a pCAL-n plasmid. Individual core histones were expressed from pET3d plasmids. A single colony from selective LB plate of the strain of bacteria (e.g. BL21, BL21 DE3 pLysS) carrying the expression vector for the protein of interest was added to 150 ml of selective 2XTY media and incubated shaking at 250 rpm, 37°C overnight. 25 ml of this starter culture was diluted in 0.5 L of selective 2XTY media and incubated shaking at 250 rpm, 37°C until OD at 600 nm reached 0.5 as determined by spectrophotometer. For the pNIC28 system, 0.5 ml of 1 M IPTG was added to 0.5 L of culture and then incubated at 30°C for 3 hr. For the pTAQ system, the same amount of IPTG was used and the culture was incubated at 18°C overnight. For the pCAL-n system, 50 µl of 1 M IPTG was added to 0.5 L of culture when OD at 600 nm reached 1.0 and cells were incubated at 18°C overnight. For the pBAD24 system, protein expression was induced using 0.1% L-arabinose solution and cells were incubated at 18°C

overnight. For the pET3d system, expression conditions were the same as for the pNIC28 system except for the expression time, which was reduced to 2 hr. After induction, cells were collected by centrifugation at 6,000 g, 4°C for 20 min and washed once in ice-cold 1X PBS. After centrifugation at 4,000 g, 4°C for 20 min, the derived cell pellet was stored at -80°C. To extract the recombinant proteins, the cell pellet was defrosted on ice and 10 ml of lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1% NP 40) supplemented with complete inhibitor cocktail (Roche) was added for each 0.5 L cell culture derived pellet. After resuspension, cells were aliquoted in 20 ml volumes and each volume was sonicated on ice at 60% amplitude using a VCX130PB sonicator (Sonics & Materials) for 30 s followed by a 30 s pause. This sonication cycle was performed 8 times for histone-containing lysates and 6 times for all other protein lysates. The lysate was then cleared by centrifugation at 12,500 rpm, 4°C for 20 min. The derived whole cell lysate was applied to the chromatographic methods described below.

2.7. Protein methods

Measuring protein concentration

Bradford assay was carried out to determine protein concentration. The stock Bradford dye (Bio-Rad) was diluted 1 in 5 in H_2O and mixed thoroughly. A series of protein volumes ranging from 1 to 10 µl was each added to 1 ml of diluted

Bradford dye and mixed thoroughly. OD at 595 nm was then measured in a plastic cuvette using an Ultrospec 3300 spectrophotometer (GE Healthcare). In parallel, a series of BSA (New England Biolabs) standards with known protein concentrations were also measured by Bradford to generate a standard curve. By plotting the OD values of the protein of interest on the standard curve, the protein concentration was then derived.

Dialysis

Proteins derived from purification preparations were generally dialysed using a sealed Spectra/Por dialysis tubing (Spectrum) with a molecular weight cut off (MWCO) range of 6,000-8,000 Da in 2 L of BC buffer (50 mM Hepes pH 7.9, 10% glycerol, 0.5 mM DTT) containing an appropriate concentration of KCl ranging from 100 to 400 mM. Dialysis was performed at 4°C overnight following which protein sample was collected, cleared of any precipitates by centrifugation at 13,200 rpm, 4°C for 20 min and stored at -80°C.

SDS-PAGE and Coomassie blue staining

Proteins to be resolved by SDS-PAGE were prepared in loading buffer (2% SDS, 100 mM Tris pH 6.8, 100 mM DTT, 10% glycerol, 0.1% bromophenol blue) and

denatured by boiling at 100°C for 5 min. For all SDS-PAGE, 0.75 mm thick gels were prepared with a resolving percentage of 29:1 acrylamide to bis-acrylamide solution (Bio-Rad), 390 mM Tris pH 8.8, 0.1% SDS, 0.1% APS, 0.04% TEMED in the resolving gel and 5% 29:1 acrylamide to bis-acrylamide solution, 129 mM Tris pH 6.8, 0.1% SDS, 0.1% APS, 0.1% TEMED in the stacking gel. Gels were cast and run using the Mini-PROTEAN Tetra or the PROTEAN II XL electrophoresis systems (Bio-Rad) in 1X TGS at 200 V constant, RT until the loading dye migrated off the bottom of the gel. Gels were stained with Coomassie blue staining solution (50% methanol, 10% acetic acid, 0.1% brilliant blue R250) for 15 min. Stained gels were washed once with H_2O and destained in 30% methanol, 10% acetic acid solution until the background staining was removed. Gels were incubated in H_2O overnight at RT and imaged the next morning.

Semi-dry transfer to nitrocellulose membrane and Western blot

A single layer of either 0.2 µm (for histone proteins) or 0.45 µm (for all other proteins) nitrocellulose membrane (Bio-Rad) was wet in semi-dry transfer buffer (39 mM glycine, 48 mM Tris, 0.037% SDS, 20% methanol) and placed on top of three layers of 0.3 mm Whatman paper pre-soaked in the same buffer. Additional three layers of pre-soaked Whatman paper were placed on top. Air bubbles were

removed by gently rolling a pipette twice in each direction over the assembled gel sandwich. Transfer was performed using the Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad) with 0.2 A constant, at 23 V maximum, RT for 1 hr 20 min. The efficiency of transfer was evaluated by staining the nitrocellulose membrane with 0.2% Ponceau S, 1% acetic acid solution for 5 min at RT. The membrane was then briefly destained using 0.1M NaOH, which was washed away with H_2O. The membrane was incubated in 5% milk, 1X PBS + 0.1% Tween 20 (PBST) for 30 min at RT. The membrane was incubated in 1:1,000 dilution of the primary antibody in 5% milk, 1X PBST at 4°C rocking overnight. In the following morning, the membrane was washed for 5 min in 1X PBST 3 times. The membrane was then incubated in 1:5,000 dilution of an ECL HRP-conjugated secondary antibody in 1% BSA, 1X PBST for 1 hr at RT. The membrane was again washed 3 times in PBST for 5 min each at RT. Equal volumes of ECL solution A (2.5 mM luminal, 396 µM p-coumaric acid, 100 mM Tris-HCl pH 8.5) and ECL solution B (5.632 mM H_2O_2, 100 mM Tris-HCl pH 8.5) were mixed together thoroughly. The ECL mixture was then applied to the entire surface of the membrane and incubated at RT for 1 min. The membrane was completely wrapped with Saran film and exposed to ECL-sensitive film within a hypercassette (GE Healthcare).

2.8. Protein purification

Chromatography media and solutions

Mono Q (1 ml): anion exchange resin.

Resource Q (1 ml): anion exchange resin.

Superdex HiLoad 16/60 200 PG (120 ml): gel filtration resin.

Sephacryl S200 (120 ml): gel filtration resin.

Anion exchange buffer: 50 mM Tris pH 8.0, 1 mM DTT, 100 mM (AE100) to 1,000 mM (AE1000) NaCl.

Gel filtration buffer: 20 mM Hepes KOH pH 7.9, 10% glycerol, 200 (GF200) or 300 mM KCl (GF300).

Denaturing gel filtration buffer (SAU1000): 7 M Urea, 20 mM sodium acetate pH 5.2, 1 M NaCl, 5 mM 2-mercaptoethanol, 1 mM EDTA.

Ion exchange chromatography

Ion exchange purifications were performed using pre-packed Mono Q or Resource Q columns based on a FPLC system (GE Healthcare) under conditions recommended by the manufacturer. Protein samples were applied to resins that

were pre-equilibrated overnight in AE100. The bound-resin was washed with 20 column volumes of AE100 before a linear gradient was applied over a 40 ml volume to reach AE1000 ionic concentration. Fractions were monitored by UV spectrometry at 280 nm and analysed by Coomassie stained SDS-PAGE. Proteins were then dialysed into an appropriate BC buffer for storage at -80°C.

Gel filtration chromatography

Proteins were applied to a pre-equilibrated gel filtration column that was either based on a FPLC system (e.g. Superdex HiLoad 16/60 200 PG) or was set up on the bench (e.g. Sephacryl S200) under conditions recommended by the manufacturer. Elution fractions were monitored by UV spectrometry at 280 nm or by Bradford assay and analysed by Coomassie stained SDS-PAGE. Proteins were then dialysed into an appropriate BC buffer for storage at -80°C.

Purification of 6XHis and Strep II tagged proteins from *E. coli* cells

The whole cell lysate was incubated with 0.5 ml IMAC Sepharose 6 Fast Flow Ni-NTA resin (GE Healthcare) rotating for 1 hr at 4°C. The resin was pre-washed 2 times in the lysis buffer described above. The lysate-resin mixture was loaded onto

a disposable 10 or 20 ml Poly-Prep chromatography column (Bio-Rad) and the unbound solution was allowed to flow through. The protein-bound resin was then washed with 40 column volumes of Ni-NTA wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH adjusted to 8.0 using NaOH). Proteins were eluted generally 5-7 times with 1 column volume of Ni-NTA elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH adjusted to 8.0 using NaOH). The fractions were pooled together and applied directly to a pre-washed StrepTactin Superflow resin column (IBA) if the recombinant protein also carried a C-terminal Strep II tag. The sample was allowed to flow through the resin 5 times to allow for the capture of the desired protein. The column was then washed with 40 column volumes of StrepTactin wash buffer (100 mM Tris-HCl pH 8, 150 mM NaCl). Proteins were then eluted 5-7 times with 1 column volume of StrepTactin elution buffer (100 mM Tris-HCl pH 8, 150 mM NaCl, 2.5 mM desthiobiotin). Protein concentrations in the eluted fractions were then measured by Bradford assay. The peak and side fractions were respectively pooled together and dialysed overnight in 2 L BC buffer (50 mM Hepes pH 7.9, 10% glycerol, 0.5 mM DTT) containing 100 mM to 400 mM KCl for long term storage at -80°C. The purity of each elution fraction was also analysed by running 1-5 µl of proteins on SDS-PAGE and successive Coomassie blue staining.

Purification of Flag-tagged proteins from Sf9 cells

The Sf9 cell pellet derived from 0.5 L expression culture was resuspended in 10 ml lysis buffer F (10 mM Tris pH 8.0, 500 mM NaCl, 4 mM MgCl2, 0.4 mM EDTA, 2 mM DTT, 20% glycerol, 0.4 mM PMSF) supplemented with complete protease inhibitor cocktail (Roche). Cells were homogenised by gently douncing with a tight fitting pestle 30 times over 30 min. The cell lysate was cleared by centrifugation at 11,000 rpm, 4°C for 10 min. The resulting supernatant was diluted with 7 ml of dilution buffer (10 mM Tris, 10% glycerol, 0.02% NP40). 250 µl of anti-Flag M2 resin (Sigma-Aldrich) was washed with 10 ml of lysis buffer collecting at 2,000 rpm for 3 min. The diluted extract was then added to the resin and incubated rotating at 4°C for 3 hr. The resin was collected by centrifugation at 2,000 rpm for 3 min and washed 3 times each with 15 ml wash buffer F (10 mM Tris pH 8.0, 150 mM NaCl, 2 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT, 15% glycerol, 0.4 mM PMSF). The washed resin was resuspended in 1 ml of wash buffer F and transferred to a 1.5 ml Eppendorf tube. The Eppendorf tube was adjusted into an open-lid 15 ml falcon tube and centrifuged at 2,000 rpm, 4°C for 2 min. The supernatant was then removed from the top of the resin, which was eluted 5 times each with 300 µl wash buffer F containing 0.2 mg/ml Flag peptide rotating at 4°C for 20 min. To analyse the sample purity, 5-10 µl of each elutions was run on SDS-PAGE and Coomassie stained. Proteins were stored at -80°C without further dialysis.

Purification of CBP tagged proteins

To purify CBP-tagged recombinant proteins, the cell pellet was resuspended in 12 ml of $CaCl_2$ binding buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM MgOAc, 1 mM imidazole, 2 mM $CaCl_2$) and lysed via sonication as described before. The lysate was cleared by centrifugation at 15,000 rpm, 4°C for 20 min. The resulting supernatant was incubated with 1 ml CaM affinity resin (Stratagene) that was pre-washed 2 times each with 9 ml $CaCl_2$ binding buffer for 3 hr at 4°C. The resin was collected by centrifugation at 1,000 rpm, 4°C for 2 min. The resin was then washed 3 times each with 10 ml $CaCl_2$ binding buffer. The resin was then applied to a 20 ml Poly-Prep chromatography column and the proteins were eluted 5 times each with 350 µl elution buffer (50 mM Tris-HCl pH 8.0, 10 mM 2-mercaptoethanol, 2 mM EGTA-OH pH 8.0, 150 mM NaCl). The eluted proteins were dialysed in 2 L BC50 (50 mM Hepes pH 7.9, 10% glycerol, 0.5 mM DTT, 50 mM KCl) containing 0.5 mM EDTA and stored at -80°C.

Purification of Taq DNA polymerase by heat precipitation and biphasic extraction

To isolate recombinant Taq Pol, the cell pellet was lysed via sonication as described before. Following sonication and centrifugation the supernatant was incubated at 75°C for 1 hr and then centrifuged again at 12,500 rpm, 4°C for 10 min. 30g

$(NH_4)_2SO_4$ was slowly added to each 100 ml supernatant while mixing. The Taq Pol was then collected by centrifugation at 12,500 rpm, 4°C for 10 min and resuspended in 10 ml buffer A (50 mM Hepes pH 7.9, 50 mM glucose, 1 mM EDTA). To remove contaminating DNA, the Taq Pol sample was adjusted to contain 10% in volume of saturated dibasic potassium phosphate pH 10.3, following which an equal volume of 100% ethanol was added and mixed vigorously. The organic phase (upper layer) was separated from the aqueous phase (lower layer) by centrifugation at 2,000 g, RT for 2 min. The upper organic phase containing the Taq Pol protein was then collected and dialysed overnight in 2 L buffer B (20 mM Hepes pH 7.9, 100 mM KCl, 0.1 mM EDTA, 0.5 mM PMSF, 10% glycerol) at 4°C. In the following morning, the protein was dialysed again in 2 L of fresh buffer B for another 4 hr. The collected protein sample was then subject to gel filtration chromatography to further refine its purity.

Purification of core histones

Histone containing cell pellet stored at -80°C was resuspended in 20 ml wash buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM benzamidine) and sonicated as described above. Inclusion bodies were collected by centrifugation at 23,000 g, 4°C for 20 min. The pellet was resuspended in 20 ml TW buffer (wash buffer with 1% Triton X100) and then washed 2 times with wash buffer. 0.5 ml DMSO was added to the pellet and mixed vigorously for 30 min at RT with a

magnetic stirrer. 2 ml unfolding buffer (7 M guanidine hydrochloride, 20 mM Tris pH 7.5, 10 mM DTT) was added and then gently mixed for 1 hr at RT, after which the solution was cleared by centrifugation at 13,200 rpm, RT for 10 min. The derived supernatant was applied onto a Sephacryl S200 gel filtration column that was equilibrated with SAU1000 buffer overnight. 1.5 ml elution fractions were collected at 1 min 15 s intervals and monitored by Bradford assay. In order to assess the eluted protein purity, 5 µl of every three or four fractions was loaded onto a 20 cm x 20 cm, 18% SDS-PAGE gel and run at 10 mA constant, RT overnight with flowing water cooling. The fractions containing monomeric histone proteins as shown by Coomassie staining of SDS-PAGE were pooled and dialysed at 4°C into 3 changes of 4 L each H_2O with 0.25 mM DTT and stored at -80°C.

2.9. Chromatin reconstitution methods

Generation of 601 DNAs

147 bp or 216 bp long 601 DNA sequences were PCR amplified from the pGEM-3Z 601 plasmid. The 342 bp long double 601 DNA sequence was PCR amplified from the pBlu2SKP 2X601 +48 plasmid. The PCR products were purified using a 1 ml Resource Q anion exchange column (GE Healthcare) over 40 ml of an increasing NaCl gradient from 0.1 M to 1 M in 50 mM Tris pH 8.0 buffer. The DNA sample was eluted in 200 µl fractions monitored by UV spectrometry at 260 nm and visualised

by agarose gel electrophoresis in 1X TAE with ethidium bromide staining. The peak fractions were pooled together and dialysed in 10 mM Tris pH 8.0 for storage at -20°C.

Installation of methyl lysine analogues (MLAs) into core histones

Installation of H3K4 tri-methyl lysine analogues into core histones was based on the published method by Simon et al (Simon et al., 2007). In brief, the pET3d plasmid was engineered to carry a H3 histone gene with K4C mutation by site directed mutagenesis. The histone was expressed and purified using the same methods described above. The concentration of the purified protein was measured by Bradford assay. Proteins were concentrated using a 20 ml Vivaspin concentrator column (Sartorius) with MWCO of 5,000 Da to a final concentration of 30 mg/ml. 250 µl of 30 mg/ml proteins were reduced in 735 µl of 0.22 µm filtered alkylation buffer (1.33 M Hepes pH 7.8, 5.32 M guanidine hydrochloride, 13.2 mM D/L-methionine) supplemented with 20 µl of 1 M DTT for 1 hr at 37°C. 100 mg of (2-bromoethyl) trimethylammonium bromide (Sigma-Aldrich) was then added to the reaction and incubated with occasional mixing at 50°C for 2 hr 30 min. After initial incubation, 10 µl of 1 M DTT was added and the reaction was allowed to proceed for an additional 2 hr 30 min. The reaction was then stopped by adding 50 µl of 14

M 2-mercaptoethanol and the sample was applied to a 5 ml HiTrap desalting column (GE Healthcare). The column was washed with H_2O containing 2 mM 2-mercaptoethanol and 0.5 ml fractions were then collected. 1 µl of eluted proteins was analysed by Bradford assay. The efficiency of the installation reaction was evaluated by Western blotting using H3K4me3-specific antibodies and the histones were stored at -80°C.

Refolding of histone octamer

Refolding of histone octamer was performed based on the published method by Dyer et al (Dyer et al., 2004). In brief, purified recombinant core histones were concentrated using a 20 ml Vivaspin concentrator column (Sartorius) with MWCO of 5,000 Da to a final concentration of 10 mg/ml. 3 mg of each of the four core histones were diluted in unfolding buffer (7 M guanidine hydrochloride, 20 mM Tris pH 7.5, 10 mM DTT) to a concentration of 2 mg/ml and incubated at RT for 1 hr. Equal weight ratios of the four core histones were mixed together and diluted to 1 mg/ml with unfolding buffer. The histone mixture was then dialysed 3 times against 4 L each refolding buffer (2 M NaCl, 10 mM Tris pH 7.5, 1 mM Na-EDTA, 5 mM 2-mercaptoethanol) overnight at 4°C. The dialysed material was concentrated using a Vivaspin concentrator column (Sartorius) to 1 ml volume at 4°C. The sample was applied to a Superdex HiLoad 16/60 200 PG column in the FPLC, which was pre-equilibrated with refolding buffer. The octamers were eluted at 0.5 ml/min in

1 ml fractions and analysed via Coomassie stained 18% SDS-PAGE to evaluate the purity and stoichiometry of the sample. The fractions containing octamers with equal stoichiometry of the four core histones were combined and stored at 4°C without freezing.

Reconstitution of nucleosome core particle

Reconstitution of nucleosome core particle was performed based on the published method by Dyer et al (Dyer et al., 2004). At least three different ratios of DNA to histone octamer were set up where 1 µg of 601 DNA was mixed with the appropriate amount of histone octamer in 2 M NaCl. 1-2 µl of the [^{32}P] end-labelled version of the same 601 DNA was added to the reaction. The reaction was initially incubated for 30 min at RT. An equal volume of 10 mM Tris-HCl pH 7.6 was then added to the reaction, which was incubated for 1 hr at RT. The reconstitution mix was then diluted with the appropriate volume of 10 mM Tris-HCl pH 7.6 at every hour to reach progressively 0.8 M, 0.67 M, 0.2 M and 0.1 M NaCl concentrations. The reconstituted nucleosome was analysed by loading a small aliquot onto 0.8% agarose gel in 0.2X TB and running the gel at 4°C. The gel was dried on DE81 chromatography paper (Whatman) using a vacuum-driven gel dryer (GE Healthcare) and autoradiography of the gel was taken using a FLA7000 phosphorimager (Fujifilm) or with a BioMax XAR film (Kodak). The reconstituted nucleosomes were stored at 4°C for up to one week without freezing.

Purification of recombinant nucleosomes by sucrose gradient sedimentation

Sucrose gradients were poured in 12.5 ml open top polyclear centrifuge tubes (Seton Scientific) designed for the SW 41 Ti ultracentrifuge rotor (Beckman Coulter) using a manual two chamber gradient former. For nucleosome purification, 5-20% sucrose gradient was poured in 100 mM Tris pH 7.6 buffer with 1 mM DTT. The reconstituted nucleosomes were loaded drop-wise on top of the sucrose gradient and centrifuged at 35,000 rpm, 4°C in a SW 41 Ti rotor for 19 hr. 500 µl size fractions were collected manually starting from the top of the gradient. The fractions were analysed by running onto 0.8% agarose gel in 0.2X TB at 4°C. The fractions containing the desired mononucleosome or dinucleosome species were pooled together and used for subsequent EMSA assays.

2.10. DNA binding experiments

Electrophoretic mobility shift assay (EMSA)

EMSA reactions were prepared in binding buffer (4% Ficoll 400, 20 mM Hepes pH 7.9, 150 mM KCl, 1 mM EDTA, 0.5 mM DTT) and up to 2 µg of protein sample was incubated with 25 ng/µl competitor poly-dAdT DNA (Sigma-Aldrich) for 10 min at RT before the [^{32}P] end-labelled 601 DNA or nucleosome probes were added. 0.125%

bromophenol blue dye was added to the reaction mixture to allow visualisation of the running band on the gel. The mixture was allowed to incubate for 20 min at RT. EMSA assays with DNA or reconstituted nucleosome probes were performed respectively on 1.3% and 0.8% agarose gels in 0.2X TB. Gels were run at 100 V constant, 4°C until the band had migrated for approximately 7 cm. The gel was then dried onto a DE81 anion exchanger paper (Whatman) via a vacuum driven gel-dryer (GE Healthcare) and autoradiography of the gel was taken using a FLA7000 phosphorimager (Fujifilm).

Surface plasmon resonance (SPR)

Surface plasmon resonance measurement was conducted using a Biacore T100 machine (GE Healthcare). Three different oligonucleotide probes containing respectively 0, 1 or 6 CpG sites were used (See Appendix for sequence details). For each probe, 5 µg of single stranded template DNA was mixed with equivalent molar amount of 5' biotinylated M13 reverse primer (Invitrogen) and incubated in 10 mM Tris pH 8.0, 50 mM NaCl for 10 min at 100°C in a heat block. After incubation, the hot block was removed from the heat source and allowed to gradually cool down to RT. The derived DNA product was incubated as template with 6 u Klenow polymerase (New England Biolabs) and 5 µl of 2 mM dNTP mix in a 20 µl reaction volume at RT. The product of the extension reaction was purified using a QIAquick Gel Extraction kit (QIAGEN) and eluted in 50 µl EB buffer. 5 pmol

of each probe was diluted in 1X coupling buffer (10 mM Hepes pH 7.9, 150 mM NaCl, 3 mM EDTA, 0.005% Tween 20) and immobilised on the surface of a flow cell in a streptavidin (SA) coated sensor chip (GE Healthcare) at a flow rate of 5 µl/min. All flow cells were then washed with 200 µM biotin in 1X binding buffer (10 mM Hepes pH 7.9, 150 mM NaCl, 1 mM DTT) to block all unbound streptavidin sites. Serial dilutions (3 µM, 1.5 µM, 750 nM, 375 nM, 187.5 nM, 0 nM) of the analyte protein were prepared in 1X binding buffer containing 100 ng/µl of poly(dAdT) DNA and 20 µl of each dilution was injected at a flow rate of 5 µl/min. Bound proteins were allowed to dissociate in 1X binding buffer for 10 min followed by an elution step with 20 µl of regeneration buffer (2.5 M NaCl, pH 5.3). For each experiment, two flow cells were measured in parallel, one coated with the 0CG control DNA probe and the second with either 1CG or 6CG binding probe. For a given concentration of the analyte, the resonance unit (RU) used to calculate the K_D is derived as the difference between the RU measured in either the 1CG or 6CG flow cell and the background RU measured in the 0CG control flow cell. The K_D value was then estimated using a Langmuir binding curve fit of the RUs calculated at each analyte concentration.

2.11. Enzyme activity assay

Histone demethylase assay by KDM2A

A previously established histone demethylase assay was used to quantify KDM2A demethylase activity (Tsukada et al., 2005). In brief, histone demethylase assay was performed in a 15 µl reaction volume if it was destined for analysis by Western blotting or in a 30 µl reaction volume if it was to be processed for formaldehyde release. 0.5 µg of recombinant or calf thymus derived histones (for Western blot analysis) or 100 kcpm of ^{3}[H]-labelled nucleosomes (for formaldehyde release processing) were incubated with 2.5 µg of recombinant KDM2A protein in demethylase buffer (50 mM Hepes pH 7.9, 23 µM $Fe(NH_4)_2(SO_4)_2$, 1 mM α-KG, 2 mM ascorbate) for 1 hr (for formaldehyde release processing) or 3 hr (for Western blot analysis) at 37°C with occasional agitation. For Western blot analysis, the reaction was loaded onto a 15% SDS-PAGE and processed as described before. For formaldehyde release processing, the reaction volume was increased to 100 µl with BC buffer containing 150 mM KCl and 30 µg of BSA per reaction. 100 µl of 25% TCA was then added to the reaction and mixed thoroughly. The reaction was then incubated on ice for 15 min and centrifuged at 13,200 rpm, 4°C for 10 min. 180 µl of the derived supernatant was added to 200 µl of conversion reagent (30% ammonium acetate, 0.6% acetic acid, 0.2% 2,4-pentanedione). The mixture was incubated at 37°C for 25 min. The converted formaldehyde was extracted using

400 µl of 1-pentanol (Sigma-Aldrich). The mixture was briefly vortexed and centrifuged at 13,200 rpm, RT for 3 min. 350 µl of the separated upper phase was added to 2 ml of Ultima Gold scintillation fluid (Perkin Elmer) in a scintillation vial and thoroughly mixed. The radioactivity was then measured using a scintillation counter.

Histone methyltransferase assay by Set2

Histone methyltransferase (HMT) assay was performed in 25 µl reaction volume in the presence of 2 µl of ^{3}[H]-labelled SAM, 5 µl of 5X HMT buffer (100 mM Tris-HCl pH 8.0, 20 mM EDTA, 5 mM PMSF, 2.5 mM DTT) and 3-5 µg of mononucleosomes previously purified from chicken erythrocytes. This ratio of reagents was maintained during scale up reactions. The reaction mixture was incubated for 2 hr at 30°C and then dialysed overnight into 2 L of histone storage buffer (10 mM Hepes-KOH pH 7.9, 10 mM KCl, 0.2 mM PMSF, 10% glycerol). In the following morning, the histones were dialysed again into 2 L of fresh histone storage buffer for 4 hr and then stored at -20°C. To visualise the labelled histones, a small aliquot of the reaction was analysed by Coomassie stained 18% SDS-PAGE. Following destaining, 5 gel volumes of EN3HANCE solution (Perkin Elmer) was added to the gel and incubated for 40 min. The gel was then washed with H_2O for exactly 10 min and dried at 65°C on top of 2 layers of 0.3 mm Whatman paper under vacuum.

Autoradiography of the gel was taken overnight at -80°C using a BioMax XAR film (Kodak).

3. Chapter Three – Enzymatic activity and DNA binding property of KDM2A

3.1. KDM2A is a H3K36 mono- and di-demethylase

The initial report where the histone demethylase function of KDM2A was appreciated, focused on detailing the chemical nature of the JmjC domain mediated reaction and its substrate specificity (Tsukada et al., 2005). The study showed the specificity of KDM2A JmjC domain activity towards mono and di-methylated forms of lysine 36 on histone H3 by applying both *in vitro* activity assay and by KDM2A overexpression in a 293T cell line. Given that this was the first published study of KDM2A demethylase function, it was important to commence our current study by testing the validity of the published results. To this end, in the first part of this chapter, an *in vitro* system was developed to generate recombinant histone and chromatin substrates carrying defined methylation properties. In the presence of these substrates, the biochemical property of KDM2A JmjC domain was analysed using both qualitative and quantitative approaches.

In addition to the N-terminal JmjC domain and the C-terminal F-box domain, two other evolutionary conserved domains are present in KDM2A. These are the zinc finger CXXC domain and the plant homeodomain (PHD) (Figure 3.2A). As previously evidenced in other DNA-interacting proteins like CFP1 (Voo et al., 2000), the CXXC domain is a DNA binding module that specifically targets non-methylated CpG dinucleotides. However, its functional role within the context of the KDM2A demethylase remains uncharacterised. Therefore, within the second part of this chapter, the functional relevance of the CXXC domain in KDM2A is addressed by applying two different methods that are routinely used to study DNA-protein interactions. The PHD domain has been reported to act as a histone binding module with specificity towards both methylated and non-methylated forms of lysine 4 on histone H3 (Li et al., 2006; Peña et al., 2006; Ruthenburg et al., 2011; Shi et al., 2006a; Voo et al., 2000). The contribution of this protein domain to KDM2A function will be discussed in chapter 5.

Expression and purification of recombinant KDM2A

The exact enzymatic substrate of KDM2A was first determined by *in vitro* activity assay with full length recombinant human KDM2A protein expressed via the baculovirus expression vector system (BEVS) (Tsukada et al., 2005). Since the first introduction of baculovirus-mediated heterologous gene expression, this method has progressively become one of the most versatile and popular systems for

recombinant protein production. BEVS combines the large scale production level typically associated with prokaryotic expression systems, with the accuracy in translating large molecular weight proteins carrying post-translational modifications that is characteristic of the mammalian expression system (Jarvis, 2009). With the purpose of generating large amounts of recombinant full length KDM2A protein for subsequent biochemical characterisation, and considering its relatively large molecular weight (approximately 133 kDa), the BEVS methodology was initially adopted in this study.

The BEVS protocol described here involves a number of key stages (Figure 3.1) (Jarvis, 2009). Firstly, a full length (aa 1-1162) N-terminal Flag-tagged human KDM2A construct that included the four previously mentioned conserved domains (i.e. JmjC, zinc finger CXXC, PHD and F-box) was cloned into a LIC (Ligation independent cloning)-adapted version of the commercial pFastBac donor plasmid where the inserted gene is flanked by two attTn7 transposon sequences (Figure 3.1). The donor plasmid was then transformed into a competent DH10Bac strain of *E. coli* that carries the baculovirus genome, denoted as bacmid, designed to contain the acceptor sites for attTn7. Within the DH10Bac cell, additional helper plasmids are responsible for encoding the bacterial transposon Tn7, which specifically recognises attTn7 sequences. Site-specific transposition then allowed for the precise transfer of the KDM2A gene into the bacmid. Because the transposition event was engineered to disrupt a local *lacZα* gene, cells carrying successfully transposed bacmid were identified as white colonies on X-gal (5-

bromo-4-chloro-indolyl-β-D-galactopyranoside)-rich culture plates in a classical blue/white screen. However, as the breakdown of X-gal into the insoluble blue coloured product is not immediate, cells carrying unaltered bacmid DNA may show delayed manifestation of blue colouring. To avoid selecting false positives, suspected white colonies were re-streaked on fresh X-gal-containing plates. After further incubation, genuine positive colonies could be in such way confirmed. In the final step, the bacmid DNA was isolated from a single successful clone and purified using standard alkaline extraction method. The correct insertion of the KDM2A gene was verified by basic PCR using M13 primer pairs, and the bacmid was subsequently transfected into the Sf9 (*Spodoptera frugiperda*) host insect cell line to initiate generation of KDM2A-specific baculovirus.

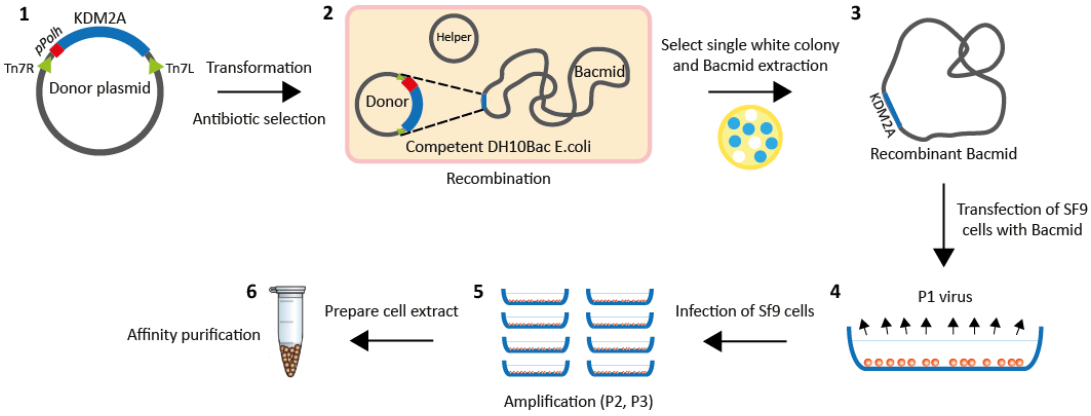


Figure 3.1 – Baculovirus expression vector system (BEVS)

Donor/shuttle vectors carrying the KDM2A gene (1) are initially transformed into DH10Bac *E. coli* cells that contain the bacmid genome (2). Site-specific transposition then ensures for the correct insertion of KDM2A into the bacmid genome. Inclusion of the foreign gene disrupts a local *lacZα* gene allowing for

blue/white selections of colonies that carry successfully modified bacmids (3). Following extraction, the bacmid is transfected into host insect cells for viral production (4). The later derived P1 virus is subject to two rounds of amplification to obtain a high titre P3 viral stock (5). P3 baculovirus is then used to induce expression of recombinant KDM2A protein that can be purified via tag-affinity (6).

The viral titre is one of the most critical factors in determining the efficiency of protein expression in BEVS. Given a fixed amount of virus inoculum per each litre of cell culture, higher viral titre generally correlates with higher protein production until a saturation point is reached. In order to maximise KDM2A-specific baculovirus titre, two rounds of amplification were carried out following transfection stage to produce P2 and P3 viral stocks. Generally, a viral plaque assay is preferred to quantitatively determine the approximate viral titre, expressed in the form of 10^X plaque forming units (pfu)/ml of inoculum (Jarvis, 2009). However, for the purpose of this study, the viral titre was semi-quantitatively inferred from the level of protein production. Indeed, at each infection step, Western blot analysis of aliquots from the same number of infected Sf9 cells indicated an incremental rise in protein expression level (data not shown). The recombinant protein was visualised using both antibodies specific for Flag and KDM2A. It was observed that, after two rounds of infection, 1 µl of the derived P3 viral stock was sufficient to induce an optimal level of protein expression when applied to a single well containing $2x10^6$ Sf9 cells (Figure 3.2B). By repeating the same procedure, a number of KDM2A JmjC, CXXC and PHD mutant-specific P3 baculoviruses was generated alongside the wild type (WT) equivalent (Figure 3.2C). A complete list of

KDM2A-specific baculoviruses with a description of each construct feature is illustrated in Table 1.

At the end of the infection period, standard Flag affinity purification protocol (Tsukada et al., 2005) was applied to isolate the recombinant protein from harvested Sf9 cell extract. Surprisingly the elution fractions contained a relatively small proportion of intact full length KDM2A protein, visible in the top region of the gel, compared to a number of lower bands likely corresponding to degradation products of KDM2A (Figure 3.2D). It is known that degraded proteins generally occur due to the presence of cellular proteases that are released into the medium during cell lysis. To counteract proteolytic activity, full spectrum protease inhibitors are routinely included in protein purification buffers during the extraction stage. However, because of their relatively short half-life and high cost, protease inhibitors are not usually supplemented in cell culture medium. To limit the level of degradation accumulated during the infection stage, the standard 72 hr incubation time was reduced to 48 hr to the expense of recombinant protein yield. Despite shortening of the infection time, considerable amounts of degraded proteins persisted in the purified fractions. Change of cell growth medium recipe or introduction of a new Sf9 cell batch also did not improve the preservation of intact full length KDM2A protein. Indeed, more than 80% of the purified KDM2A protein was in the degraded form and consequently unsuitable for further biochemical studies. In light of this result, an alternative expression system was needed.

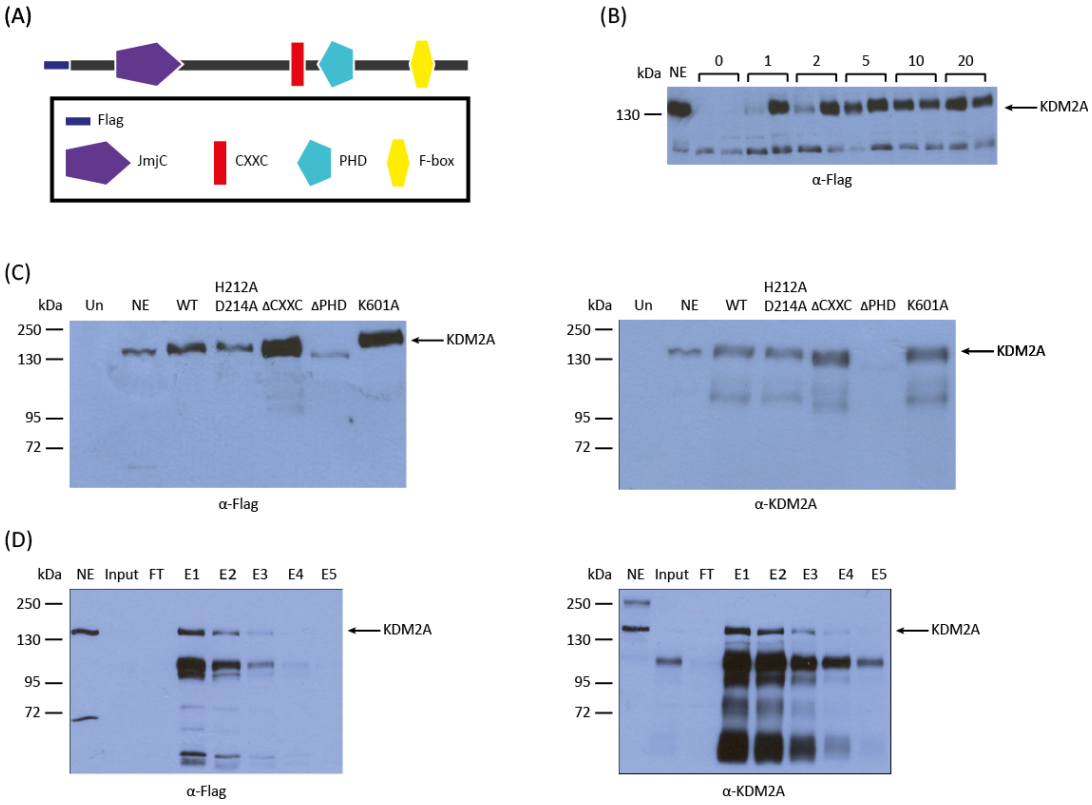


Figure 3.2 – Expression and purification of full length KDM2A via BEVS.

(A) Full length KDM2A protein contains four conserved domains: the JmjC catalytic domain, the zinc finger CXXC domain, the PHD domain and the F-box. (B) The P3 virus was titrated by applying 0 to 20 µl of virus per $2x10^6$ Sf9 cells. Flag-specific antibodies were used to evaluate the level of recombinant protein production. Shown here is a representative Western blot of KDM2A WT P3 virus titration assay. This is from one of two KDM2A WT expression trials attempted in this study. (C) Flag and KDM2A-specific antibodies were used to assess the expression level of KDM2A wild type (WT) and mutant constructs including H212A D214A, ΔCXXC, ΔPHD and K601A. A detailed description of each KDM2A baculovirus construct is given in Table 1. The α-Flag and α-KDM2A Western blots summarise the expression efficiency of each of the KDM2A baculovirus constructs generated in this study. Except for KDM2A WT construct, all other KDM2A mutant P3 viruses have been generated once. Un, uninfected. NE represents nuclear extract containing full length Flag-tagged KDM2A derived from mammalian 293T cell. Here it is used as a Western blotting control and molecular weight marker for full length KDM2A protein. (D) α-Flag and α-KDM2A Western blots showing the elution fractions of Flag-affinity purification where the full length KDM2A can be seen on top and a number of degradation products below. This is from one of two KDM2A WT purification trials attempted in this study. A detailed description of the Flag-affinity

purification protocol used is given in section 2.8 of Materials and Methods. FT, flow through.

Previous efforts by other members of our group have attempted to express full length KDM2A protein via a prokaryotic approach. For this purpose, an N-terminal 6xHis and C-terminal Strep II tagged KDM2A construct was initially cloned into a LIC-modified pNIC28 plasmid and transformed into a BL21 strain of *E. coli* cells. The presence of two distinct tags at both ends of the construct is designed to ensure for the isolation of full length protein after two consecutive affinity purification steps.

Given that bacterial protein synthesis machinery is not evolutionary fine-tuned to process large molecular weight proteins like KDM2A, it was surprising to notice that a good level of KDM2A expression could be achieved in transformed BL21 cells upon 3 hr induction with 0.5 ml of 1 M IPTG (Isopropyl β-D-1-thiogalactopyranoside) per 0.5 L cell culture (Rob Klose, personal communication).

A double affinity purification protocol routinely used in our laboratory involves firstly the application of the extracted cell lysate to a 0.5 ml Ni-NTA resin column and elution at 250 mM imidazole concentration (Figure 3.3A). A pool of the eluted fractions is subsequently loaded onto an equivalent size StrepTactin resin column. The Ni-NTA resin has the advantage of providing high recombinant protein binding capacity and is particularly suitable for the initial protein “capturing” stage.

However, a number of histidine-rich or metal-chelating bacteria proteins are notoriously found in Ni-NTA elutions and can affect the overall purity of the recombinant protein (Bolanos-Garcia and Davies, 2006). Therefore a subsequent protein "polishing" step using the StrepTactin resin, which displays much higher recombinant protein binding selectivity, is required.

Despite the high level of protein expression observed, early attempts to purify full length recombinant KDM2A by applying the double affinity protocol were unsuccessful. Indeed, fractions eluted from the Ni-NTA column failed to bind to the following StrepTactin resin with most of the KDM2A protein carried over in the flow through (Rob Klose, personal communication). As StrepTactin resin is designed for protein purification under physiological condition, the buffers employed should not disrupt tag-resin interaction. Instead, loss of binding can occur if the C-terminal tag is physically inaccessible due to spontaneous folding of this protein end. Alternatively, the tag could be absent as a result of early protein synthesis termination or post-translational proteolytic cleavage.

In the absence of any structural information regarding potential folding features at KDM2A C-terminal end, it is difficult to rationally engineer a functional affinity tag. Instead, a series of C-terminally-truncated versions of the original construct was cloned, expressed and extracted from transformed BL21 cells in order to systematically determine the largest functional C-terminal Strep II tagged-construct that could be obtained under the prokaryotic expression system. After

multiple trials, a 747 aa construct, named KDM2A6 and containing the JmjC, CXXC and PHD domains but lacking the F-box C-terminal region (Figure 3.3C), was observed to stably interact with the StrepTactin resin and elute in the presence of 2.5 mM desthiobiotin, a binding competitor of the Strep II tag (Figure 3.3B). Quantitative assessment of the proteins eluted from the StrepTactin column via Bradford assay at OD 595 nm provided an estimation of the average protein yield to be around 0.5 mg per litre of cell culture.

Because the construct described above includes both the catalytic domain and the putative DNA/chromatin binding modules, it represented the best attainable candidate for subsequent enzymatic activity and chromatin binding analysis. The KDM2A6 version is hereafter considered as the “WT” protein and all the mutant species characterised in this study were derived from this specific construct (Table 1).

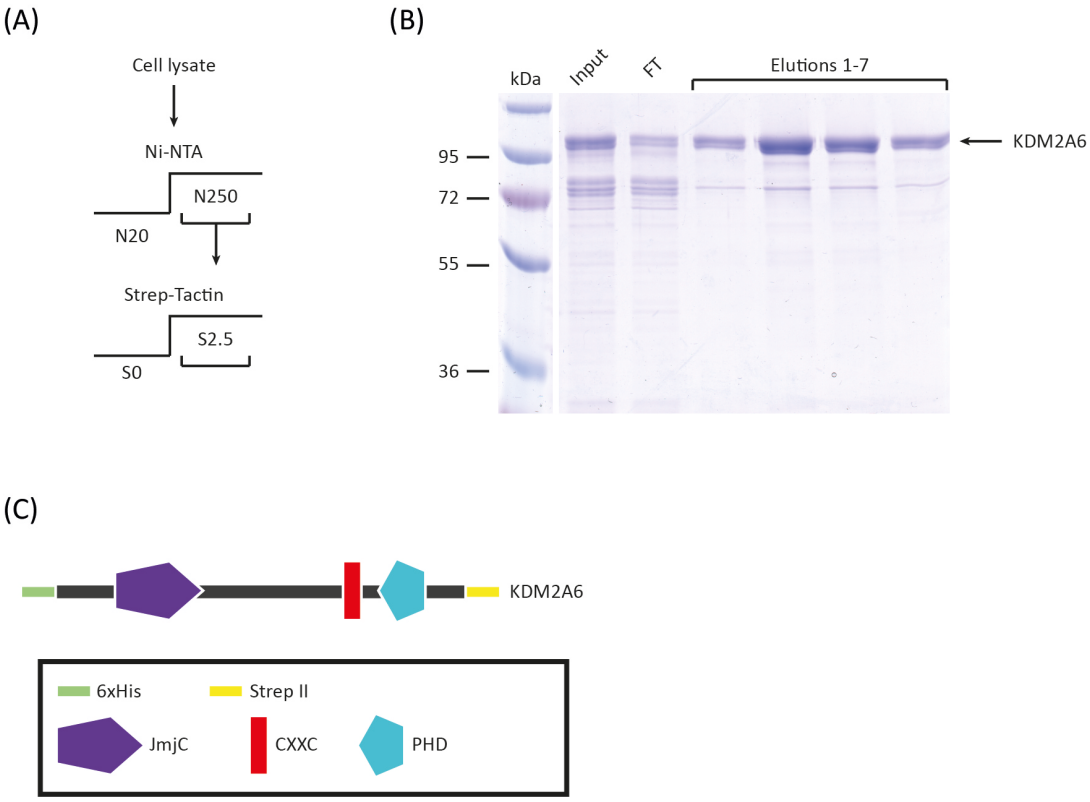


Figure 3.3 – Purification of KDM2A6

(A) Double affinity purification of KDM2A6 expressed in BL21 cells involves first capturing by Ni-NTA column followed by StrepTactin resin-based polishing. The concentrations of the respective binding competitors imidazole and desthiobiotin used during wash and elution stages are indicated. A detailed description of the double-affinity purification protocol used is given in section 2.8 of Materials and Methods. (B) Coomassie stained 8% SDS-PAGE showing a number of elutions derived from the StrepTactin column. An estimated over 80% pure recombinant KDM2A can be achieved. The level of recombinant protein purity shown in this representative gel is consistently observed in more than three repeat purifications performed in this study for KDM2A6. The same level of purity can also be reproducibly applied to all KDM2A6-derived mutant proteins. FT, flow through. (C) Schematic illustrating the conserved JmjC, CXXC and PHD domains in KDM2A6.

KDM2A6 removes H3K36 mono- and di-methylation from histones *in vitro*

To verify that the KDM2A6 construct described above is indeed a *bona fide* WT enzyme, an *in vitro* activity assay was carried out in the presence of both calf thymus-derived and synthetically modified H3 histones carrying all three H3K36 methylation states under previously published demethylase assay conditions (Tsukada et al., 2005). The resulting product of the enzymatic reaction was probed with antibodies specific for non-methylated, H3K36me1, H3K36me2 and H3K36me3 histones and visualised by Western blotting. In agreement with previously published data, the presence of 2.5 μg of KDM2A6 caused complete depletion of H3K36me1 and H3K36me2 state in 0.5 μg of both calf thymus and synthetically derived H3 histones (Figure 3.4). However, probing of the same reaction products using H3K36me3-specific antibody failed to generate detectable chemiluminescent signal, even in the control demethylase reactions where KDM2A6 was absent. In a number of replicate experiments, inconsistent results were observed with the above batch of H3K36me3-specific antibody (data not shown). Owing to a lack of reproducible results for the H3K36me3-specific blot, it is not possible in this experiment to evaluate KDM2A6 activity towards H3K36me3 substrates. Therefore although published results indicate the specificity of this enzyme towards mono and di-methylated forms of H3K36, the experiments here do not unequivocally demonstrate the specificity of my enzyme preparation

towards these substrates. This could be more rigorously addressed using alternative more direct measurements of the reaction products, for example by using mass spectrometry based analysis of H3K36 methylation.

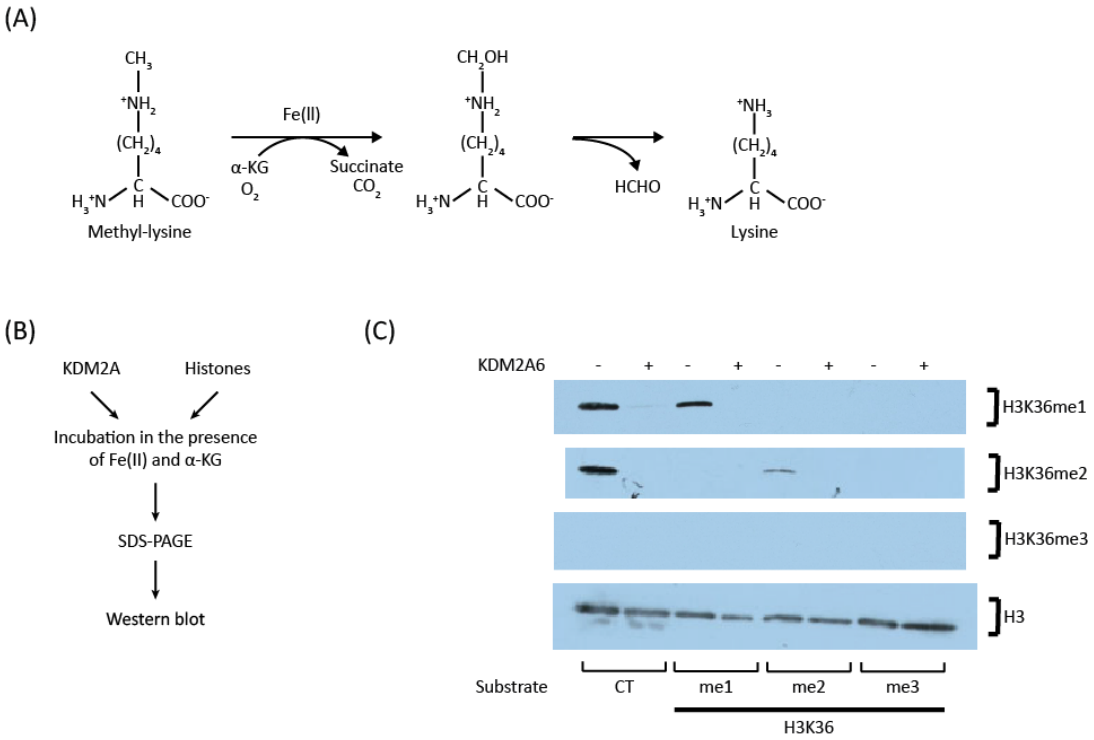


Figure 3.4 – KDM2A6 removes H3K36me1 and H3K36me2 from calf thymus and recombinant histones

(A) Demethylation by the JmjC domain occurs via a hydroxylation reaction that requires Fe(II) and α-KG as cofactors and generates formaldehyde as by-product. (B) Recombinant KDM2A6 was incubated with either calf thymus (CT) extracted or synthetically mono-, di- and tri- H3K36 methylated histones in the presence of the cofactors. The reaction was allowed to proceed for three hours at 37°C to ensure saturation of the enzymatic product, which can then be separated on SDS-PAGE and analysed by Western blotting. (C) The modified histones were probed using antibodies specific for K36me1 (first gel from the top), K36me2 (second gel) and K36me3 (third gel) modified and all methyl state (bottom gel) H3 histones. The H3K36me3-specific antibody failed to generate any signal. The assay was repeated

more than three times with inconsistent results with respect to the effects on H3K36me3. Therefore, this experiment is not informative with regard to KDM2A6 activity towards H3K36me3 substrates. Consequently, its expected specificity towards mono and di-methylated substrates cannot be commented on. It can also be noted that signal intensities from the H3K36me2-specific antibody are different between reactions containing calf thymus histones or synthetically di-H3K36 methylated histones in the respective no-enzyme control lanes. This is possibly due to discrepancies in the amount of H3K36me2-modified histone estimated for the two substrates.

Western blotting is a method that is inherently dependent on the quality of the antibodies used and can provide only a semi-quantitative readout of KDM2A demethylase activity. An alternative, more quantitative measurement of demethylase activity is through detection of formaldehyde that accumulates over time, as the level of release of this by-product is stochiometrically related to the number of methyl groups removed (Figure 3.4A). Previous detection of formaldehyde has relied on labelling of the histone methyl group with radioactive isotopes like ^{3}[H] which is transferred to the formaldehyde molecule as a consequence of demethylation (Tsukada et al., 2005). Recently, a radioisotope-independent assay for the detection of formaldehyde in solution has been developed which takes advantage of the formaldehyde dehydrogenase (FDH) enzyme (Lizcano et al., 2000). This protein converts formaldehyde to formic acid in a NAD^{+} dependent manner. As NAD^{+} is reduced to NADH, the latter can be measured fluorometrically at OD 340nm (Figure 3.5A).

Recombinant FDH proteins derived from the *P. putida* strain of bacteria are commercially available but are expensive and impure. The molecular size of full length FDH protein (approximately 42 kDa) makes it well suited for cloning into the previously utilised pNIC28 vector and expression by BL21 cells. The construct was engineered to carry an N-terminal 6xHis tag, enabling purification via Ni-NTA chromatography (Fig 3.5B and C). The purity of the Ni-NTA collected fractions was further refined by subjecting the protein sample to a 1 ml Mono Q anion exchange column, following which a single protein species of the expected molecular weight was observed (Figure 3.5B and D).

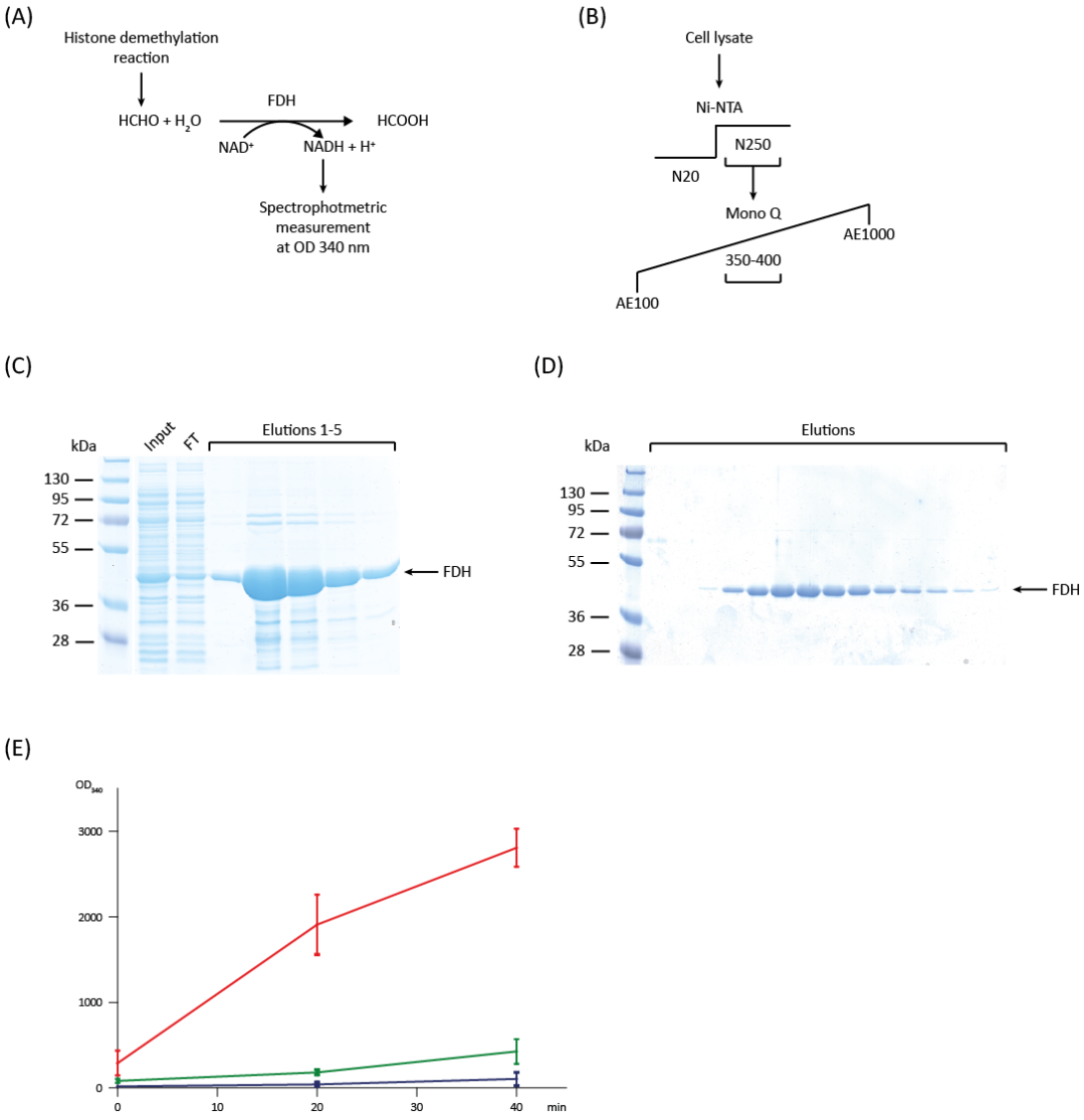


Figure 3.5 – Formaldehyde dehydrogenase (FDH)-coupled demethylase assay

(A) FDH converts formaldehyde generated from the demethylation reaction to formic acid by reducing NAD^+ to NADH. The latter can be spectrophotometrically measured at OD 340 nm. (B) Recombinant FDH expressed in BL21 cells was first isolated via Ni-NTA column followed by an ion exchange (Mono Q) polishing step. The concentration of imidazole in the wash and elution stage of Ni-NTA purification is indicated. The salt gradient applied in the Mono Q and the salt concentration estimated in FDH elution fractions is indicated. (C) Elevated amounts of FDH can be expressed and captured by the Ni-NTA resin as evidenced by SDS-PAGE of the first five elutions. FT, flow through. (D) These were pooled together and applied to the Mono Q anion-exchange column to obtain single band-pure FDH. (E) FDH-coupled demethylase assay was carried out in the absence (blue line) and

presence (green line) of KDM2A6. The same demethylation reaction condition used for Western blot analysis as detailed in section 2.11 of Materials and Methods was used. For each reaction, 0.5 µg of calf thymus histones was used as substrate. Red line corresponds to FDH-dependent NADH fluorescence signal in the presence of 100 mM HCHO. This serves as a positive control for FDH activity. Error bars correspond to standard errors of the means (SEM) derived from three replicate assays.

To quantify KDM2A6 activity using the FDH-coupled assay, the same demethylation reaction conditions described for the above Western blot analysis (section 2.11 of Materials and Methods) were reproduced. At the end of the standard 3 hr enzymatic incubation, 0.3 µg of Mono Q-purified FDH protein together with 6 µl of 5 mM NAD^+ was added to each reaction. Fluorescent measurements were taken at 0 min, 20 min, 40 min time points (Figure 3.5E).

It was immediately apparent that despite the purified FDH enzyme being fully functional, as evidenced by the high fluorescence signal in the presence of formaldehyde (Figure 3.5E, red line), only a modest increase above background level of fluorescence signal could be detected when KDM2A6 was present in the reaction (Figure 3.5E, green line). Importantly, the same degree of fluorescent signal was observed in the presence of the control enzyme Rph1 (data not shown), a JmjC domain containing and H3K36me2/me3-specific demethylase from budding yeast (Klose et al., 2007). Therefore, the weakness of fluorescence signal observed is unlikely due to poor KDM2A6 demethylase activity. Instead, it is possible that the amount of formaldehyde accumulated as a by-product of JmjC domain catalysis is

rather small and probably close to the detection limit of the FDH assay. It was decided that such sensitivity issues would make comparative analysis between WT and mutant KDM2A protein activity difficult. Therefore, an alternative system was required to quantify KDM2A activity with better sensitivity.

KDM2A removes H3K36 methylation from nucleosomes *in vitro*

As mentioned earlier, methods of measuring demethylase activity-associated formaldehyde release previously relied on enzymatic ^{3}[H]-labelling of the modified histone lysine residue. A radioactivity-based assay offers two major advantages over other methods for measuring enzymatic activity in general and H3K36 demethylase activity in particular. Firstly, it provides unparalleled detection sensitivity making it particularly suited for *in vitro* assays in which the level of measured product is limited. Secondly, because the yeast H3K36 methyltransferase Set2 is able to catalyse the addition of methyl groups exquisitely in a nucleosomal context (Morris et al., 2005; Strahl et al., 2002), it allows for the introduction of Set2-modified nucleosome as a more biologically relevant substrate for testing KDM2A demethylase activity.

To generate recombinant Set2 enzyme, a gift plasmid pCAL-n carrying the *S. pombe* full length *set2* gene was obtained from Brian Strahl's laboratory (Morris et al., 2005). This construct has been previously engineered to contain an N-terminal

calmodulin-binding peptide (CBP) tag to facilitate extraction via calmodulin (CaM) affinity resin (Figure 3.6A). CaM interacts with the CBP tag in the presence of Ca^{2+}, a process that occurs naturally in muscle cells from which the CBP tag was originally derived. Given the stability of the interaction and the fact that Ca^{2+} is the only requirement for binding, the CaM-CBP system is particularly favoured for purification of recombinant proteins under very mild conditions. Indeed, 2 mM Ca^{2+} was sufficient to allow CBP-tagged Set2 binding to the CaM resin. The Set2 protein was subsequently eluted off the column by addition of 2 mM Ca^{2+} chelating agent EGTA (Figure 3.6A).

In order to test the histone methyltransferase activity of the purified Set2 and to titrate the optimal concentration of enzyme to be used, a series of aliquots containing up to 10 µg of purified Set2 was incubated each with 3 µg of chicken erythrocyte-derived mononucleosomes in the presence of ^{3}[H]-labelled SAM. Any unincorporated SAM was subsequently dialysed away and the methylated products were separated by SDS-PAGE. The histones were then visualised by autoradiography to assess the amount of ^{3}[H]-containing methyl-groups installed (Figure 3.6B and C). It was observed that 2.5 µg of recombinant Set2 was able to optimally methylate 3 to 5 µg of chicken nucleosomes. This ratio was maintained during subsequent scale up reaction to generate methylated nucleosome substrate for KDM2A activity assay.

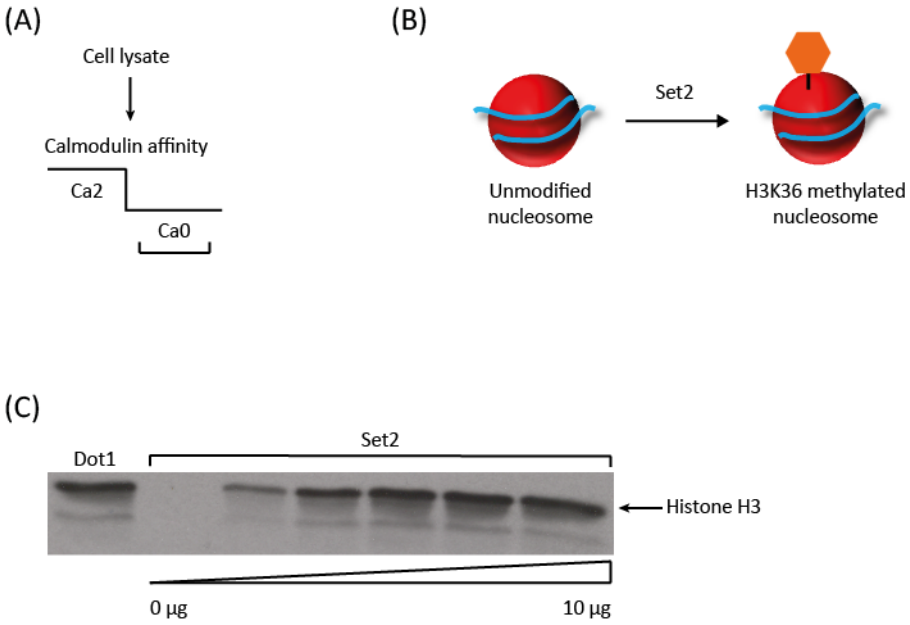


Figure 3.6 – Histone H3K36 methyltransferase assay by recombinant yeast Set2

(A) Recombinant CBP-tagged yeast Set2 expressed in *E. coli* was captured by CaM affinity resin in the presence of 2 mM Ca^{2+} and subsequently eluted in the absence of Ca^{2+}. (B) *S. pombe* Set2 mediates the installation of all three methylation states on H3K36 residue specifically in a nucleosomal context (Morris et al., 2005). (C) Autoradiography of Set2 methylated histone H3 showing Set2 concentration-dependent increase in methylated H3 substrate. A detailed description of the methyltransferase assay is given in section 2.11 of Materials and Methods. Dot1 is an H3K79-specific histone methyltransferase enzyme and is used as a positive control in the assay here.

To measure KDM2A6 demethylase activity on Set2-labelled nucleosomes, 2.5 µg of recombinant KDM2A6 was first incubated with 100 kcpm of ^{3}[H]-labelled nucleosome under the same demethylase assay conditions described before except for the incubation time, which was reduced to 1 hr (Figure 3.7A). Following

incubation, the formaldehyde-containing soluble fraction was isolated and processed for scintillation counting.

In clear contrast to the FDH-coupled assay, a strong ^{3}[H] signal was readily observed when KDM2A6 was present in the reaction, demonstrating that the enzymatic activity towards H3K36 methylated histone substrates is conserved in a nucleosomal context (Figure 3.7B). To verify that the observed demethylase activity is indeed attributable to KDM2A6 and more importantly that such activity is dependent on an intact JmjC domain, a catalytic mutant construct (KDM2A6 H212A) was generated and its demethylase activity compared to that of the WT protein (Figure 3.7B and C). Substitution of the first histidine residue, within the aforementioned His-X-Asp/Glu-X_n-His motif that is responsible for Fe(II) coordination, induced a significant decrease in KDM2A demethylase activity (Figure 3.7B). Nevertheless, the H212A mutant still retained more than 50% of WT protein activity suggesting that although its capacity to coordinate Fe(II) was diminished, it has not been totally abolished. To investigate further, 0.25 mM of the chelating agent EDTA was added to the reaction at the time of incubation to sequester free Fe(II) ions. As expected, following EDTA addition, the mutant KDM2A lost all enzymatic activity likely due to its inability to compete for Fe(II) binding (Figure 3.7B). In contrast, the WT protein still displayed partial demethylase activity. This is probably due to the binding affinity of the JmjC domain active site being sufficiently strong to retain Fe(II) derived from cell culture medium. Preservation of

Fe(II) binding is, however, unlikely to occur when the active site is mutated as in KDM2A6 H212A.

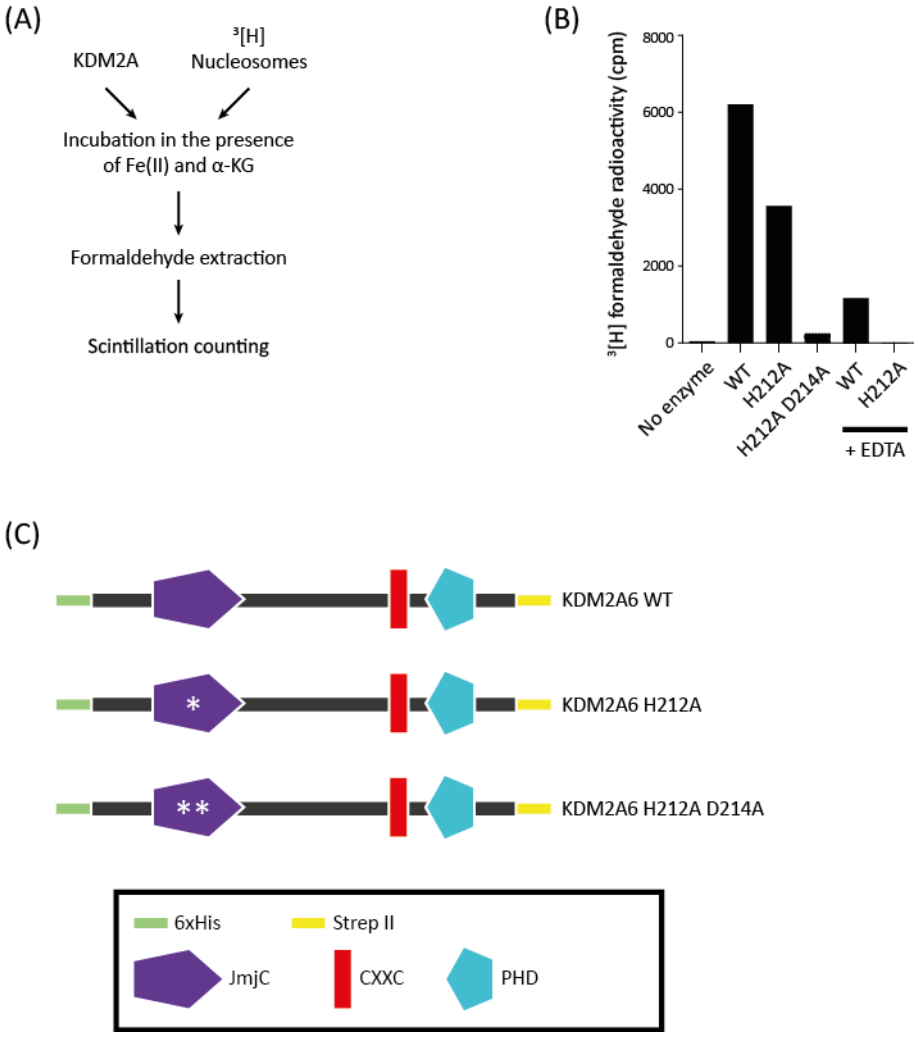


Figure 3.7 – KDM2A removes H3K36 methylation from nucleosomes *in vitro*

(A) Recombinant KDM2A6 was incubated together with ^{3}[H]-labelled nucleosomes in the presence of the cofactors and under the same experimental conditions described before for Western blot analysis except for the incubation time, which was limited to 1 hr. The protein components were then precipitated by trichloroacetic acid (TCA) and the remaining formaldehyde-containing soluble fraction was processed for scintillation counting. Details of the experimental

procedure are given in section 2.11 of Materials and Methods. (B) The WT KDM2A protein displayed strong demethylase activity towards Set2-modified nucleosomes while the catalytic mutant H212A also retained partial activity. This activity was however completely abolished in the presence of 0.25 mM EDTA indicating that the mutation had indeed weakened the capacity to coordinate Fe(II), which is essential for catalytic activity. The introduction of a second adjacent mutation, D214A, led to full ablation of demethylase activity in the absence of EDTA. This specific demethylase assay in the presence and absence of EDTA was performed once. (C) Schematic illustrating the main conserved domains in KDM2A6 WT and catalytic mutant proteins. Each single point mutation is indicated with an asterisk.

To generate a catalytically deficient protein that is completely devoid of any demethylase activity, a second mutation was introduced on the background of the H212A mutation to generate KDM2A6 H212A D214A double mutant. When subject to activity assay, the concomitant substitution of both histidine and aspartate residues within the Fe(II)-binding motif resulted into total abrogation of demethylase activity in the absence of EDTA (Figure 3.7B).

3.2. KDM2A specifically recognises non-methyl CpGs through the zinc finger CXXC domain

KDM2A binds preferentially to non-methylated CpG DNA *in vitro*

Knowledge of the enzymatic activity does not immediately translate into an understanding of the protein biology, as additional factors need to be considered, including the cellular context where the specific activity takes place. Within the context of the KDM2A protein, the catalytic JmjC domain exists in combination with highly conserved CXXC and PHD domains that have known DNA or histone-binding roles in other chromatin enzymes. Each of these domains can potentially contribute to KDM2A biology.

It was previously reported that, in HeLa cells, KDM2A colocalises with HP1 at heterochromatic regions where it contributes to the maintenance of a repressed chromatin state (Frescas et al., 2008). Interestingly, it was observed that KDM2A proteins bearing mutated zinc finger CXXC domains failed to localise to pericentric heterochromatin suggesting that this putative DNA binding module may target KDM2A to specific genomic regions *in vivo*. Considering the well characterised binding specificity of the CXXC domain for non-methylated CpG DNA observed in other CXXC-containing proteins like CFP1 (Voo et al., 2000), it was argued that

KDM2A is likely to recognise via the CXXC domain CGI sequences within the heterochromatin domains.

The conclusions from the study described above somewhat conflict with the generally accepted idea that CGIs predominantly associate with open, H3K4me3-rich, euchromatic regions, typically at the 5' end of genes (Illingworth and Bird, 2009). To reconcile these apparently conflicting observations, an understanding of the molecular mechanism that mediates KDM2A interaction with DNA is imperative. Given the correlation between KDM2A genomic loci and CGIs, the zinc finger CXXC domain is an obvious candidate to examine. Therefore, to investigate the biochemical properties of the CXXC domain in relation to KDM2A targeting specificity observed *in vivo*, an *in vitro* DNA binding assay based on the electrophoretic mobility shift (EMSA) method was first devised.

Historically, EMSA assays have been extensively used to characterise the affinity of nucleic acid-interacting proteins for particular nucleotide sequences (Garner and Revzin, 1981). The method is based on the principle that protein associated-DNA/RNA complexes migrate slower than free, unbound DNA/RNA probes in a gel. This migration difference is usually denoted as an upward "shift" of the bound complex with respect to the free material.

Previous SELEX-based *in vitro* studies of nucleosome-DNA interaction have uncovered a number of synthetic DNA sequences that display high nucleosome positioning capacity (Lowary and Widom, 1998). Interestingly, among those

oligonucleotides, a particular sequence, named 601, has been noticed to carry 14 CpG dinucleotides within a 147 bp interval (See sequence details in Appendix). The 1.1 obs/exp CpG density and 59% CG content of this artificial sequence is characteristic of nucleotide features that CGI regions display *in vivo* (Illingworth and Bird, 2009). Because of the sequence similarity to CGIs, the 601 sequence represents an ideal probe to test KDM2A binding to these genomic elements *in vitro*.

For this purpose, a 147 bp long, non-methylated 601 sequence was end-labelled with ^{32}[P] and an aliquot corresponding to approximately 5,000 cpm/µl of such labelled probe was incubated with 2 µg of KDM2A6 in the presence of 0.25 µg of poly dAdT. The latter serves as a binding competitor to reduce background noise associated with non-specific protein-DNA interactions due to the inherent "stickiness" of the reaction components. Following incubation, any complex species were separated by native 1.3% agarose gel and an autoradiography was taken. As shown in Figure 3.8C (Left panel), KDM2A6 displayed very strong binding to the 601 DNA evidenced by the formation of a slow migrating DNA-protein complex (i.e. the shift) and complete depletion of the free DNA probe (bottom of the gel). Importantly, this interaction is not dependent upon demethylase activity as evidenced by the fact that the double mutant H212A D214A was as proficient as the WT protein in binding DNA (3.8B and C, left panel).

To investigate whether the observed DNA interaction is, as hypothesised, due to the zinc finger CXXC domain activity, a deletion mutant of this domain (KDM2A6 ΔCXXC) (Figure 3.8B) was assayed for DNA binding under the same EMSA conditions. It was observed that, in the absence of the CXXC domain, KDM2A interaction with the 601 DNA was ablated, in fitting with the initial hypothesis that the KDM2A CXXC domain acts as a DNA-binding module (Figure 3.8C left panel). However, it is reasonable to argue that by deleting a significant portion of the protein, one can potentially undermine its overall structural stability, which may lead to an indirect loss of function that is not strictly related to the activity exerted by the deleted region. Instead, site-directed mutagenesis of one or a few key functional residues is generally preferred to minimise structural influences.

Two previous structural studies of the MLL1 CXXC domain gave a detailed account of the DNA binding interphase involved in CpG base pair recognition within this protein domain (Allen et al., 2006; Cierpicki et al., 2009). By referring to these reports, a sequence alignment analysis of CXXC domains derived from a group of DNA binding proteins including MLL1 was performed to identify the functionally equivalent residues that are likely to be conserved in KDM2A (Figure 3.8A).

In the MLL1 CXXC domain, residues R1182 to C1188 are known to form an extended loop structure that can insert into the major groove of the DNA helix. In particular, the four residues I1184, K1185, K1186 and Q1187 cooperate to specifically interrogate two CG pairs mainly through formation of hydrogen bonds

(Cierpicki et al., 2009). As revealed by the alignment, these key residues were found to be highly conserved in KDM2A suggesting that they may play a similar function in this protein.

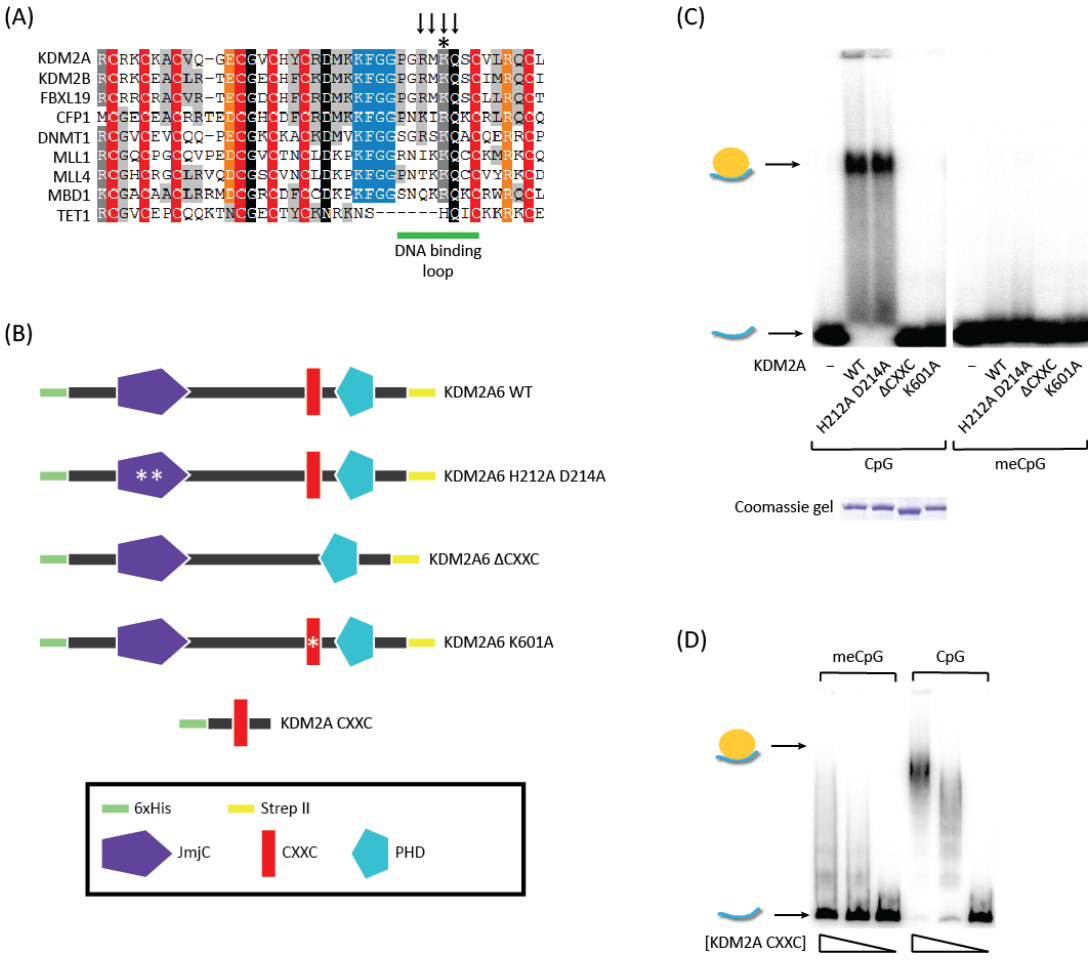


Figure 3.8 – KDM2A preferentially binds to non-methylated CpG DNA

(A) Sequence alignment showing the high conservation of the CXXC domain among previously characterised DNA binding proteins including CFP1, DNMT1, MLL1, MLL4, MBD1 and TET1. Highlighted in red are the eight cysteines responsible for the coordination of two zinc ions and serve to dictate the overall fold of the CXXC domain. In orange are the arginine and aspartate/glutamate that are known to form together a salt bridge. This interaction is important for the stability of the fold.

Also shown in blue is the conserved KFGG motif that is believed to induce the turn observed in the crescent fold. Finally, underlined in green is the DNA binding interface where the key residues that mediate CG base pair recognition are indicated with arrows. The K601A mutation has been marked with an asterisk. TET1 is the only protein in this group whose CXXC domain does not carry DNA binding activity (Frauer et al., 2011). Indeed, the CG-recognition interface, the salt bridge and the KFGG motif are all absent. FBXL19 appears to be a product of KDM2A gene duplication and deletion. It lacks the N-terminal JmjC domain but carries a fully conserved CXXC domain. (B) Schematic illustrating the KDM2A constructs whose DNA binding properties were tested by EMSA. Each point mutation is represented by an asterisk. (C) The identities of the migrating species in the EMSA gel are illustrated through carton representation on the left. KDM2A6 (yellow oval) WT protein binds strongly to the 601 DNA (short blue thread) in a CXXC domain dependent fashion since deletion or point mutation of one of the key binding residues in that domain leads to a complete abrogation of binding. Given the number of CpGs in the 601 DNA sequence, more than one binding events are predicted to occur. Indeed, in the lanes corresponding to KDM2A6 WT and H212A D214A, the lack of any intermediate complex species in the gel is indicative of saturation of all available binding sites on the 601 DNA probe under the concentration of KDM2A6 protein used. Given the CXXC domain specificity for non-methylated CpGs, this interaction is dependent on the methylation status of the DNA probe. The DNA probes in the left panel carry non-methylated CpGs while the ones in the right panel carry methylated CpGs (meCpG). The Coomassie gel underneath the left panel shows that an equal amount of protein (WT or mutant) was used in each binding reaction. A detailed description of the EMSA protocol is given in section 2.10 of Materials and Methods. Detailed sequence of the 147 bp 601 DNA probe is shown in the Appendix. Shown here is a representative of two replicate assays. (D) A KDM2A construct carrying the CXXC domain alone is sufficient to mediate DNA binding. The DNA probes used are the same ones in (C). In the left and right panels, binding to methylated and non-methylated probes are shown respectively. In this assay three decreasing concentrations of KDM2A CXXC protein were used. In the right panel, middle lane, corresponding to an intermediate concentration of KDM2A CXXC, several unsaturated complex species are visible. In contrast, under saturating concentration of protein (right panel, left lane), one predominant complex species is seen. Shown here is a representative of two replicate assays.

One of these residues, K601, the equivalent of K1186 in MLL1, was mutated and its role in DNA binding tested by EMSA (Figure 3.8B). As shown in figure 3.8C (Left

panel), alanine substitution of this particular lysine residue led to a complete abrogation of DNA interaction that is comparable to that observed in the case of a whole CXXC domain deletion. Importantly, because K601 is not involved in determining the overall folding of the CXXC domain (Allen et al., 2006), the lack of DNA binding observed is not a consequence of change in domain structure.

In light of this result, it is reasonable to conclude that, at least *in vitro*, KDM2A binding to DNA is mediated by the CXXC domain, in particular the K601 residue that plays an essential role in CpG recognition. In further support of this finding, CpG methylation of the 601 DNA probe by the *E. coli* DNA methyltransferase M.SssI prevents DNA binding by all KDM2A constructs tested (Figure 3.8C, right panel). Moreover, EMSA assays using a KDM2A construct that carries only the CXXC domain indicate that this domain alone is sufficient for DNA binding and other protein domains are unlikely to participate in this interaction (Figure 3.8D).

In a compartmentalised cellular environment, the targeting mechanism of an enzyme directly determines its local concentration at a given genomic locus. This in turn, together with the local availability of the preferred target substrate, can influence the biological outcome of the enzyme activity. Therefore, having established the function of KDM2A CXXC domain in DNA recognition, it was important to assess whether such binding is necessary for KDM2A demethylase activity in a chromatin context.

Using the previously described radioactivity-based formaldehyde release assay, the effect of the CXXC domain function on JmjC-mediated demethylase activity was assessed in the presence of Set2-methylated nucleosomes. Interestingly, both deletion and mutation of the CXXC domain (Figure 3.9A) induced a decrease in demethylase activity although in different measures (Figure 3.9B). While the K601A mutant retained much of the original catalytic activity, deletion of the CXXC domain caused an approximately 80% fall in JmjC activity. Thus the data appears to suggest that a functional CXXC domain does contribute to an efficient demethylase activity in KDM2A. This outcome was unexpected considering that large amounts of enzyme were used in the demethylase reaction. Moreover, as the assay conditions are set to measure the final amount of formaldehyde accumulated at the endpoint of the reaction, the assay does not provide information regarding its kinetics. It is therefore possible that the observed effect of the CXXC domain may not be related to a deficiency of the JmjC catalytic activity per se but instead due to a decrease in the rate of formaldehyde accumulation (e.g. because of weaker substrate interaction). Also, given the discrepancies in activity observed between CXXC deletion and point mutation mutants, the contribution of the CXXC domain to demethylation may not be simply limited to substrate binding. In conclusion, the exact nature of the relationship between CXXC domain DNA binding and JmjC histone demethylase activity need to be further investigated.

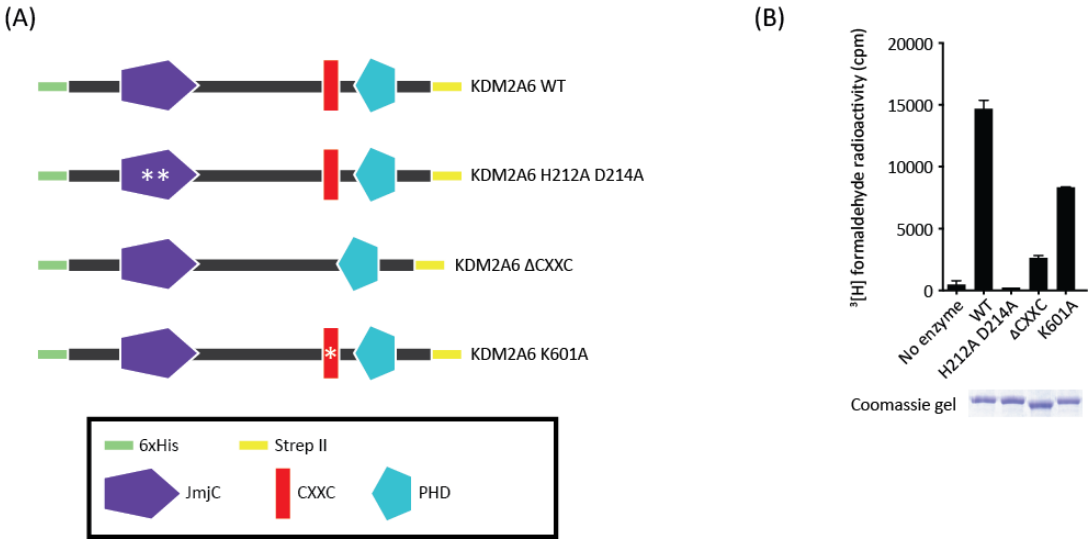


Figure 3.9 – Efficient KDM2A demethylase activity is dependent on a functional CXXC domain

(A) Schematic illustrating the KDM2A constructs used to test for CXXC domain contribution to demethylase activity. An asterisk represents each point mutation. (B) The same experimental conditions as in Figure 3.7B were applied here, except for a fresh batch of Set2-labelled nucleosomes was used as substrate. Although the same amount of substrate in terms of ^{3}[H] radioactivity (cpm) is controlled for each reaction, the absolute formaldehyde radioactivity measured in the presence of WT KDM2A is higher in Figure 3.9B than in Figure 3.7B. We have noticed similar absolute differences in formaldehyde release in these types of experiments previously and this is generally related to the efficiency of formaldehyde extraction which is dependent on the freshness of the conversion reagent or the efficiency with which the substrate has been labelled by the Set2 enzyme. Therefore, absolute values between replicate assays with different batches of labelled substrate may not be comparable in absolute radioactive formaldehyde release (cpm). However, values between constructs within the same assay are comparable. K601A mutation of the CXXC domain partially affected demethylase activity while complete removal of this domain drastically reduced the capacity to demethylate nucleosomes. H212A D214A serves as a negative control for the demethylase assay here. The Coomassie gel underneath the graph shows that an equal amount of protein (WT or mutant) was used in each demethylase reaction. Error bars correspond to standard errors of the means (SEM) derived from two replicate assays.

Quantification of KDM2A CXXC domain binding affinity by surface plasmon resonance

A variety of biophysical tools exist nowadays to examine DNA-protein interaction (Dey et al., 2012). One of the relatively recent techniques that allow real time measurement of protein binding affinity and kinetics is surface plasmon resonance (SPR).

The underlying physical principle of SPR is rather complex (Figure 3.10) (Schasfoort and Tudos, 2008). In brief, when a beam of polarised light that travels through a medium of high refractive index (e.g. a prism) encounters a medium of low refractive index, it is totally reflected, a phenomenon known as total internal reflection. However, the electromagnetic field component of the light can still penetrate over a short distance into the low reflective medium generating an evanescent wave whose intensity attenuates exponentially as it propagates away (approx. 300 nm) from the interphase. When a thin layer of noble metal (e.g. gold) is inserted at this interface, the electron oscillations, also known as plasmons, of the metal conductor can be coupled (i.e. resonate) with that of the photons of the polarised light to enhance the intensity of the evanescent wave. This phenomenon is known as surface plasmon resonance. Importantly, this resonance results in a loss of energy so that not the entire incident light is reflected now. Such reduction in reflected light intensity is greatest at a specific incident angle, denoted as θ.

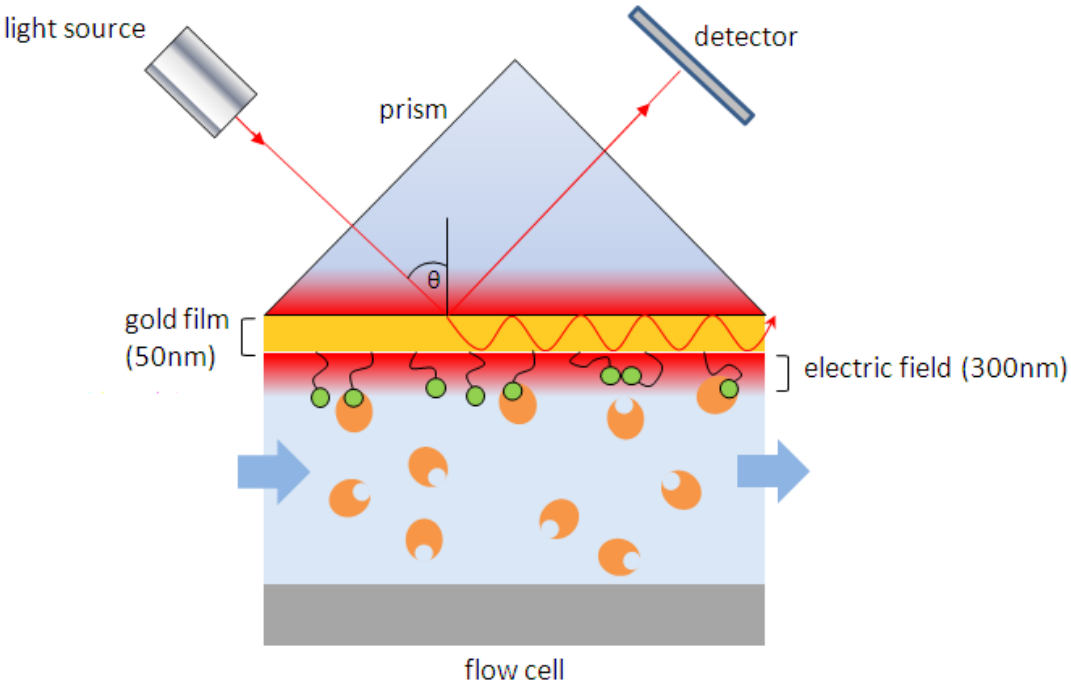


Figure 3.10 – Surface plasmon resonance

The physical principle of SPR and its general application for measuring DNA-protein interaction is illustrated here. A detailed description is presented in the main text. The green filled circles represent the ligand while the orange-coloured "C" shapes represent the analyte.

In practical application, the face of the metal film that is in contact with the low refractive medium forms the ceiling of a sensor chip chamber. This surface is usually pre-coated with streptavidin to allow immobilisation of biotinylated DNA probe (referred as the ligand). The DNA-binding protein (referred as the analyte) is injected into the buffer that constantly flows over the DNA coated surface. As DNA-protein complexes form and accumulate at the surface, the increase in the

local concentration of molecules results in an increase of the refractive index, which is registered as a change of the θ angle by the photo-detector. This is normally measured in resonance units (RUs). As this change is monitored in real time, the shift in RUs can be plotted against time to generate a sensorgram (Figure 3.11).

SPR is a particularly useful technique for determining DNA-protein binding affinity and kinetics. In particular, by injecting a dilution series of the analyte and by measuring the level of binding at equilibrium for each analyte concentration, the affinity constant can be calculated using the Langmuir binding isotherm (Schasfoort and Tudos, 2008).

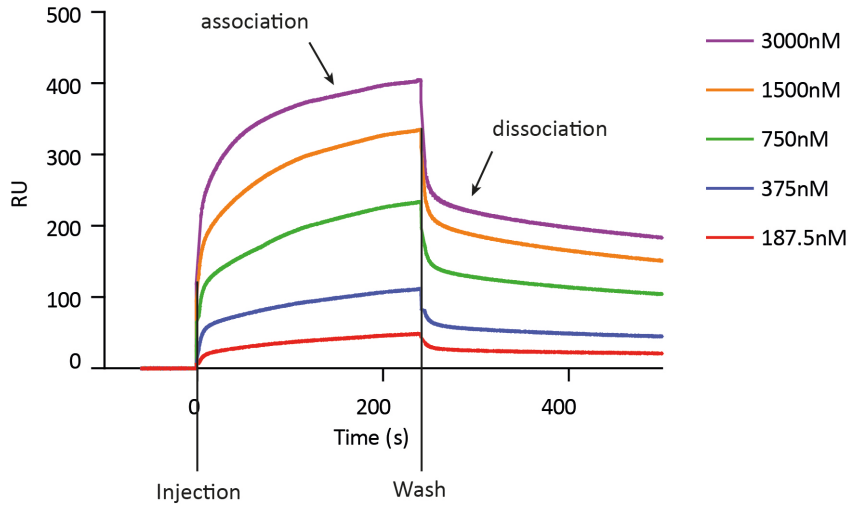


Figure 3.11 – Sensorgram of KDM2A binding to the DNA probe

A typical plot of KDM2A association and dissociation curves when interacting with the 1CG or 6CG DNA probe. A dilution series of KDM2A6 WT input sample ranging from 187.5 nM to 3000 nM was used for each experimental run. Sample injection

and wash time points are indicated, as well as the association and dissociation phases of binding.

In order to measure the binding affinity of KDM2A in relation to CpG-containing DNA, a 75 bp random sequence of DNA was designed to contain either one or six equal distant CpGs (1CG or 6CG) (See the Appendix for sequence details). A stock of ssDNA of each of these two probes was chemically synthesised and converted to dsDNA through Klenow extension using a biotinylated version of the M13 reverse primer. Saturating amount of each of the two dsDNA probes was then applied to the streptavidin-coated surface of a single flow cell in the SPR sensor chip. A separate DNA probe derived from the same 75 bp sequence which contains no CpGs was also included in a third flow cell to serve as a control for background noise (i.e. non-specific protein-DNA interactions). Unbound DNA probes were flushed away with running buffer. Upon injection of KDM2A6 into the flow cell chambers, a 240 s time lag was set before the protein was allowed to dissociate by flooding the chambers with running buffer (Figure 3.11). This was designed to allow for an equilibrium binding state to be reached. For the purpose of this study, a full experimental run is defined as five consecutive association and dissociation cycles where dilution concentrations of KDM2A6 ranging from 187.5 nM up to 3000 nM were individually applied, each in the presence of 100 ng/µl of the non-specific binding competitor dAdT. Given the intensity of the binding measured at equilibrium in RU (calculated as the difference between the RU in the binding flow

cell, i.e. 1CG or 6CG, and the RU in the 0CG control flow cell) and the known concentration of the input analyte, the affinity constant was calculated using the default Biacore evaluation program in the form of K_D (dissociation constant).

The obtained K_D values for both 1CG and 6CG probes were 0.64 µM and 0.56 µM respectively, which is slightly higher than the previously published K_D for the MLL1 CXXC domain (Figure 3.12A and B) (Birke et al., 2002). Whether this truly reflects a difference in binding affinity inherent to the two constructs would need to be further evaluated. Initially this could be achieved by repeating the experiment for KDM2A CXXC and MLL1 CXXC proteins under the same assay conditions.

According to a one to one binding model, at equilibrium state, the number of binding sites available (i.e. whether 1CG or 6CG probe is used) does not affect the K_D, which is an indication of the strength of the interaction and not a measurement of the number of bound analyte (which is instead given by the RU value). Structural data shows that a single CXXC domain interacts with one single CpG in a one to one relationship (Cierpicki et al., 2009). It is therefore expected that the K_D measured for KDM2A, whether exposed to 1 CpG or 6 CpGs, would remain the same.

Indeed, the absolute RU measured at each KDM2A6 concentration for the 6CG probe is higher than in the 1CG probe indicating that more KDM2A proteins are bound to the 6CG DNA than the 1CG equivalent. However, the close similarity between the 1CG and 6CG affinity values suggest that KDM2A interaction with CpG DNA is consistent with the Langmuir model of a monovalent analyte where each

binding event is independent as opposed to cooperative (Schasfoort and Tudos, 2008).

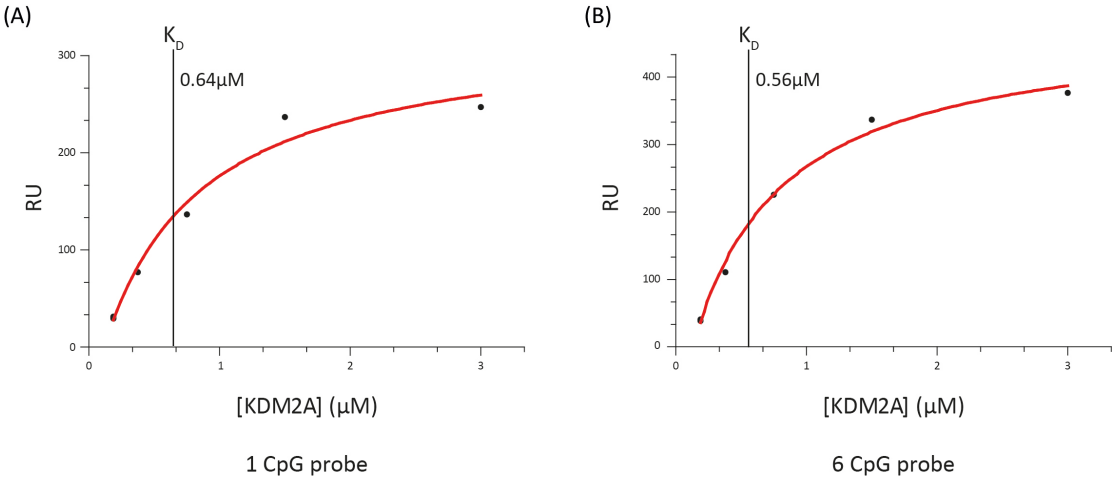


Figure 3.12 – K_D measurement of KDM2A determined by SPR

(A) The K_D value of KDM2A6 binding affinity for the 1 CpG containing probe was determined by SPR. The graph shows non-linear fitting (in red) of the absorption values (in RU) at equilibrium with the Langmuir binding isotherm. The point of intercept between the vertical line and the curve represents the K_D. Data is from one representative experiment of two replicate assays. (B) The same measurement was performed as in (A) in relation to the DNA probe carrying 6 CpGs. Data is from one representative experiment of two replicate assays.

3.3. Discussion

KDM2A displays mono- and di-methylated H3K36 demethylase activity against both free histones and chromatin

The baculovirus-based expression vector system is suitable for large scale production of mammalian recombinant proteins (Jarvis, 2009) as the insect protein synthesis machinery has evolved to efficiently process large molecular weight proteins carrying complex post-translational modifications. However, like any other cell-based expression system, the recombinant protein is inevitably subject to a variable degree of degradation. Depending on the extent of interference to normal cell biology, in some proteins this degradation can be rather severe. Indeed, overexpression of KDM2A was observed to cause early stage cell toxicity that cannot be attributed to baculovirus infection alone. A large proportion of expressed KDM2A protein in Sf9 insect cells was found in the degraded form.

Alternative attempts to generate full length recombinant KDM2A in bacteria were also undermined by the inefficiency of the prokaryotic system to synthesise and post-translationally modify large molecular weight proteins like KDM2A. Due to such limitations, only a truncated version of the full length KDM2A protein, KDM2A6, was amenable to expression and double affinity purification. Luckily, KDM2A6 includes the three domains, JmjC, CXXC and PHD that are the focus of this

functional study. Instead, the omitting C-terminal F-box is beyond the scope of the current investigation.

In this chapter the demethylase activity and substrate specificity of recombinant KDM2A was evaluated *in vitro* using three different methods. Western blot analysis of KDM2A activity in the presence of endogenous purified or recombinant histone proteins indicate an enzymatic function towards mono- and di-methylated H3K36 in agreement with both previous *in vitro* and *in vivo* observations (Tsukada et al., 2005). However, due to the inconsistency of the H3K36me3 signal, it is not possible on this occasion to comment on KDM2A capacity to act also on the tri-methylated substrate or whether the enzyme is only specific towards mono and di-methylated forms. In these assays, it is important to note that Western blotting is only suitable as qualitative readout of enzymatic activity and its utility is further hampered by a reliance on the quality of the antibodies employed. This makes it difficult to use as a means for quantitative analysis for enzymatic activity. Future experiments using mass spectrometry to assess KDM2A demethylase activity on differentially methylated H3 histone tail peptides will be better suited in clarifying the substrate specificity of this enzyme and for experiments designed to measure time dependent reaction rates.

A FDH-mediated fluorometric assay was recently developed for the detection of methylamine deamination reactions (Lizcano et al., 2000). As formaldehyde is also the by-product of JmjC domain catalysed reaction, the assay was believed to

provide an attractive, radioisotope-independent method of measuring demethylase activity. It was however observed that the rate-limiting step in this enzyme-coupled assay is the JmjC mediated release of formaldehyde. Considering the limited recombinant input material, in this study the FDH assay did not provide sufficient sensitivity to distinguish KDM2A-dependent enzymatic activity above background noise. Nonetheless, given KDM2A binding capability towards chromatin (illustrated in chapter 4), it would be interesting to compare KDM2A activity towards Set2-methylated nucleosomes as opposed to calf thymus histones under the FDH assay condition. Previous work suggests that KDM2A is active on core histones, mono and oligo-nucleosomes with no apparent substrate preference but this observation was dependent on semi-quantitative Western blotting results (Tsukada et al., 2005). A better alternative would be to quantitatively measure KDM2A activity towards non-radiolabeled, Set2 only-modified recombinant nucleosomes. However, at the time the FDH assay was carried out in this study, the procedure for the *in vitro* reconstitution of nucleosome was still under optimisation and the efficiency of Set2 methylation on recombinant nucleosomes also remained to be tested. In the future, the FDH assay will be applied to Set2-methylated recombinant nucleosomes to investigate any enzymatic preference for chromatin substrates.

One of the most sensitive methods for measuring enzymatic activity involves labelling of the substrate molecule with radioisotopes. In many instances, these biomolecules cannot be chemically synthesised and enzyme-mediated installation

is the most convenient and straightforward way. To specifically methylate K36 residue on histone H3, recombinant yeast Set2 was used (Morris et al., 2005). Set2 is only enzymatically active on nucleosome substrate so this also provides a good opportunity to test KDM2A activity on chromatin.

KDM2A was observed to readily demethylate Set2-methylated nucleosomes and this activity was dependent on a functional JmjC domain. However, because the yeast Set2, in contrast to the human homolog SETD2, deposits all three methylation states on histone H3K36 residue (Youdell et al., 2008), the observed radioactivity is a reflection of the total demethylase activity on all available H3K36 substrates. In order to discriminate between each methylation state, Western blot analysis of the demethylated nucleosome products using methylation state-specific antibodies would be required.

Previous data indicate that H3K36me3 is refractory to KDM2A demethylase activity but the molecular basis for such selectivity between the di- and tri-methylated states remains to be defined. In the absence of any structural information of the KDM2A JmjC domain in complex with its histone substrate, studies aimed at understanding the substrate selectivity were limited to primary sequence comparisons with closely related JmjC demethylases. For example, recent alignment of the catalytic cores from KDM2A and KDM4A, for which structural data is available, revealed no apparent sequence conservation except for the region that is predicted to mediate methyl-lysine recognition in KDM4A (Chen et al., 2006).

However, KDM4A displays only weak affinity for H3K36me2 while it mainly targets H3K36me3. Whether due to a lack of affinity or an incompatible binding pocket, it seems that the structural basis for discrimination between di-and tri-methylation is rather subtle and further structural data is needed to uncover it.

KDM2A binds preferentially to CpG DNA via the CXXC domain

Recent *in vivo* studies of KDM2A function suggested preferential targeting to CpG-rich sequences where the zinc finger CXXC domain was hypothesised to mediate DNA binding (Frescas et al., 2008). In this study, using sequence defined CGI-like DNA probes and EMSA assays, the DNA binding function of the CXXC domain in KDM2A was for the first time confirmed. Like other CXXC domains, the KDM2A CXXC domain displays specific affinity for non-methylated CpG dinucleotides. Importantly, sequence alignment with previously characterised CXXC domains and site-directed mutagenesis has revealed the K601 residue in this domain to play an essential role in mediating CpG recognition.

Interestingly, it was observed that, at least *in vitro*, the efficiency of KDM2A demethylase activity appears to be dependent on a functional CXXC domain as both CXXC deletion and point mutant proteins displayed marked reduction in catalytic activity. It is possible that CXXC-mediated DNA binding may facilitate targeting of KDM2A demethylase activity towards a nucleosomal substrate. This

would be in agreement with the ability of KDM2A CXXC domain to interact with non-methylated CpG-containing linker DNA as evidenced in chapter 4. Indeed, the chicken erythrocyte-derived mononucleosomes used for Set2 methylation were previously isolated using partial MNase digestion and are expected to carry varying length of linker DNA (between 5 to 55 bp) (Rob Klose, personal communication). Importantly, the CXXC domain contribution may not affect the capability per se of KDM2A to remove H3K36 methylation from nucleosomes but instead improve the rate at which demethylation occurs (e.g. as a result of a more stable enzyme-substrate interaction). In chapter 4, KDM2A CXXC domain has been shown to preferentially interact with the 216 bp recombinant nucleosome as opposed to the shorter, linker-less 147 bp form. It would be therefore interesting to evaluate KDM2A demethylase activity towards these two different chromatin substrates using the Set2 methylation system described. Moreover, to measure potential effects on the rate of the reaction, a time-course experiment comparing WT and CXXC-deficient enzymes will be required.

On the other hand, the above result is somewhat surprising considering that in a simple artificial *in vitro* setting designed to measure enzymatic equilibrium end point with large amounts of enzyme, the probability of enzyme-substrate interaction should be very high due to molecular crowding even in the absence of a specific targeting mechanism. Nonetheless, initial enzyme titration experiments with both WT and CXXC mutant proteins confirmed the necessity of a functional DNA targeting module for efficient demethylation to occur (data not shown). This

is further supported by the fact that addition of the non-specific endonuclease benzonase in the demethylase reaction severely affected activity of all KDM2A constructs tested (data not shown).

Because it was observed that deletion of the CXXC domain induces a more severe deficiency than K601A point mutation alone, it can be argued that some sort of structural element is also likely to be involved. Supporting structural data is required to evaluate whether DNA binding by the CXXC domain induces any conformational change in adjacent protein regions that may affect the way the JmjC active site engages its histone substrate.

In a compartmentalised cellular environment, the DNA binding activity of the CXXC domain is likely to be essential for controlling local KDM2A concentration and directing its demethylase function to presumed target foci of non-methylated DNA where the specific demethylase activity is required. Indeed, recent IF and ChIP experiments in the laboratory showed the CXXC domain to be required for KDM2A targeting to non-methylated CGIs in mouse ES cells (Blackledge et al., 2010). By applying genome-wide ChIP sequencing analysis, it was also demonstrated that KDM2A presents a much broader genome coverage than suggested by previous literature and its localisation correlates almost perfectly with sites of CGIs (Blackledge et al., 2010).

KDM2A is a monovalent DNA-binding protein

Finally in this chapter, the affinity strength of the CXXC domain in relation to one or multiple CpG sites was measured in real time by SPR. In both instances, KDM2A was observed to display a binding affinity that is slightly higher than previously published value for the MLL1 CXXC domain (33 nM K_D) (Birke et al., 2002). There are several confounding factors that may lead to the observed discrepancy. Firstly, the difference could be partly inherent to the two constructs. Sequence alignment suggests a high conservation of the key residues responsible for domain folding and base-specific recognition exists between KDM2A CXXC and MLL1 CXXC. However, the protein sequence outside the CG-interacting region is much less conserved, although structural data indicate that these amino acids also contribute to the overall stability of the binding (Cierpicki et al., 2009). Secondly, the DNA probes used in the SPR assays are different. Although the presence of non-methylated CpG is the only requirement for CXXC-DNA interaction, it cannot be excluded that binding affinity may vary according to the surrounding DNA sequence context. Thirdly, the methods used to derive the K_D values are different. The MLL1 CXXC K_D was calculated from the ratio of dissociation rate (k_{off}) and the association rate (k_{on}) by applying kinetic measurements. Conversely, the KDM2A CXXC K_D was directly derived in this study from measurements of the binding affinity at equilibrium state for a series of analyte concentrations. Data collected for binding kinetic analysis (i.e. on/off rates) are different from those measured at

equilibrium end points since different experimental settings apply. Also, the two strategies are prone to different sources of artefact (e.g. mass transport limit as opposed to erroneous estimate of active ligand concentration) (Schasfoort and Tudos, 2008). Lastly, the MLL1 CXXC construct used in the published study carries a relatively bulky GST (glutathione S-transferase) tag (Birke et al., 2002). The GST protein has the ability to form homodimers (Parker et al., 1990). Therefore, it is possible that the K_D calculated for MLL1 CXXC is actually the result of a bivalent binding event, which is not comparable to the monovalent model proposed for KDM2A CXXC in this study.

Since KDM2A preferentially localises at CGIs *in vivo*, and given that CGIs by definition are characterised by elevated CpG density, it is reasonable to speculate that a positive cooperative binding mode may apply where a successful binding event by one KDM2A molecule favours adjacent binding events by additional KDM2A molecules until saturation of all local binding sites or depletion of the pool of KDM2A molecules is reached. The SPR data presented here indicate that this is not the case since KDM2A binds to a stretch of DNA with equal affinity whether it contains one or multiple CpG sites. This observation suggests that, instead of a cooperative model, it is likely that a single KDM2A molecule interacts with a single CpG pair in a monovalent fashion where each binding event is independent. Cooperative DNA binding is characteristic of certain transcription factors where interaction among the activator/co-activator pairs leads to synergistic induction of target gene promoter activity (Goto et al., 2002). It is tempting to speculate, given

the binding behaviour, that KDM2A may not display transcription factor-like function, where fast mobilisation to defined genomic targets and tight regulation of local concentration is required. Instead, a basal but constant activity associated with genome-wide chromatin maintenance function may be better applied to KDM2A. However, it might be too simplistic to conclude that the *in vivo* distribution of this protein is only a reflection of CpG density (i.e. the higher the number of CpGs, the more KDM2A localise to the region), as one would predict from the monovalent binding model. As discussed in chapter 4, additional factors in the genome are likely to be at play.

One limitation of the SPR assay is the fact that no information can be deduced with regard to the number of analyte that is bound to each ligand. Previous DNase I footprinting analysis using the MLL1 CXXC domain indicates that this domain can protect up to two nucleotides on each side of the bound CpG (i.e. a 6 bp footprint) (Birke et al., 2002). Considering that the CpGs in the 6CG DNA probe are equally separated by seven nucleotides, it is expected that up to six KDM2A molecules could be bound simultaneously to a single 6CG ligand. However, due to possible steric hindrance effect derived from protein domains adjacent to the CXXC region in KDM2A and in the absence of information about the structural relationship between different domains in this protein, the actual number of KDM2A molecule that can bind at the same time to a single 6CG probe will need to be determined empirically in future studies via DNase I footprinting assay.

4. Chapter Four – KDM2A chromatin binding specificity

4.1. Assembly of recombinant chromatin by the salt dilution method

Chromatin exists under different forms. As mentioned earlier, at gross level, euchromatin and heterochromatin domains can be distinguished across the eukaryotic genome displaying contrasting structural properties that are known to impact on local protein accessibility and diffusion rate. In chapter 3, by creating a purposely simplified *in vitro* context, the CXXC domain was shown to be the main DNA targeting module in KDM2A. However, *in vivo*, depending on the nature of the local chromatin environment, restrictions are likely to be present that may limit how the KDM2A CXXC domain interacts with local DNA sequences. Although recent studies have uncovered at atomic detail the precise interactions between the CXXC domain and a CpG-containing DNA substrate (Allen et al., 2006; Cierpicki et al., 2009), these structural data do not provide information in relation to the mechanism by which such DNA binding occurs in a chromatin context. To understand how different chromatin states influence KDM2A interaction with

binding sites, a detailed knowledge of how the KDM2A CXXC domain interfaces with a strand of DNA packaged within a nucleosome is essential.

This chapter is aimed at addressing the above question by taking advantage of a completely recombinant *in vitro* chromatin binding system. By introducing chromatin substrates with defined properties, their mode of interaction with KDM2A was examined. A detailed account of the *in vitro* chromatin assembly process is given in the first part of the chapter followed by presentation of KDM2A chromatin binding data.

Purification of recombinant Taq DNA polymerase

Over recent years, a number of biochemical tools were developed to enable the *in vitro* assembly of recombinant chromatin complexes that recapitulate the physiological chromatin template on which gene regulatory activities take place. The importance of these studies is evidenced by the achievement of the first fine map detailing histone-DNA interactions occurring within the nucleosomal core (Luger et al., 1997a).

The *in vitro* system used in our study is based on the salt dialysis gradient reconstitution method first pioneered around the same time when nucleosomes were first discovered (Oudet et al., 1975). The principle of the salt dialysis method relies on the fact that H2A/H2B dimers and H3/H4 tetramers form a stable

octameric complex at high salt concentrations (Luger et al., 1999). By progressively dialysing the salt away, octamers can be gently dissociated back into tetramers and dimers. If a DNA template is introduced, by exploiting the differential affinity of the DNA molecule for H3/H4 tetramer and H2A/H2B dimer at specific salt concentrations, the dissociated octamer can be stepwise transferred and reassembled onto the DNA template. If the reconstitution reaction is small enough, the salt dilution gradient can be substituted with a series of discrete dilution steps without compromising the final efficiency of the reconstitution.

Prior to setting up chromatin reconstitution reactions, all necessary recombinant DNA and histone constituents were first individually generated and purified in this study. Given the strong nucleosome positioning property of the previously described synthetic 601 DNA sequence, it represented the ideal template for *in vitro* nucleosome reconstitution. Initially, a gift plasmid pGEM3Z carrying the 601 sequence was obtained from Jonathan Widom's laboratory. By applying traditional PCR using a custom primer pair, a 147 bp long 601 DNA was synthesised. However, due to the high cost associated with setting up large-scale PCR reactions, use of commercially derived Taq DNA polymerase (Taq Pol) was not economically affordable. Therefore, in order to express recombinant Taq polymerase in the laboratory, a gift plasmid carrying the *taq* gene from *T. aquaticus* was obtained from Dr Adele Marston and expressed in *E. coli*.

The thermal resistant property of Taq Pol can be exploited for its easy isolation from the pool of host cell proteins (Lawyer et al., 1993). Indeed, following gentle overnight expression, the Taq Pol containing cell lysate was incubated at 75°C for 1 hr to induce precipitation of all endogenous bacterial proteins that become denatured at such high temperature. Following centrifugation, the Taq Pol-rich supernatant part could then be collected (Figure 4.1A). Given the high affinity of the Taq Pol active site for DNA substrates, a notorious challenge in recombinant Taq Pol purification is the presence of contaminating host genomic and plasmid DNA. Although total removal of these contaminants is often difficult and nonessential, large amounts of residual DNA can potentially affect the efficiency and purity of subsequent PCR products. Therefore, to remove any DNA molecules from the Taq Pol sample, a rapid biphasic extraction method was utilised which exploits the relative insolubility of DNA in ethanol. Consequently, DNA contaminants were trapped in a lower aqueous phase while the Taq Pol protein was retained in the upper salt-saturated organic phase (Louwrier, 1999).

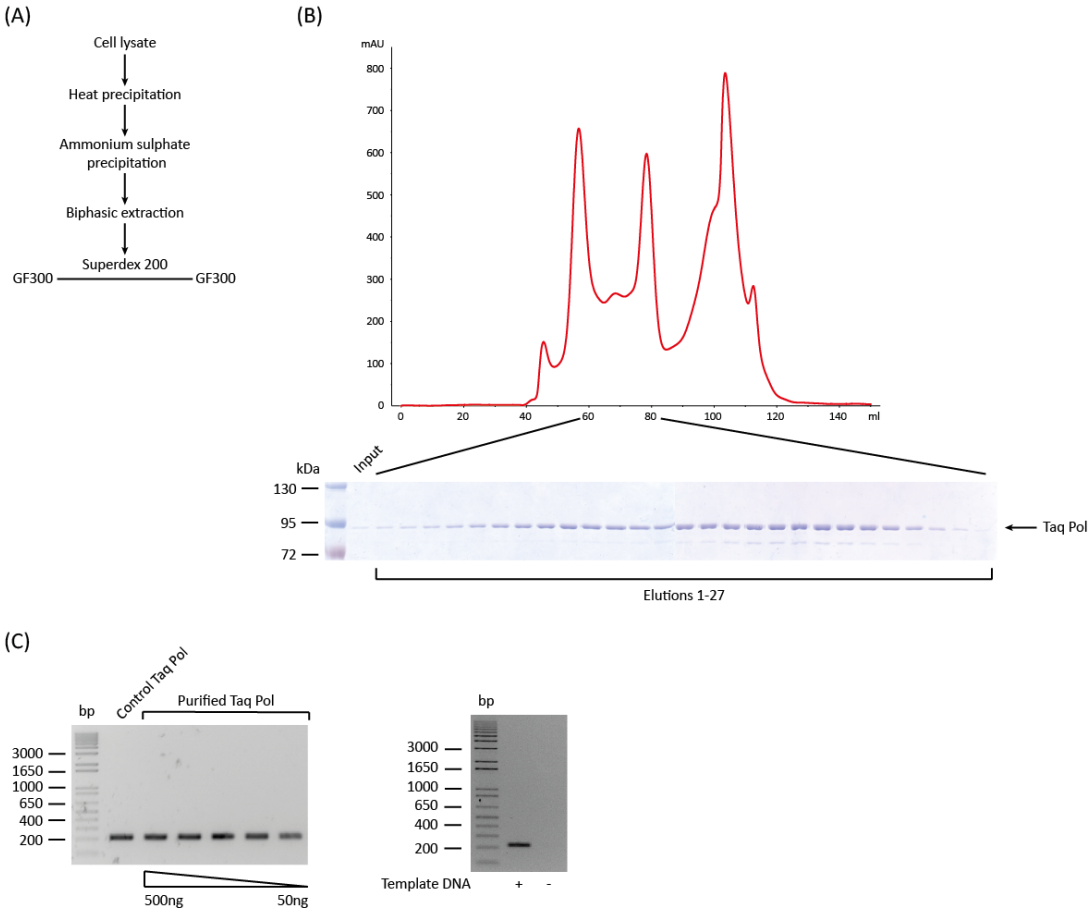


Figure 4.1 – Purification of Taq DNA polymerase

(A) Recombinant Taq Pol was expressed in BL21 cells overnight. The derived cell lysate was heated to 75°C to precipitate all host cell proteins. The remaining supernatant was treated with ammonium sulphate to isolate the Taq Pol protein. To eliminate contaminating DNA, a biphasic extraction protocol was applied to separate Taq Pol and DNA in organic and aqueous phase respectively. Details of experimental procedure are given in section 2.8 of Materials and Methods. (B) Following dialysis, the Taq Pol sample was further "polished" via gel filtration chromatography with elutions monitored by UV spectrometry at 280 nm and analysed by Coomassie staining on SDS-PAGE. The indicated elutions carrying a protein band corresponding to the predicted molecular weight of Taq Pol were pooled together. Several other elution peaks were visible in the gel filtration profile, possibly corresponding to DNA contaminants. Shown here is a representative gel filtration elution trace and corresponding SDS-PAGE of one of two repeat purifications of Taq Pol performed in this study. (C) The capacity of the isolated protein sample to enable DNA synthesis was tested by standard PCR in comparison to a commercial control Taq Pol (Fermentas) (left panel). A series of concentrations of purified protein sample was used to determine the optimal

amount of enzyme for PCR. A plasmid carrying a 216 bp 601 sequence was used as template. As expected, a single PCR product was seen to migrate slightly above 200 bp in an agarose gel. That the observed PCR product was indeed the 216 bp 601 sequence was further confirmed by the fact that, in the absence of the 601-containing template, no such PCR product can be seen (right panel, right lane). The titration experiment was performed for each batch of Taq Pol purified.

The high amount of salt in the Taq Pol preparation derived from the biphasic extraction step was easily dialysed away. To further refine protein purity, the dialysed Taq Pol sample was applied to a Superdex 200 gel filtration column and fractions of each of the elution peaks monitored by UV spectrometry at 280 nm were analysed by Coomassie staining on SDS-PAGE (Figure 4.1B). In gel filtration chromatography, the input protein mixture is fractionated with individual protein species eluting from the column in an orderly fashion according to its size and shape. This property facilitated the identification of a number of elutions carrying a single protein band that corresponded to the expected molecular weight of Taq Pol (approx 95 kDa). To establish the identity of the observed Coomassie stained band, the isolated fractions were pooled together and tested for their capacity to support DNA synthesis under standard PCR conditions (Figure 4.1C). A dilution range from 500 ng to 50 ng of the purified protein was applied and compared to 1 µg of commercially derived Taq Pol enzyme (Fermentas) in order to titrate and determine the optimal amount of the presumed Taq Pol to be used. It was observed that the purified protein enabled PCR amplification of a 216 bp long 601 DNA sequence as efficiently as the commercial Taq Pol protein confirming that the

isolated protein was indeed the Taq Pol enzyme. In particular, 200 ng of the purified Taq Pol protein was sufficient to sustain efficient DNA synthesis within a 50 µl PCR reaction. This stoichiometry was maintained during subsequent scale up PCR generation of the 601 DNA.

Generation of the 601 nucleosome positioning sequence DNA

With a large stock of functional Taq polymerase available, 96 of 50 µl volume PCR reactions were initially set up to generate the same 147 bp 601 DNA sequence described in previous EMSA assays in chapter 3 (Figure 4.2A). Considering the large volume of the resulting PCR pool (approx. 4.8 ml), in order to isolate the relevant amplified product, standard extraction kit-based approach would be too laborious, costly and probably inefficient. Instead, given the negative charged backbone of a DNA molecule, the PCR products can be easily trapped by an anion exchange resin that contains, for example, quaternary amines (Q-resin). Indeed, the pool of 147 bp 601 PCR DNA was loaded onto a 1 ml Resource Q column. By applying a salt elution gradient, bound DNA was eluted within a 500 to 600 mM NaCl range where a sharp peak was detected by UV spectrometry at 260 nm (Figure 4.2B). The fractions corresponding to this elution peak were analysed by ethidium bromide staining on 1% agarose gel, which revealed DNA bands of the expected length (i.e. 147 bp). The analysed fractions were pooled together and dialysed in 10 mM Tris buffer (pH 8.0) for storage.

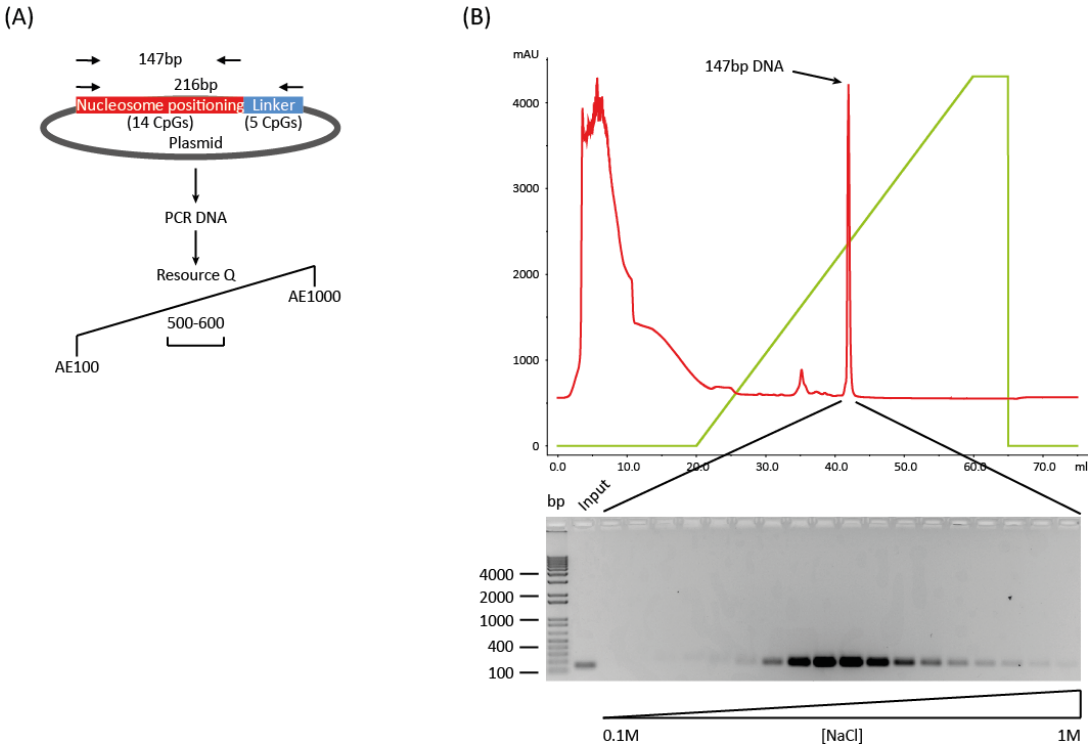


Figure 4.2 – Purification of the 601 DNA sequence synthesised by PCR

(A) A 147 bp or 216 bp 601 DNA sequence was PCR amplified from the pGEM3Z 601 plasmid template using custom primer pairs (See Appendix for sequence details of 601 DNA). The PCR product was then applied to a 1 ml Resource Q anion exchange column and eluted within approximately 500 to 600 mM salt range. The number of CpGs in the nucleosomal and linker regions is stated within brackets. Details of the experimental procedure are given in section 2.9 of Materials and Methods. (B) Profile of the eluted 601 DNA peak was visualised by UV spectrometry at 260 nm (in red). The fractions comprised within this peak were analysed by ethidium bromide staining on 1% agarose gel. The profile in green represents the 0.1 M to 1 M salt gradient applied to the Resource Q column. Shown here is one representative of three repeat purifications of 147 bp 601 DNA performed in this study.

Purification of recombinant M.SssI and CpG methylation of the 601 DNA

In chapter 3, CpG-methylated 601 DNA probe for EMSA was generated using commercially obtained (New England Biolabs) CpG-specific bacterial DNA methyltransferase M.SssI (Figure 4.3C). However, in this chapter, generation of a large stock of fully methylated 601 DNA for the purpose of nucleosome reconstitution requires large scale *in vitro* DNA methylation reaction. Due to the prohibitive cost associated with the use of commercially-derived M.SssI enzyme, it was decided to express this enzyme in the laboratory. To this end, an N-terminal 6xHis tagged version of the M.SssI gene from *Spiroplasma sp* was cloned into a pBAD24 vector suitable for protein expression in *E. coli*. The choice of this particular vector system was dictated by the cell toxicity derived from overexpressing the M.SssI protein due to inevitable interference with the host DNA methylation program. The L-arabinose responsive araBAD promoter system is particularly suitable for expression of toxic recombinant proteins like M.SssI because of the tightly regulated induction system which ensures minimum undesired basal level of recombinant protein expression in the absence of the transcription activator (Hirsh and Schleif, 1977). To further safeguard the host genome, the pBAD24 M.SssI plasmid was transformed into a particular strain of *E. coli* defective in the methylation-specific restriction endonucleases mcrBC, mcrA

and mrr (Kelleher and Raleigh, 1991; Raleigh and Wilson, 1986) to prevent restriction system-mediated cleavage of host DNA.

After gentle overnight expression, the M.SssI protein was purified from cell lysate via standard Ni-NTA affinity method (Figure 4.3A) and the collected elutions analysed by Coomassie staining on SDS-PAGE (Figure 4.3B). Indeed, a major protein band corresponding to the expected molecular size of M.SssI (approx 45 kDa) was visible with only minor contaminating species. Given that the Ni-NTA derived sample was estimated to be over 90% pure, further protein polishing steps were deemed not necessary. A small aliquot of the elution pool was tested for its capacity to methylate a 216 bp long 601 DNA sequence *in vitro*. As evidenced by the restriction digest assay of the methylated product with the methylation sensitive restriction enzyme HhaI, the M.SssI-methylated DNA was fully resistant to cleavage by HhaI as opposed to the non-methylated DNA. This result confirms that the isolated protein is indeed the M.SssI enzyme carrying CpG-specific DNA methyltransferase activity (Figure 4.3D). A further titration assay was able to determine that a total of 2 µg of the purified M.SssI was sufficient to mediate complete methylation of 1 µg of 216 bp 601 DNA where each DNA molecule is known to contain a total of 19 CpG sites (Appendix). This ratio was maintained during subsequent scale up methylation reactions for the generation of methylated 601 DNA.

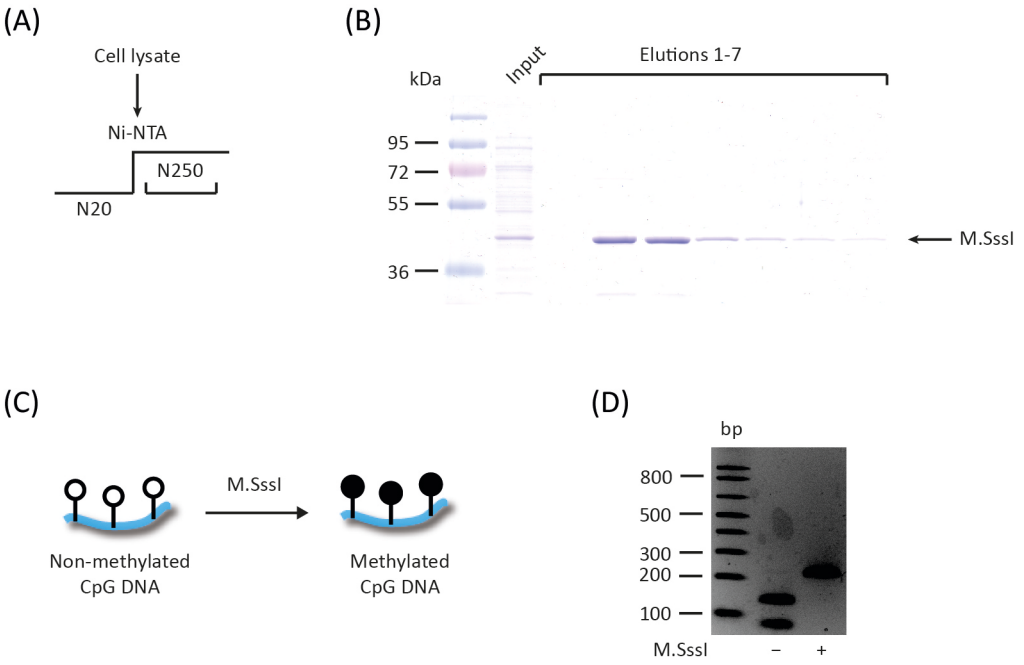


Figure 4.3 – Purification of M.SssI and CpG-DNA methylation

(A) M.SssI was purified from *E. coli* cell lysate using a single step Ni-NTA-based approach. (B) The Ni-NTA elutions were stained by Coomassie dye revealing a major protein band corresponding to the expected molecular weight of M.SssI. Shown here is the gel corresponding to a single purification of M.SssI performed in this study. (C) Schematic illustrating the CpG DNA methylation function of M.SssI. Open lollipops represent sites of non-methylated CpG while closed lollipops are the methylated equivalents. (D) The purified M.SssI was tested for its ability to methylate a 216 bp 601 DNA. When subject to digest by the methylation sensitive restriction enzyme HhaI, the M.SssI-methylated DNA was fully resistant to HhaI cleavage indicating that the DNA was efficiently methylated (right lane). This methylation protection assay was performed each time a DNA probe was methylated by M.SssI.

Expression and purification of recombinant core histones

Concomitant to the generation of the 601 DNA template for nucleosome reconstitution, each of the four core histone genes derived from *X. laevis* were

individually expressed in *E. coli*. Because of the inherent affinity of histones for DNA substrates, their overexpression can be extremely toxic to the host cell. Consequently, a variant DE3 pLysS version of the standard BL21 *E. coli* strain was chosen as host for histone expression. The pLysS plasmid in these cells encodes for the T7 lysozyme, which specifically inhibits the T7 RNA polymerase on which the T7 expression system is dependent on. This is designed to enforce a tightened control over the induction system and to minimise undesired expression of the recombinant protein resulting from basal level of T7 RNA Pol activity in the absence of IPTG (Studier et al., 1990).

The expression level of recombinant histones within such designed system was observed to be extremely high, reaching milligram quantities for each litre of cell culture. As often occurs when very high local concentration of foreign proteins is reached within the host cell, overexpression of each of the four core histones have been observed to cause precipitation of the histones and formation of aggregates known as inclusion bodies (Luger et al., 1997b; Marston, 1986). Given the insoluble nature of these inclusion bodies, a denaturing purification method based on treatment with chemicals like guanidine hydrochloride was required to extract monomeric histone proteins from these aggregates (Figure 4.4A). Because of the incomplete breakdown of inclusion bodies even under concentrations of up to 7M guanidine hydrochloride, the monomeric histones need to be isolated from higher mass aggregate products by applying the denatured mixture to a Sephacryl S200 gel filtration column equilibrated in 1 M urea buffer. The histone elution peak was

detected by Bradford assay analysis of each individual fraction and aliquots of these were then analysed by Coomassie staining on SDS-PAGE to identify the elutions carrying mostly monomeric histone proteins (Figure 4.4B). Finally, the pool of combined elutions was extensively dialysed against H_2O in order to remove all denaturing agents in the elution buffer. The derived histone sample was then concentrated to approx 10 mg/ml for storage.

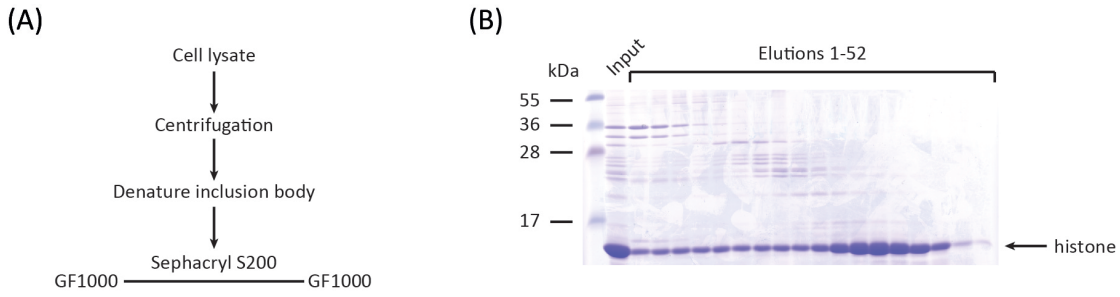


Figure 4.4 – Purification of recombinant core histones

(A) Following expression, the histone-rich inclusion bodies were isolated from the cell lysate by centrifugation. The inclusion bodies were subsequently broken down using high concentrations of the denaturing agent guanidine hydrochloride and the denatured products separated by gel filtration chromatography. Details of the experimental procedure are given in section 2.8 of Materials and Methods. (B) The elution fractions containing mostly monomeric histone proteins were identified by Coomassie staining of SDS-PAGE and pooled together. The multiple bands seen above the monomeric histone population are likely oligomeric histone aggregates derived from incomplete breakdown of inclusion bodies. Show here is a representative gel of the elution profile observed for all four core histone purifications in this study.

Reconstitution and purification of histone octamer

The strategy for the *in vitro* reconstitution of the four recombinant core histones into a single octamer complex involves essentially a reverse process of that occurring during nucleosome reconstitution (Figure 4.5A). Indeed, stoichiometric amounts of each core histones were first incubated under denaturing conditions to prevent formation of non-physiological aggregates. When a homogenous mixture was reached, this was then subject to extensive dialysis against 2 M NaCl. During the dialysis process, the high amount of salt ions in solution favours the spontaneous assembly first of H3/H4 tetramers and H2A/H2B dimers followed by formation of the octameric species. Since the process is non-enzymatically driven and dependent on the accuracy of the stoichiometry of the input histone samples, this approach to reconstitute histone octamers cannot provide 100% efficiency. Therefore, correctly assembled octamers need to be separated from histone dimers and higher order aggregates via gel filtration fractionation over a Superdex 200 column (Figure 4.5B). Two major protein elution peaks were visible by UV spectrometry at 280 nm. To investigate their identities, fractions corresponding to the elution peaks were analysed by Coomassie stained SDS-PAGE. It was immediately apparent that the early elution peak contained correctly assembled histone octamers as evidenced by the exact stoichiometry of the four histone components. Instead, the second peak corresponded to unincorporated/excess

H2A/H2B dimers. Only the octamer-containing fractions were then collected and stored without further processing required.

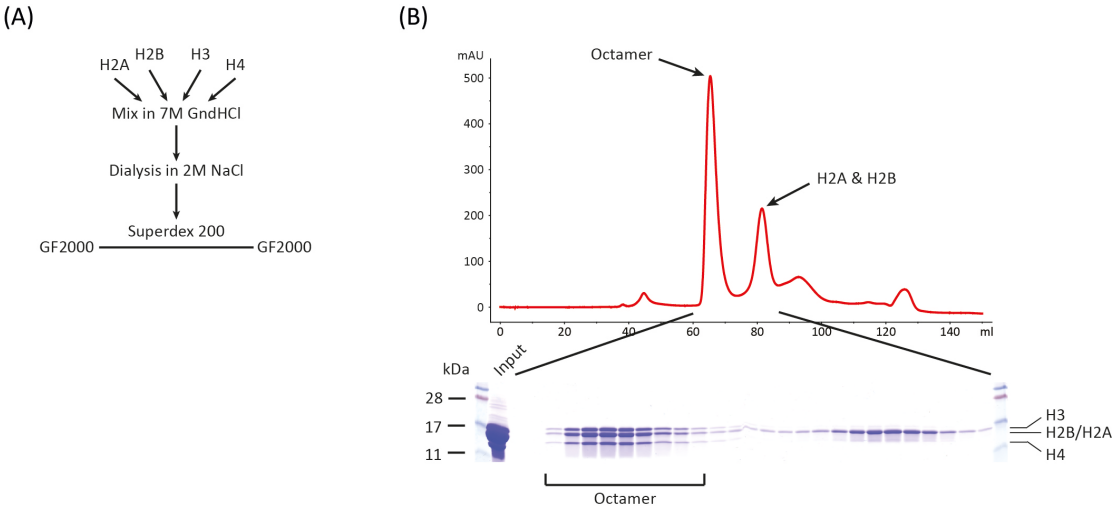


Figure 4.5 – Reconstitution of histone octamer

(A) Stoichiometric amounts of each of the four core histones were mixed together first under denaturing conditions and then subject to dialysis against high salt buffer. The octameric species formed during the salt dialysis process can then be separated using standard gel filtration chromatography. Details of the experimental procedure are given in section 2.9 of Materials and Methods. (B) The derived elution fractions containing properly assembled histone octamer can be identified via Coomassie stained SDS-PAGE by assessing the stoichiometry of the histone components. The second elution peak observed corresponds to excess H2A/H2B dimers. Shown here is a representative gel filtration profile of two replicate reconstitution experiments.

Reconstitution and purification of nucleosomes

With all the constituent reagents in place, the step dilution variant (also known as salt jumping) of the salt dialysis gradient method described at the beginning of this

chapter was performed to first assemble a 147 bp 601 DNA containing nucleosome that was aimed to recapitulate the canonical nucleosome core particle whose structure has been previously solved (Luger et al., 1997a). Previous nucleosome reconstitution studies indicate that the precise stoichiometry of the input DNA versus histones is one of the most critical factors in determining the efficiency of the reconstitution. Ideally, a molar ratio of one DNA molecule per each histone octamer applies. Under conditions where the amount of histone octamer is limiting, the DNA template will be undersaturated and nucleosomes not properly formed while even a 10% excess of histone octamer relative to DNA template can cause large accumulation of higher order, non-physiological aggregates (Carey et al., 2009). However, because very accurate assessment of input histone octamer concentration by Bradford assay is difficult, the optimal ratio between DNA and histones need to be determined empirically by setting up a range of trial ratios (Figure 4.6B).

Therefore, reconstitution reactions with DNA : histone mass ratios ranging from 1:1.8 to 1:2.4 were set up by initially incubating the 147 bp 601 DNA template and the histone octamers together under 2 M NaCl solution (Figure 4.6A). The salt concentration was then progressively diluted at each hour down to a final concentration of 0.1 M NaCl. Previous biochemical studies have shown that nucleosomes preferentially form around 0.6 M salt (Hansen et al., 1991). The subsequent dilution to a lower salt concentration was meant to stabilise the assembled nucleosomes for short-term storage.

In order to analyse the efficiency of each nucleosome reconstitution reaction, the reconstitution products were separated using native 0.8% agarose electrophoresis. It was observed that reconstituted nucleosomes are poorly stained by ethidium bromide dye (data not shown), likely because the presence of histones interferes with intercalation of the ethidium bromide molecule into the DNA double helix. Consequently, to facilitate visualisation of the gel migrating species, approximately 5000 cpm of ^{32}[P]-end labelled 147 bp 601 DNA was added to each reaction at the start of the reconstitution. By taking autoradiographs of the agarose gel, the optimal reconstitution reaction defined as the one most abundant in slow-migrating nucleosome species could then be identified (Figure 4.6B). Following a number of nucleosome reconstitution trials, it was noticed that even with an empirically determined optimal DNA: histone ratio, the salt dilution method does not provide 100% nucleosome reconstitution efficiency and some residual, unincorporated DNA is always present in the final reconstitution reaction. Since the presence of free DNA substrates can potentially complicate subsequent analysis of KDM2A interaction specifically with nucleosome substrates, a sucrose gradient fractionation method was devised to remove the contaminating free DNA. For this purpose, the optimal reconstitution reaction product was loaded onto 5% to 25% sucrose gradient and centrifuged at 35,000 rpm overnight to allow equilibration of sedimentation to occur. The fractions of the sedimented gradient were analysed on native 0.8% agarose gel and the ones containing only

nucleosome species were pooled and stored without further processing (Figure 4.6C).

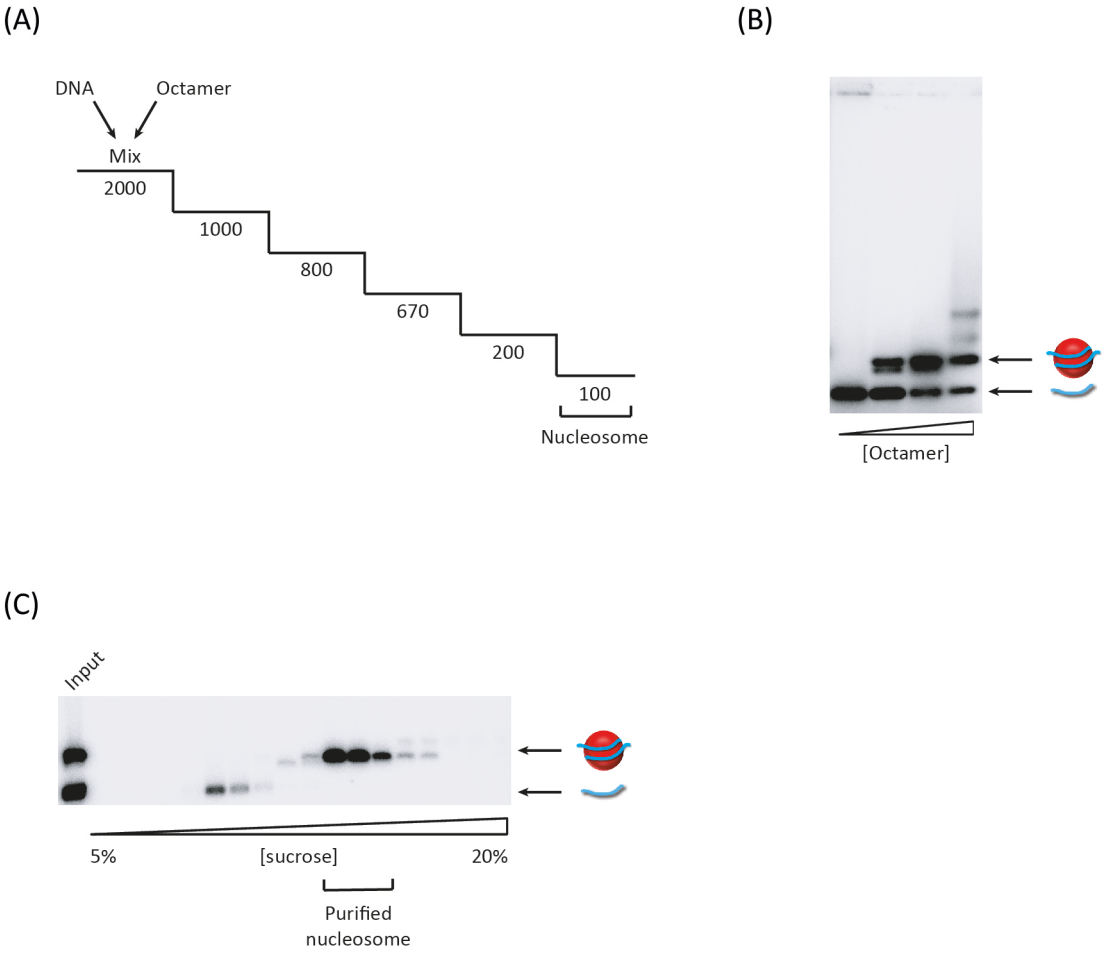


Figure 4.6 – Reconstitution and purification of nucleosome core particle

(A) Mass ratios of 601 DNA and histone octamer were incubated initially at 2 M NaCl after which the salt was diluted stepwise at each hour down to a final 0.1 M NaCl concentration. (B) Because each reconstitution reaction was spiked with a small amount of ^{32}[P]-labelled 601 DNA, the assembled products can be separated using native agarose electrophoresis and then visualised by autoradiography. The reconstitution reaction carrying the highest level of nucleosome species could be in such way identified. The cartoon representations of nucleosome (red ball) and free DNA (short blue thread) on the right of the gel indicate the identity of the

corresponding bands in the gel. (C) To remove any undesired, non-incorporated, free DNA, the products of the reconstitution reaction were purified through a 5-20% sucrose gradient. Only the nucleosome-containing fractions (the slower migrating species near the bottom of the gradient) were then pooled together and used for subsequent chromatin assays. Details of the experimental procedure are given in section 2.9 of Materials and Methods. Show here is a representative reconstitution and purification of more than three experiments performed in this study.

4.2. KDM2A binding to chromatin

KDM2A preferentially binds to non-methylated CpG dinucleotides in the linker region of a 216 bp mononucleosome

The canonical nucleosome core particle, where 147 bp of DNA is fully wrapped around an octameric core, represents the basic unit of chromatin organisation in the genome. Consequently, it is the simplest form of chromatin substrate from which to start studying the mechanism by which KDM2A interacts with chromatin in an *in vitro* context.

To test KDM2A binding affinity towards an *in vitro* assembled canonical nucleosome, the same EMSA-based approach described in chapter 3 was used. Indeed, up to 2 µg of the WT and K601A mutant recombinant KDM2A6 protein was each incubated with approximately 5000 cpm of non-methylated 147 bp 601 DNA containing mononucleosomes under the same experimental conditions adopted

for the previously described DNA binding assay. A control reaction containing only histone-free 601 DNA as probe was also included to confirm that the WT KDM2A protein tested is indeed capable of DNA binding.

Following separation of the binding reaction products via native 0.8% agarose gel electrophoresis, an autoradiography of the gel surprisingly revealed complete absence of interaction between both the WT and K601A KDM2A and the 147 bp mononucleosome substrate (Figure 4.7A). Importantly, a full shift was observed in the DNA only control lane suggesting that although capable of binding to DNA, KDM2A was somehow refractory to interaction with the 147 bp nucleosome. Several repeat EMSA assays, each using freshly reconstituted 147 bp mononucleosomes as binding substrates, yielded the same negative result suggesting that the observed result is unlikely to be explained as a stochastic event. Therefore, it was concluded that KDM2A is unable to interact with a canonical 147 bp mononucleosome under the binding conditions tested.

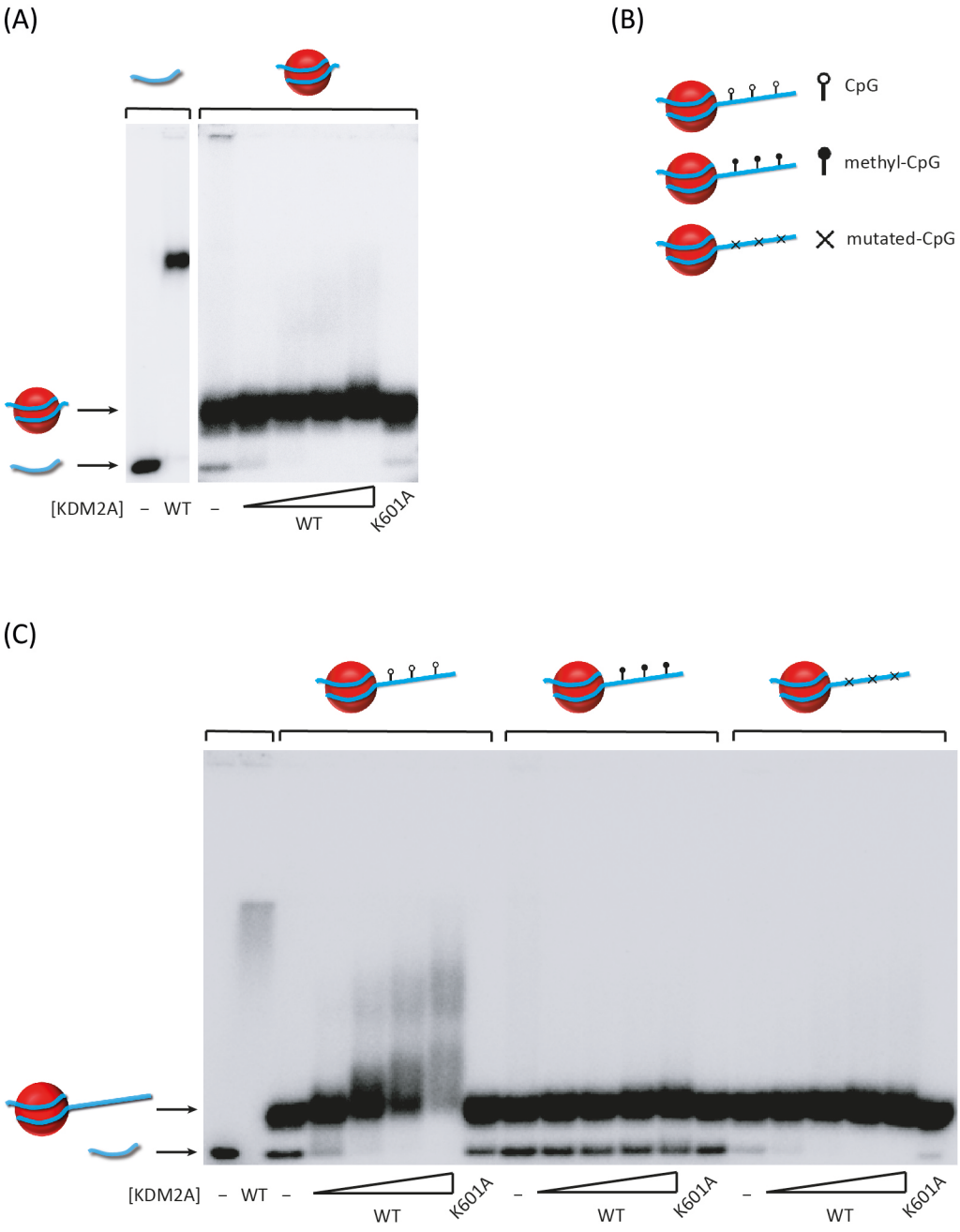


Figure 4.7 – KDM2A binding to mononucleosome

(A) WT and K601A KDM2A proteins were incubated with the 147 bp *in vitro* reconstituted nucleosome under the same EMSA condition described for the DNA-binding assay detailed in section of Materials and Methods. Although WT KDM2A was observed to display efficient DNA-binding capacity, the lack of any shift in the nucleosome containing lanes suggests that it does not interact with the 147 bp nucleosome under the binding conditions tested. The 147 bp DNA is non-

methylated and contains 14 CpGs. Sequence details are given in the Appendix. Shown here is a representative EMSA of three replicate assays. (B) Schematic depicting a 216 bp mononucleosome that carries 69 bp of CpG-containing linker DNA. Methylated and mutant variants are also shown. Open-lollipops represent non-methylated CpG sites, closed-lollipops indicate the methylated equivalents and x-marks stand for mutated CpGs. (C) KDM2A binds to the 216 bp nucleosome as evidenced by the protein concentration dependent shift of the bound chromatin complex and the concomitant depletion of the unbound nucleosomes. Importantly, such interaction occurs via CXXC domain-mediated recognition of the CpG sites within the linker DNA as evidenced by the abrogation of binding when these linker CpGs were purposely mutated. Shown here is one representative EMSA of two replicate assays.

However, it can be argued that, in living cells, nucleosomes are not found in isolation but form chromatin fibres where each nucleosome is spaced from its neighbour by a variable length of linker DNA. Therefore, in order to generate a more biologically relevant chromatin substrate, a 216 bp long 601 DNA-containing mononucleosome was reconstituted using the same method described for the shorter nucleosome species. As well as carrying the same 147 bp 601 core DNA sequence, the 216 bp nucleosome contains an extra 69 bp linker DNA region which extends from the nucleosome core (Figure 4.7B top). Importantly, this linker region contains 5 CpG sites.

When binding affinities of the WT and K601A KDM2A proteins towards the 216 bp mononucleosome were respectively tested by EMSA, in contrast to the 147 bp mononucleosome binding assay, a clear WT KDM2A concentration dependent formation of complex with chromatin was observed (Figure 4.7C). This result is evidenced by the increasing intensity in the gel shift correlated by increasing

depletion of the free, unbound 216 bp mononucleosomes. In contrast, the K601A mutant was unable to interact with this substrate indicating that the observed interaction is CXXC domain-dependent. In further support of this argument, nucleosomes carrying methylated 601 DNA were also seen to be refractory to KDM2A interaction (Figure 4.7B middle and C). Since the only difference between the 147 bp and the 216 bp nucleosome species lies in the 69 bp linker DNA, it was reasonable to hypothesise that the observed chromatin interaction occurs within this particular portion of the 216 bp nucleosome. To test whether this is the case, all four CpG sites within the linker DNA were mutated by PCR amplification using a mismatch primer pair and the derived mutant 216 bp 601 DNA was incorporated into nucleosomes. Indeed, when the mutant 216 bp mononucleosome that is specifically deprived of CpG sites in the linker region (Figure 4.7B bottom) was introduced into the same EMSA assay, no KDM2A binding events were observed (Figure 4.7C) indicating that KDM2A interaction with the 216 bp mononucleosome is indeed mediated by CXXC domain-dependent recognition of CpG sites within the linker DNA region.

The 69 bp linker DNA carries 5 CpGs, separated by a minimum of 6 nucleotides from each other. Given the footprint of around 6 bp previously estimated for the MLL1 CXXC domain (Birke et al., 2002), it is reasonable to expect that more than one KDM2A molecules can bind simultaneously to the 69 bp linker region. Indeed, with increasing concentration of KDM2A protein in the binding reaction, at least two species of complex can be observed (Figure 4.7 C, left panel, non-methylated

substrate). The slower migrating of the two species is likely to harbour more than one KDM2A molecules, although the exact number cannot be deduced from the gel. Further DNase I footprinting assays on KDM2A-bound 216 bp nucleosome would be needed to determine the location and number of bound molecules for each species.

KDM2A binds preferentially to non-methylated CpG dinucleotides in the linker region between two nucleosomes.

Since the above results revealed that linker DNA is critical for CXXC-dependent KDM2A interaction with mononucleosomes *in vitro*, it would be interesting to know whether such binding property is valid within a more complex, multi-nucleosome environment that more closely resembles chromatin found *in vivo*. Therefore, in order to generate a recombinant dinucleosome substrate for EMSA binding assay, a gift plasmid pBlu2SKP 2X601 +48 carrying two 147 bp 601 DNA sequences separated by 48 bp of CpG-containing linker region was obtained from Tom Owen Hugh's laboratory (See the Appendix for sequence details). From this template, a 342 bp long DNA sequence capable of precisely accommodating two nucleosomes was amplified by PCR using a custom primer pair and subsequently purified in the same way as for the 147 bp and 216 bp 601 DNAs (Figure 4.8A). By

applying the same reconstitution methodology used for generating mononucleosomes, a 342 bp long 2x601 dinucleosome was then assembled *in vitro* (Figure 4.8C). Indeed, with increasing amounts of histone octamer relative to a fixed amount of 2x601 DNA, a shift towards a higher proportion of fully assembled dinucleosomes (corresponding to the slowest migrating species in the gel) relative to sub-stoichiometric dinucleosomes (migrating just below the full dinucleosomes) was observed. This is an indication of occupancy at available 601 nucleosome positioning sites reaching saturation. Therefore, it is expected that the final reconstitution product isolated mainly consists of fully assembled dinucleosomes. Nonetheless, the exact positions of the two nucleosomes on the DNA template cannot be inferred from the gel and future DNase I footprinting assay will be required to address this question. Moreover, to minimise potential interference in EMSA caused by contaminating sub-stoichiometric dinucleosome species, limited digestion of the reconstituted material with a combination of 5' and 3' exonucleases would also be beneficial.

EMSA analysis of WT and K601A KDM2A affinity for the 2x601 dinucleosome showed that, as expected, KDM2A specifically interacts with CpG sites within the 48 bp linker region in a CXXC domain dependent manner (Figure 4.8B left panel). Indeed, targeted mutation of these CpGs was sufficient to fully abrogate KDM2A binding to the dinucleosome (Figure 4.8B right panel). Because the difference in molecular weight between KDM2A-bound and unbound dinucleosome is not as remarkable as in the case of the mononucleosome EMSA assay, the observed gel

shift is not very pronounced. In order to clearly demonstrate that KDM2A is able to bind to dinucleosomes, a KDM2A-specific antibody was included in the binding reaction during incubation. The presence of the antibody caused the formation of a larger antibody-bound KDM2A-dinucleosome complex that in the agarose gel corresponded to what is generally defined as a "supershift". Importantly, such "supershift" could not be observed when the linker CpG mutant dinucleosome was included as the binding substrate (Figure 4.8B). This result further confirmed that indeed KDM2A binding is specifically targeted to the linker region between two nucleosomes.

The inclusion of a protein-specific antibody in the EMSA is useful to confirm that the observed shift is protein-dependent and not derived from contaminating factors in the sample preparation. Nonetheless, it does not allow discrimination between the identities of the protein-bound substrates when a heterogeneous group of binding probes is present. Instead, as a complementary approach to the supershift experiment, KDM2A preferential binding substrate can be confirmed by adding non-radioactive dinucleosomes with or without CpGs in the linker to a reaction containing labelled non-methylated linker CpG dinucleosomes. By comparing the ability of each of the two non-labelled probes to compete with the radioactive dinucleosomes for KDM2A binding, the substrate preference of the protein can be thus deduced.

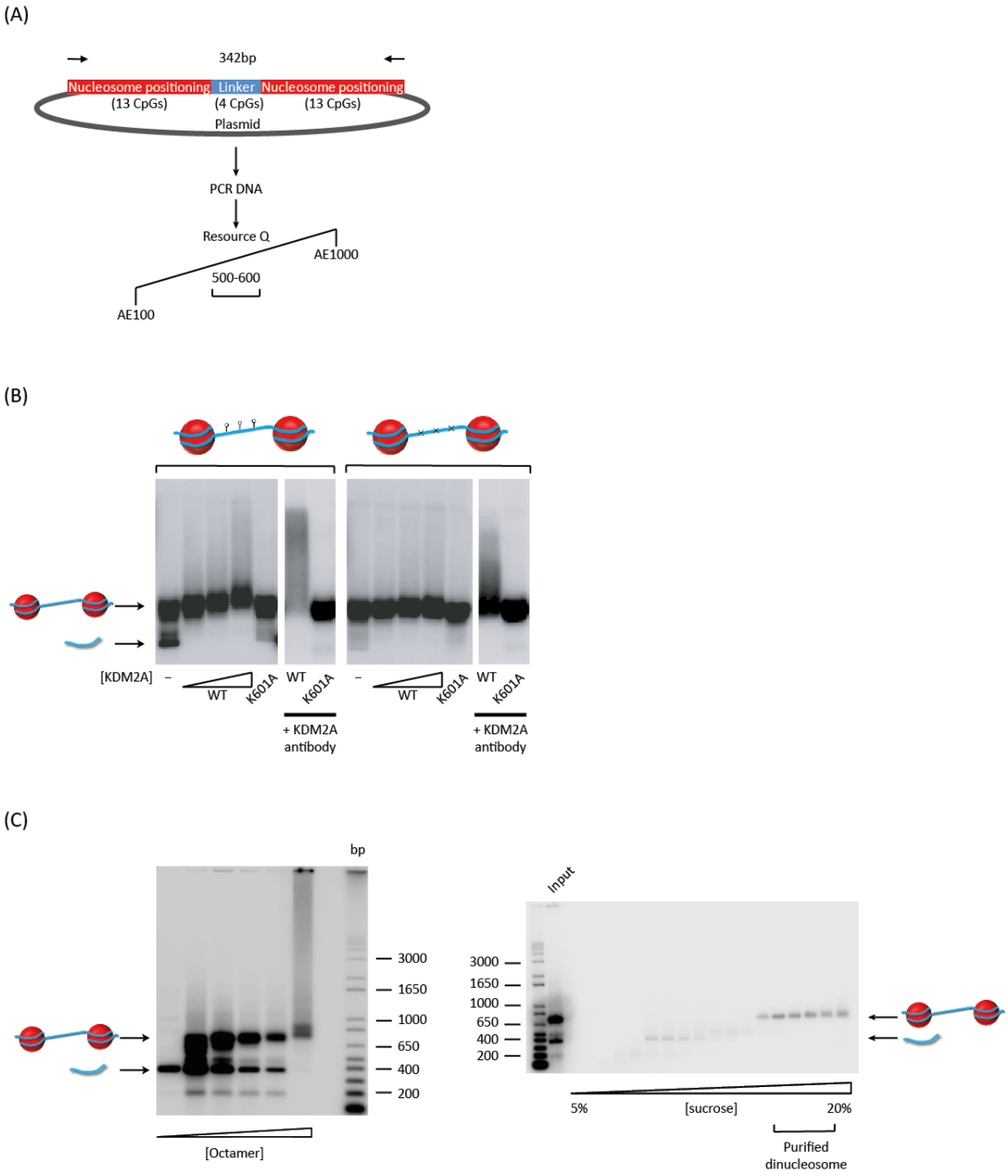


Figure 4.8 – KDM2A binding to dinucleosome

(A) A 342 bp DNA containing two 601 nucleosome positioning sequences spaced by a 48 bp linker region was PCR amplified from a template plasmid. The PCR products were purified using the same ion exchange chromatography approach used in Figure 4.2A. The 2x601 DNA was then used as template to reconstitute dinucleosomes *in vitro* (C). The numbers of CpGs in the two nucleosomal sequences and in the middle linker region is stated within brackets. (B) KDM2A was shown to interact with the dinucleosome by EMSA. In particular, this binding relied on a functional CXXC domain and the presence of CpGs within the linker region

between nucleosomes. This result was confirmed by the introduction of a KDM2A-specific antibody which "supershifted" the KDM2A-chromatin complex. The identities of the migrating species in the gel are shown by cartoon representations of the dinucleosome (two red balls spaced by a middle blue linker) and free DNA (short blue thread). In the left panel, dinucleosomes with non-methylated CpGs in the linker (open-lollipops in the linker) were used as substrates. In the right panel, dinucleosomes whose linker CpGs have been mutated (x in the linker) were used. Shown here is a representative EMSA of two replicate experiments. (C) The dinucleosomes were reconstituted using the same approach shown in Figure 4.6B. An increasing range of histone octamer was included in the reconstitution reaction to empirically identify the DNA to octamer ratio that yields the highest proportion of dinucleosomes as final products. Under sub-optimal concentrations of histone octamers, a duplet product can be observed migrating slower than free DNA in the gel (left gel, second lane from the left). The higher of the two bands is likely to correspond to fully reconstituted dinucleosomes while the lower band possibly consists of sub-stoichiometric dinucleosomes (e.g. species where only one of the two 601 sequences is occupied by histone octamer). Any contaminating free DNA was purified away by sucrose gradient (right gel) as similarly shown in Figure 4.6C for mononucleosomes. Shown here is one representative reconstitution and sucrose gradient of more than three experiments performed in this study.

4.3. Discussion

KDM2A preferentially interacts with linker DNA *in vitro*

One of the greatest advantages of an *in vitro* biochemical system lies in its simplicity where ideally a handful of defined factors can be introduced allowing the initial hypothesis to be tested without the presence of confounding variables that may complicate interpretation of the final experimental result. For this reason, *in vitro* biochemical tools are well suited for investigating the mechanistic detail of interactions between biomolecules.

To date, the zinc finger CXXC domain has been shown in a number of nuclear proteins to act as a DNA binding module and mediate genomic targeting *in vivo* (Blackledge et al., 2010; Frauer et al., 2011; Ma et al., 1993; Voo et al., 2000). However, a detailed knowledge of the mechanism by which such localisation is achieved within a chromatin context is lacking. In this chapter, by taking advantage of a recombinant chromatin system that features nucleosome substrates with defined physical and chemical properties, it was revealed for the first time how this particular zinc finger domain interfaces with chromatin. The above *in vitro* EMSA experiments have demonstrated that KDM2A requires the presence of CpG-containing linker DNA in order to bind to chromatin. This requirement was shown to be valid in both mononucleosomes carrying overhanging linker DNA and dinucleosomes. Recent genome wide nucleosome mapping studies suggest that the average linker length in humans is around 38 bp (Barski et al., 2007; Schones et al., 2008). This value is relatively similar to the 48 bp linker DNA found in the *in vitro* dinucleosome. Therefore, considering that the entire dinucleosome DNA sequence is highly enriched in CpGs (1.1 obs/exp and 56% GC content), it can be argued that this recombinant chromatin substrate closely recapitulates the particular chromatin environment that KDM2A is likely to encounter at human CGIs *in vivo*.

Why does KDM2A only bind to linker DNA?

The finding that the zinc finger CXXC domain can only interact with CpGs contained within the histone-free linker DNA is novel but not surprising given that nucleosomes are well known to inhibit DNA binding by many transcription factors in the cell. It is possible that the inability of the CXXC domain to bind to nucleosome core DNA is due to structural properties that are inherent to the way the CXXC domain approaches the DNA double helix. Analysis of recently published structures of the MLL1 and DNMT1 CXXC domains in complex with short stretches of substrate DNA (Cierpicki et al., 2009; Song et al., 2010) suggests that although discrimination of the CpG base pair is mediated by an extended surface loop region that inserts into the major groove of the DNA helix, many other electrostatic and hydrophobic interactions occur throughout the whole of the CXXC domain and are responsible for the stability of the DNA binding. In particular, a number of residues have been found that make contact with the minor groove of the DNA helix (Cierpicki et al., 2009). Site-directed mutagenesis of these residues has been observed to cause a 4 to 5 fold reduction in DNA binding affinity. Therefore, it seems that the CXXC domain requires interaction with both major and minor grooves of the DNA helix to achieve a stable binding. Considering the structure of a nucleosome core particle (Luger et al., 1997a), two reasons can be advanced that may provide a speculative explanation for the inability of the CXXC domain to access DNA sequences buried into the nucleosome core.

Firstly, while the DNA is wound around the octamer core, significant physical distortions of the DNA helix occurs generating a torsional stress that affects the direction towards which major and minor grooves are faced. Such alteration in orientation, which is not present in the more flexible linker DNA region, can limit the way the CXXC domain can access both major and minor DNA grooves at the same time.

Secondly, structural data has revealed that side chains from mostly arginine residues of the core histones can approach the inward facing major and minor grooves contributing to the stability of the nucleosome core particle (Luger et al., 1997a). Therefore, it is likely that the presence of these histone side chains may sterically clash with residues from the CXXC domain as it attempts to access the DNA helix.

How does KDM2A interact with *in vivo* chromatin?

KDM2A requirement for linker DNA in order to bind to chromatin is not limited to recombinant nucleosome substrates. In agreement with the *in vitro* results presented in this chapter, recent sucrose gradient sedimentation assays of chromatin isolated from mouse ES cells in the laboratory have shown that KDM2A exclusively co-sediments with the poly-nucleosome fraction while being completely absent from the fractions carrying mononucleosomes whose linker

DNA has been fully digested away (Zhou et al., 2011). Moreover, ChIP analysis of known KDM2A-targeted CGI promoter sites has shown that KDM2A is only found in chromatin carrying intact linker DNA. Together, these results demonstrate that KDM2A requires linker DNA for chromatin interaction *in vivo*.

Several factors so far known to influence KDM2A targeting across the genome can be summarised. Firstly, current data suggests that the CXXC domain is the main DNA binding module in KDM2A. Given that CpG dinucleotides are the only sequence requirement for such interaction and because each CXXC domain binding event is independent with no positive cooperativity involved, it would be expected that the concentration of KDM2A at any particular genomic locus is determined by the local density of CpGs. Secondly, because the CXXC domain can only bind to non-methylated CpGs, the methylation status of the target DNA is also important. These two binding requirements fit nicely with the unique properties displayed by CGIs, which are therefore, not surprisingly, preferential target sites for KDM2A in the genome (Blackledge et al., 2010).

The observation in this chapter that KDM2A is targeted only to CpG sites within linker DNA regions indicates that nucleosome occupancy represents an additional important factor in dictating genome-wide localisation of this protein. Therefore it is useful to consider how much of the genome might actually be accessible and contain non-methylated CpG dinucleotides for KDM2A to associate with. Firstly, CGIs are known to constitute less than 1% of the entire genome (Illingworth and

Bird, 2009). As mentioned earlier, the average genome GC content is estimated to be approximately 41% (Lander et al., 2001). Consequently, assuming equal proportions of G and C, the expected CpG frequency is given by 0.2 X 0.2 = 0.04. Nevertheless, the observed genome average CpG frequency is only 0.008 (thus obs/exp ratio of 0.2) due to under-representation of the cytosine base (Josse et al., 1961). This frequency is around 15 times higher in CGIs (i.e. 0.12), and corresponds to roughly 1 CpG in 10 bp (Bird et al., 1985; Illingworth et al., 2008). Importantly, CpGs appear to follow a random distribution across CGIs with no apparent consensus motifs (Bird et al., 1985). Therefore, assuming a normal distribution of CpGs, one would expect that a maximum of 0.1 x 1% = 0.1% of the genome to have non-methylated CpGs for KDM2A binding. Genome-wide nucleosome mapping shows that the average linker DNA length in humans is approximately 38 bp (Barski et al., 2007; Schones et al., 2008). However, studies show that at transcription regulatory elements (i.e. promoters and enhancers), where CGIs are predominantly found, the average linker length decreases to around 30 bp (Valouev et al., 2011). This is likely to be a consequence of chromatin remodeler-dependent nucleosome spacing (Schwarzbauer et al., 2012) and deficiency in the H1 linker histone at CGIs (Tazi and Bird, 1990). Indeed, the presence of H1 histone *in vivo* correlates with longer nucleosome repeat length (Fan et al., 2005). By applying the above 30 bp linker value, it can be estimated that at CGIs 17% of DNA sequence is found in linker region and therefore accessible to KDM2A. Combining this percentage with the frequency of CpGs in CGIs, which in turn represent less

than 1% of the genome, one can reasonably conclude from this simple calculation that a mere 0.017% of the genome is potentially targeted by KDM2A. Obviously there are many other factors and caveats that could alter the real number of CpGs that could be occupied by KDM2A in vivo. Nevertheless this calculation provides a useful way of demonstrating that nucleosome occupancy can limit the number of accessible KDM2A binding sites through a simple mode of nucleosome-based exclusion.

Interestingly, CGIs correlate with regions of DNase I hypersensitive sites suggesting that an open chromatin structure exists (Thurman et al., 2012), which may favour KDM2A accessibility to these genomic regions. Moreover, CGIs also highly correlate with TSSs, which are known to harbour strongly phased nucleosomes (Valouev et al., 2011). The precise position of these nucleosomes will in turn influence the probability a particular adjacent linker sequence is accessible and consequently the likelihood of KDM2A been observed at that defined sequence. In contrast, the bulk, non-CGI portion of the genome is usually characterised by low CpG density, pervasive DNA methylation, more compact and restricting higher order chromatin and random distribution of nucleosomes. All these elements act together to reduce the probability that at any given time, foci containing high number of KDM2A molecules can be observed across these non-CGI regions.

5. Chapter Five – Role of the PHD domain in KDM2A

5.1. KDM2A PHD domain is not required for binding to chromatin

Expression and purification of KDM2A PHD domain

In chapters 3 and 4, the zinc finger CXXC domain was uncovered as the main chromatin binding module in KDM2A and a detailed account of its binding properties in relation to defined DNA and chromatin substrates reconstituted *in vitro* was described. Interestingly, located a few amino acids away from the KDM2A CXXC domain is another zinc-binding domain, the PHD finger. This well conserved protein domain was first characterised in the *Arabidopsis* DNA-binding protein HAT3.1 based on sequence similarity with the homeodomain of the maize ZMHOX1a protein (Schindler et al., 1993). A key feature that became immediately apparent was the regular spacing of the cysteines and histidines, which led to the initial prediction that this protein domain may carry metal ion-binding capacity. Since this discovery, the PHD domain has been observed in a large number of nuclear proteins across different organisms including yeast, fly and human (Bienz, 2006). The most commonly found PHD domain motif comprises of a conserved

Cys4-His-Cys3 sequence that serves to coordinate two zinc ions and contribute to the establishment of an overall domain structure that closely resembles that of the RING domain (Schindler et al., 1993). Nonetheless, the PHD domain lacks the E2-ligase interacting surface that is usually found in RING-domain proteins and therefore does not display E3 ubiquitin ligase catalytic activity (Bienz, 2006).

Contrary to early suggestions of a role in DNA-binding, recent crystal structures of the ING2 subunit of the SIN3a HDAC histone deacetylase complex and of the BPTF subunit of the NURF chromatin remodelling complex have both revealed that the PHD domain acts as a H3K4me3-specific histone binding module (Li et al., 2006; Shi et al., 2006a). Additional structural evidence suggests that this domain can also recognise the non-methylated form of the histone H3 tail at lysine 4 in further support of the idea that the PHD domain has evolved to interpret histone modifications (Lan et al., 2007). Importantly, the PHD domain has been observed to cooperate with other chromatin binding modules. For example, in *S. cerevisiae*, the PHD domain of the Rco1 subunit in the Rpd3s HDAC complex has been shown to stabilise Rpd3s-nucleosome interaction and facilitate H3K36 binding by the chromodomain of the neighbouring Eaf3 subunit (Li et al., 2007). The phenomenon of multivalent chromatin is widespread among different nuclear protein complexes (Ruthenburg et al., 2007). Indeed, such combinatorial reading of more than one histone modification offers significant advantages over single protein interactions both in terms of the overall thermodynamics of the bound complex and the kinetics of binding. Given the close proximity of the CXXC and PHD domains in

KDM2A, it is tempting to speculate that the PHD domain may play a role in aiding CXXC domain recognition of chromatin substrate. In particular, because the PHD domains of other previously characterised chromatin binders have been demonstrated to specifically associate with H3K4me3 (Li et al., 2006; Shi et al., 2006a), a modification that is preferentially enriched at CGIs where KDM2A also predominantly localises to (Blackledge et al., 2010), it is reasonable to hypothesise that the PHD domain in KDM2A may also function to recognise this particular chromatin mark.

Therefore, to understand the biological function of this domain in KDM2A, an N-terminal 6xHis and C-terminal Strep II tagged-KDM2A construct carrying only the PHD domain was cloned and expressed in *E. coli* using the pNIC28 vector system described before (Figure 5.1C). The expressed recombinant protein was subject to standard Ni-NTA and StrepTactin double affinity purification (Figure 5.1A). To further refine sample purity, the combined elution from the StrepTactin resin was applied to a Superdex 200 gel filtration column (Figure 5.1A). UV spectrometry at 280 nm registered a single elution peak from the column and subsequent Coomassie staining confirmed the presence of a major protein band that corresponded to the expected molecular weight of KDM2A PHD domain (16 kDa) (Figure 5.1B). According to SDS-PAGE, an estimated > 90% purity was achieved in the Superdex 200 column elution with only residual contaminating degradation products present that were beyond the resolving capacity of this specific gel

filtration resin. This purified recombinant PHD construct was used for all following domain characterisation experiments.

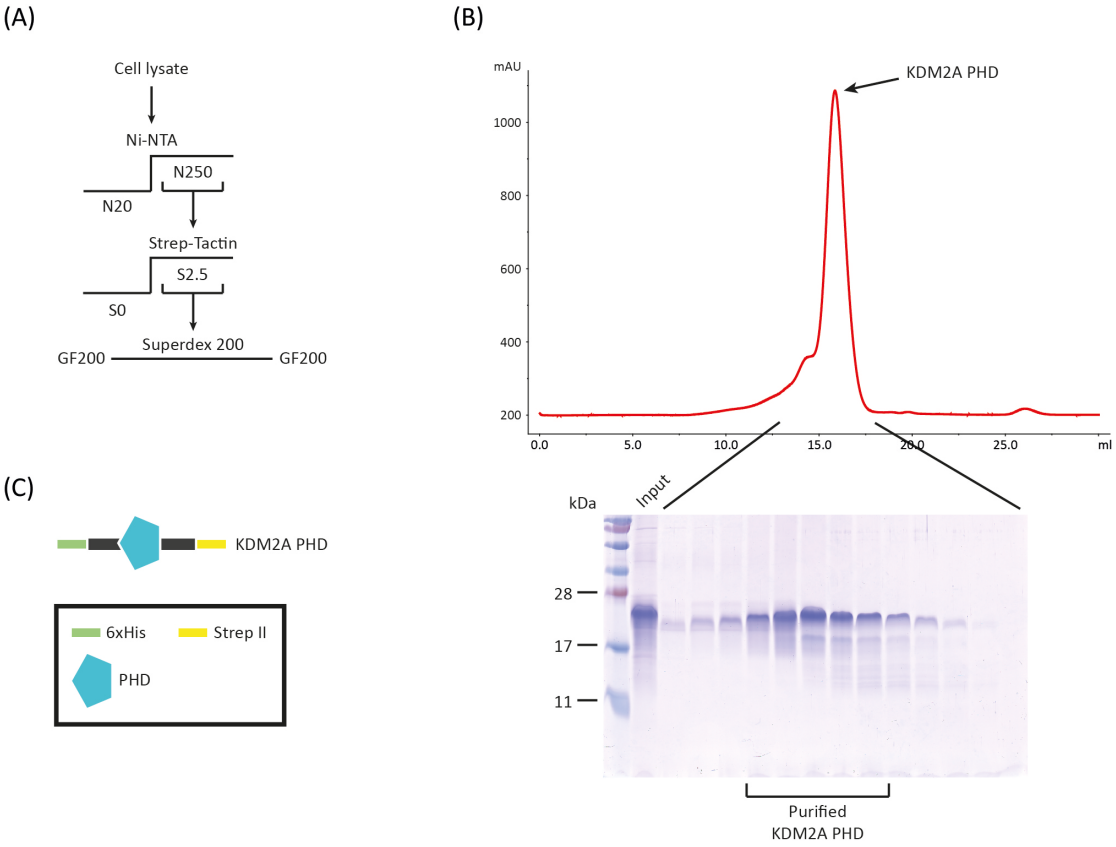


Figure 5.1 – Purification of KDM2A PHD domain

(A) Following expression of KDM2A PHD protein in BL21 *E. coli* cells, the cell lysate was subject to standard double affinity purification via Ni-NTA and StrepTactin columns as shown for KDM2A6 constructs in Figure 3.3A. The derived sample was further "polished" by gel filtration chromatography. Details of the experimental procedure are given in section 2.8 of Materials and Methods. (B) A single elution peak from the gel filtration column was visible with the associated protein species confirmed to be of the expected molecular size of KDM2A PHD by Coomassie-stained SDS-PAGE. Shown here is one representative gel of two repeat purifications performed in this study. (C) Schematic illustrating the KDM2A PHD construct carrying N-terminal 6xHis and C-terminal Strep II tags.

KDM2A PHD domain does not bind specifically to any known histone modification

Historically, investigation of protein-ligand binding affinity in relation to histone substrates has mainly relied on pull-down assays using immobilised recombinant histone tail peptides (Wysocka, 2006). However, given that pull-down assays are essentially based on tag affinity purification, the strength of protein-ligand interaction is inevitably influenced by the salt condition under which binding is tested. Consequently, depending on the arbitrary value of the salt concentration applied during pull-down, contrasting results can be observed. Indeed, pull-down assays are notoriously prone to artefact formation due to the difficulty in optimising the signal to noise ratio and any observed interactions generally need to be confirmed by alternative methods. Moreover, only one potential substrate can be tested at a time making this technique extremely laborious and unsuited for large-scale screening of interactions.

Recently, high-throughput proteomic techniques such as the peptide microarray-based approach have been developed to overcome this limitation (Nady et al., 2008; Rothbart et al., 2012). The principle of the peptide array is essentially the same of that exploited by an ELISA (enzyme-linked immunosorbent assay). Indeed, protein-specific antibodies are utilised to detect any stable protein-ligand interaction over a vast array of binding baits that are pre-immobilised over a supporting matrix. This method has the advantage of providing an unbiased

approach where a wide range of binding substrates can be tested under the same experimental condition at once.

In order to identify potential histone modifications that may be specifically recognised by the KDM2A PHD domain, a collaboration was set up with Cheryl Arrowsmith's laboratory where a comprehensive screening of the purified KDM2A PHD domain binding affinity towards approximately 600 of the currently known modifications on all core histones was performed using a custom-made peptide microarray. Unfortunately, none of these modifications were observed to associate specifically with the KDM2A PHD domain under the assay condition tested suggesting that this particular protein domain is not likely to function as a histone-binding module (data not shown).

However, it can be argued that a limited fragment of histone tail is unlikely to fully recapitulate the physiological substrate that the KDM2A PHD domain interfaces with *in vivo* as possible interactions with nucleosome core regions may also be important for substrate recognition. Therefore, in order to investigate the binding properties of the PHD domain in relation to more biologically relevant chromatin substrates, the same *in vitro* nucleosome reconstitution approach that enabled detailed characterisation of the KDM2A CXXC domain function was applied here.

Installation of methyl analogues on histone H3 lysine 4

Given the laborious preparation and processing steps required to assemble nucleosomes with defined physical and chemical properties *in vitro*, a comprehensive screening study where distinct nucleosomes each carrying a specific histone modification are tested in individual binding assays would be time-consuming and impractical. Since H3K4me3 distribution highly correlates with CGIs where KDM2A localisation is observed *in vivo* (Blackledge et al., 2010), it was obvious to investigate whether the KDM2A PHD domain could be involved in the recognition of H3K4 tri-methylated nucleosomes.

To this end, a recently published chemical methylation method which allows for the *in vitro* site-specific installation of methyl groups on histone tails was chosen to generate recombinant H3K4me3 histones for subsequent incorporation into *in vitro* assembled nucleosomes (Simon et al., 2007). This semi-synthetic approach is based on the alkylation of a precisely positioned cysteine residue into N-methylated aminoethylcysteine, an analogue of methyl-lysine (Figure 5.2A). Indeed, the derived residue side chain is chemically and structurally almost identical to that of a natural methyl-lysine with length difference of only ~0.28 Å. By applying this method, tri-methyl-lysine analogues were specifically generated using purposely K4C mutated recombinant H3 histones. The close structural resemblance of this product to *bona fide* H3K4me3 was demonstrated by Western blot using H3K4me3-specific antibody, which was able to discriminate this H3K4me3

analogue from the non-methylated H3K4 counterpart (Figure 5.2B). This specific H3K4me3 histone was then incorporated into 216 bp 601 DNA-containing nucleosomes using the reconstitution procedure described before (data not shown).

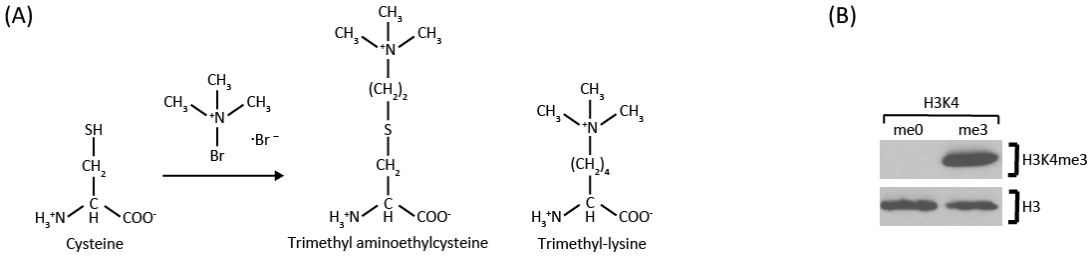


Figure 5.2 – Installation of H3K4 trimethyl-lysine analogue

(A) Alkylation by (2-bromoethyl)-trimethylammonium bromide of the cysteine that replaced the original lysine residue (H3K4) generates a tri-methylated form of aminoethylcysteine, which is structurally almost equivalent to a natural trimethyl-lysine (shown on the right for comparison). (B) This similarity was verified by Western blot using an H3K4me3-specific antibody, which was able to distinguish between the methyl-analogue (right lane) and the non-methylated lysine residue (left lane). Shown here is a representative blot of two repeat alkylation assays performed in this study.

Neither KDM2A PHD domain or H3K4 trimethylation is required for KDM2A binding to chromatin

With the above synthetically H3K4 methylated 216 bp mononucleosome, the same previously described EMSA-based binding assay was used to analyse KDM2A PHD affinity for this particular chromatin substrate. With the purpose of serving as a

negative control for binding, a mutant version of the PHD domain construct carrying a C645A mutation, which was shown in other PHD domain-containing proteins to affect both structure and function, was engineered (Figure 5.3A and E) (Capili et al., 2001; Kalkhoven et al., 2002). The effect of this mutation on binding capacity in relation to both non-methylated and H3K4me3 nucleosomes was then examined. As evidenced by EMSA, no apparent difference in binding affinity was observed between the WT and the C645A mutant PHD domain with respect to both DNA and non H3K4-methylated nucleosome substrate suggesting that the PHD domain does not contribute to KDM2A DNA binding (Figure 5.3B). Importantly, both KDM2A PHD proteins were observed to bind with equal efficiency to the H3K4me3 modified nucleosome (Figure 5.3C). This is supported by the fact that even at non-saturating concentrations of protein, the same extent of retardation of the migrating complex for KDM2A WT and C645A can been seen. It can be argued however that, if a minor binding preference of the PHD domain for H3K4me3 exists, this may not be apparent by directly comparing the degree of retardation in the gel. An alternative approach in future studies would be to evaluate the ability of non-labelled H3K4me3 substrate to compete with labelled-H3K4me0 equivalents and vice versa. The substrate preference could then be deduced by assessing the extent by which binding of the labelled-probe is inhibited. Regardless of whether a subtle preference exists or not, it is reasonable to conclude at least that H3K4me3 does not constitute as a major binding substrate

for the PHD domain in KDM2A. This result is also in agreement with the above peptide array data.

In order to understand the inability of the KDM2A PHD domain to interact with modified histones, sequence alignment of PHD domains derived from known H3K4me3 (ING2, BPTF, PHF8) and H3K4me0 (BHC80, DNMT3L) binding proteins was carried out and compared to that of KDM2A (Figure 5.3D). A highly conserved cysteine-rich sequence was immediately apparent in all proteins suggesting an overall structural homology is likely to exist. Nonetheless, previously published crystal structures of the aligned PHD domains have revealed that a handful of strategically positioned aromatic residues play a critical role in the specific interrogation of the histone H3 lysine 4 side chain. These residues are instead absent in the KDM2A protein in further support of the idea that the PHD domain of this particular protein has not evolved to recognise histone substrates.

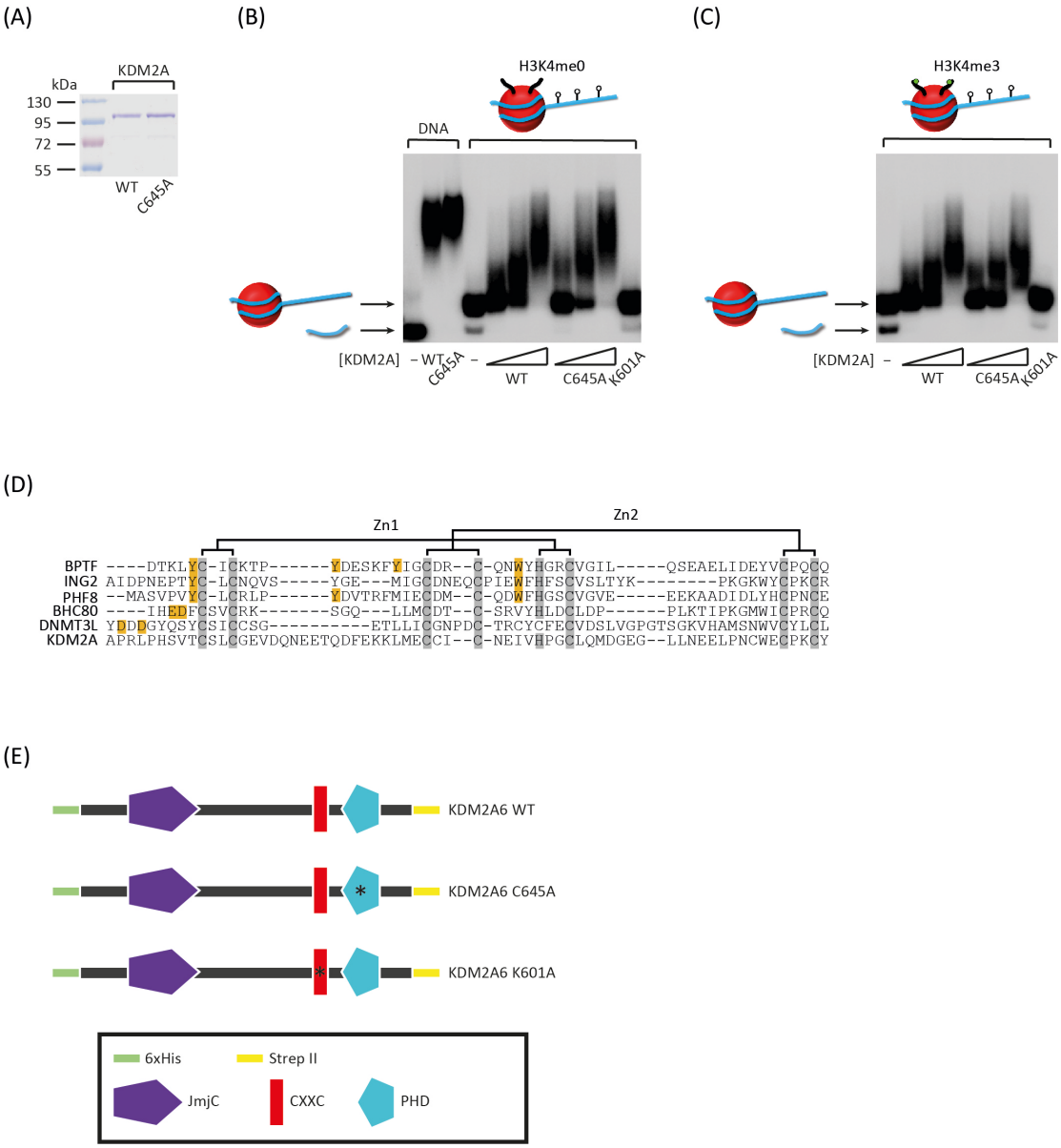


Figure 5.3 – KDM2A PHD domain does not bind to H3K4me3 nucleosomes

(A) SDS-PAGE showing the C645A mutant construct of the KDM2A PHD domain (right lane). (B) The PHD domain is not involved in DNA recognition as evidenced by the same binding efficiency in both WT and mutant proteins. The same EMSA conditions applied in the assay shown in Figure 4.7C were used here. The nucleosomes carry the same non-methylated linker CpG (open lollipops)-containing 216 bp 601 sequence previously described. (C) The KDM2A PHD domain does not function as H3K4me3-specific binding module as shown by the lack of difference in binding affinity between WT and mutant proteins towards nucleosomes incorporating H3K4 trimethyl-analogues. The methyl analogue is

depicted as a green dot on black histone tail in the cartoon representation of the nucleosome. Gels shown in (B) and (C) are representatives of two replicate experiments. (D) Sequence alignment of the PHD domain from known H3K4me3 and H3K4me0 interacting proteins and KDM2A. The alignment indicates that KDM2A PHD has a conserved cysteine/histidine (highlighted in grey)-dependent structure. This is evident from the conserved pairing of the seven cysteines and one histidine that are responsible for the coordination of two zinc ions. Nonetheless, it is clear that KDM2A PHD lacks the specific aromatic residues (highlighted in orange) that are important for histone recognition. (E) Schematic illustrating the domain architecture of the KDM2A constructs used for the EMSA experiments in (B) and (C). The point mutation is represented by an asterisk.

5.2. Discussion

The KDM2A PHD domain does not have histone binding function

In this chapter, through both unbiased peptide array screening approach and targeted nucleosome binding assays with defined histone modification, it was shown that the KDM2A PHD domain does not carry chromatin interacting activity. In particular, the lack of binding towards H3K4me3 nucleosomes indicates that genome-wide targeting of KDM2A is likely to be independent of the presence of this active transcription mark. Consequently, it can be speculated that the biological function of KDM2A *in vivo* is most likely not directly associated with transcription activity in fitting with the idea of KDM2A acting as a basal chromatin maintenance factor.

However, recent proteome analyses of a number of PHD fingers in yeast indicate that the binding affinity of PHD domains for chromatin substrates can present K_D values above the 100 μM range and therefore beyond the sensitivity of the peptide array method (Shi et al., 2006b). Indeed, it cannot be completely ruled out that the KDM2A PHD domain may display very weak binding affinity for histone marks other than H3K4 methylation. However, even if such weak interaction was proven to be real, its biological significance would remain to be argued.

PHD domain targeting to histone modifications other than H3K4 methylation has also been reported. For example, a couple of yeast PHD domain proteins including Ecm5 and Nto1 have been shown in recent studies to carry limited binding preference for H3K36 methylation (Shi et al., 2006b). Other nuclear proteins like the human DPF3b has been shown to interact with acetylated H3K14 via a tandem PHD finger (Zeng et al., 2010). Interestingly, structural modelling of PHD domains of Ecm5 and Nto1 by comparison with published H3K4-interacting PHD domain structures suggests that the aromatic binding surface may represent a unified way of trapping tri-methyl groups in general while adjacent protein sequences are important for dictating specificity towards a particular lysine residue. Given that these aromatic residues are completely absent from KDM2A, it is unlikely that the PHD domain in this protein act as a histone binder.

It is interesting to note that Jhd1, the yeast orthologue of the mammalian KDM2A protein, carries a functional PHD domain that has been shown to mediate

H3K4me3 recognition (Fang et al., 2007; Shi et al., 2006b). Importantly, this domain aids in targeting of the H3K36 demethylase activity specifically at gene promoters in a genome without DNA methylation in a similar fashion to the role played by the CXXC domain in KDM2A. Therefore, an attractive scenario emerges where the advent of the DNA methylation system in mammalian genome and the increasing reliance on CXXC-domain dependent genomic recruitment may have caused a progressive loss in selective pressure throughout evolution which ultimately led to the abrogation of the histone-binding function of the PHD domain in KDM2A.

Possible functions of the KDM2A PHD domain

PHD fingers are also known to mediate protein-protein interactions (Musselman and Kutateladze, 2011) and it can be therefore speculated that the KDM2A PHD domain may indeed be responsible for association with partner proteins. However, to date no evidence of KDM2A being in complex with other nuclear proteins has been found.

Alternatively, given the structural similarity of the PHD domain to the RING domain, it would also be interesting to investigate the possibility of KDM2A possessing ubiquitin-ligase activity by performing a comparative analysis of the recently published crystal structure of KDM2A PHD domain (PDB ID: 4BBQ) with

documented RING domains. In particular, it would be important to see whether the E2 ligase-binding surface that is characteristic of RING domains is conserved in KDM2A.

6. Chapter Six – General discussion and future directions

6.1. Chromatin accessibility is important for KDM2A localisation

The initial drive that promoted the study presented here lies in our limited understanding of the mechanistic details that enable most chromatin modifying enzymes to acquire precise and specific genomic targeting within a complex and dynamic chromatin landscape. The rapid advent of high-throughput massively parallel sequencing technology over recent years has undoubtedly facilitated an increasingly detailed mapping of epigenetic marks and chromatin-associated factors across the genome (Park, 2009). Although such descriptive studies are important for deciphering gene regulatory networks that contribute to essential biological processes, they lack in the capability to provide a mechanistic basis that underlies the observed pattern of distribution. Therefore, in this study, an *in vitro* biochemical system based on recombinant chromatin substrates with well defined physical and chemical properties was first developed. Such system was then utilised to uncover the molecular mechanism by which the recently discovered histone demethylase KDM2A protein interacts with chromatin. Evidence is

provided here that suggests this particular enzyme preferentially associates with non-methylated CpG containing DNA sequences via a zinc finger CXXC domain. In particular, this interaction is inhibited when the DNA substrate is incorporated within a nucleosome. As shown by solved crystal structures, such limitation is likely dictated by the requirement of the CXXC domain to simultaneously access both faces of the DNA double helix in order to establish a stable binding complex (Cierpicki et al., 2009).

This reliance on histone-free DNA sequences implies that DNA accessibility governed by local chromatin structure can considerably restrict KDM2A localisation *in vivo*. Genome wide profiling of KDM2A distribution via ChIP-Seq indicates that it preferentially localises at CGIs and targeting to these genomic loci is strictly dependent on a functional CXXC domain (Blackledge et al., 2010). Given the specificity of this protein domain for non-methylated CpGs, KDM2A is naturally attracted to CGIs because of the elevated density of non-methylated CpGs compared with the remaining bulk of the genome where pervasive CpG methylation prevails. By performing both sucrose gradient fractionation of MNase digested global chromatin and ChIP profiling at known KDM2A CGI target gene loci, it was recently demonstrated in mouse ES cells that a clear restriction of KDM2A binding sites to linker DNA regions also occurs *in vivo* (Zhou et al., 2011). Indeed, this binding property is likely to play a specific role in aiding KDM2A localisation to CGIs. This is supported by observations in mouse ES cells which lack all three DNMTs and detectable DNA methylation where it was observed that KDM2A is

retained at established CGI regions without redistribution to newly created non-methylated CpGs elsewhere in the genome (Zhou et al., 2011). This result suggests that absence of methylation is not the only determinant for KDM2A binding to DNA *in vivo*. Instead, local chromatin structure is likely to be also a contributing factor. CGIs notoriously overlap with DNase I hypersensitive sites, which are thought to represent a relatively open and accessible chromatin environment (Thurman et al., 2012). This is not surprising given that CGIs are predominantly found at the 5' end of genes and correlate with H3K4me3 enriched euchromatin domains. In contrast, other parts of the genome can be subject to higher order chromatin organisation mediated by the linker histone H1. This translates into a more compacted chromatin conformation where linker DNAs are less accessible for interaction (Zlatanova et al., 1994). Therefore, KDM2A presence may be limited to CGIs in part because chromatin organisation at other genomic regions is unfavourable for binding and under normal circumstances methylated on CpG dinucleotides.

It is interesting to note that previous biochemical analysis of chromatin composition at CGIs has revealed a specific depletion of histone H1 at these genomic elements (Tazi and Bird, 1990). Given the relative lack of specificity of histone H1 interaction with DNA substrates, it can be speculated that the absence of histone H1 at CGIs may be a consequence of active displacement process. Past biochemical studies have shown that deacetylation of histone tails in particular at H4K16 is critical for nucleosome mobilisation mediated by chromatin remodelers

like ACF/ISWI and chromatin compaction (Shogren-Knaak, 2006). Since open state chromatin is generally acetylated, this particular modification may affect histone H1 binding at CGIs. Alternatively, the histone variant H3.3 which is enriched at TSS has also been shown to inhibit H1 histone binding to chromatin (Braunschweig et al., 2009). A third scenario involves exclusion of the H1 histone by direct binding competition of other linker DNA targeting proteins like KDM2A.

Future research work in the laboratory will aim at addressing these possibilities by taking advantage of an *in vitro* chromatin system where recombinant H1 histone can be incorporated into reconstituted chromatin substrates. To this end, a previously published ATP-dependent chromatin assembly approach that relies on the histone chaperone NAP1 and the chromatin remodeler ACF/ISWI will be utilised to reconstitute regularly spaced nucleosome arrays (Fyodorov and Kadonaga, 2003). Importantly, a CpG-rich DNA template similar to the 601 sequence illustrated in chapter 3 and 4 will be chosen for nucleosome reconstitution to closely recapitulate the chromatin environment observed *in vivo*. By providing with these *in vitro* chromatin templates, the respective binding affinities of KDM2A and histone H1 will be compared through competition assay in EMSA. Alternatively, histone H1 can be pre-assembled onto chromatin to generate higher order structure where cross-fibre interactions are likely to be present and linker DNAs less accessible. Such compacted chromatin organisation is believed to be characteristic of heterochromatic domains *in vivo* (Dillon, 2004). Given its requirement for accessible linker DNA, KDM2A would be expected to display low

affinity towards this type of chromatin. Therefore, by analysing KDM2A binding behaviour, it will be possible to test the idea that the presence of higher order chromatin structure alone, irrespective of the underlying DNA sequence, is a major impediment to KDM2A localisation in the genome.

Perhaps even more interestingly, acetylated histones and histone variants like H3.3, which are characteristic of TSSs, can also be incorporated either individually or together into the reconstituted chromatin. Again, by evaluating changes in KDM2A binding affinity, it will be possible to assess potential roles these chromatin features may play in influencing KDM2A targeting to CGIs.

6.2. KDM2A contributes to a transcriptional permissive chromatin environment at CGIs

Having illustrated the molecular mechanism that enables KDM2A targeting at CGIs, questions remain regarding the exact contribution of this particular histone demethylase enzyme towards CGI function. Previous ChIP analysis at CGIs has shown that the level of H3K36me2 is significantly depleted at KDM2A-bound regions while suppression of KDM2A activity via RNAi was able to restore the methylation level indicating that KDM2A plays a role in regulating the abundance of this particular histone modification at CGIs (Blackledge et al., 2010). Removal of

H3K36me2 inevitably also limits the presence of H3K36me3, as the di-methylated state is a necessary biological substrate for SETD2 mediated conversion. Given that H3K36me2 is thought to function in part by recruiting the Rpd3S histone deacetylase complex to repress transcription initiation (Carrozza et al., 2005; Li et al., 2009a; Youdell et al., 2008), it is rather tempting to speculate that KDM2A may antagonise this pathway by creating a chromatin environment that is permissive to transcription initiation.

In a similar fashion to KDM2A, it was recently demonstrated that the CXXC domain-containing protein CFP1 also co-localises with CGI sequences (Thomson et al., 2010). Although CFP1 does not carry any chromatin modifying activity per se, it can associate with the histone H3 lysine 4 (H3K4) methyltransferase complex (SET1 complex) and target tri-methylation of H3K4 (H3K4me3) to CGIs (Lee, 2005). Importantly, reduction of CFP1 presence at CGIs results in consequent depletion of H3K4me3 in the same regions. Given that CGIs highly correlate with gene promoters, the CFP1-targeted increase in local H3K4me3 concentration may in turn attract the NURF chromatin remodelling complex and the SAGA histone acetyltransferase complex to reinforce an open accessible chromatin architecture that favours transcription factor and RNA Pol II binding (Vermeulen et al., 2010; Wysocka et al., 2006). Consistent with this scenario, promoter-proximal pausing of RNA Pol II is a widespread phenomenon observed at many TSS, suggesting that the unique chromatin environment at CGIs may facilitate constitutive binding of the RNA Pol II without productive transcription elongation occurring (Core et al., 2008;

Seila et al., 2008). The advantage of harbouring a paused RNA Pol II may be explained by recent studies in murine macrophages where it was observed that rapid induction of CGI promoter containing immediate early genes occurred independently of SWI/SNF chromatin remodelling as opposed to non CGI-containing genes which required ATP-dependent remodelling prior to expression (Ramirez-Carrozzi et al., 2009).

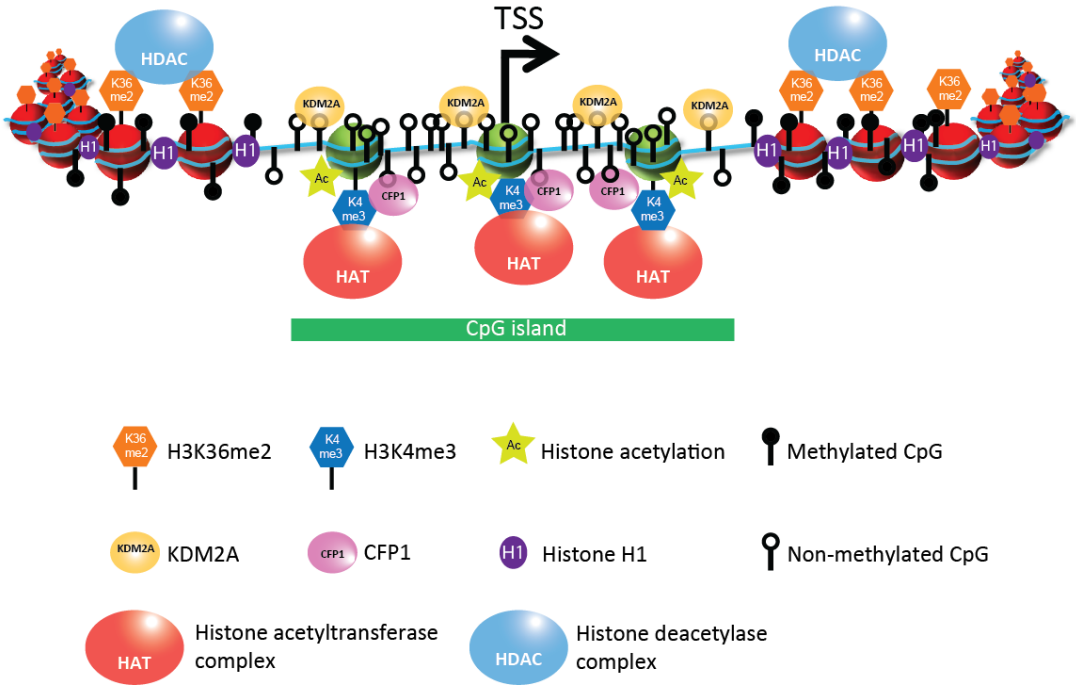


Figure 6.1 – zinc finger CXXC proteins KDM2A and CFP1 contribute to an open and transcriptionally permissive environment at CGIs

In a mammalian genome, the zinc finger CXXC domain proteins KDM2A and CFP1 specifically localise to CGIs (Blackledge et al., 2010; Thomson et al., 2010), which are predominantly found at promoters of most genes (TSS, transcription start site) (Deaton and Bird, 2011). Their targeting occurs through specific recognition by the CXXC domain of non-methylated CpG dinucleotides that are available in the linker DNA between nucleosomes (Zhou et al., 2011). Access to these sequences is likely

to be facilitated by the open chromatin architecture present at CGIs. This is in contrast to the surrounding non-CGI regions, which display a relatively more condensed, higher order chromatin organisation. Such chromatin domains are maintained through the nucleosome "cross-linking" function of the linker histone H1 and a lack of histone acetylation derived from the action of HDAC homologous to the yeast Rpd3S complex (e.g. HDAC1 and 2 of the Sin3a complex). It is believed that recruitment of the HDAC occurs via binding to H3K36me2 in a similar fashion to the Rpd3S-dependent pathway in yeast (Carrozza et al., 2005). Therefore, the specific removal of H3K36me2 by KDM2A is likely to prevent localisation of HDACs at CGIs via the above mechanism. At the same time, CFP1 mediates the specific deposition of H3K4me3 by the SET1 complex. This modification can in turn be recognised by HAT complexes like SAGA (Vermeulen et al., 2010). The consequent presence of acetylation (in particular H4K16ac) is known to disfavour chromatin compaction (Shogren-Knaak, 2006). Therefore, the cooperative action of KDM2A and CFP1 at CGIs contribute to the reinforcement of an accessible local chromatin environment that is permissive to transcription initiation.

The high degree of sequence conservation of the CXXC-domains among the nuclear proteins listed in Figure 3.8A implies they are likely to localise at CGIs via the same targeting mechanism found in KDM2A. In particular, recent ChIP sequencing data in the laboratory indicates that the KDM2A paralogue, KDM2B shares a similar genome-wide distribution preference at CGIs (Farcas et al., 2012). However, in contrast to KDM2A, KDM2B has been observed to form part of a variant PRC1 complex (Farcas et al., 2012; Gao et al., 2012; Gearhart et al., 2006; Sanchez et al., 2007). Interestingly, the KDM2B CXXC domain has been shown to be required for targeting of PRC1 to a subset of CGI promoters that appear to function as mammalian equivalents of *Drosophila* PREs. (Farcas et al., 2012). The resulting "bivalent" promoters contain both H3K4me3 derived from the SET1 enzyme complex and H3K27me3 deposited by PRC2 (Bernstein et al., 2006). These genes

are considered to be poised between active transcription and stable repression. Therefore, a model is emerging where the chemical and structural properties of CGI chromatin is specifically interpreted by the CXXC domain to bring together via a shared mechanism multiple chromatin modifiers. The convergence of different enzymatic activities cooperatively shapes a unique chromatin environment at CGIs that is permissive to nucleation of activities that can then function to activate or silence gene expression.

As mentioned earlier, two homologues of the mammalian KDM2A protein, dKDM2 and Jhd1 can be found in *D. melanogaster* and *S. cerevisiae* respectively. Interestingly, unlike vertebrates, the genomes of both these organisms are known to be deficient in DNA methylation (Bird and Taggart, 1980; Proffitt et al., 1984). Consequently, CpG dinucleotides are found at the expected statistical frequency and the whole genome is CGI-like. In such a scenario, a CXXC-domain dependent chromatin recognition mechanism like the one illustrated for mammalian KDM2A would not be sufficient to ensure correct targeting of the enzymatic activity to specific genomic loci. This raises the intriguing question of how the *Drosophila* dKDM2 and the yeast Jhd1 protein achieve specific localisation in their respective genomes. As stated before, the overall domain architecture between KDM2A and dKDM2 is well conserved. Closer comparison of the respective CXXC domains indicates that the structural cysteines and the CpG-interacting residues are highly conserved between the two species (data not shown) suggesting that a functional conservation of the CXXC domain in dKDM2 may also exists. Therefore, in the

absence of other contributing factors, one would expect dKDM2 to bind pervasively via the CXXC domain over the entire genome with local density likely influenced by chromatin accessibility. Instead, studies show that dKDM2 is part of the dRAF repressive complex, which contains a number of PcG proteins (Lagarou et al., 2008). Therefore, it appears that the demethylase activity of dKDM2 may instead be specifically recruited to Polycomb responsive elements (PREs) by taking advantage of the PcG targeting mechanism. Interestingly, it was recently shown that the mammalian paralogue of KDM2A, KDM2B, also interacts with Polycomb proteins indicating that the single KDM2 orthologue in flies may be more related functionally to KDM2B than KDM2A (Farcas et al., 2012). Sequence alignment of KDM2A and dKDM2 PHD domains shows that all the zinc-coordinating cysteines are conserved in dKDM2 although, like in KDM2A, the aromatic residues that have been shown in other PHD domains to mediate H3K4 binding are absent (data not shown). In lack of literature information regarding this domain in dKDM2, its function remains to be investigated.

In contrast to dKDM2, the yeast homologue Jhd1 does not carry a zinc finger CXXC domain and does not appear to associate stably with any nuclear partner proteins (Fang et al., 2007). However, like its human and fly homologues, Jhd1 contains a PHD domain, found N-terminally to the catalytic JmjC domain. Interestingly, close inspection of this domain sequence has revealed that, not only the structural cysteines are conserved, but also the existence of two tryptophan residues can be observed (data not shown). These occupy conserved positions with respect to the

aromatic residues that bind H3K4me3 in the PHD domains of ING2 and BPTF (Li et al., 2006; Shi et al., 2006a) and therefore may have functional relevance. This finding is in agreement with recent biochemical evidence of Jhd1 PHD domain binding to H3K4 trimethylated peptide *in vitro* (Shi et al., 2006b). Considering that Jhd1 shows conserved demethylase activity towards H3K36me2 (Fang et al., 2007; Kim and Buratowski, 2007), and given that this modification in yeast is believed to inhibit transcription via recruitment of the Rpd3s deacetylase complex (Carrozza et al., 2005; Li et al., 2009a; Youdell et al., 2008), it would be rather tempting to speculate that Jhd1 may be recruited via PHD domain-binding of H3K4me3 to active gene promoters, where, like KDM2A, it can contribute to the maintenance of a local chromatin environment that is permissive to transcription initiation (Fang et al., 2007).

References

Aapola, U., Lyle, R., Krohn, K., Antonarakis, S.E., and Peterson, P. (2001). Isolation and initial characterization of the mouse Dnmt3l gene. Cytogenet Cell Genet *92*, 122-126.

Adkins, M.W., Howar, S.R., and Tyler, J.K. (2004). Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast PHO5 and PHO8 genes. Mol Cell *14*, 657-666.

Agalioti, T., Chen, G., and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. Cell *111*, 381-392.

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572-576.

Alberts, B. (2002). Molecular biology of the cell, 4th ed / Bruce Alberts ... [et al.] edn (New York: Garland Science).

Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J.*, et al.* (2008). A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. Cell *134*, 599-609.

Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., and Warren, A.J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. EMBO J *25*, 4503-4512.

Allis, C.D., Jenuwein, T., and Reinberg, D. (2006). Epigenetics (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet *23*, 185-188.

Andersen, E.C., and Horvitz, H.R. (2007). Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. Development *134*, 2991-2999.

Anderson, J.D., and Widom, J. (2000). Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. J Mol Biol *296*, 979-987.

Angelov, D., Lenouvel, F., Hans, F., Muller, C.W., Bouvet, P., Bednar, J., Moudrianakis, E.N., Cadet, J., and Dimitrov, S. (2004). The histone octamer is invisible when NF-kappaB binds to the nucleosome. J Biol Chem *279*, 42374-42382.

Antequera, F., Tamame, M., Villanueva, J.R., and Santos, T. (1984). DNA methylation in the fungi. J Biol Chem *259*, 8033-8036.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W., and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell *86*, 263-274.

Bai, L., Charvin, G., Siggia, E.D., and Cross, F.R. (2010). Nucleosome-depleted regions in cell-cycle-regulated promoters ensure reliable gene expression in every cell cycle. Dev Cell *18*, 544-555.

Bai, L., and Morozov, A.V. (2010). Gene regulation by nucleosome positioning. Trends Genet *26*, 476-483.

Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J., and Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. EMBO J *28*, 3785-3798.

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J Biol Chem *280*, 17732-17736.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. Trends Genet *27*, 454-464.

Barbaric, S., Fascher, K.D., and Horz, W. (1992). Activation of the weakly regulated PHO8 promoter in S. cerevisiae: chromatin transition and binding sites for the positive regulatory protein PHO4. Nucleic Acids Res *20*, 1031-1038.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. Cell *143*, 470-484.

Bazett-Jones, D.P. (1992). Electron spectroscopic imaging of chromatin and other nucleoprotein complexes. Electron Microsc Rev *5*, 37-58.

Beck, D.B., Oda, H., Shen, S.S., and Reinberg, D. (2012). PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. Genes Dev *26*, 325-337.

Bedford, M.T., and Clarke, S.G. (2009). Protein arginine methylation in mammals: who, what, and why. Mol Cell *33*, 1-13.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482-485.

Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A., and Schubeler, D. (2008). Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster. Mol Cell Biol *28*, 3401-3409.

Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., van Leeuwen, F., Bell, S.P., Imhof, A.*, et al.* (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. EMBO J *26*, 4974-4984.

Bender, L.B., Suh, J., Carroll, C.R., Fong, Y., Fingerman, I.M., Briggs, S.D., Cao, R., Zhang, Y., Reinke, V., and Strome, S. (2006). MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line. Development *133*, 3907-3917.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R.*, et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol *203*, 971-983.

Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci U S A *80*, 5559-5563.

Bienz, M. (2006). The PHD finger, a nuclear protein-interaction domain. Trends Biochem Sci *31*, 35-40.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res *8*, 1499-1504.

Bird, A.P., and Taggart, M.H. (1980). Variable patterns of total DNA and rDNA methylation in animals. Nucleic Acids Res *8*, 1485-1497.

Bird, A.P., and Wolffe, A.P. (1999). Methylation-induced repression--belts, braces, and chromatin. Cell *99*, 451-454.

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res *30*, 958-965.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. Molecular Cell *38*, 179-190.

Bolanos-Garcia, V.M., and Davies, O.R. (2006). Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochim Biophys Acta *1760*, 1304-1313.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Muller, S., Eils, R., Cremer, C., Speicher, M.R.*, et al.* (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol *3*, e157.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.

Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell *64*, 1123-1134.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311-322.

Braunschweig, U., Hogan, G.J., Pagie, L., and van Steensel, B. (2009). Histone H1 binding is inhibited by histone variant H3.3. The EMBO Journal *28*, 3635-3645.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.

Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nature Immunology *4*, 235-240.

Bryant, G.O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., and Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. PLoS Biol *6*, 2928-2939.

Byvoet, P., Shepherd, G.R., Hardin, J.M., and Noland, B.J. (1972). The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. Arch Biochem Biophys *148*, 558-567.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.

Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R.D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell *87*, 1249-1260.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Capili, A.D., Schultz, D.C., Rauscher, I.F., and Borden, K.L. (2001). Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. EMBO J *20*, 165-177.

Carey, M., Li, B., and Workman, J.L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Mol Cell *24*, 481-487.

Carey, M., Peterson, C.L., and Smale, S.T. (2009). Transcriptional regulation in eukaryotes : concepts, strategies, and techniques, 2nd edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C*., et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P*., et al.* (2005). Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. Cell *123*, 581-592.

Cenciarelli, C., Chiaur, D.S., Guardavaccaro, D., Parks, W., Vidal, M., and Pagano, M. (1999). Identification of a family of human F-box proteins. Current biology : CB *9*, 1177-1179.

Chedin, F., Lieber, M.R., and Hsieh, C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A *99*, 16916-16921.

Chen, Z., Zang, J., Whetstine, J., Hong, X., Davrazou, F., Kutateladze, T.G., Simpson, M., Mao, Q., Pan, C.H., Dai, S*., et al.* (2006). Structural insights into histone demethylation by JMJD2 family members. Cell *125*, 691-702.

Cho, H., Orphanides, G., Sun, X., Yang, X.J., Ogryzko, V., Lees, E., Nakatani, Y., and Reinberg, D. (1998). A human RNA polymerase II complex containing factors that modify chromatin structure. Mol Cell Biol *18*, 5355-5363.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Cierpicki, T., Risner, L.E., Grembecka, J., Lukasik, S.M., Popovic, R., Omonkowska, M., Shultis, D.D., Zeleznik-Le, N.J., and Bushweller, J.H. (2009). Structure of the MLL CXXC domain–DNA complex and its functional role in MLL-AF9 leukemia. Nature Structural & Molecular Biology *17*, 62-68.

Cikala, M., Alexandrova, O., David, C.N., Proschel, M., Stiening, B., Cramer, P., and Bottger, A. (2004). The phosphatidylserine receptor from Hydra is a nuclear protein with potential Fe(II) dependent oxygenase activity. BMC Cell Biol *5*, 26.

Clissold, P.M., and Ponting, C.P. (2001). JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2beta. Trends Biochem Sci *26*, 7-9.

Clouaire, T., and Stancheva, I. (2008). Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? Cell Mol Life Sci *65*, 1509-1522.

Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X., Stallcup, M.R., and Cheng, X. (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. Nat Struct Mol Biol *15*, 245-250.

Cooper, D.N., Taggart, M.H., and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. Nucleic Acids Res *11*, 647-658.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Cross, S.H., Meehan, R.R., Nan, X., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet *16*, 256-259.

de Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I.*, et al.* (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nat Struct Mol Biol *18*, 977-983.

De Smet, C., Lurquin, C., Lethe, B., Martelange, V., and Boon, T. (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol Cell Biol *19*, 7327-7335.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & Development *25*, 1010-1022.

Defossez, P.A., and Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. Prog Mol Biol Transl Sci *101*, 377-398.

DeLange, R.J., Fambrough, D.M., Smith, E.L., and Bonner, J. (1969). Calf and pea histone IV. 3. Complete amino acid sequence of pea seedling histone IV; comparison with the homologous calf thymus histone. J Biol Chem *244*, 5669-5679.

Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., and Rani, V. (2012). DNA-protein interactions: methods for detection and analysis. Mol Cell Biochem *365*, 279-299.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491-496.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. Journal of Biological Chemistry *285*, 26114-26120.

Dillon, N. (2004). Heterochromatin structure and function. Biology of the cell / under the auspices of the European Cell Biology Organization *96*, 631-637.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature.

Dodge, J.E. (2004). Inactivation of Dnmt3b in Mouse Embryonic Fibroblasts Results in DNA Hypomethylation, Chromosomal Instability, and Spontaneous Immortalization. Journal of Biological Chemistry *280*, 17986-17991.

Duerre, J.A., and Lee, C.T. (1974). In vivo methylation and turnover of rat brain histones. J Neurochem *23*, 541-547.

Duns, G., van den Berg, E., van Duivenbode, I., Osinga, J., Hollema, H., Hofstra, R.M., and Kok, K. (2010). Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. Cancer Res *70*, 4287-4291.

Durant, M., and Pugh, B.F. (2007). NuA4-directed chromatin transactions throughout the Saccharomyces cerevisiae genome. Mol Cell Biol *27*, 5327-5335.

Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol *375*, 23-44.

Ehrlich, M. (2003). The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clinical Immunology *109*, 17-28.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res *10*, 2709-2721.

Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., and Elgin, S.C. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc Natl Acad Sci U S A *87*, 9923-9927.

Eltsov, M., Maclellan, K.M., Maeshima, K., Frangakis, A.S., and Dubochet, J. (2008). Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. Proc Natl Acad Sci U S A *105*, 19732-19737.

Engeholm, M., de Jager, M., Flaus, A., Brenk, R., van Noort, J., and Owen-Hughes, T. (2009). Nucleosomes can invade DNA territories occupied by their neighbors. Nat Struct Mol Biol *16*, 151-158.

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

Fang, J., Hogan, G.J., Liang, G., Lieb, J.D., and Zhang, Y. (2007). The Saccharomyces cerevisiae Histone Demethylase Jhd1 Fine-Tunes the Distribution of H3K36me2. Molecular and Cellular Biology *27*, 5055-5065.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W.*, et al.* (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. eLife *1*, e00205.

Fedor, M.J., Lue, N.F., and Kornberg, R.D. (1988). Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast. J Mol Biol *204*, 109-127.

Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Current biology : CB *12*, 1052-1058.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science *330*, 622-627.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I.*, et al.* (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res *22*, 2399-2408.

Ferguson-Smith, A.C., and Greally, J.M. (2007). Epigenetics: perceptive enzymes. Nature *449*, 148-149.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398-402.

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

Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature *438*, 1181-1185.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574-1577.

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., Wang, M., Qin, W., Soding, J., Spada, F*., et al.* (2011). Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. PLoS ONE *6*, e16627.

Frescas, D., Guardavaccaro, D., Bassermann, F., Koyama-Nasu, R., and Pagano, M. (2007). JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. Nature *450*, 309-313.

Frescas, D., Guardavaccaro, D., Kuchay, S.M., Kato, H., Poleshko, A., Basrur, V., Elenitoba-Johnson, K.S., Katz, R.A., and Pagano, M. (2008). KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. Cell Cycle *7*, 3539-3547.

Furuhashi, H., Takasaki, T., Rechtsteiner, A., Li, T., Kimura, H., Checchi, P.M., Strome, S., and Kelly, W.G. (2010). Trans-generational epigenetic regulation of C. elegans primordial germ cells. Epigenetics Chromatin *3*, 15.

Fussner, E., Ching, R.W., and Bazett-Jones, D.P. (2011). Living without 30nm chromatin fibers. Trends in Biochemical Sciences *36*, 1-6.

Fyodorov, D.V., and Kadonaga, J.T. (2003). Chromatin assembly in vitro with purified recombinant ACF and NAP-1. Methods Enzymol *371*, 499-515.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell *45*, 344-356.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol *196*, 261-282.

Garner, M.M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res *9*, 3047-3060.

Gaudet, F. (2003). Induction of Tumors in Mice by Genomic Hypomethylation. Science *300*, 489-492.

Gaydos, L.J., Rechtsteiner, A., Egelhofer, T.A., Carroll, C.R., and Strome, S. (2012). Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. Cell reports *2*, 1169-1177.

Ge, R., Wang, Z., Zeng, Q., Xu, X., and Olumi, A.F. (2011). F-box protein 10, an NF-kappaB-dependent anti-apoptotic protein, regulates TRAIL-induced apoptosis through modulating c-Fos/c-FLIP pathway. Cell Death Differ *18*, 1184-1195.

Gearhart, M.D., Corcoran, C.M., Wamstad, J.A., and Bardwell, V.J. (2006). Polycomb Group and SCF Ubiquitin Ligases Are Found in a Novel BCOR Complex That Is Recruited to BCL6 Targets. Molecular and Cellular Biology *26*, 6880-6889.

Gelbart, M.E., Larschan, E., Peng, S., Park, P.J., and Kuroda, M.I. (2009). Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. Nat Struct Mol Biol *16*, 825-832.

Gil, J., and Peters, G. (2006). Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. Nat Rev Mol Cell Biol *7*, 667-677.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A Role for Snf2-Related Nucleosome-Spacing Enzymes in Genome-Wide Nucleosome Organization. Science *333*, 1758-1760.

Goriely, S., Demonte, D., Nizet, S., De Wit, D., Willems, F., Goldman, M., and Van Lint, C. (2003). Human IL-12(p35) gene activation involves selective remodeling of a single nucleosome within a region of the promoter containing critical Sp1-binding sites. Blood *101*, 4894-4902.

Goto, N.K., Zor, T., Martinez-Yamout, M., Dyson, H.J., and Wright, P.E. (2002). Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. J Biol Chem *277*, 43168-43174.

Gowher, H., and Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. J Mol Biol *309*, 1201-1208.

Gowher, H., Leismann, O., and Jeltsch, A. (2000). DNA of Drosophila melanogaster contains 5-methylcytosine. EMBO J *19*, 6918-6923.

Gowher, H., Stockdale, C.J., Goyal, R., Ferreira, H., Owen-Hughes, T., and Jeltsch, A. (2005). De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. Biochemistry *44*, 9899-9904.

Grant, P.A., Duggan, L., Cote, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F.*, et al.* (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev *11*, 1640-1650.

Gregory, P.D., Schmid, A., Zavari, M., Lui, L., Berger, S.L., and Horz, W. (1998). Absence of Gcn5 HAT activity defines a novel state in the opening of chromatin at the PHO5 promoter in yeast. Mol Cell *1*, 495-505.

Gu, W., Szauter, P., and Lucchesi, J.C. (1998). Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of Drosophila melanogaster. Developmental genetics *22*, 56-64.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. Cell *130*, 77-88.

Guo, Junjie U., Su, Y., Zhong, C., Ming, G.-l., and Song, H. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. Cell *145*, 423-434.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

Han, Z., Liu, P., Gu, L., Zhang, Y., Li, H., Chen, S., and Chai, J. (2007). Structural basis for histone demethylation by JHDM1. Frontier Science *1*, 52-61.

Hansen, J.C., van Holde, K.E., and Lohr, D. (1991). The mechanism of nucleosome assembly onto oligomers of the sea urchin 5 S DNA positioning sequence. J Biol Chem *266*, 4276-4282.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445-458.

Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.

Hausinger, R.P. (2004). FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. Crit Rev Biochem Mol Biol *39*, 21-68.

Hayakawa, T., Ohtani, Y., Hayakawa, N., Shinmyozu, K., Saito, M., Ishikawa, F., and Nakayama, J.-i. (2007). RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation. Genes to Cells *0*, 070606122915002-???

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M*., et al.* (2010). Nucleosome dynamics define transcriptional enhancers. Nature Genetics *42*, 343-347.

He, J., Kallin, E.M., Tsukada, Y.-i., and Zhang, Y. (2008). The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15Ink4b. Nature Structural & Molecular Biology *15*, 1169-1175.

Hebbes, T.R., Clayton, A.L., Thorne, A.W., and Crane-Robinson, C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. EMBO J *13*, 1823-1830.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A*., et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301-304.

Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochem Biophys Res Commun *52*, 504-510.

Hirsh, J., and Schleif, R. (1977). The araC promoter: transcription, mapping and interaction with the araBAD promoter. Cell *11*, 545-550.

Hjelm, R.P., Kneale, G.G., Sauau, P., Baldwin, J.P., Bradbury, E.M., and Ibel, K. (1977). Small angle neutron scattering studies of chromatin subunits in solution. Cell *10*, 139-151.

Ho, M.S., Ou, C., Chan, Y.R., Chien, C.T., and Pi, H. (2008). The utility F-box for protein destruction. Cell Mol Life Sci *65*, 1977-2000.

Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. Science *325*, 626-628.

Hon, G., Wang, W., and Ren, B. (2009). Discovery and annotation of functional chromatin signatures in the human genome. PLoS computational biology *5*, e1000566.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. J Biol Chem *268*, 305-314.

Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S.V., Rahman, M.A., Finne, K., Aalen, R.B., and Aasland, R. (2011). The CW domain, a new histone recognition module in chromatin proteins. EMBO J *30*, 1939-1952.

Horn, P.J., and Peterson, C.L. (2001). The bromodomain: a regulator of ATP-dependent chromatin remodeling? Front Biosci *6*, D1019-1023.

Horowitz, R.A., Agard, D.A., Sedat, J.W., and Woodcock, C.L. (1994). The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. J Cell Biol *125*, 1-10.

Hu, M., Sun, X.J., Zhang, Y.L., Kuang, Y., Hu, C.Q., Wu, W.L., Shen, S.H., Du, T.T., Li, H., He, F.*, et al.* (2010). Histone H3 lysine 36 methyltransferase Hypb/Setd2 is

required for embryonic vascular remodeling. Proc Natl Acad Sci U S A *107*, 2956-2961.

Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2000). Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. J Mol Biol *304*, 355-370.

Iguchi-Ariga, S.M., and Schaffner, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes & Development *3*, 612-619.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J.*, et al.* (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol *6*, e22.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands – 'A rough guide'. FEBS Letters *583*, 1713-1720.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLoS Genetics *6*, e1001134.

Imhof, A., Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D.P. (2010). Global Chromatin Architecture Reflects Pluripotency and Lineage Commitment in the Early Mouse Embryo. PLoS ONE *5*, e10531.

Iyer, L.M., Abhiman, S., and Aravind, L. (2011). Natural History of Eukaryotic DNA Methylation Systems. Prog Mol Biol Transl Sci *101*, 25-104.

Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J *14*, 2570-2579.

Izban, M.G., and Luse, D.S. (1992). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. J Biol Chem *267*, 13647-13655.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Jahner, D., Stuhlmann, H., Stewart, C.L., Harbers, K., Lohler, J., Simon, I., and Jaenisch, R. (1982). De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature *298*, 623-628.

Janzer, A., Stamm, K., Becker, A., Zimmer, A., Buettner, R., and Kirfel, J. (2012). The H3K4me3 histone demethylase Fbxl10 is a regulator of chemokine expression, cellular morphology and the metabolome of fibroblasts. J Biol Chem.

Jarvis, D.L. (2009). Baculovirus-insect cell expression systems. Methods Enzymol *463*, 191-222.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S., and Jones, P.A. (2009). Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Mol Cell Biol *29*, 5366-5376.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-251.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nature Reviews Genetics *10*, 161-172.

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

Johnson, S.M., Tan, F.J., McCullough, H.L., Riordan, D.P., and Fire, A.Z. (2006). Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin. Genome Res *16*, 1505-1516.

Johnson, T.B., and Coghill, R.D. (1925). Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. J Am Chem Soc *47*, 2838-2844.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet *19*, 187-191.

Jorgensen, H.F., Ben-Porath, I., and Bird, A.P. (2004). Mbd1 Is Recruited to both Methylated and Nonmethylated CpGs via Distinct DNA Binding Domains. Molecular and Cellular Biology *24*, 3387-3395.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol Cell *20*, 971-978.

Josse, J., Kaiser, A.D., and Kornberg, A. (1961). Enzymatic synthesis of deoxyribonucleic acid. VIII. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid. J Biol Chem *236*, 864-875.

Jurkowska, R.Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., and Jeltsch, A. (2008). Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic Acids Res *36*, 6656-6663.

Kalkhoven, E., Teunissen, H., Houweling, A., Verrijzer, C.P., and Zantema, A. (2002). The PHD type zinc finger is an integral part of the CBP acetyltransferase domain. Mol Cell Biol *22*, 1961-1970.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2008). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kelleher, J.E., and Raleigh, E.A. (1991). A novel activity in Escherichia coli K-12 that directs restriction of DNA modified at CG dinucleotides. J Bacteriol *173*, 5220-5223.

Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J.*, et al.* (2005). Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep *7*, 397-403.

Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520-527.

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proceedings of the National Academy of Sciences *108*, 13564-13569.

Kim, T., and Buratowski, S. (2007). Two Saccharomyces cerevisiae JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation. J Biol Chem *282*, 20827-20835.

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993b). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512-520.

Kireeva, M.L., Hancock, B., Cremona, G.H., Walter, W., Studitsky, V.M., and Kashlev, M. (2005). Nature of the nucleosomal barrier to RNA polymerase II. Mol Cell *18*, 97-108.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol Cell Biol *25*, 3305-3316.

Klose, R.J., Gardner, K.E., Liang, G., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2007). Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in Saccharomyces cerevisiae? Mol Cell Biol *27*, 3951-3961.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. Nature Reviews Genetics *7*, 715-727.

Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I., and Bird, A.P. (2005). DNA Binding Selectivity of MeCP2 Due to a Requirement for A/T Sequences Adjacent to Methyl-CpG. Molecular Cell *19*, 667-678.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol *2*, E171.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet *41*, 376-381.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kottakis, F., Polytarchou, C., Foltopoulou, P., Sanidas, I., Kampranis, S.C., and Tsichlis, P.N. (2011). FGF-2 regulates cell proliferation, migration, and angiogenesis through an NDY1/KDM2B-miR-101-EZH2 pathway. Mol Cell *43*, 285-298.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693-705.

Koyama-Nasu, R., David, G., and Tanese, N. (2007). The F-box protein Fbl10 is a novel transcriptional repressor of c-Jun. Nat Cell Biol *9*, 1074-1080.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M.*, et al.* (2003a). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell *11*, 721-729.

Krogan, N.J., Keogh, M.C., Datta, N., Sawa, C., Ryan, O.W., Ding, H., Haw, R.A., Pootoolal, J., Tong, A., Canadien, V.*, et al.* (2003b). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. Mol Cell *12*, 1565-1576.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C.*, et al.* (2003c). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Molecular and Cellular Biology *23*, 4207-4218.

Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. Adv Genet *70*, 27-56.

Kuo, A.J., Cheung, P., Chen, K., Zee, B.M., Kioi, M., Lauring, J., Xi, Y., Park, B.H., Shi, X., Garcia, B.A.*, et al.* (2011). NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. Mol Cell *44*, 609-620.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A

ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes & Development *22*, 2799-2810.

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature *448*, 718-722.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W*., et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langmore, J.P., and Paulson, J.R. (1983). Low angle x-ray diffraction studies of chromatin structure in vivo and in isolated nuclei and metaphase chromosomes. J Cell Biol *96*, 1120-1131.

Larschan, E., Alekseyenko, A.A., Gortchakov, A.A., Peng, S., Li, B., Yang, P., Workman, J.L., Park, P.J., and Kuroda, M.I. (2007). MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. Mol Cell *28*, 121-133.

Larschan, E., Bishop, E.P., Kharchenko, P.V., Core, L.J., Lis, J.T., Park, P.J., and Kuroda, M.I. (2011). X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature *471*, 115-118.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095-1107.

Laurent, B.C., Yang, X., and Carlson, M. (1992). An essential Saccharomyces cerevisiae gene homologous to SNF2 encodes a helicase-related protein in a new family. Mol Cell Biol *12*, 1893-1902.

Lawyer, F.C., Stoffel, S., Saiki, R.K., Chang, S.Y., Landre, P.A., Abramson, R.D., and Gelfand, D.H. (1993). High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. PCR Methods Appl *2*, 275-287.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73-84.

Lee, D.Y., Teyssier, C., Strahl, B.D., and Stallcup, M.R. (2005a). Role of protein methylation in regulation of transcription. Endocr Rev *26*, 147-170.

Lee, J.H. (2005). CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. Journal of Biological Chemistry *280*, 41725-41731.

Lee, J.S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007a). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell *131*, 1084-1096.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005b). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature *437*, 432-435.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007b). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet *13*, 233-245.

Levy, A., and Noll, M. (1981). Chromatin fine structure of active and repressed genes. Nature *289*, 198-203.

Li, B., Gogol, M., Carey, M., Lee, D., Seidel, C., and Workman, J.L. (2007). Combined Action of PHD and Chromo Domains Directs the Rpd3S HDAC to Transcribed Chromatin. Science *316*, 1050-1054.

Li, B., Jackson, J., Simon, M.D., Fleharty, B., Gogol, M., Seidel, C., Workman, J.L., and Shilatifard, A. (2009a). Histone H3 Lysine 36 Dimethylation (H3K36me2) Is Sufficient to Recruit the Rpd3s Histone Deacetylase Complex and to Repress Spurious Transcription. Journal of Biological Chemistry *284*, 7970-7976.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362-365.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. Nat Struct Mol Biol *12*, 46-53.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature *442*, 91-95.

Li, Y., Trojer, P., Xu, C.F., Cheung, P., Kuo, A., Drury, W.J., 3rd, Qiao, Q., Neubert, T.A., Xu, R.M., Gozani, O.*, et al.* (2009b). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. J Biol Chem *284*, 34283-34295.

Liang, G., He, J., and Zhang, Y. (2012). Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. Nat Cell Biol *14*, 457-466.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, C.H., Li, B., Swanson, S., Zhang, Y., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2008). Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the Drosophila KDM4A demethylase. Mol Cell *32*, 696-706.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D., and Felsenfeld, G. (2001). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. Science *293*, 2453-2455.

Lizcano, J.M., Unzeta, M., and Tipton, K.F. (2000). A spectrophotometric method for determining the oxidative deamination of methylamine by the amine oxidases. Anal Biochem *286*, 75-79.

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell *49*, 203-210.

Louwrier, A. (1999). Nucleic acid removal from Taq polymerase preparations using an aqueous/organic biphasic system. Biotechniques *27*, 444-445.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol *276*, 19-42.

Lu, T., Jackson, M.W., Singhi, A.D., Kandel, E.S., Yang, M., Zhang, Y., Gudkov, A.V., and Stark, G.R. (2009a). Validation-based insertional mutagenesis identifies lysine demethylase FBXL11 as a negative regulator of NF B. Proceedings of the National Academy of Sciences *106*, 16339-16344.

Lu, T., Jackson, M.W., Wang, B., Yang, M., Chance, M.R., Miyagi, M., Gudkov, A.V., and Stark, G.R. (2009b). Regulation of NF- B by NSD1/FBXL11-dependent reversible lysine methylation of p65. Proceedings of the National Academy of Sciences *107*, 46-51.

Lu, X., Simon, M.D., Chodaparambil, J.V., Hansen, J.C., Shokat, K.M., and Luger, K. (2008). The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. Nat Struct Mol Biol *15*, 1122-1124.

Lucio-Eterovic, A.K., Singh, M.M., Gardner, J.E., Veerappan, C.S., Rice, J.C., and Carpenter, P.B. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. Proc Natl Acad Sci U S A *107*, 16952-16957.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science *327*, 996-1000.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997a). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luger, K., Rechsteiner, T.J., Flaus, A.J., Waye, M.M., and Richmond, T.J. (1997b). Characterization of nucleosome core particles containing histone proteins made in bacteria. J Mol Biol *272*, 301-311.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol *119*, 1-16.

Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in Drosophila melanogaster. Nature *408*, 538-540.

Ma, Q., Alder, H., Nelson, K.K., Chatterjee, D., Gu, Y., Nakamura, T., Canaani, E., Croce, C.M., Siracusa, L.D., and Buchberg, A.M. (1993). Analysis of the murine All-1 gene reveals conserved domains with human ALL-1 and identifies a motif shared with DNA methyltransferases. Proc Natl Acad Sci U S A *90*, 6350-6354.

Maeshima, K., Hihara, S., and Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo? Curr Opin Cell Biol *22*, 291-297.

Mancini, D.N., Singh, S.M., Archer, T.K., and Rodenhiser, D.I. (1999). Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. Oncogene *18*, 4108-4119.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

Marin, I., and Baker, B.S. (2000). Origin and evolution of the regulatory gene male-specific lethal-3. Molecular biology and evolution *17*, 1240-1250.

Marston, F.A. (1986). The purification of eukaryotic polypeptides synthesized in Escherichia coli. Biochem J *240*, 1-12.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y.*, et al.* (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253-257.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008a). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res *18*, 1073-1083.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C.*, et al.* (2008b). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

McDowall, A.W., Smith, J.M., and Dubochet, J. (1986). Cryo-electron microscopy of vitrified chromosomes in situ. EMBO J *5*, 1395-1402.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436-439.

Millar, C.B. (2002). Enhanced CpG Mutability and Tumorigenesis in MBD4-Deficient Mice. Science *297*, 403-405.

Mirsky, A.E., and Ris, H. (1947). The Chemical Composition of Isolated Chromosomes. J Gen Physiol *31*, 7-18.

Mirsky, A.E., and Ris, H. (1951). The composition and structure of isolated chromosomes. J Gen Physiol *34*, 475-492.

Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. Cell *128*, 787-800.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell *30*, 755-766.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development *99*, 371-382.

Moreira, J.M., and Holmberg, S. (1998). Nucleosome structure of the yeast CHA1 promoter: analysis of activation-dependent chromatin remodeling of an RNA-polymerase-II-transcribed gene in TBP and RNA pol II mutants defective in vivo in response to acidic activators. EMBO J *17*, 6028-6038.

Morris, S.A., Shibata, Y., Noma, K., Tsukamoto, Y., Warren, E., Temple, B., Grewal, S.I., and Strahl, B.D. (2005). Histone H3 K36 methylation is associated with transcription elongation in Schizosaccharomyces pombe. Eukaryot Cell *4*, 1446-1454.

Morse, R.H. (2007). Transcription factor access to promoter elements. J Cell Biochem *102*, 560-570.

Murray, K. (1964). The Occurrence of Epsilon-N-Methyl Lysine in Histones. Biochemistry *3*, 10-15.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat Genet *39*, 1507-1511.

Musselman, C.A., and Kutateladze, T.G. (2011). Handpicking epigenetic marks with PHD fingers. Nucleic Acids Res *39*, 9061-9071.

Nady, N., Min, J., Kareta, M.S., Chedin, F., and Arrowsmith, C.H. (2008). A SPOT on the chromatin landscape? Histone peptide arrays as a tool for epigenetic research. Trends Biochem Sci *33*, 305-313.

Nakanishi, S., Lee, J.S., Gardner, K.E., Gardner, J.M., Takahashi, Y.H., Chandrasekharan, M.B., Sun, Z.W., Osley, M.A., Strahl, B.D., Jaspersen, S.L.*, et al.* (2009). Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1. J Cell Biol *186*, 371-377.

Nakayama, J., Klar, A.J., and Grewal, S.I. (2000). A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis. Cell *101*, 307-317.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

Ng, S.S., Kavanagh, K.L., McDonough, M.A., Butler, D., Pilka, E.S., Lienard, B.M., Bray, J.E., Savitsky, P., Gileadi, O., von Delft, F.*, et al.* (2007). Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. Nature *448*, 87-91.

Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G., and Burley, S.K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. Proc Natl Acad Sci U S A *93*, 4862-4867.

Nimura, K., Ura, K., Shiratori, H., Ikawa, M., Okabe, M., Schwartz, R.J., and Kaneda, Y. (2009). A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. Nature *460*, 287-291.

Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A.S., Imamoto, N., Ishikawa, T.*, et al.* (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. EMBO J *31*, 1644-1653.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C.D., Tempst, P., and Reinberg, D. (2002a). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev *16*, 479-489.

Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R.*, et al.* (2002b). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Mol Cell *9*, 1201-1213.

Noll, M., and Kornberg, R.D. (1977). Action of micrococcal nuclease on chromatin and the location of histone H1. J Mol Biol *109*, 393-404.

Oda, H., Hubner, M.R., Beck, D.B., Vermeulen, M., Hurwitz, J., Spector, D.L., and Reinberg, D. (2010). Regulation of the histone H4 monomethylase PR-Set7 by CRL4(Cdt2)-mediated PCNA-dependent degradation during DNA damage. Mol Cell *40*, 364-376.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet *19*, 219-220.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Olins, A.L., Senior, M.B., and Olins, D.E. (1976). Ultrastructural features of chromatin nu bodies. J Cell Biol *68*, 787-793.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D*., et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

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

Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. EMBO J *19*, 6141-6149.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes & Development *10*, 1991-2002.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet *10*, 669-680.

Parker, M.W., Lo Bello, M., and Federici, G. (1990). Crystallization of glutathione S-transferase from human placenta. J Mol Biol *213*, 221-222.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P*., et al.* (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394-397.

Pedersen, M.T., and Helin, K. (2010). Histone demethylases in development and disease. Trends in Cell Biology *20*, 662-671.

Peña, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R., and Kutateladze, T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature *442*, 100-103.

Perales, R., Zhang, L., and Bentley, D. (2011). Histone occupancy in vivo at the 601 nucleosome binding element is determined by transcriptional history. Mol Cell Biol *31*, 3485-3496.

Perlmann, T. (1992). Glucocorticoid receptor DNA-binding specificity is increased by the organization of DNA in nucleosomes. Proc Natl Acad Sci U S A *89*, 3884-3888.

Perlmann, T., and Wrange, O. (1988). Specific glucocorticoid receptor binding to DNA reconstituted in a nucleosome. EMBO J *7*, 3073-3079.

Pfau, R., Tzatsos, A., Kampranis, S.C., Serebrennikova, O.B., Bear, S.E., and Tsichlis, P.N. (2008). Members of a family of JmjC domain-containing oncoproteins immortalize embryonic fibroblasts via a JmjC domain-dependent process. Proc Natl Acad Sci U S A *105*, 1907-1912.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev *20*, 2922-2936.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517-527.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat Biotechnol *28*, 1057-1068.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS Genet *8*, e1002717.

Proffitt, J.H., Davie, J.R., Swinton, D., and Hattman, S. (1984). 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol *4*, 985-988.

Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., and Xu, R.M. (2010). The Structure of NSD1 Reveals an Autoregulatory Mechanism Underlying Histone H3K36 Methylation. Journal of Biological Chemistry *286*, 8361-8368.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc Regulates Transcriptional Pause Release. Cell *141*, 432-445.

Rahman, S., Sowa, M.E., Ottinger, M., Smith, J.A., Shi, Y., Harper, J.W., and Howley, P.M. (2011). The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. Mol Cell Biol *31*, 2641-2652.

Raleigh, E.A., and Wilson, G. (1986). Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proc Natl Acad Sci U S A *83*, 9070-9074.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. Cell *138*, 114-128.

Rao, B., Shibata, Y., Strahl, B.D., and Lieb, J.D. (2005). Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. Mol Cell Biol *25*, 9447-9459.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D*., et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Rechtsteiner, A., Ercan, S., Takasaki, T., Phippen, T.M., Egelhofer, T.A., Wang, W., Kimura, H., Lieb, J.D., and Strome, S. (2010). The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. PLoS Genet *6*.

Reid, J.L., Moqtaderi, Z., and Struhl, K. (2004). Eaf3 regulates the global pattern of histone acetylation in Saccharomyces cerevisiae. Mol Cell Biol *24*, 757-764.

Reinberg, D., and Sims, R.J., 3rd (2006). de FACTo nucleosome dynamics. J Biol Chem *281*, 23297-23301.

Reinke, H., and Horz, W. (2004). Anatomy of a hypersensitive site. Biochim Biophys Acta *1677*, 24-29.

Reinke, H., and Hörz, W. (2003). Histones Are First Hyperacetylated and Then Lose Contact with the Activated PHO5 Promoter. Molecular Cell *11*, 1599-1607.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P.V. (2007). Critical DNA Binding Interactions of the Insulator Protein CTCF: A SMALL NUMBER OF ZINC FINGERS MEDIATE STRONG BINDING, AND A SINGLE FINGER-DNA INTERACTION CONTROLS BINDING AT IMPRINTED LOCI. Journal of Biological Chemistry *282*, 33336-33345.

Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. (2003). Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. Mol Cell *12*, 1591-1598.

Robinson, P.J.J. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. Proceedings of the National Academy of Sciences *103*, 6506-6511.

Robzyk, K., Recht, J., and Osley, M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science *287*, 501-504.

Roh, T.-y., Ngau, W.C., Cui, K., Landsman, D., and Zhao, K. (2004). High-resolution genome-wide mapping of histone modifications. Nature Biotechnology *22*, 1013-1016.

Roh, T.Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev *19*, 542-552.

Rothbart, S.B., Krajewski, K., Strahl, B.D., and Fuchs, S.M. (2012). Peptide microarrays to interrogate the "histone code". Methods Enzymol *512*, 107-135.

Routh, A., Sandin, S., and Rhodes, D. (2008). Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proceedings of the National Academy of Sciences *105*, 8872-8877.

Ruiz-Garcia, A.B., Sendra, R., Pamblanco, M., and Tordera, V. (1997). Gcn5p is involved in the acetylation of histone H3 in nucleosomes. FEBS Lett *403*, 186-190.

Ruthenburg, Alexander J., Li, H., Milne, Thomas A., Dewell, S., McGinty, Robert K., Yuen, M., Ueberheide, B., Dou, Y., Muir, Tom W., Patel, Dinshaw J.*, et al.* (2011). Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. Cell *145*, 692-706.

Ruthenburg, A.J., Li, H., Patel, D.J., and David Allis, C. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nature Reviews Molecular Cell Biology *8*, 983-994.

Sadeh, R., and Allis, C.D. (2011). Genome-wide "re"-modeling of nucleosome positions. Cell *147*, 263-266.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol *7*, 437-447.

Sanchez, C., Sanchez, I., Demmers, J.A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol Cell Proteomics *6*, 820-834.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986). Sequence periodicities in chicken nucleosome core DNA. J Mol Biol *191*, 659-675.

Saxonov, S. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences *103*, 1412-1417.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.

Schasfoort, R.B.M., and Tudos, A.J. (2008). Handbook of surface plasmon resonance (Cambridge: Royal Society of Chemistry).

Schindler, U., Beckmann, H., and Cashmore, A.R. (1993). HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. Plant J *4*, 137-150.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schreiber, S.L., and Bernstein, B.E. (2002). Signaling network model of chromatin. Cell *111*, 771-778.

Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J.*, et al.* (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev *18*, 1263-1271.

Schuettengruber, B., Martinez, A.M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nat Rev Mol Cell Biol *12*, 799-814.

Schwarzbauer, K., Bodenhofer, U., and Hochreiter, S. (2012). Genome-wide chromatin remodeling identified at GC-rich long nucleosome-free regions. PLoS One *7*, e47924.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., Wang, J.-P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Sekinger, E.A., Moqtaderi, Z., and Struhl, K. (2005). Intrinsic Histone-DNA Interactions and Low Nucleosome Density Are Important for Preferential Accessibility of Promoter Regions in Yeast. Molecular Cell *18*, 735-748.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. Annu Rev Biochem *76*, 75-100.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R.*, et al.* (2006a). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96-99.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H.A., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G.E., Fingerman, I.M., Briggs, S.D.*, et al.* (2006b). Proteome-wide Analysis in Saccharomyces cerevisiae Identifies Several PHD Fingers as Novel Direct and Selective Binding Modules of Histone H3 Methylated at Either Lysine 4 or Lysine 36. Journal of Biological Chemistry *282*, 2450-2455.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell *19*, 857-864.

Shogren-Knaak, M. (2006). Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. Science *311*, 844-847.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Simon, M.D., Chu, F., Racki, L.R., de la Cruz, C.C., Burlingame, A.L., Panning, B., Narlikar, G.J., and Shokat, K.M. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. Cell *128*, 1003-1012.

Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res *14*, 6711-6719.

Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2010). Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation. Science *331*, 1036-1040.

Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. Mol Cell *36*, 245-254.

Steiner, N.C., Hahnenberger, K.M., and Clarke, L. (1993). Centromeres of the fission yeast Schizosaccharomyces pombe are highly variable genetic loci. Mol Cell Biol *13*, 4578-4587.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev *64*, 435-459.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F*., et al.* (2002). Set2 Is a Nucleosomal Histone

H3-Selective Methyltransferase That Mediates Transcriptional Repression. Molecular and Cellular Biology *22*, 1298-1306.

Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z., and Cedar, H. (2009). Developmental programming of CpG island methylation profiles in the human genome. Nature Structural & Molecular Biology *16*, 564-571.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol *185*, 60-89.

Studitsky, V.M., Clark, D.J., and Felsenfeld, G. (1994). A histone octamer can step around a transcribing polymerase without leaving the template. Cell *76*, 371-382.

Studitsky, V.M., Kassavetis, G.A., Geiduschek, E.P., and Felsenfeld, G. (1997). Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. Science *278*, 1960-1963.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem *279*, 27816-27823.

Sugiyama, T., Cam, H.P., Sugiyama, R., Noma, K., Zofall, M., Kobayashi, R., and Grewal, S.I. (2007). SHREC, an effector complex for heterochromatic transcriptional silencing. Cell *128*, 491-504.

Sun, Z.W., and Allis, C.D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104-108.

Syntichaki, P., Topalidou, I., and Thireos, G. (2000). The Gcn5 bromodomain co-ordinates nucleosome remodelling. Nature *404*, 414-417.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L*., et al.* (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930-935.

Takahashi, K., Murakami, S., Chikashige, Y., Niwa, O., and Yanagida, M. (1991). A large number of tRNA genes are symmetrically located in fission yeast centromeres. J Mol Biol *218*, 13-17.

Takahata, S., Yu, Y., and Stillman, D.J. (2009). FACT and Asf1 regulate nucleosome dynamics and coactivator binding at the HO promoter. Mol Cell *34*, 405-415.

Tanaka, Y., Okamoto, K., Teye, K., Umata, T., Yamagiwa, N., Suto, Y., Zhang, Y., and Tsuneoka, M. (2010). JmjC enzyme KDM2A is a regulator of rRNA transcription in response to starvation. The EMBO Journal *29*, 1510-1522.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025-1040.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909-920.

Thastrom, A., Lowary, P.T., Widlund, H.R., Cao, H., Kubista, M., and Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. J Mol Biol *288*, 213-229.

Thomas, J.O., and Kornberg, R.D. (1975). An octamer of histones in chromatin and free in solution. Proc Natl Acad Sci U S A *72*, 2626-2630.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D*., et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-1086.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B*., et al.* (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75-82.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics *10*, 442.

Tims, H.S., Gurunathan, K., Levitus, M., and Widom, J. (2011). Dynamics of nucleosome invasion by DNA binding proteins. J Mol Biol *411*, 430-448.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Res *18*, 1084-1091.

Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T., and Sedgwick, B. (2002). Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature *419*, 174-178.

Trewick, S.C., McLaughlin, P.J., and Allshire, R.C. (2005). Methylation: lost in hydroxylation? EMBO Rep *6*, 315-320.

Trojer, P., Li, G., Sims, R.J., 3rd, Vaquero, A., Kalakonda, N., Boccuni, P., Lee, D., Erdjument-Bromage, H., Tempst, P., Nimer, S.D*., et al.* (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. Cell *129*, 915-928.

Tsukada, Y.-i., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2005). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tsukiyama, T., Palmer, J., Landel, C.C., Shiloach, J., and Wu, C. (1999). Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae. Genes Dev *13*, 686-697.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.

Tweedie, S., Charlton, J., Clark, V., and Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol *17*, 1469-1475.

Tzatsos, A., Paskaleva, P., Lymperi, S., Contino, G., Stoykova, S., Chen, Z., Wong, K.K., and Bardeesy, N. (2011). Lysine-specific demethylase 2B (KDM2B)-let-7-enhancer of zester homolog 2 (EZH2) pathway regulates cell cycle progression and senescence in primary cells. J Biol Chem *286*, 33061-33069.

Tzatsos, A., Pfau, R., Kampranis, S.C., and Tsichlis, P.N. (2009). Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus. Proc Natl Acad Sci U S A *106*, 2641-2646.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K*., et al.* (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res *18*, 1051-1063.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature *474*, 516-520.

van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745-756.

Vanyushin, B.F., Tkacheva, S.G., and Belozersky, A.N. (1970). Rare bases in animal DNA. Nature *225*, 948-949.

Venter, U., Svaren, J., Schmitz, J., Schmid, A., and Horz, W. (1994). A nucleosome precludes binding of the transcription factor Pho4 in vivo to a critical target site in the PHO5 promoter. EMBO J *13*, 4848-4855.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G*., et al.* (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967-980.

Vettese-Dadey, M., Grant, P.A., Hebbes, T.R., Crane- Robinson, C., Allis, C.D., and Workman, J.L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. EMBO J *15*, 2508-2518.

Vogelmann, J., Valeri, A., Guillou, E., Cuvier, O., and Nollmann, M. (2011). Roles of chromatin insulator proteins in higher-order chromatin organization and transcription regulation. Nucleus *2*, 358-369.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol *20*, 2108-2121.

Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., and Wolffe, A.P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet *23*, 62-66.

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nature Reviews Molecular Cell Biology *13*, 115-126.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.

Wang, T., Chen, K., Zeng, X., Yang, J., Wu, Y., Shi, X., Qin, B., Zeng, L., Esteban, M.A., Pan, G.*, et al.* (2011). The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. Cell Stem Cell *9*, 575-587.

Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J.R., 3rd, and Jia, S. (2009). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. Mol Cell *33*, 428-437.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes & Development *2*, 1136-1143.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics *39*, 457-466.

Whitehouse, I., and Tsukiyama, T. (2006). Antagonistic forces that position nucleosomes in vivo. Nat Struct Mol Biol *13*, 633-640.

Widom, J. (2001). Role of DNA sequence in nucleosome stability and dynamics. Q Rev Biophys *34*, 269-324.

Widom, J., and Klug, A. (1985). Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. Cell *43*, 207-213.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.

Winston, J.T., Koepp, D.M., Zhu, C., Elledge, S.J., and Harper, J.W. (1999). A family of mammalian F-box proteins. Current biology : CB *9*, 1180-1182.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T., and Migeon, B.R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X

chromosome: implications for X-chromosome inactivation. Proc Natl Acad Sci U S A *81*, 2806-2810.

Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. J Biol Chem *278*, 34739-34742.

Woodcock, C.L., Frado, L.L., and Rattner, J.B. (1984). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J Cell Biol *99*, 42-52.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nature Reviews Molecular Cell Biology *11*, 607-620.

Wysocka, J. (2006). Identifying novel proteins recognizing histone modifications using peptide pull-down assay. Methods *40*, 339-343.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P.*, et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev *17*, 654-663.

Xu, W., Edmondson, D.G., and Roth, S.Y. (1998). Mammalian GCN5 and P/CAF acetyltransferases have homologous amino-terminal domains important for recognition of nucleosomal substrates. Mol Cell Biol *18*, 5659-5669.

Yamasaki-Ishizaki, Y., Kayashima, T., Mapendano, C.K., Soejima, H., Ohta, T., Masuzaki, H., Kinoshita, A., Urano, T., Yoshiura, K., Matsumoto, N.*, et al.* (2007). Role of DNA methylation and histone H3 lysine 27 methylation in tissue-specific imprinting of mouse Grb10. Mol Cell Biol *27*, 732-742.

Yang, H., Pesavento, J.J., Starnes, T.W., Cryderman, D.E., Wallrath, L.L., Kelleher, N.L., and Mizzen, C.A. (2008). Preferential dimethylation of histone H4 lysine 20 by Suv4-20. J Biol Chem *283*, 12085-12092.

Yang, M., Gocke, C.B., Luo, X., Borek, D., Tomchick, D.R., Machius, M., Otwinowski, Z., and Yu, H. (2006). Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell *23*, 377-387.

Yang, X.J., and Seto, E. (2003). Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. Curr Opin Genet Dev *13*, 143-153.

Yang, Y., Lu, Y., Espejo, A., Wu, J., Xu, W., Liang, S., and Bedford, M.T. (2010). TDRD3 is an effector molecule for arginine-methylated histone marks. Mol Cell *40*, 1016-1023.

Yoder, J.A., Soman, N.S., Verdine, G.L., and Bestor, T.H. (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol *270*, 385-395.

Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev *22*, 3422-3434.

Youdell, M.L., Kizer, K.O., Kisseleva-Romanova, E., Fuchs, S.M., Duro, E., Strahl, B.D., and Mellor, J. (2008). Roles for Ctk1 and Spt6 in Regulating the Different Methylation States of Histone H3 Lysine 36. Molecular and Cellular Biology *28*, 4915-4926.

Young, N.L., Dimaggio, P.A., and Garcia, B.A. (2010). The significance, development and progress of high-throughput combinatorial histone code analysis. Cell Mol Life Sci *67*, 3983-4000.

Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R., and Hunt, J.F. (2006). Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. Nature *439*, 879-884.

Yuan, W., Xie, J., Long, C., Erdjument-Bromage, H., Ding, X., Zheng, Y., Tempst, P., Chen, S., Zhu, B., and Reinberg, D. (2009). Heterogeneous nuclear ribonucleoprotein L Is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity in vivo. J Biol Chem *284*, 15701-15707.

Zee, B.M., Britton, L.M., Wolle, D., Haberman, D.M., and Garcia, B.A. (2012). Origins and formation of histone methylation across the human cell cycle. Mol Cell Biol *32*, 2503-2514.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet *39*, 1512-1516.

Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature *466*, 258-262.

Zhang, F., and Boothby, M. (2006). T helper type 1-specific Brg1 recruitment and remodeling of nucleosomes positioned at the IFN-gamma promoter are Stat4 dependent. J Exp Med *203*, 1493-1505.

Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat Struct Mol Biol *15*, 381-388.

Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q., and Ding, J. (2006). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Res *34*, 6621-6628.

Zhang, X., Yang, Z., Khan, S.I., Horton, J.R., Tamaru, H., Selker, E.U., and Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell *12*, 177-185.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W.*, et al.* (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Research *38*, 4246-4253.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nature Structural & Molecular Biology *16*, 847-852.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev *13*, 1924-1935.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev *15*, 2343-2360.

Zhou, J.C., Blackledge, N.P., Farcas, A.M., and Klose, R.J. (2011). Recognition of CpG Island Chromatin by KDM2A Requires Direct and Specific Interaction with Linker DNA. Molecular and Cellular Biology *32*, 479-489.

Zhu, J., He, F., Hu, S., and Yu, J. (2008). On the nature of human housekeeping genes. Trends Genet *24*, 481-484.

Zhu, W.G., Srinivasan, K., Dai, Z., Duan, W., Druhan, L.J., Ding, H., Yee, L., Villalona-Calero, M.A., Plass, C., and Otterson, G.A. (2003). Methylation of Adjacent CpG Sites Affects Sp1/Sp3 Binding and Activity in the p21Cip1 Promoter. Molecular and Cellular Biology *23*, 4056-4065.

Zlatanova, J., Leuba, S.H., Yang, G., Bustamante, C., and van Holde, K. (1994). Linker DNA accessibility in chromatin fibers of different conformations: a reevaluation. Proc Natl Acad Sci U S A *91*, 5277-5280.

Appendix

Construct	Features and functional properties
KDM2A	Full length human KDM2A aa 1-1162 (Figure 3.2).
KDM2A6	C-terminally truncated version (aa 1-747) of full length KDM2A. This construct contains the JmjC, CXXC and PHD domains (Figure 3.3). It carries JmjC-mediated H3K36 demethylase activity (Figure 3.7) and CXXC-dependent binding to non-methylated CpG DNA (Figure 3.8).
H212A	H212A substitution is present in the JmjC active site. In KDM2A6 H212A, this specific mutation partially abrogates JmjC demethylase activity (Figure 3.7) but does not affect binding to non-methylated CpG DNA (data not shown).
H212A D214A	H212A and D214A substitutions are present in the JmjC active site. In KDM2A6 H212A D214A, the mutations completely abrogate JmjC demethylase activity (Figure 3.7) but do not affect binding to non-methylated CpG DNA (Figure 3.8).

ΔCXXC	Deletion of the zinc finger CXXC domain (Δ aa 563-609). In KDM2A6 ΔCXXC, the deletion of this domain completely abrogates KDM2A binding to non-methylated CpG DNA (Figure 3.8) and partially abrogates JmjC demethylase activity (Figure 3.9).
KDM2A CXXC	This construct contains only the zinc-finger CXXC domain of KDM2A. It binds to non-methylated CpG DNA as efficiently as KDM2A6 (Figure 3.8).
K601A	K601A substitution in the DNA binding interface of the CXXC domain. In KDM2A6 K601A, the point mutation completely abrogates KDM2A binding to non-methylated CpG DNA (Figure 3.8).
ΔPHD	Deletion of the PHD domain (Δ aa 619-666) (Figure 3.2).
C645A	C645A substitution in the PHD domain. In KDM2A6 C645A, the point mutation does not affect KDM2A binding to non-methylated CpG DNA (Figure 5.3).

Table 1

The above table summarises all the wild type (WT) and derived mutant constructs generated in this study. The prokaryotic constructs include: KDM2A6, H212A, H212A D214A, ΔCXXC, KDM2A CXXC, K601A and C645A. The baculovirus constructs include: KDM2A full length, H212A D214A, ΔCXXC, ΔPHD and K601A. The functional properties of the constructs described are valid within the *in vitro* conditions tested.

EMSA probe sequences

147bp 601 DNA (59% GC content; 1.1 obs/exp):

GGCCGCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCA
CCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTA
GTCTCCAGGCACGTGTCAGATATATAC

216bp 601 DNA (59% GC content; 1.0 obs/exp):

GGCCGCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCA
CCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTA
GTCTCCAGGCACGTGTCAGATATATACATCCTGTGCATGTATTGAACAGCGACCTTGCCG
GTGCCAGTCGGATAGTGTTCCGAGCTCCCGACCTGC

342bp 2x601 (48bp linker) DNA (56% GC content; 1.1 obs/exp):

CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAG
GCACGTGTCAGATATATACATCCTGTGCATGTATTGAACAGCGACCTGAATTCTGGACCCT
ATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCT
CTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTA
CTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGT

The CpGs are in red and the underlined sequences are nucleosomal DNA as assessed experimentally in previous studies (Engeholm et al., 2009). The %GC

content is calculated as: number of Gs + Cs / total number of nucleotides. The obs/exp ratio is calculated as: (number of CpGs x total number of nucleotides) / (number of Cs x number of Gs) (Gardiner-Garden and Frommer, 1987).

Biacore probe sequences

0CG:

TCCATTGATCTCCCAGAGCCTGCAAGAGGCCCAACCAGGCAGACTTGGGGCTAG
GTCCCAGGAAACAGCTATGAC

1CG:

TCCATTGATCTCCCAGAGCCTGCAAGACCGGCAACCAGGCAGACTTGGGGCTAG
GTCCCAGGAAACAGCTATGAC

6CG:

TCCACGGATCTCCCGGAGCCTGCGAGAGGCCCGACCAGGCCGACTTGGGCGTA
GGTCCCAGGAAACAGCTATGAC

The CpGs are in red and the M13 reverse primer hybridisation site is in green.

Publications

Molecular Cell
Article

CpG Islands Recruit a Histone H3 Lysine 36 Demethylase

Neil P. Blackledge,[1] Jin C. Zhou,[1,3] Michael Y. Tolstorukov,[2,3] Anca M. Farcas,[1] Peter J. Park,[2] and Robert J. Klose[1,*]
[1]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK
[2]Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA
[3]These authors contributed equally to this work
*Correspondence: rob.klose@bioch.ox.ac.uk
DOI 10.1016/j.molcel.2010.04.009

SUMMARY

In higher eukaryotes, up to 70% of genes have high levels of nonmethylated cytosine/guanine base pairs (CpGs) surrounding promoters and gene regulatory units. These features, called CpG islands, were identified over 20 years ago, but there remains little mechanistic evidence to suggest how these enigmatic elements contribute to promoter function, except that they are refractory to epigenetic silencing by DNA methylation. Here we show that CpG islands directly recruit the H3K36-specific lysine demethylase enzyme KDM2A. Nucleation of KDM2A at these elements results in removal of H3K36 methylation, creating CpG island chromatin that is uniquely depleted of this modification. KDM2A utilizes a zinc finger CxxC (ZF-CxxC) domain that preferentially recognizes nonmethylated CpG DNA, and binding is blocked when the CpG DNA is methylated, thus constraining KDM2A to nonmethylated CpG islands. These data expose a straightforward mechanism through which KDM2A delineates a unique architecture that differentiates CpG island chromatin from bulk chromatin.

INTRODUCTION

In lower eukaryotes, RNA polymerase II-transcribed promoters are typically compact and contain transcription factor binding sites closely linked to defined polymerase engagement sites (Fuda et al., 2009; Juven-Gershon et al., 2008). In many higher eukaryotes, including humans, a more prevalent class of promoter has evolved that is contained within CpG islands (Bird et al., 1985; Takai and Jones, 2002).

CpG islands were originally biochemically identified using methylation-sensitive restriction enzymes that specifically release nonmethylated regions of the genome from bulk genomic DNA (Bird et al., 1985; Cooper et al., 1983). First called HpaII tiny fragment islands (HTF islands) because of the restriction enzyme used to isolate them (Bird, 1986), these were later renamed CpG islands based on the observation that they corresponded to contiguous nonmethylated regions of the genome that usually had higher than average levels of CpG dinucleotides and GC content (Bird et al., 1985; Bird, 1986; Cooper et al., 1983). These attributes of CpG islands are directly related to their nonmethylated state, as they escape the DNA methylation induced C-to-T transition mutations that accumulate over evolutionary time as a result of 5 methyl-cytosine deamination and imperfect repair (Bird, 1980; Salser, 1977; Tykocinski and Max, 1984). Closer analysis of experimentally identified CpG islands revealed that these elements often overlapped with annotated transcription start sites of genes, suggesting they may contribute to promoter function (Bird et al., 1985; Gardiner-Garden and Frommer, 1987; Tykocinski and Max, 1984). Based on the elevated frequency of CpG and GC content of experimentally isolated CpG islands, computational algorithms have exploited these properties to predict CpG islands genome-wide (Gardiner-Garden and Frommer, 1987; Takai and Jones, 2002). From these predictions it is apparent that gene regulatory elements that are contained within CpG islands encode normal core promoter elements with transcription factor binding sites but differ from compact promoters of lower eukaryotes in that they tend to range in size from a few hundred base pairs to several kilobases, utilize dispersed transcriptional start sites, and are generally more permissive to transcriptional initiation than non-CpG island promoters (Core et al., 2008; Illingworth and Bird, 2009; Juven-Gershon et al., 2008; Ramirez-Carrozzi et al., 2009; Seila et al., 2008; Takai and Jones, 2002). Although the majority of human genes are associated with CpG islands, their role in promoter architecture and function remains poorly understood despite over 20 years of active investigation.

A DNA binding domain, called the ZF-CxxC domain, has been isolated that specifically recognizes nonmethylated CpG dinucleotides in vitro (Voo et al., 2000). The possibility that such a domain may bind CpG islands in vivo is intriguing but has not been tested (Glaser et al., 2006, 2009; Voo et al., 2000). Interestingly, the ZF-CxxC domain is found in proteins involved in regulating chromatin modifications including components of the DNA methylation system (Jorgensen et al., 2004; Pradhan et al., 2008), the H3K4 methyltransferase system (Birke et al., 2002; Lee and Skalnik, 2005), and recently identified H3K36 demethylases (Tsukada et al., 2006). Therefore one potential role for CpG islands could be as a nucleation site for chromatin-modifying proteins that help define a chromatin environment that differentiates these regulatory regions from bulk genomic chromatin.

Based on this possibility, here we investigate the recently identified ZF-CxxC domain containing H3K36me2 demethylase

Reprinted from Molecular Cell, 38, Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J, CpG islands recruit a histone H3 lysine 36 demethylase, 179-190, Copyright (2010), with permission from Elsevier.

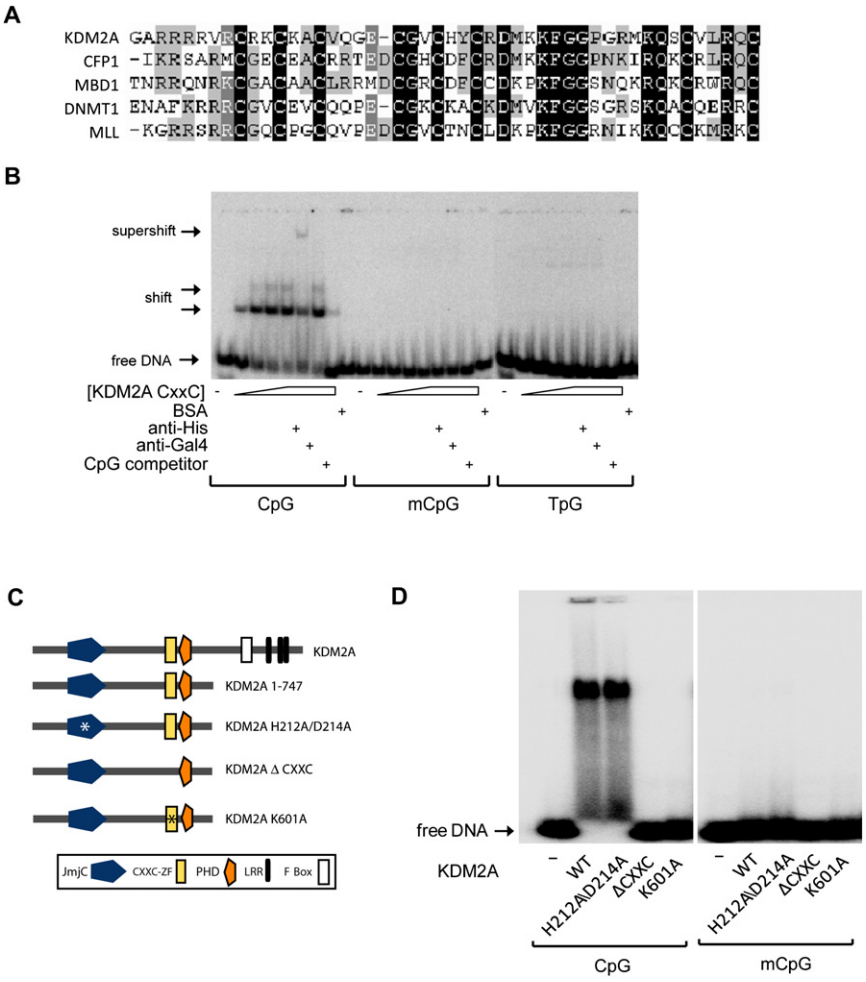


Figure 1. KDM2A Is a Nonmethylated CpG Binding Protein

(A) Multiple sequence alignment of the ZF-CxxC domain from KDM2A, CFP1, MBD1, DNMT1, and MLL.

(B) A recombinant KDM2A ZF-CxxC domain binds nonmethylated CpG DNA in a protein concentration-dependent manner (left panel), and this association is abrogated by CpG methylation (middle panel) or mutation of the CpGs (right panel) as assessed by EMSA analysis. Binding to nonmethylated CpG is specific, as the interaction is competed with unlabeled competitor DNA and an antibody against the recombinant ZF-CxxC domain supershifts the complex.

(C) A schematic illustrating recombinant KDM2A 1–747 and mutant versions of KDM2A, indicating the domain architecture of the protein including the Jumonji C (JmjC), zinc finger CxxC (ZF-CxxC), plant homeodomain-like (PHD), leucine-rich repeat (LRR), and F box (F box) domains.

(D) KDM2A binds nonmethylated CpG dinucleotides (left panel, WT), and methylation blocks this binding (right panel, WT) as assessed by EMSA. Mutation of the catalytic JmjC domain (H212A/D214A) does not affect DNA binding, but deletion of the ZF-CxxC domain (ΔCxxC) or a point mutation in the predicted DNA binding face of the ZF-CxxC domain (K601A) abrogates DNA binding.

enzyme KDM2A and demonstrate that it binds to nonmethylated CpG islands leading to a depletion of H3K36me2 over these elements. The intrinsic DNA binding specificity of KDM2A provides a simple yet elegant mechanism for recognizing CpG island DNA independently of either transcription factors or the transcriptional state of the associated gene. This permits KDM2A to directly impact chromatin architecture and differentiate CpG island chromatin from bulk chromatin.

RESULTS

The KDM2A ZF-CxxC Domain Binds Nonmethylated CpG DNA

To understand if CpG islands are a nucleation site for ZF-CxxC domain proteins, we focused on the KDM2 H3K36 demethylase enzymes (Tsukada et al., 2006). These members of the ZF-CxxC domain protein family are particularly interesting, as there is relatively little understanding of how recently identified histone demethylase enzymes are targeted to chromatin substrates (Klose and Zhang, 2007). In mammals there are two KDM2 enzymes. We focused on KDM2A, as KDM2B has a nucleolar localization signal and is concentrated in this specialized compartment (Frescas et al., 2007). Sequence homology analysis revealed that the KDM2A ZF-CxxC domain is conserved with other functional ZF-CxxC domain-containing proteins (Figure 1A), but its capacity to interact with nonmethylated CpG DNA has not been tested. A recombinant KDM2A protein encompassing the ZF-CxxC domain specifically recognized a DNA probe containing two nonmethylated CpGs in an electrophoretic mobility shift assay (EMSA), and binding was abrogated when the CpGs were mutated or methylated (Figure 1B). Importantly, binding to nonmethylated CpG DNA was also detected using a larger version of the protein that included the catalytic Jumonji C (JmjC) domain on longer DNA probes with CpG content similar to that found in CpG islands (Figures 1C and D). The DNA binding specificity of KDM2A was ZF-CxxC domain dependent, as a mutant version of KDM2A lacking the ZF-CxxC domain (ΔCXXC) or with a point mutation in the proposed DNA binding face (K601A) (Allen et al., 2006; Cierpicki et al., 2010) was unable to interact with DNA (Figure 1D). Together these observations demonstrate that KDM2A is a nonmethylated CpG binding protein.

To quantify the interaction of KDM2A with nonmethylated CpG containing DNA, a probe containing one CpG dinucleotide was immobilized on a solid support, and surface plasmon resonance was utilized to determine the dissociation constant (see Figure S1 available online). The K_D for KDM2A on this sequence was 0.64 μM, which is similar to the previously measured affinities for a related ZF-CxxC domain interacting with CpG DNA (Allen et al., 2006; Birke et al., 2002; Cierpicki et al.,

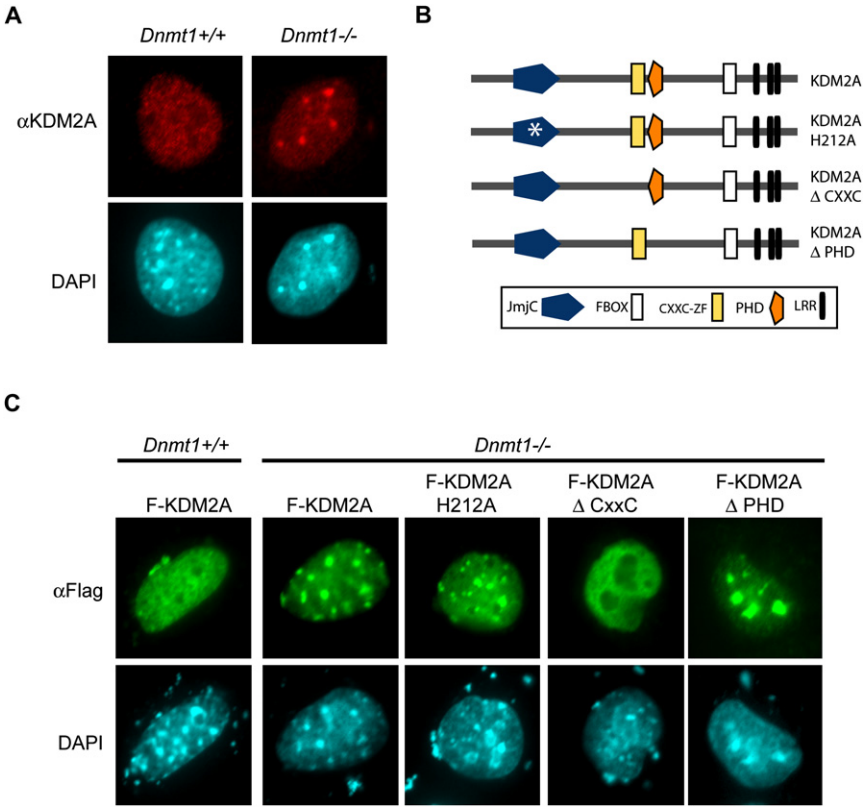


Figure 2. KDM2A Is Targeted to Nonmethylated DNA In Vivo

(A) Indirect immunofluoresence with anti-KDM2A antibodies (red) and DAPI staining of DNA (blue). Endogenous KDM2A staining is found throughout the nucleoplasm in *Dnmt*+/+ MEFs (upper left panel). DAPI bright-staining foci mark regions of densely methylated pericentromeric heterochromatin (lower left panel). In *Dnmt1*$^{-/-}$ MEFs almost devoid of CpG methylation, KDM2A localizes to the hypomethylated pericentromeric DAPI bright-staining repeat regions (compare upper and lower right panels).

(B) A schematic indicating the epitope-tagged wild-type and mutant versions of KDM2A.

(C) Indirect immunofluorescence analyzing localization of exogenous KDM2A expressed in *Dnmt1*$^{+/+}$ (left panel) and *Dnmt1*$^{-/-}$ MEFs (right panels). Deletion of the ZF-CxxC domain (ΔCxxC), but not mutation of the catalytic domain (H212A) or deletion of the PHD domain (ΔPHD), abrogates localization of KDM2A to hypomethylated pericentromeric regions in *Dnmt1*$^{-/-}$ MEFs.

2010). Because KDM2A will encounter CpG island sequences in vivo that have contiguous regions of DNA containing multiple nonmethylated CpG dinucleotides, a second probe based on the one CpG-containing sequence was engineered by substitution mutations to contain six evenly spaced CpG sites. Interestingly, the K_D for KDM2A binding to this probe was 0.56 μM and matched very closely the affinity for the probe containing only one CpG dinucleotide. These observations suggest KDM2A specifically interacts with individual nonmethylated CpG dinucleotides and does not rely on multiple CpGs for binding. This observation is in agreement with a recently published structural study indicating that the ZF-CxxC domain interrogates only one CpG dinucleotide when interacting with DNA (Cierpicki et al., 2010).

KDM2A Is Recruited via a ZF-CxxC Domain to Nonmethylated DNA In Vivo

DNA binding analyses in vitro clearly demonstrated that KDM2A recognizes nonmethylated DNA. For KDM2A to recognize CpG islands in vivo, the ZF-CxxC domain would need to target KDM2A to nonmethylated CpGs in native chromatin. To study the chromatin binding and localization properties of endogenous KDM2A, a KDM2A-specific antibody was generated (Figures S2A–S2C). Using this antibody, KDM2A localization was analyzed in mouse embryonic fibroblasts (MEFs) by indirect immunofluorescence. As observed previously with epitope-tagged KDM2A (Frescas et al., 2008; Tsukada et al., 2006), endogenous KDM2A localized to the nucleus and was broadly distributed throughout the nucleoplasm (Figure 2A and Figure S2D). Normal mouse fibroblasts contain large repetitive tracks of densely methylated DNA at pericentromeric heterochromatin that are therefore refractory to binding of ZF-CxxC domain-containing proteins (Jorgensen et al., 2004). In MEFs deficient for the maintenance DNA methyltransferase 1 (*Dnmt1*), these regions of the genome lose DNA methylation (Figure S2J) and become concentrated foci of nonmethylated DNA (Jorgensen et al., 2004; Lande-Diner et al., 2007). In *Dnmt1*-deficient MEFs, KDM2A signal was concentrated at nonmethylated pericentromeric heterochromatin DAPI bright foci in addition to its normal nucleoplasmic staining, indicating that the protein is recruited to nonmethylated DNA in chromatin (Figure 2A and Figure S2E). Importantly, targeting of KDM2A to nonmethylated DNA relied on the ZF-CxxC domain, as an epitope-tagged version of KDM2A (Figure 2B) with a deletion of this domain remained nucleoplasmic in *Dnmt1*-deficient MEFs (Figure 2C and Figure S2H), whereas mutation of other conserved domains in KDM2A had no effect on localization to pericentromeric regions (Figures 2B and 2C and Figures S2F, S2G, and S2I). These observations demonstrate that the ZF-CxxC domain targets KDM2A to nonmethylated chromatin in vivo.

KDM2A Binds Nonmethylated CpG Island DNA

In the context of a normal cell, the majority of nonmethylated CpG DNA is found in contiguous regions referred to as CpG islands (Bird et al., 1985; Bird, 1986). KDM2A had a clear preference for nonmethylated DNA in vitro and was targeted to nonmethylated pericentromeric DNA in *Dnmt1*-deficient cells. Based on these observations, we sought to examine if KDM2A specifically associated with nonmethylated CpG islands in normal cells. Chromatin immunoprecipitation (ChIP) followed by quantitative PCR (ChIP-qPCR) showed that in mouse embryonic stem cells (ESCs) KDM2A was bound to the CpG island containing *Gnas*

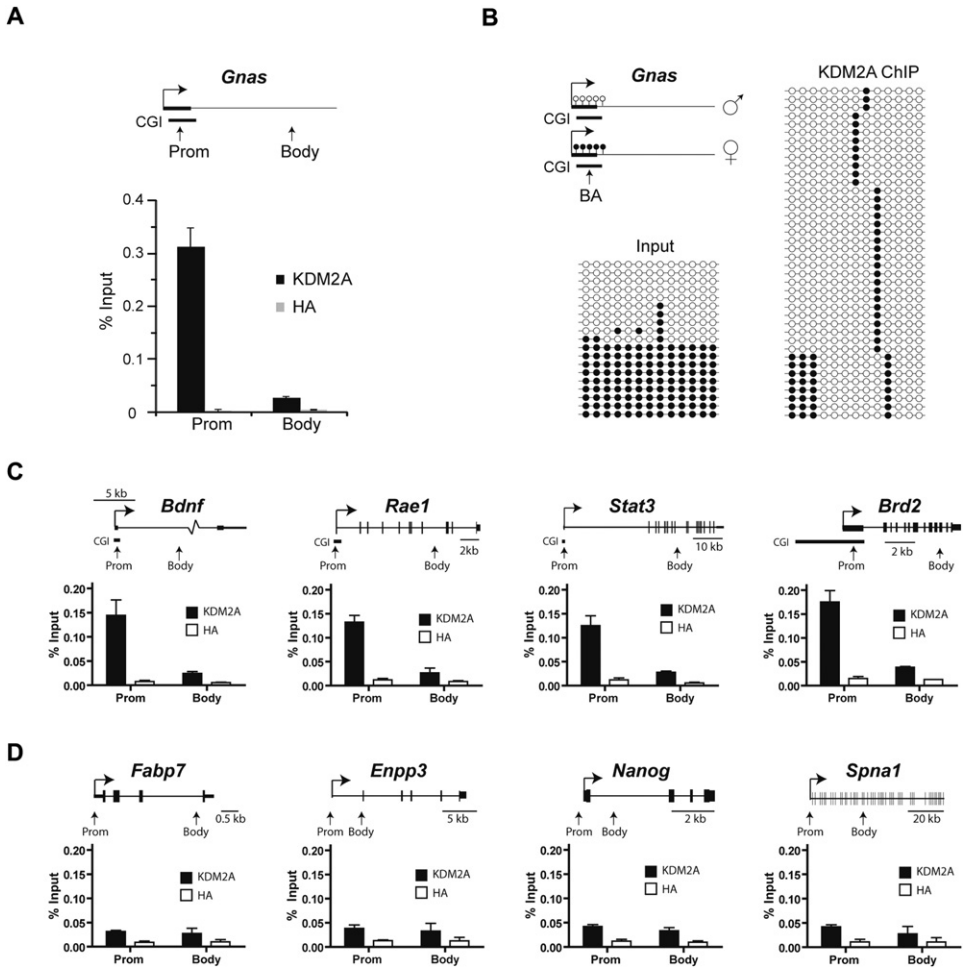


Figure 3. KDM2A Binds to Nonmethylated CpG Islands

(A) A schematic of the imprinted *Gnas* gene, which has a CpG island (CGI) promoter (upper panel). KDM2A binds the *Gnas* CpG island as indicated by ChIP with a KDM2A-specific antibody followed by qPCR with promoter and body primer sets (lower panel). ChIP with an HA-specific antibody was used as a negative control. Data are from two biological replicates, and error bars represent SEM.

(B) The *Gnas* CpG island is methylated on the maternal allele and nonmethylated on the paternal allele (upper panel) with the bisulfite PCR amplicon indicated (BA). Bisulfite sequencing performed on ChIP input DNA (left panel) and KDM2A ChIP DNA (right panel) indicates that KDM2A preferentially binds the nonmethylated paternal allele. Empty and filled circles represent nonmethylated and methylated CpG dinucleotides, respectively.

(C) KDM2A binds to other CpG island promoters. A schematic of each gene is shown, and ChIP data are presented as in (A).

(D) KDM2A is not enriched at genes with non-CpG island promoters. ChIP data are presented as in (A).

promoter but not to the body of the gene (Figure 3A). The *Gnas* gene is subject to an allele-specific epigenetic imprinting process that results in dense methylation of the CpG island on the maternal allele while the paternal allele remains nonmethylated (Figure 3B). Bisulfite DNA sequencing confirmed that this configuration was maintained in our ESC lines (Figure 3B), as approximately 50% of alleles were methylated and 50% were nonmethylated. Strikingly, when the DNA isolated in the KDM2A ChIP was analyzed by bisulfite sequencing, this material was specifically enriched from the nonmethylated allele. This demonstrates that KDM2A associates specifically with the nonmethylated CpG island of the paternal allele in vivo (Figure 3B). Furthermore, when several other CpG island genes were analyzed by ChIP-qPCR, KDM2A was specifically enriched at CpG island promoters, but not at the corresponding gene bodies (Figure 3C and Figure S3). Similar ChIP-qPCR analysis of several genes with non-CpG island promoters showed no KDM2A enrichment (Figure 3D and Figure S3). Together these observations suggest that KDM2A is preferentially targeted to nonmethylated CpG islands.

KDM2A Associates with CpG Islands Genome-wide

To assess whether KDM2A association with CpG islands in ESCs is a genome-wide phenomenon, we exploited chromatin immunoprecipitation coupled to massively parallel sequencing (ChIP-seq). After KDM2A-bound DNA was aligned to the mouse genome, a striking overlap was observed between KDM2A binding and algorithm-predicted CpG island elements. This

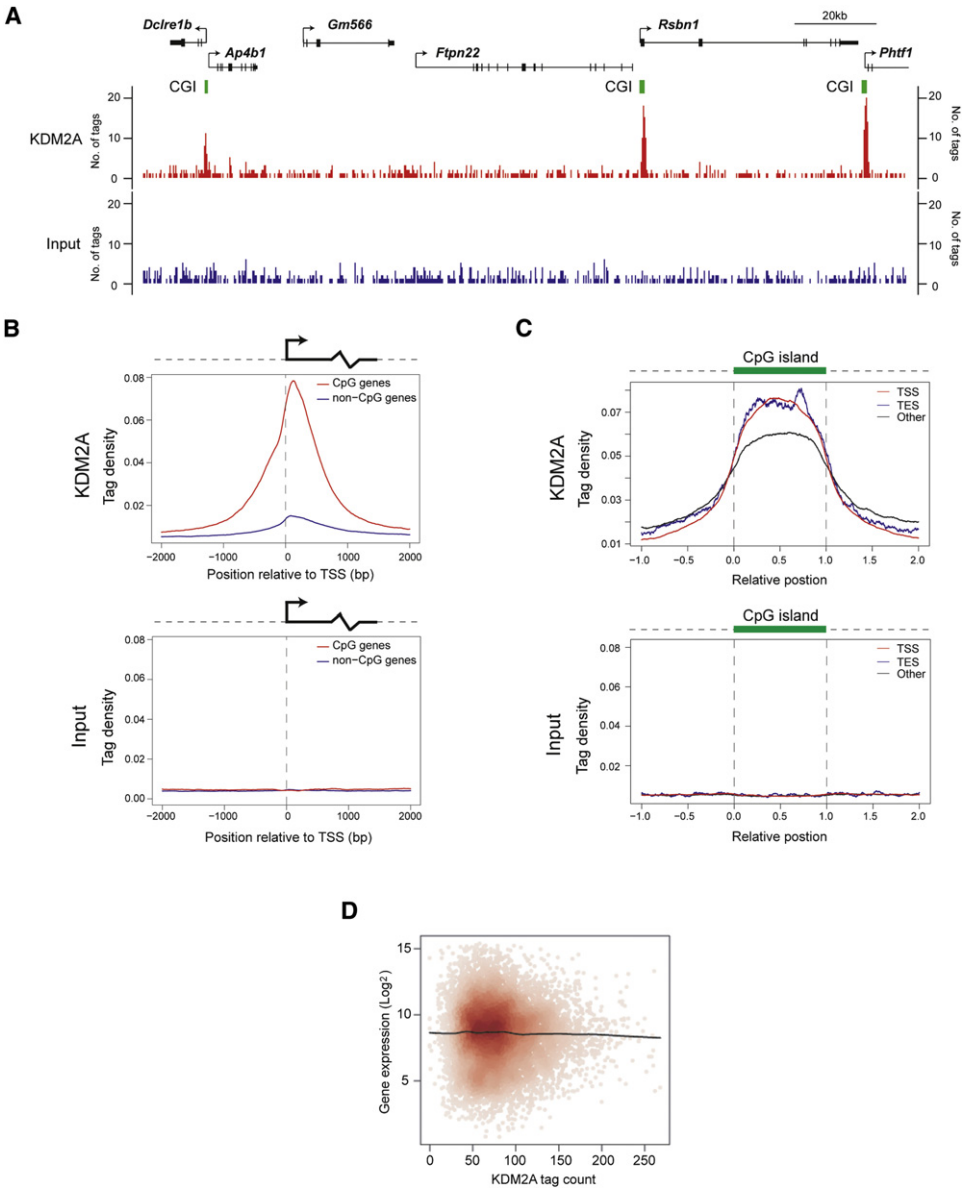


Figure 4. KDM2A Binds to CpG Islands Genome-wide

(A) KDM2A ChIP-seq analysis of an approximately 180 kb region of chromosome 3 (chr3:103598131–103782151) in mouse ESCs (top panel, red) with comparative analysis from input material (lower panel, blue). Annotated genes in this region are illustrated above the sequence trace with arrows indicating transcription start sites, vertical black bars corresponding to exons, and CpG islands represented by green bars.

(B) Average KDM2A tag density (upper panel) and ChIP input tag density (lower panel) at CpG island-associated transcription start sites (TSS) (red line) compared to non-CpG island TSS (blue line). Density profiles were calculated by averaging tag counts at each position in a 4 kb region around the TSS for all promoters in a set.

(C) Average KDM2A tag density (upper panel) and ChIP input tag density (lower panel) across CpG islands at TSS (red line), transcription end sites (TES) (blue line), and other genomic CpG island locations (black line). Tag density is plotted relative to the CpG island coordinates, where the CpG island start and end correspond to positions 0 and 1, respectively.

(D) KDM2A binding does not correlate with gene expression levels, as indicted by comparison of gene expression level versus KDM2A tag count in the TSS proximal regions of all CpG island genes. Red dots correspond to individual genes, and a black line representing loess regression is shown to indicate the trend in data distribution. The gene expression levels were estimated based on published microarray data for mouse ESCs (GEO accession number GSE8024) (Mikkelsen et al., 2007).

was apparent when we focused on contiguous regions of the genome containing both CpG island and non-CpG island genes with significant KDM2A enrichment only observed at CpG island elements (Figure 4A and Figures S4A–S4D). This remarkable specificity held true when KDM2A tag density was analyzed over all transcription start sites in the genome with specific

KDM2A tag density enrichment at bioinformatically defined CpG islands and negligible tag density over non-CpG island promoters (Figure 4B). Importantly, there was no tag density enrichment at either promoter type when the input chromatin sample was analyzed in the same way (Figure 4B). In addition to CpG islands that are associated with gene promoters, some CpG islands are found at transcription end sites or separate from either annotated transcription start or end sites. Similar to the enrichment seen at CpG island promoter regions, KDM2A tag density was enriched over CpG island elements regardless of whether or not they were promoter associated (Figure 4C). Importantly, localization of KDM2A to CpG islands appears to depend on the presence of underlying nonmethylated CpG DNA sequence and not simply as a secondary consequence of engagement of the transcriptional machinery, as KDM2A-specific enrichment was observed at both expressed and nonexpressed CpG island genes (Figures S4E–S4G).

Based on ChIP-seq analysis, KDM2A occupies greater than 90% of CpG islands and over 95% of transcription start site-associated CpG islands as defined by the CpG island prediction algorithm (Table S1). Importantly, these regions encompass the majority of high-magnitude KDM2A binding clusters (Figure S4H, Figure 4A, Figures S4A–S4D). Algorithm-based peak finding also identified a large fraction of low-magnitude KDM2A-enriched regions that occurred outside of bioinformatically annotated CpG islands (Table S1). To investigate whether these binding events occurred at nonmethylated regions of the genome, a series of non-CpG island KDM2A ChIP-seq peaks (both promoter associated and non-promoter associated) were analyzed by bisulfite sequencing (Figures S4L–S4Q). The analyzed promoter-associated non-CpG island KDM2A peaks correspond to contiguous regions of nonmethylated DNA (Figures S4L and S4M), indicating they are likely CpG islands that fall below the criteria for inclusion in the algorithm-defined CpG island set. This suggests that a proportion of bona fide nonmethylated CpG islands has been excluded from the CpG islands prediction set, as has been observed previously (Illingworth et al., 2008). This is also in fitting with the observation that non-CpG island KDM2A peaks tend to exhibit GC content and observed/expected CpG ratios greater than that of bulk genomic DNA (Figures S4J and S4K). Other KDM2A sites tended to be of very low magnitude, and bisulfite sequencing verified that these regions only contain small amounts of mosaically nonmethylated DNA (Figures S4N–S4Q). Therefore, in fitting with the observation that the majority of nonmethylated DNA is found in CpG islands, strong KDM2A binding corresponds to annotated CpG islands (Figure S4H), and KDM2A tag density is more concentrated at strong as opposed to weak CpG islands (Figure S4I).

Together these observations demonstrate that KDM2A recognizes CpG islands through an intrinsic DNA binding capacity as opposed to relying on nucleation through transcription factors, as do other histone demethylases and most chromatin-modifying enzymes. Because KDM2A does not rely on the transcriptional regulatory machinery to interact with DNA sequences, one would envisage that binding and therefore function at CpG island genes was independent of the underlying transcriptional state of the associated gene. This is supported by the fact that KDM2A enrichment was detected at over 95% of transcription start sites that contain CpG islands, but tag density did not correlate with gene expression levels of these CpG island genes in mouse ESCs (Figure 4D). Furthermore, KDM2A binds both expressed and nonexpressed CpG island genes (Figures S4E–S4G). These observations are particularly important as direct binding of KDM2A to CpG island DNA suggests a function for this protein that does not directly result in transcriptional activation or repression per se.

CpG Islands Are Specifically Depleted of H3K36me2

A potential role for KDM2A at CpG islands could be to mark these regions of the genome by specifically depleting H3K36 methylation. KDM2A is a histone H3 lysine 36 (H3K36) demethylase that preferentially removes the dimethyl (me2), and to a lesser extent the monomethyl (me1), modification state (Tsukada et al., 2006). To explore if H3K36 methylation was depleted at CpG island promoters, ChIP was carried out at a series of CpG island and non-CpG island genes, and H3K36me1, -me2, and -me3 modifications were analyzed. Remarkably, the levels of H3K36me1 and -me2 were significantly depleted in the promoter regions of CpG island genes when compared to the promoters of non-CpG island genes and the body of both classes of genes (Figure 5A and Figures S5A and S5B). Depletion of H3K36me2 was most clearly observed at CpG island promoters, in line with the previously characterized enzymatic preference of KDM2A toward this modification state (Tsukada et al., 2006) and our observations by ChIP-seq that KDM2A binds CpG islands genome-wide. To understand if the H3K36me2 depletion at CpG island promoters corresponded precisely to regions of KDM2A binding, KDM2A and H3K36me2 ChIP profiles were compared over a series of tiled amplicons covering genes with CpG island and non-CpG island promoters (Figures 5B and 5C). Importantly, at CpG island-containing promoters, KDM2A binding peaked over the CpG island region and precisely corresponded to the region of chromatin that showed depletion of H3K36me2 (Figure 5B). In clear contrast, non-CpG island promoters had no KDM2A and lacked obvious H3K36me2 depletion (Figure 5C). Interestingly, over the body of expressed genes, depletion of H3K36me2 was also observed, and this corresponded to the appearance of the H3K36me3 state (Figure 5B). This may represent conversion of H3K36me2 to the -me3 modification state, in agreement with previous observations in ESCs that H3K36me3 occurs exclusively over the body of actively transcribed genes (Mikkelsen et al., 2007) and the recent discovery in fly and mouse that the H3K36me2 and -me3 modification states are placed by different enzymes (Bell et al., 2007; Edmunds et al., 2008). Surprisingly, we observed relatively constant levels of H3K36me2 upstream, over the body, and downstream of genes in all regions analyzed. This suggested that H3K36me2 may be more broadly distributed in chromatin of higher eukaryotes than previously realized and parallels published mass spectrometry data that indicated that up to 40% of mammalian cellular histone H3 was characterized by the H3K36me2 state (Garcia et al., 2008; Peters et al., 2003; Robin et al., 2007). Because KDM2A was targeted to CpG islands based on nonmethylated DNA sequence, it appears to function in a unique and generic manner to remove H3K36me2 from island regions and epigenetically differentiate these

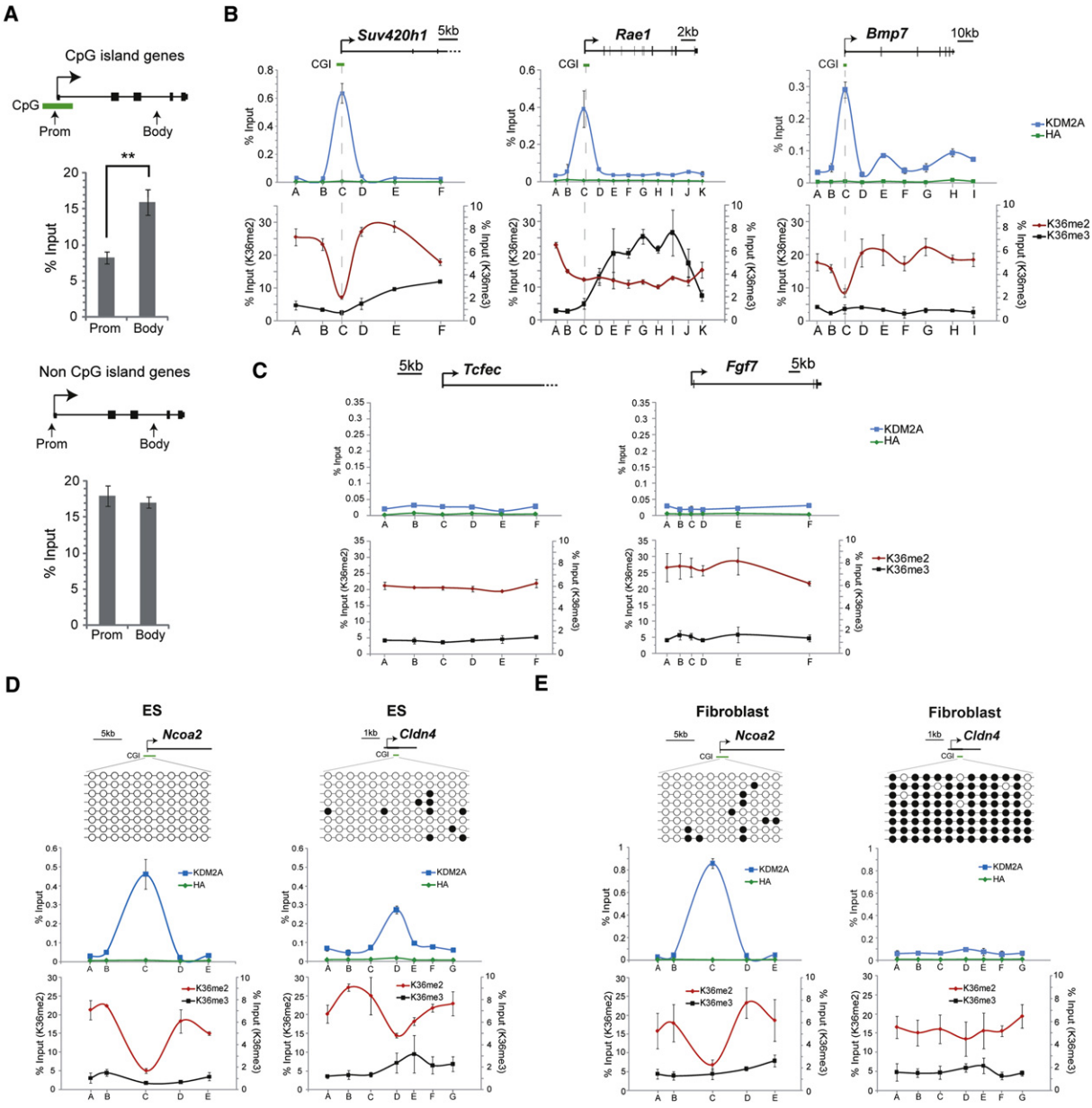


Figure 5. Depletion of H3K36me2 at KDM2A-Bound CpG Islands

(A) H3K36me2 is depleted at KDM2A-bound CpG islands as indicated by the average H3K36me2 ChIP enrichments for a series of CpG island (upper panel) and non-CpG island genes (lower panel). For each type of gene, qPCR primer sets specific to promoter and body regions were used. For each primer set, H3K36me2-specific enrichment was normalized to histone H3 occupancy. Each data set represents the average of nine individual genes (see Figure S5). At CpG island genes, K36me2 is significantly depleted at the promoter relative to the body region (**$p < 0.01$), whereas at genes with non-CpG island promoters there is no significant difference in K36me2 between promoter and body regions ($p > 0.05$).

(B and C) (B) Comparison of KDM2A (upper panel, blue line) and H3K36me2/me3 (lower panel, red and black lines) ChIP profiles across tiled CpG island genes and (C) non-CpG island genes, indicating specific enrichment of KDM2A and depletion of H3K36me2 at CpG islands. Individual genes are depicted to scale above each panel. Anti HA antibody (green line) is included as a negative control in the KDM2A ChIP experiments. Enrichment values for H3K36me2 (left-hand y axis) and me3 (right-hand y axis) are normalized to histone H3 occupancy.

(D and E) (D) Comparison of DNA methylation by bisulfite sequencing (upper panel), KDM2A occupancy by ChIP (middle panel), and H3K36me2/me3 levels by ChIP (bottom panel) at *Ncoa2* and *Cldn4* in mouse ESCs and (E) mouse fibroblasts. Empty and filled circles (upper panel) represent nonmethylated and methylated CpG dinucleotides, respectively. KDM2A ChIP and H3K36me2/me3 ChIP are represented as described for (B) and (C). DNA methylation of the *Cldn4* CpG island in fibroblasts blocks KDM2A binding and abrogates the depletion of H3K36me2 over the CpG island region that is observed in ESCs, whereas the *Ncoa2* CpG island remains nonmethylated in both cell types, retains KDM2A, and has a depletion of H3K36me2 at the CpG island. All ChIP data represent at least two biological replicates, and error bars indicate SEM.

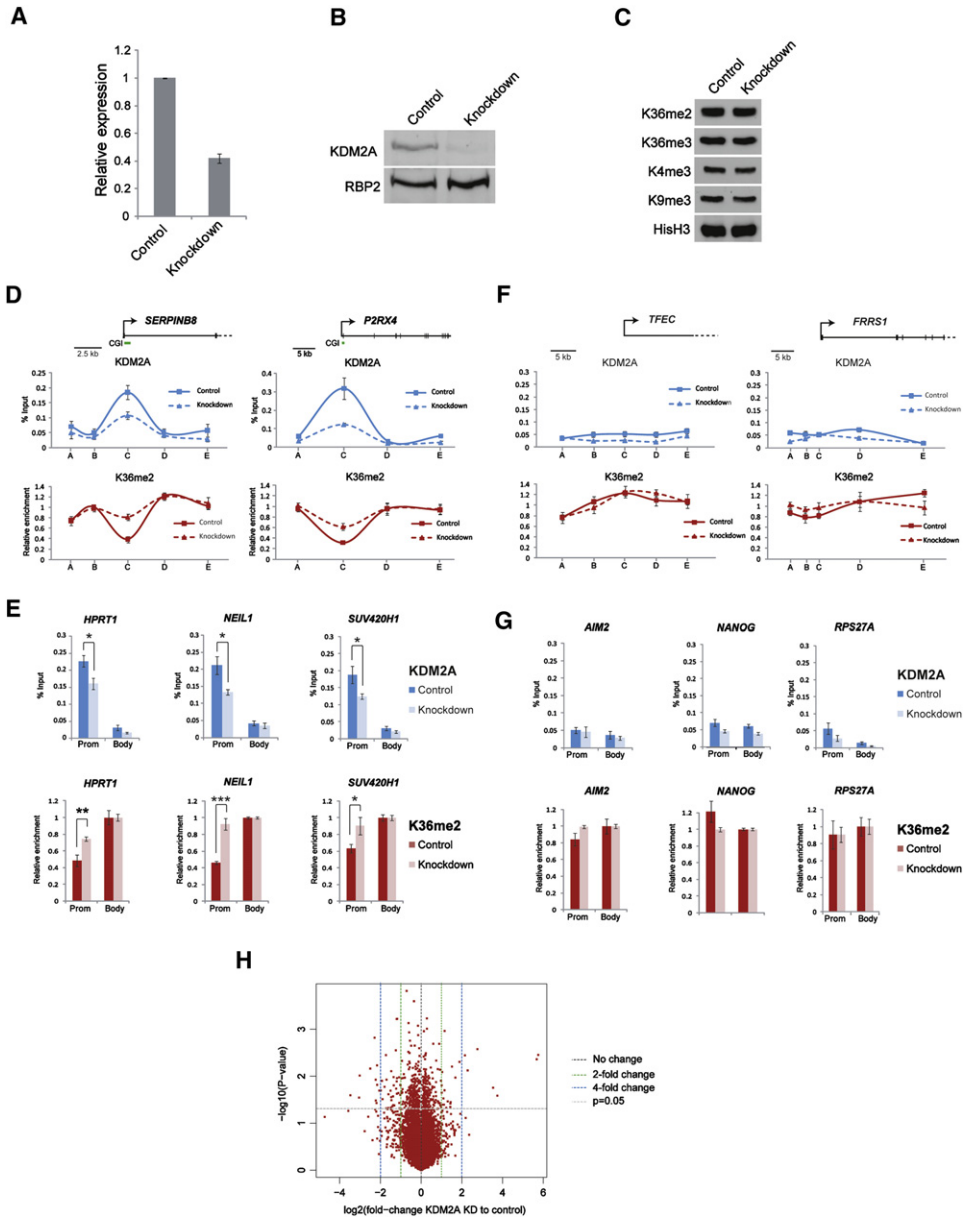


Figure 6. Depletion of KDM2A Causes Increases in H3K36me2 at CpG Islands

(A) A KDM2A knockdown cell line displays an approximately 60% decrease in KDM2A mRNA levels compared to the control cell line.

(B) Western blot with an antibody against KDM2A shows depletion of KDM2A protein levels in the KDM2A knockdown cell line compared to control cell line. Western blot with an RBP2 antibody demonstrates equal loading.

(C) KDM2A knockdown cells exhibit no gross differences in levels of H3K36me2, H3K36me3, H3K4me3, or H3K9me3 relative to control cells. Western blot with a histone H3 antibody demonstrates equal loading.

(D) Shown is comparison of KDM2A (upper panel blue lines) and H3K36me2 (lower panel red lines) ChIP profiles across tiled CpG island genes in control (solid lines) and KDM2A knockdown cell line (dashed lines). KDM2A knockdown cells exhibit specific loss of KDM2A enrichment at CpG islands, resulting in a CpG island-specific gain of H3K36me2. Individual genes are depicted to scale above each panel.

(E) Comparison of KDM2A (upper panel, blue bars) and H3K36me2 (lower panel, red bars) ChIP at promoter and body regions of CpG island genes in control (dark blue and dark red) and KDM2A knockdown cells (light blue and light red). At CpG island regions, the KDM2A knockdown cell line shows a significant loss of KDM2A-specific enrichment compared to control cells (*$p < 0.05$). Loss of KDM2A at CpG islands correlates with a significant H3K36me2 increase in knockdown cells compared to control (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

(F and G) Non-CpG island genes were analyzed as in (D) and (E). At genes with non-CpG island promoters, KDM2A knockdown cells show no significant difference in KDM2A or K36me2 enrichment compared to control cells ($p > 0.05$). All ChIP data represent at least two biological replicates, and error bars indicate SEM.

(H) Volcano plot comparison of expression values in the KDM2A knockdown and control cell lines reveals subtle but significant changes in gene expression. Fold change in expression values (KDM2A knockdown relative to control) and the statistical significance of the differences (two sided t test) were computed for each

regulatory elements from bulk chromatin and non-CpG island promoters.

Depletion of H3K36me2 Is Lost when KDM2A Binding Is Inhibited by CpG Island Methylation

KDM2A binding to CpG DNA in vitro is blocked by DNA methylation, and it preferentially associates with nonmethylated CpG islands in vivo (Figures 1B, 1D, and 3B). Although a clear attribute of CpG islands is their capacity to remain free of DNA methylation whether the associated gene is active or repressed, as cells differentiate during development a small number of CpG island promoters undergo de novo DNA methylation through a poorly understood mechanism. Because DNA methylation blocks KDM2A binding, this permitted chromatin analysis at a defined gene either in the presence or absence of KDM2A. To this end, we examined H3K36me2 at a control gene with a CpG island that is nonmethylated in both ESCs and differentiated fibroblasts (*Ncoa2*) and a gene that has a CpG island that acquires DNA methylation in fibroblasts (*Cldn4*) (Figures 5D and 5E). In mouse ESCs and fibroblasts, the *Ncoa2* CpG island promoter was nonmethylated, bound KDM2A, and showed specific depletion of H3K36me2 over the CpG island (Figures 5D and 5E). In stark contrast, while the nonmethylated *Cldn4* CpG island also showed KDM2A binding and H3K36me2 depletion in ESCs, acquisition of CpG island DNA methylation in differentiated fibroblasts correlated with the absence of KDM2A binding and the appearance of H3K36me2 at the CpG island (Figure 5E, right panel). Therefore, KDM2A binding correlates with CpG island-specific depletion of H3K36me2. Strikingly, this observation also highlights that cell-type-specific epigenetic changes in CpG island methylation can regulate KDM2A binding and thus impact promoter-specific H3K36 modification.

Depletion of KDM2A Causes Increases in H3K36me2 at CpG Islands

Despite exhaustive attempts to knock down KDM2A in a number of different ES and primary cell lines using an array of RNAi-mediated approaches, KDM2A has proved to be highly refractory to depletion, suggesting it may contribute to an essential process in these cell types. To try and circumvent this limitation, a stably expressed shRNA-mediated approach was used to deplete KDM2A in human cervical carcinoma cells. Approximately 60% knockdown was achieved at the RNA level as assessed by RT-PCR (Figure 6A), and clear depletion of the protein was observed when assessed by western blot analysis (Figure 6B). To understand if KDM2A depletion affected global levels of H3K36me2, bulk cellular histone was isolated and histone methylation levels analyzed by western blot (Figure 6C). No obvious global changes in H3K36me2 or other histone methylation marks were observed, in fitting with the contention that H3K36me2 is an abundant modification and KDM2A-mediated depletion is restricted to a small fraction of the genome. To understand if KDM2A contributes to the depletion of H3K36me2 at CpG islands, ChIP-qPCR was used to analyze KDM2A and H3K36me2 occupancy in the control and KDM2A knockdown cell lines (Figures 6D–6G). Importantly, in the control line, KDM2A occupancy was restricted to CpG island sequences and H3K36me2 was depleted at CpG islands, as observed in mouse ES and fibroblast cells (Figures 6D and 6E). In the KDM2A knockdown cell line, we observed a reduction in KDM2A ChIP signal specifically at CpG islands genes (Figures 6D and 6E). The reduction of KDM2A at CpG island genes resulted in a significant increase in H3K36me2 levels specifically at the CpG island. Importantly, this reduction in KDM2A occupancy and increase in H3K36me2 in the KDM2A knockdown line was not due to alteration in the underlying DNA methylation profiles due to perturbed KDM2A levels (Figures S6A and S6B). In contrast, we did not observe any significant increase in the levels of H3K36me2 at non-CpG island genes (Figures 6F and 6G). Together these observations demonstrate that depletion of KDM2A results in encroachment of H3K36me2 into CpG islands regions and indicate a causative role for KDM2A demethylase activity in defining depletion of this mark at CpG islands.

When the transcription levels of CpG island genes were analyzed comparing the control and KDM2A knockdown line by RT-PCR, there was no general correlation between the depletion of KDM2A and upregulation or downregulation of steady-state transcript levels (Figure S6C). This is in fitting with the observation in mouse ESCs that KDM2A does not specifically bind to expressed or nonexpressed CpG island genes and argues against KDM2A playing a role as a traditional activator or repressor. We did observe some subtle changes in gene expression between the control and KDM2A KD cell line by RT-PCR (Figure S6C). This prompted an analysis of the genome-wide expression profiles in the control and KDM2A knockdown cell lines by microarray-based gene expression profiling (Figure 6H). Interestingly, this analysis revealed statistically significant changes in gene expression, but the majority of these were subtle and less than 2-fold. KDM2A potentially associates with up to two-thirds of genes through association with CpG islands, but we did not observe any clear correlation between these gene expression changes and CpG island or non-CpG island genes (data not shown), suggesting that any CpG island-specific changes are indiscernible from downstream secondary systemic effects. As elaborated on in the discussion, removal of H3K36me2 by KDM2A as part of the systems that define CpG island chromatin architecture may contribute to the transcriptional competency of CpG island genes. Consistent with this possibility, depletion of KDM2A causes subtle changes in gene expression that appear to be tolerated by these transformed cells. One possible explanation for the fact that KDM2A depletion has been unattainable in primary nontransformed cells could be that these cell types are unable to tolerate the aggregated effect of the gene expression perturbations resulting from KDM2A depletion, even though these are moderate in magnitude. Nevertheless, these observations suggest that KDM2A is not functioning as a potent CpG island-specific transcriptional activator or repressor but instead defines chromatin architecture

transcript interrogated on the array (red dots). Vertical dashed black line represents zero change in expression, vertical dashed green lines represent a 2-fold change in expression, and vertical dashed blue lines represent a 4-fold change in expression. Transcript data points above the horizontal dashed gray line show a statistically significant change in expression between KDM2A knockdown and control cell lines ($p < 0.05$).

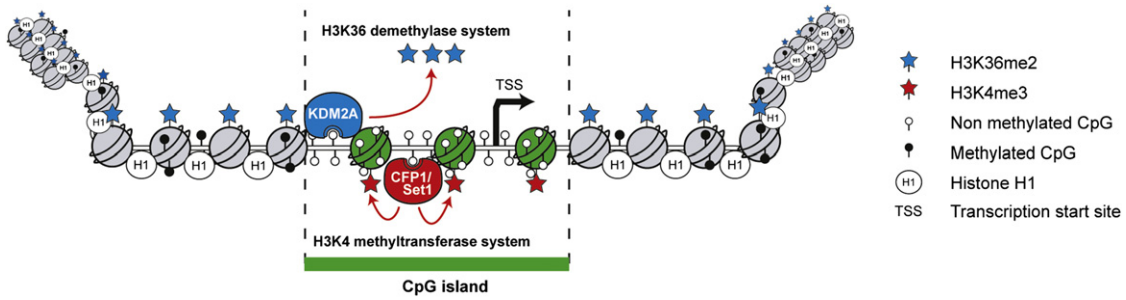


Figure 7. Unique Chromatin Architecture at CpG Islands

CpG islands utilize nonmethylated CpG dinucleotides to recruit the ZF-CxxC domain protein KDM2A, depleting these regions of H3K36me2 and creating a chromatin environment that is unique from bulk non-CpG island chromatin. CpG island chromatin is also modified by H3K4me3, and this relies on targeting of the Set1 complex by the ZF-CxxC domain containing CFP1 protein (A. Bird, personal communication). Another facet of CpG island chromatin is depletion of histone H1. Because CpG islands are usually part of regulatory regions or associated with transcription start sites, this unique chromatin architecture may function to differentiate these regulatory regions from bulk chromatin in large complex mammalian genomes.

at CpG islands through enzymatic removal of H3K36me2, a process that may indirectly contribute to normal transcriptional profiles.

DISCUSSION

Although the existence of CpG islands has been known for over two decades, the mechanisms through which they contribute to genome function have remained poorly understood. Here we provide compelling evidence that CpG islands are actively recognized by KDM2A binding to nonmethylated CpG DNA through a ZF-CxxC domain (Figures 1–3). Using ChIP-seq we show that KDM2A binds to 90% CpG island elements genome-wide and that this nucleation event is independent of the transcriptional status of the associated gene (Figure 4). We demonstrate that KDM2A occupancy at CpG islands imposes a unique H3K36me2-depleted chromatin signature that differentiates these regulatory elements from bulk chromatin, and we provide functional data demonstrating that depletion of KDM2A results in H3K36me2 spreading into CpG island elements (Figures 5 and 6). Together these observations provide insight into how CpG islands can impact chromatin architecture at regulatory elements by utilizing their underlying DNA sequence as a nucleation site for a chromatin-modifying enzyme (Figure 7). Therefore, these data suggest a central function of CpG island DNA may be to impose a chromatin architecture that differentiates CpG island chromatin from bulk chromatin, highlighting these important regulatory regions within large and complex mammalian genomes. In support of this contention, CpG island chromatin is also enriched in H3K4me3 regardless of whether the associated gene is expressed (Bernstein et al., 2006; Guenther et al., 2007). Intriguingly, Set1 and Mll H3K4 methyltransferase complexes that trigger this modification also have ZF-CxxC domains (Birke et al., 2002; Lee and Skalnik, 2005), suggesting they may use a similar mechanism as KDM2A to target CpG islands (Figure 7). This possibility is supported by evidence that the ZF-CxxC domain-containing protein, CFP1, which is a component of the Set1 H3K4 methyltransferase complex, is targeted to CpG island elements genome-wide, where it deposits H3K4me3 methylation in a manner that is independent of the transcriptional state of the associated gene (A. Bird, personal communication). These observations regarding CFP1 nucleation and H3K4me3 perfectly mirror our discovery that KDM2A binds CpG islands, resulting in enzymatic depletion of H3K36me2 independently of the transcriptional state of the associated gene. Together these data indicate that ZF-CxxC domain-containing proteins are important mediators of CpG island chromatin architecture and highlight the fact that CpG islands as a DNA-encoded genetic element are functioning to directly impact the epigenetic state of surrounding chromatin.

Interestingly, the processes that impose a unique chromatin environment at CpG islands are presumably even more complex. For example, from biochemical assays it is known that CpG island chromatin is specifically depleted of linker histone H1 (Tazi and Bird, 1990) (Figure 7). Given that CpG island chromatin architecture appears to form irrespective of transcriptional state, an interesting question is what impact this chromatin environment has on the function of associated genes. It is known that H3K36me2 and H1 can have an inhibitory effect on transcription initiation (Carrozza et al., 2005; Cheung et al., 2002; Levine et al., 1993; Li et al., 2009; Strahl et al., 2002; Youdell et al., 2008), and H3K4me3 is generally associated with processes positively contributing to transcription (Kouzarides, 2007). Therefore, one possibility is that unique CpG island chromatin architecture reinforced by several intersecting chromatin-regulating pathways may define regions of the genome that are more permissive to nucleation of the transcriptional machinery, effectively differentiating CpG islands from bulk chromatin and highlighting regulatory regions of the genome. The inherent complexity of the CpG chromatin signature makes this a difficult hypothesis to directly examine experimentally, but this concept is indirectly supported by the observation in genome-wide run-on transcription assays that CpG island promoters sustain nonproductive transcriptional initiation events in both sense and antisense directions even in the absence of activated directional transcription, whereas non-CpG island genes fail to show this property (Core et al., 2008; Seila et al., 2008). An important implication of this hypothesis is that CpG island chromatin architecture would provide an environment that is permissive to transcription but not drive productive directional

transcriptional output, a process that requires transcription factor binding and concerted gene activation mechanisms. This type of CpG island-specific transcriptional competence has recently been shown to contribute to the induction kinetics of CpG island-containing genes in activated macrophages (Ramirez-Carrozzi et al., 2009).

In conclusion, we have provided a mechanistic link between CpG island elements and nucleation of a histone demethylase and demonstrated that CpG islands function to define cellular chromatin landscapes at these regulatory elements. A corollary of this observation from a genome evolution standpoint is that ZF-CxxC domain recognition of CpG island elements may also impose an additional selective pressure over evolutionary time to maintain the nonmethylated state of CpG island elements. Based on our understanding of CpG island function presented here, a future challenge will be to understand if chromatin modifications at CpG islands impact the transcriptional machinery and how non-CpG island promoters differ in the absence of these mechanisms.

EXPERIMENTAL PROCEDURES

Cell Culture

Cell types used in the study were cultured as described in the Supplemental Experimental Procedures.

Chromatin Immunoprecipitation

For KDM2A, ChIP cells were fixed for 1 hr in 2 mM EGS, followed by 15 min in 1% formaldehyde. For histone modification, ChIP cells were fixed for 10 min in 1% formaldehyde alone. A detailed ChIP protocol, including information about specific antibodies used for ChIP, is described in the Supplemental Experimental Procedures.

ChIP Sequencing

ChIP material was sequenced using a Solexa 2G instrument. Sequencing data are available at GEO accession number GSE21202. Sequencing data were analyzed as described in the Supplemental Experimental Procedures.

Immunofluoresence

Cells were fixed and immunostained with specific antibodies, as described in the Supplemental Experimental Procedures. Photographs were captured with an AxioSkop fluorescent microscope (Zeiss).

Protein Expression and Purification

DNA constructs used for protein expression are described in the Supplemental Experimental Procedures. Expression and purification of KDM2A ZF-CxxC domain and KDM2A 1–747 constructs were performed as previously described (Klose and Bird, 2004) except for minor modifications described in the Supplemental Experimental Procedures.

Electrophoretic Mobility Shift Assay and Surface Plasmon Resonance

Detailed protocols for EMSA and surface plasmon resonance experiments, including information about probes, are described in the Supplemental Experimental Procedures.

Microarray and Data Analysis

RNA labeling and hybridization to 4 × 44 K human gene expression microarrays (Agilent Technologies, Inc., Santa Clara, CA) were carried out by Oxford Gene Technology (Oxford, UK). Hybridizations were performed for two biological replicates for each sample. Microarray data are available at GEO accession number GSE21202. Microarray data were analyzed as described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

ChIP sequencing and microarray expression data are available at GEO accession number GSE21202.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, one table, Supplemental Experimental Procedures, and Supplemental References and can be found with this article at doi:10.1016/j.molcel.2010.04.009.

ACKNOWLEDGMENTS

We would like to thank the Oxford Wellcome Trust Centre for Human Genetics Genomics Facility for Solexa Sequencing, Doug Vernimmen and Doug Higgs for advice on ChIP, David Staunton for help with SPR, Sarah Threlfell for statistical advice, and Neil Brockdorff and Tatyana Nesterova for mouse ESC reagents and advice. We would like to thank Adrian Bird and John Thompson for sharing unpublished observations prior to publication and fruitful discussion, and Rod Bremner for discussion and critical evaluation of the manuscript. This work was supported by a Wellcome Trust CDA fellowship to R.J.K., a Medical Research Council (MRC) PhD studentship to J.C.Z., a Cancer Research UK (CRUK) PhD studentship to A.M.F., and U.S. National Institutes of Health grant R01GM082798 to P.J.P.

Received: January 11, 2010
Revised: February 26, 2010
Accepted: April 6, 2010
Published: April 22, 2010

REFERENCES

Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., and Warren, A.J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. EMBO J. *25*, 4503–4512.

Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., van Leeuwen, F., Bell, S.P., Imhof, A., et al. (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. EMBO J. *26*, 4974–4984.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. *8*, 1499–1504.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91–99.

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res. *30*, 958–965.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., and Workman, J.L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581–592.

Cheung, E., Zarifyan, A.S., and Kraus, W.L. (2002). Histone H1 represses estrogen receptor alpha transcriptional activity by selectively inhibiting receptor-mediated transcription initiation. Mol. Cell. Biol. *22*, 2463–2471.

Cierpicki, T., Risner, L.E., Grembecka, J., Lukasik, S.M., Popovic, R., Omonkowska, M., Shultis, D.D., Zeleznik-Le, N.J., and Bushweller, J.H. (2010).

Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat. Struct. Mol. Biol. *17*, 62–68.

Cooper, D.N., Taggart, M.H., and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. Nucleic Acids Res. *11*, 647–658.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J. *27*, 406–420.

Frescas, D., Guardavaccaro, D., Bassermann, F., Koyama-Nasu, R., and Pagano, M. (2007). JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. Nature *450*, 309–313.

Frescas, D., Guardavaccaro, D., Kuchay, S.M., Kato, H., Poleshko, A., Basrur, V., Elenitoba-Johnson, K.S., Katz, R.A., and Pagano, M. (2008). KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. Cell Cycle *7*, 3539–3547.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Garcia, B.A., Thomas, C.E., Kelleher, N.L., and Mizzen, C.A. (2008). Tissue-specific expression and post-translational modification of histone H3 variants. J. Proteome Res. *7*, 4225–4236.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–282.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.L., and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development *133*, 1423–1432.

Glaser, S., Lubitz, S., Loveland, K.L., Ohbo, K., Robb, L., Schwenk, F., Seibler, J., Roellig, D., Kranz, A., Anastassiadis, K., and Stewart, A.F. (2009). The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. Epigenetics Chromatin *2*, 5.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands–'a rough guide'. FEBS Lett. *583*, 1713–1720.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., et al. (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol. *6*, e22. 10.1371/journal.pbio.0060022.

Jorgensen, H.F., Ben-Porath, I., and Bird, A.P. (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol. Cell. Biol. *24*, 3387–3395.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter—the gateway to transcription. Curr. Opin. Cell Biol. *20*, 253–259.

Klose, R.J., and Bird, A.P. (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. J. Biol. Chem. *279*, 46490–46496.

Klose, R.J., and Zhang, Y. (2007). Regulation of histone methylation by demethylimination and demethylation. Nat. Rev. Mol. Cell Biol. *8*, 307–318.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A.G., Rechavi, G., and Cedar, H. (2007). Role of DNA methylation in stable gene repression. J. Biol. Chem. *282*, 12194–12200.

Lee, J.H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J. Biol. Chem. *280*, 41725–41731.

Levine, A., Yeivin, A., Ben-Asher, E., Aloni, Y., and Razin, A. (1993). Histone H1-mediated inhibition of transcription initiation of methylated templates in vitro. J. Biol. Chem. *268*, 21754–21759.

Li, B., Jackson, J., Simon, M.D., Fleharty, B., Gogol, M., Seidel, C., Workman, J.L., and Shilatifard, A. (2009). Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. J. Biol. Chem. *284*, 7970–7976.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol. Cell *12*, 1577–1589.

Pradhan, M., Esteve, P.O., Chin, H.G., Samaranayke, M., Kim, G.D., and Pradhan, S. (2008). CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry *47*, 10000–10009.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114–128.

Robin, P., Fritsch, L., Philipot, O., Svinarchuk, F., and Ait-Si-Ali, S. (2007). Post-translational modifications of histones H3 and H4 associated with the histone methyltransferases Suv39h1 and G9a. Genome Biol. *8*, R270.

Salser, W. (1977). Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harb. Symp. Quant. Biol. *42*, 985.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol. Cell. Biol. *22*, 1298–1306.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc. Natl. Acad. Sci. USA *99*, 3740–3745.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909–920.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811–816.

Tykocinski, M.L., and Max, E.E. (1984). CG dinucleotide clusters in MHC genes and in 5′ demethylated genes. Nucleic Acids Res. *12*, 4385–4396.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol. Cell. Biol. *20*, 2108–2121.

Youdell, M.L., Kizer, K.O., Kisseleva-Romanova, E., Fuchs, S.M., Duro, E., Strahl, B.D., and Mellor, J. (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. Mol. Cell. Biol. *28*, 4915–4926.

MCB
Journals.ASM.org

Recognition of CpG Island Chromatin by KDM2A Requires Direct and Specific Interaction with Linker DNA

Jin C. Zhou, Neil P. Blackledge, Anca M. Farcas, and Robert J. Klose

Department of Biochemistry, Oxford University, Oxford, United Kingdom

Up to 70% of human genes are associated with regions of nonmethylated DNA called CpG islands (S. Saxonov, P. Berg, and D. L. Brutlag, Proc. Natl. Acad. Sci. U. S. A. 103:1412–1417, 2006). Usually associated with the 5′ end of genes, CpG islands are thought to impact gene expression. We previously demonstrated that the histone demethylase KDM2A is specifically recruited to CpG islands to define a unique chromatin architecture and highlight gene regulatory regions in large and complex mammalian genomes. This targeting relies on a zinc finger CXXC DNA binding domain (ZF-CXXC), but how this demethylase interfaces with CpG island chromatin *in vivo* remains unknown. Here we demonstrate, using defined chromatin templates *in vitro* and chromatin profiling *in vivo*, that nucleosomes are a major barrier to KDM2A binding and that CpG islands are directly interpreted by the ZF-CXXC domain through specific interaction with linker DNA. Furthermore, KDM2A appears to be constrained to CpG islands not only by their nonmethylated state but also by a combination of methylated DNA and nucleosome occlusion elsewhere in the genome. Our observations suggest that both DNA sequence and chromatin structure are defining factors in interpreting CpG island chromatin and translation of the CpG signal. More generally, these features of CpG island recognition suggest that chromatin structure and accessibility play a major role in defining how transcription factors recognize DNA and regulatory elements genome-wide.

Genomic DNA information in eukaryotes is organized within the nucleus by packaging into repeated units called nucleosomes that consist of roughly 147 bp of DNA wrapped around an octamer of histone proteins. Packaging of DNA into nucleosomes permits compaction of the genome into the relatively small confines of the nucleus and can also significantly impact the function of DNA sequences with which it associates. For example, specific positioning of nucleosomes over gene regulatory elements can inhibit initiation of transcription (25–27) and nucleosomes in transcribed regions of genes can act as a barrier to transcriptional elongation (14, 15, 31, 36). More recently it has become clear that modification of both DNA and histones can provide an additional layer of information that is specifically interpreted and alters the function of surrounding regions of the genome (21, 43). In the case of histone proteins, marking of N-terminal tails by a diverse complement of posttranslational modifications can lead to direct effects on chromatin packing (38) and specific nucleation of effector proteins that translate these modification signals into a functional outcome (1, 8, 22, 43). The DNA component of the nucleosome can also be modified on the five position of the cytosine base by enzymatic addition of a methyl group which occurs mostly within the context of CpG dinucleotides (20). This methylated dinucleotide signal acts as a binding site for proteins containing a methyl-CpG binding domain (MBD) to create transcriptionally repressive chromatin states. Together these chromatin modification systems appear to have important roles in regulating the function of transcription and other nuclear processes.

Since the majority of CpG dinucleotides in the mammalian genome are methylated, a significant effort has been placed on understanding how the DNA methylation signal impacts genome function. Interestingly, there are short contiguous regions of the genome that remain free of DNA methylation and are characterized by increased CpG content and GC percentage (2). These elements are called CpG islands and are found associated with up to 70% of genes (37), but how they remain free of DNA methylation and function remain poorly understood (3). Based on their association with gene promoter elements, it has been hypothesized that CpG islands contribute to gene regulation, but how this is achieved mechanistically remains enigmatic. Recently we and others have shown that the zinc Finger CXXC DNA binding domain (ZF-CXXC)-containing proteins can specifically recognize nonmethylated CpG islands, where they recruit chromatin-modifying enzymes to create a CpG island-specific chromatin modification profile (4, 45). This targeting mechanism appears to rely on both specific recognition of nonmethylated CpG dinucleotides by the ZF-CXXC domain and inhibition of binding to other CpG dinucleotides outside CpG islands by DNA methylation. This simple yet elegant system relies on hardwired DNA sequence and epigenetic DNA methylation profiles to create unique chromatin architecture at CpG islands in contributing to gene regulatory potential.

The majority of our understanding of CpG island recognition by ZF-CXXC proteins comes from *in vitro* studies on naked DNA and descriptive work from genome wide chromatin immunoprecipitation sequencing (ChIP-seq) analysis (4, 45). Given the intimate relationship between CpG island function and chromatin, we sought to investigate how ZF-CXXC proteins interpret CpG islands in a chromatin context. To this end we have taken advantage of a completely recombinant chromatin reconstitution sys-

Received 26 September 2011 **Returned for modification** 21 October 2011
Accepted 4 November 2011
Published ahead of print 14 November 2011
Address correspondence to Robert J. Klose, rob.klose@bioch.ox.ac.uk.
J. C. Zhou and N. P. Blackledge contributed equally to this work.
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/MCB.06332-11
The authors have paid a fee to allow immediate free access to this article.

Reprinted from Molecular and Cellular Biology, 32, Zhou, J.C., Blackledge, N.P., Farcas, A.M., and Klose, R.J., Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA, 479-489, Copyright (2012), with permission from American Society for Microbiology.

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

tem to define how the ZF-CXXC protein KDM2A interacts with CpG island chromatin. Surprisingly, KDM2A has a very specific requirement for recognition of nucleosome-free linker DNA *in vitro*, and this binding mode is required for specific nucleation of KDM2A at CpG islands *in vivo*. Furthermore, we demonstrate that KDM2A binding is seemingly restricted to CpG islands even in the absence of genome methylation, suggesting that both DNA methylation and nucleosome occupancy function to restrict the promoter CpG island recognition system to these specific regions of the genome.

MATERIALS AND METHODS

DNA constructs. The vector used for prokaryotic expression of His-KDM2A-StrepII (AA 1-747) and mutant derivatives were generated as previously described (4). A mutation in the plant homeodomain (PHD domain) was introduced using the QuikChange mutagenesis XL kit (Stratagene).The pBlu2SKP 2×601 +48 plasmid (a gift from Tom Owen-Hughes) was amplified by mismatch PCR to mutate all CpG dinucleotides in the 48-bp linker region (pBlu2SKP 2×601 +48 0CG L).

Protein expression and purification. Expression and purification of KDM2A was performed as previously described (4). In brief, cells were lysed by sonication and the lysate cleared by centrifugation. The resulting lysate was loaded onto a prewashed Ni-NTA (nitrilotriacetic acid) column (GE Healthcare) and the eluate was further purified using a prewashed StrepTactin Superflow column (IBA). Purified protein was dialyzed overnight in BC100 (50 mM HEPES, pH 7.9, 100 mM KCl, 10% glycerol, 0.5 mM dithiothreitol [DTT]) and stored at −80°C until use.

Reconstitution and purification of nucleosomes. Reconstitution of nucleosomes was performed following a previously described protocol (9). In brief, recombinant *Xenopus* histones were expressed in *Escherichia coli* and purified via Sephacryl S-200 gel filtration (GE Healthcare). Stoichiometric amounts of each core histone were incubated together under high-salt conditions, and the resulting histone octamer was purified using a Superdex 200 gel filtration column (GE Healthcare). DNAs (147 bp and 216 bp) carrying high-affinity nucleosome positioning sequences (601 sequence) were PCR amplified from the pGEM-3Z 601 plasmid (a gift from Tom Owen-Hughes and Jonathan Widom). The 378-bp DNA carrying two 601 sequences and a 48-bp linker region was PCR amplified from the pBlu2SKP 2×601 +48 plasmid. All PCR products were then purified using a Resource Q anion-exchange column (GE Healthcare). Purified PCR DNA was then end labeled with [γ-^{32}P]ATP (Perkin Elmer) using T4 polynucleotide kinase (Fermentas). Unincorporated label was removed using a nucleotide removal kit (Qiagen). Equimolar ratios of purified labeled DNA and octamers were mixed together in 2 M NaCl and diluted stepwise with 10 mM Tris-HCl, pH 7.6, to reach a final concentration of 100 mM NaCl. The reconstituted nucleosome was then purified on a 5 to 20% sucrose gradient, and the resulting fractions were analyzed on a 0.8% agarose gel. Fractions containing only reconstituted mono- or dinucleosomes were then pooled together and used for the electrophoretic mobility shift assay (EMSA).

EMSA. EMSA reactions were assembled as previously described (4). Briefly, purified, labeled nucleosome substrates were incubated with the KDM2A protein in the presence of poly(dA-dT) competitor DNA for 20 min at room temperature and then loaded onto a 0.8% agarose gel in 0.2× Tris-borate (TB). The gel was run at 4°C, dried onto a DE81 anion exchanger paper (Whatman), and exposed to a phosphorimager screen.

MNase chromatin immunoprecipitation. For chromatin preparation, V6.5 mouse embryonic stem (ES) cells were fixed for 10 min in 0.37% formaldehyde and then quenched by the addition of glycine to a final concentration of 125 mM. Cells were then resuspended in 1 ml RSB buffer (10 mM Tris-HCl, pH 8, 10 mM NaCl, 3 mM $MgCl_2$) before being lysed by the addition of RSB buffer supplemented with 0.1% NP-40. Nuclei were then collected by centrifugation at 1,500 × *g* for 5 min, resuspended in RSB supplemented with 10 mM $CaCl_2$ and complete protease inhibitors (Roche) at a density of approximately 5×10^7 cells per ml, and divided into 1-ml aliquots. To each 1-ml aliquot, 200 U of micrococcal-nuclease (MNase; Worthington) was added, and the reaction mixtures were then incubated at 37°C for 2.5 min (polynucleosome chromatin preparations) or 1 h (mononucleosome chromatin preparations). The reactions were stopped by adding 20 mM EDTA, and 0.5% SDS was added to ensure complete lysis of nuclei. Chromatin immunoprecipitations were performed overnight at 4°C with approximately 3 μg of antibody and 100 μl of chromatin (corresponding to 5×10^6 cells) diluted in 900 μl of ChIP dilution buffer (1% Triton X-100, 1 mM EDTA, 20 mM Tris-HCl, pH 8, and 150 mM NaCl). Antibodies used for ChIP were anti-KDM2A (4) and anti-histone H3K4me3 (histone trimethylated at lysine 4; Abcam ab8580). Antibody-bound proteins were isolated on protein A agarose beads (Repligen), washed extensively, and eluted, and cross-links were reversed according to the Upstate protocol. Samples were then sequentially treated with RNase and proteinase K before being purified on a PureLink PCR microcolumn (Invitrogen). Real-time quantitative PCR (qPCR) was performed using Sybr green (Quantace) on a Rotor-Gene 6000 (Corbett). Primer sets used for qPCR are available upon request.

Sucrose gradient separation of native nucleosome preparations. Native nucleosome preparations and sucrose gradients were performed as previously described (16), with minor modifications. Briefly, approximately 1×10^8 purified nuclei from V6.5 cells were resuspended in 1 ml RSB buffer supplemented with 0.25 M sucrose, 3 mM $CaCl_2$, and complete protease inhibitors. Digestions were performed with 10 units (polynucleosome preparation) or 500 units (mononucleosome preparation) of MNase for 15 min at 37°C, and reactions were then stopped by adding EDTA and EGTA, each to a final concentration of 3.5 mM. After centrifugation at 5,000 rpm for 5 min, the supernatant (S1) was collected. The remaining nuclear pellet was resuspended in 300 μl of nucleosome release buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 0.5 mM EDTA, and protease inhibitor cocktail) and gently mixed at 4°C for 1 h. Following centrifugation at 5,000 rpm for 5 min, the second supernatant (S2) was collected. Nucleosome preparations comprising 700 μl S1 and 300 μl S2 were then fractionated through an 11-ml sucrose density gradient (5%-to-25% sucrose, 10 mM Tris-HCl, pH 7.4, 0.25 mM EDTA, 300 mM NaCl) at 30,000 rpm for 16 h at 4°C. Each 500-μl fraction had 50 μl removed for DNA purification and agarose gel analysis, and the remainder of each fraction was concentrated by trichloroacetic acid precipitation and used for protein analysis by Western blotting. Antibodies used for Western blotting were anti-KDM2A (4), anti-histone H3 (Ab1791), anti-histone H1 (Millipore AE-4), and anti-DNMT3B (Enzo Life Sciences 52A1018).

RESULTS

KDM2A binds extranucleosomal DNA. We recently demonstrated that recombinant KDM2A binds to nonmethylated CpG dinucleotides via its ZF-CXXC DNA binding domain (4). The capacity of KDM2A to recognize the nonmethylated fraction of the genome results in nucleation of KDM2A at CpG island elements *in vivo*, where it demethylates histone H3 lysine 36 (H3K36). These experiments demonstrated that KDM2A could bind nonmethylated CpG on naked DNA templates, but given that KDM2A is a chromatin-modifying enzyme, we were interested to investigate how KDM2A interrogates more physiological nucleosome substrates and understand if DNA alone is sufficient for KDM2A nucleation on chromatin.

To address these questions, we first used a fully recombinant nucleosome reconstitution system based on the 601 nucleosome positioning sequence (28). We chose to use the 601 nucleosome positioning sequence, as it will house a histone octamer at a well-defined position on the DNA template and recapitulates both the CpG content and GC nucleotide composition characteristics of mammalian CpG islands. Histones were assembled on a 147-bp 601 sequence by salt dilution, and the reconstituted nucleosomes

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

were purified by density gradient centrifugation, yielding a highly pure mononucleosome fraction (data not shown). KDM2A was then used in an electromobility shift assay (EMSA) comparing binding to the naked 147-bp 601 sequence or the 147-bp 601 mononucleosome. As observed previously, KDM2A bound robustly to the naked CpG-containing DNA sequence (Fig. 1A, left panel). However, addition of KDM2A did not lead to a mobility shift of the mononucleosome (Fig. 1A, right panel). This interesting observation suggests that deposition of histones on DNA may occlude KDM2A binding to the CpG dinucleotides. The lack of binding to the 147-bp nucleosome suggested that perhaps KDM2A may require extranucleosomal DNA for efficient binding to chromatin. To test this possibility, we generated a second 216-bp 601 mononucleosome that has an additional 60 bp of DNA that extend out from the nucleosome positioning sequence (Fig. 1B and C). KDM2A was incubated with the 216-bp nucleosome and analyzed by EMSA. The addition of extranucleosomal DNA to the mononucleosome resulted in a dose-dependent shift of the nucleosome substrate by KDM2A. Importantly, this interaction was dependent on DNA binding, as a point mutation in the ZF-CXXC domain abrogates interaction with the 216-bp nucleosome.

We have previously demonstrated that methylation of CpG dinucleotides abrogates KDM2A binding. To determine whether KDM2A binding to the 216-bp mononucleosome was dependent on nonmethylated CpG DNA, we reconstituted nucleosomes where all CpG dinucleotides were *in vitro* methylated prior to reconstitution (Fig. 1B and C). KDM2A binding to the methylated substrate was completely abrogated, as analyzed by EMSA, indicating that KDM2A binding to the 216-bp nucleosome is dependent on nonmethylated CpG. The main difference between the 147- and 216-bp nucleosomes is the presence of extranucleosomal DNA. This suggests that KDM2A likely recognizes nucleosome-free DNA in the context of a mononucleosome. To investigate this possibility, a 216-bp nucleosome was reconstituted in which only the CpG dinucleotides in the extranucleosomal DNA were mutated (Fig. 1B and C). Despite the fact that the nucleosome-occupied portion of the 216-bp nucleosome has an abundance of nonmethylated CpG dinucleotides, specific mutation of CpGs in the extranucleosomal DNA was sufficient to abrogate binding. Together, these observations demonstrate that KDM2A specifically interrogates extranucleosomal DNA in the 216-bp nucleosome.

KDM2A binding is restricted to linker DNA. KDM2A will never encounter an isolated mononucleosome *in vivo*, but instead extranucleosomal DNA will largely occur between nucleosomes on the chromatin polymer. Therefore, to examine whether KDM2A can bind to linker DNA between nucleosomes, we reconstituted dinucleosomes which consist of two 601 nucleosome positioning sequences separated by 48 bp of linker DNA (10). The only extranucleosomal DNA in this substrate exists between the positioned nucleosomes. KDM2A was then presented with this substrate in an EMSA (Fig. 1D). Much like the 216-bp nucleosome, KDM2A bound efficiently to the dinucleosome substrate. The retardation of the KDM2A/dinucleosome complex (Fig. 1D) compared to the KDM2A/mononucleosome complex (Fig. 1C) is lesser in magnitude, as the relative increase in mass due to KDM2A binding to the KDM2A/dinucleosome complex is smaller. To ensure that this modest mobility shift was dependent on KDM2A, we used a DNA binding mutant that abrogated the

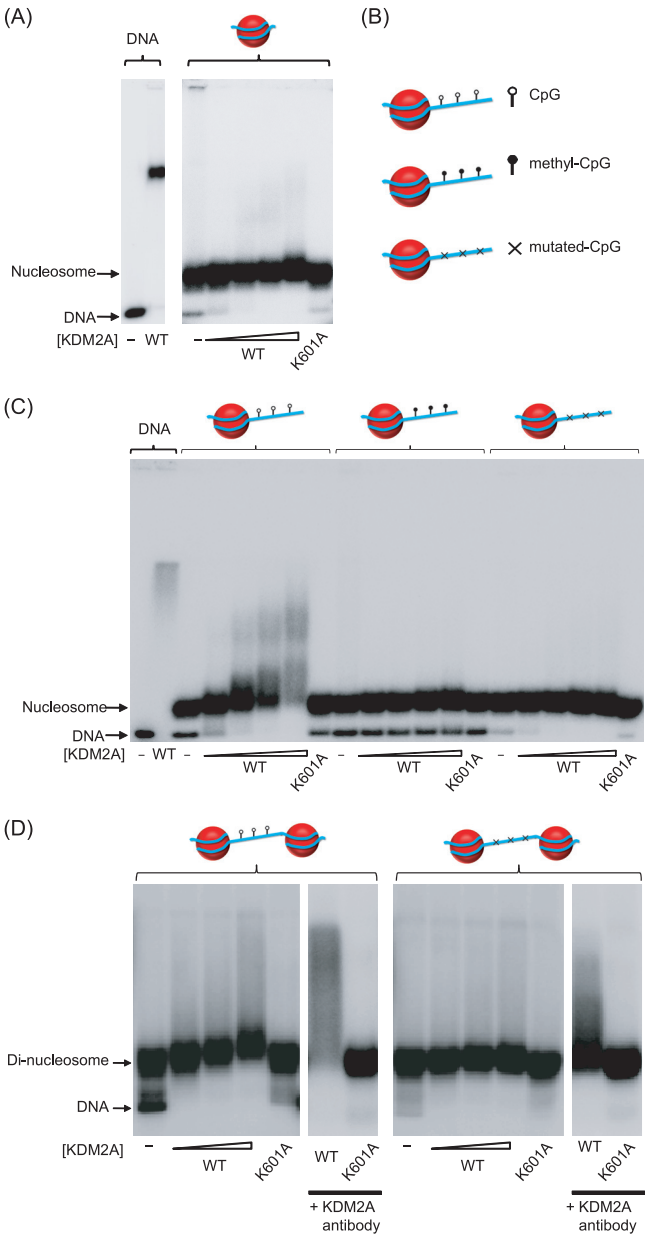


FIG 1 KDM2A interacts specifically with linker DNA *in vitro*. (A) KDM2A efficiently binds to the naked 147-bp 601 nucleosome positioning DNA (left-hand panel) but binding is completely inhibited when a nucleosome is positioned on this same DNA sequence (right-hand panel), as analyzed by EMSA. The K601A mutant version of KDM2A lacks DNA binding activity. The positions of free DNA and nucleosomal DNA are indicated to the left of the panel with arrows. (B) Nucleosomes were reconstituted using a DNA substrate with 216 bp of DNA. The cartoon indicates nucleosomes in which the extranucleosomal DNA has CpG dinucleotides (top, open lollypop), methylated CpG dinucleotides (middle, closed lollypop), or mutated CpG dinucleotides (bottom, with x). (C) KDM2A interacts specifically with nucleosomes that have CpG dinucleotides in the extranucleosomal DNA but not in the same fragments when methylated or mutated. (D) Dinucleosome substrates that have 48 bp of linker DNA were reconstituted with the CpG dinucleotides in the linker DNA either intact (left panels) or mutated (right panels). KDM2A bound specifically to the dinucleosome with intact linker CpG dinucleotides, as indicated by the protein concentration-dependent nucleosome mobility shift and capacity to supershift this species with a KDM2A-specific antibody (left-hand panels). Mutation of the CpG dinucleotide in the linker DNA resulted in a loss of binding (right-hand panels).

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

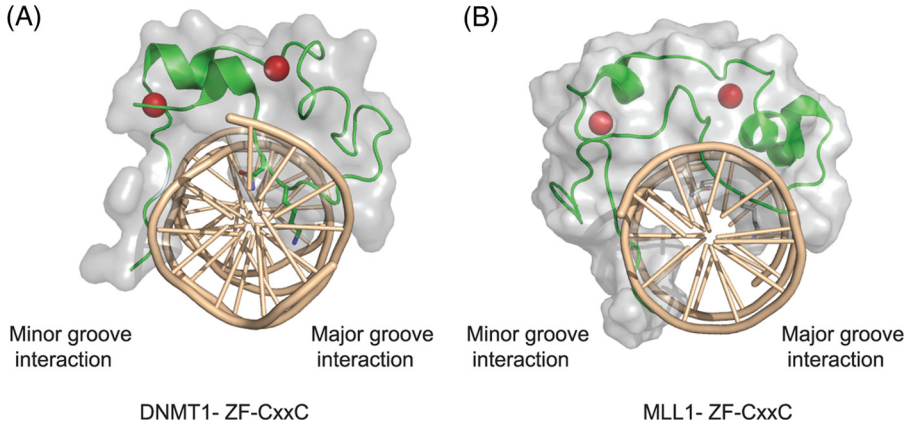


FIG 2 ZF-CXXC domain proteins bound to nonmethylated CpG DNA. (A) A cartoon depicting the structure of the DNMT1 ZF-CXXC domain bound to nonmethylated CpG DNA (Protein Data Bank [PDB] number, 3PT6). The green ribbon indicates the protein backbone and the residues that specifically interact with CpG dinucleotides in the major groove. The DNA is colored light brown and is viewed looking down the DNA double helix. The gray shading indicates the protein surface, and the red spheres correspond to the two zinc ions coordinated by the ZF-CXXC domain. (B) The same representation as in panel A of the MLL1 ZF-CXXC domain bound to nonmethylated CpG DNA (PDB number, 2KKF). From both panels it is apparent that the ZF-CXXC domain interacts specifically with CpG DNA in the major groove and interacts with minor-groove DNA on the opposite side of the DNA double helix.

mobility shift and antibodies against KDM2A to supershift the KDM2A/nucleosome complex. To verify that KDM2A interacts with the linker region, a dinucleosome was reconstituted in which the CpG dinucleotides in the linker region were mutated. Importantly, this results in a loss of nucleosome binding (Fig. 1D, right-hand panel). The very small amount of residual binding in the dinucleosome lacking linker CpGs (Fig. 1D right hand panels) appears to result from a fraction of the reconstituted dinucleosome being imperfectly phased over 601 nucleosome positioning sequences (9), resulting in accessible CpG dinucleotides. This mode of interaction is supported by the fact that the ZF-CXXC mutant (K601A) does not cause a mobility shift of either dinucleosome substrate (Fig. 1D). Together, these observations demonstrate that KDM2A interacts in the linker region between nucleosomes.

Our *in vitro* binding studies with reconstituted nucleosomes indicate that KDM2A interacts with linker DNA. To try and understand why nucleosomes might be a barrier to interaction of the ZF-CXXC domain with DNA, the structures of two recently solved ZF-CXXC/CpG DNA structures (MLL1 and DNMT1) were analyzed (Fig. 2) (7, 41). In both cases the ZF-CXXC domain forms a crescent-shaped fold that coordinates two zinc ions. Side chain residues in a single loop at the tip of the zinc finger interrogate the major groove of DNA where they recognize CpG bases and are sterically occluded by the presence of methyl CpG. Interestingly, in both of these structures a distinct portion of the ZF-CXXC domain also wraps around the DNA and interrogates the minor groove on the opposite side of the double helix. Mutations that disrupt this interaction also affect DNA binding, suggesting that simultaneous interaction with major- and minor-groove DNA is required for efficient DNA binding (7). In the context of a nucleosome, when the major groove is facing outwards, the minor groove points inwards and is often engaged by arginine residues from the histone octamer (29). This observation likely explains the lack of KDM2A binding to nucleosomal DNA, as access to information in both the major and minor groove would be sterically constrained in nucleosomal DNA compared to naked DNA. Therefore, it seems likely that ZF-CXXC domain proteins are occluded from binding nucleosomal DNA by the inability to easily interface with both major- and minor-groove DNA.

Neither H3K4me3 nor the KDM2A PHD domain is required for 216-bp nucleosome binding. It has recently been demonstrated that the plant homeodomains (PHD domains) in many proteins can function as lysine or methylated lysine binding modules (43, 50). KDM2A has a highly conserved PHD domain immediately adjacent to the ZF-CXXC DNA binding domain. The close proximity of the PHD and DNA binding domains suggested that perhaps together these domains may form a chromatin binding interface that interrogates both DNA and histone. To test this possibility, we generated a mutation in a conserved cysteine within the KDM2A PHD domain (C645A) that has been shown in other PHD domains to disrupt the structure and function (Fig. 3A) (5, 19). Although this mutation disrupts the PHD domain, the DNA binding capacity of the mutant protein was unaffected (Fig. 3B). The C645A mutant protein was then used in an EMSA to examine the effect of this mutation on substrate binding (Fig. 3B). As with binding to naked DNA, the KDM2A PHD mutant did not affect interaction with the 216-bp nucleosome, suggesting that any potential interaction with histone by the PHD domain does not significantly contribute to binding. Our reconstitution system uses unmodified recombinant histones. Although some PHD proteins interact with unmodified lysines, others interact specifically with methylated lysine residues. We have recently shown that KDM2A binds specifically to CpG islands *in vivo*. Interestingly, the majority of CpG islands are modified on the tail of histone H3 at position 4 (H3K4me3) (45), suggesting that perhaps the KDM2A PHD domain could function in concert with ZF-CXXC-mediated recognition of CpG DNA to create a high-affinity CpG island binding module. To specifically test this possibility, we used a recently developed semisynthetic technique to install H3K4me3 into our recombinant histone H3 (Fig. 3C) (39). This histone was then incorporated into a 216-bp nucleosome to create a homogenously H3K4me3 modified nucleosome. This substrate was used in an EMSA with either the KDM2A wild-type (WT) or PHD mutant protein (Fig. 3D). In both instances KDM2A bound to the modified nucleosome with the same efficiency as the unmodified

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

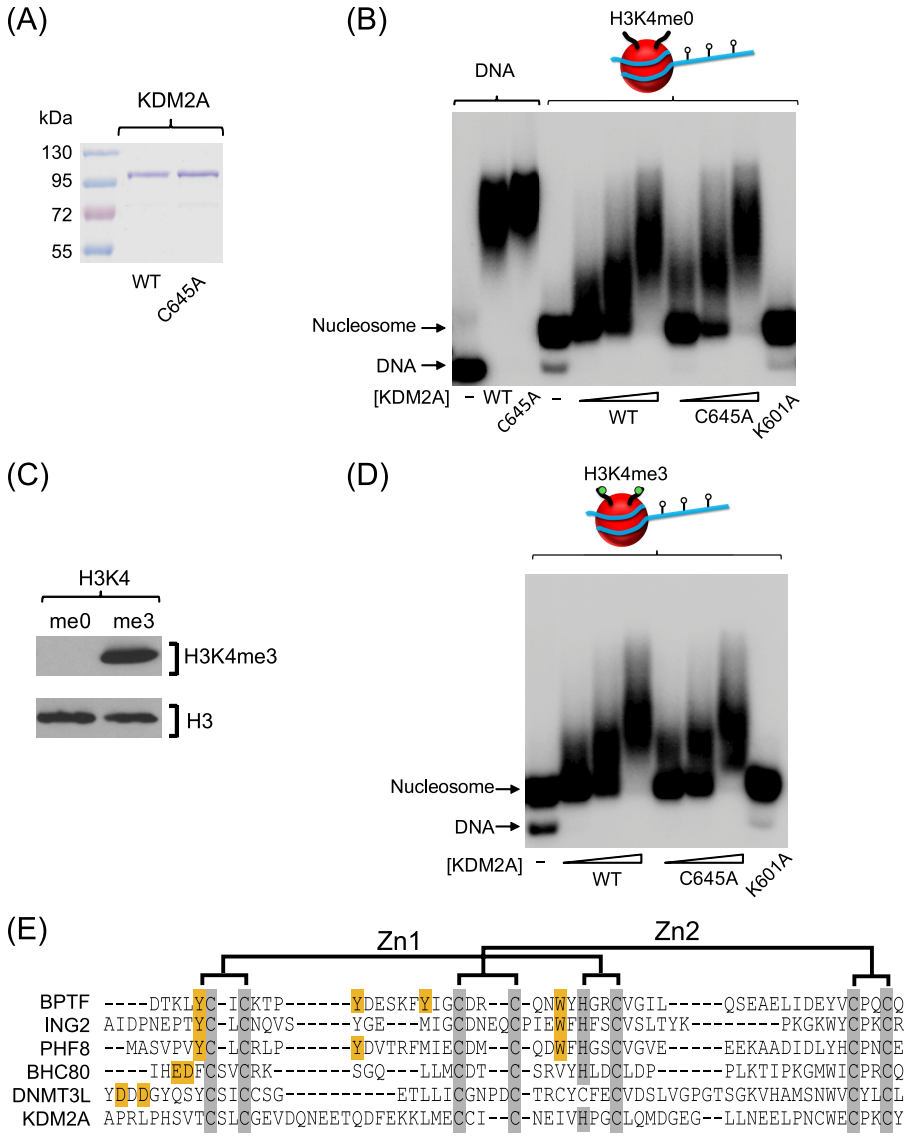


FIG 3 The KDM2A PHD domain does not contribute significantly to nucleosome binding. (A) KDM2A proteins encoding a wild-type (WT) or mutant PHD domain (C645A) were expressed and purified. (B) Both the WT and C645A KDM2A proteins bind to naked 216-bp nucleosome positioning DNA with the same level of efficiency, as determined by EMSA (left-hand lanes). Mutating the PHD domain did not inhibit binding to the 216-bp mononucleosome, whereas mutation of the ZF-CXXC DNA biding domain (K601A) completely abrogated binding. (C) Histone H3K4me3 was installed specifically into histone H3 and the incorporation verified by Western blotting of the recombinant histone with antibodies against H3K4me3 or histone H3. (D) Mononucleosomes (216 bp) were reconstituted with histones containing H3K4me3 as indicated by the green dots on the nucleosome cartoon above the EMSA panels. The addition of H3K4me3 to the mononucleosome did not increase binding of KDM2A, nor did mutation of the PHD domain inhibit binding to the nucleosome, indicating that H3K4me3 does not contribute significantly to nucleosome recognition by KDM2A. (E) Multiple sequence alignment of PHD domains that bind either to H3K4me3 (BPTF, ING2, and PHF8) or H3K4me0 (BHC80 and DNMT3L) and KDM2A. Gray boxes indicate conserved zinc-coordinating cysteines/histidines, and orange boxes indicate methyl-lysine- or lysine-interacting residues. Although KDM2A has conserved zinc-coordinating residues, lysine interaction residues are absent.

substrate and mutation of the PHD domain did not impact this binding. Therefore, nucleosome binding by KDM2A does not appear to rely on the function of the PHD domain nor on H3K4me3.

As the KDM2A PHD domain did not appear to contribute to nucleosome binding using the substrates tested, we hypothesized that the KDM2A PHD domain may interact with histone modifications on other residues of H3 or other core histones. To test this possibility, we applied the KMD2A PHD domain to a peptide array containing 600 different variations of modified and unmodified peptides corresponding to all four core histones (30). Perhaps somewhat surprisingly, in this screen we observed no specific binding of the KDM2A PHD domain to any peptide on the array, suggesting that the function of this domain is not to recognize the histone component of the nucleosome (unpublished observations). Given that the structure of several PHD domains bound to their cognate substrates has been solved, we carried out a multiple sequence alignment of the KDM2A PHD domain, comparing it to PHD domains that bind either nonmethylated or methylated lysine residues (Fig. 3E). Importantly, the structural cysteines required to form the PHD fold are conserved among all PHD do-

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

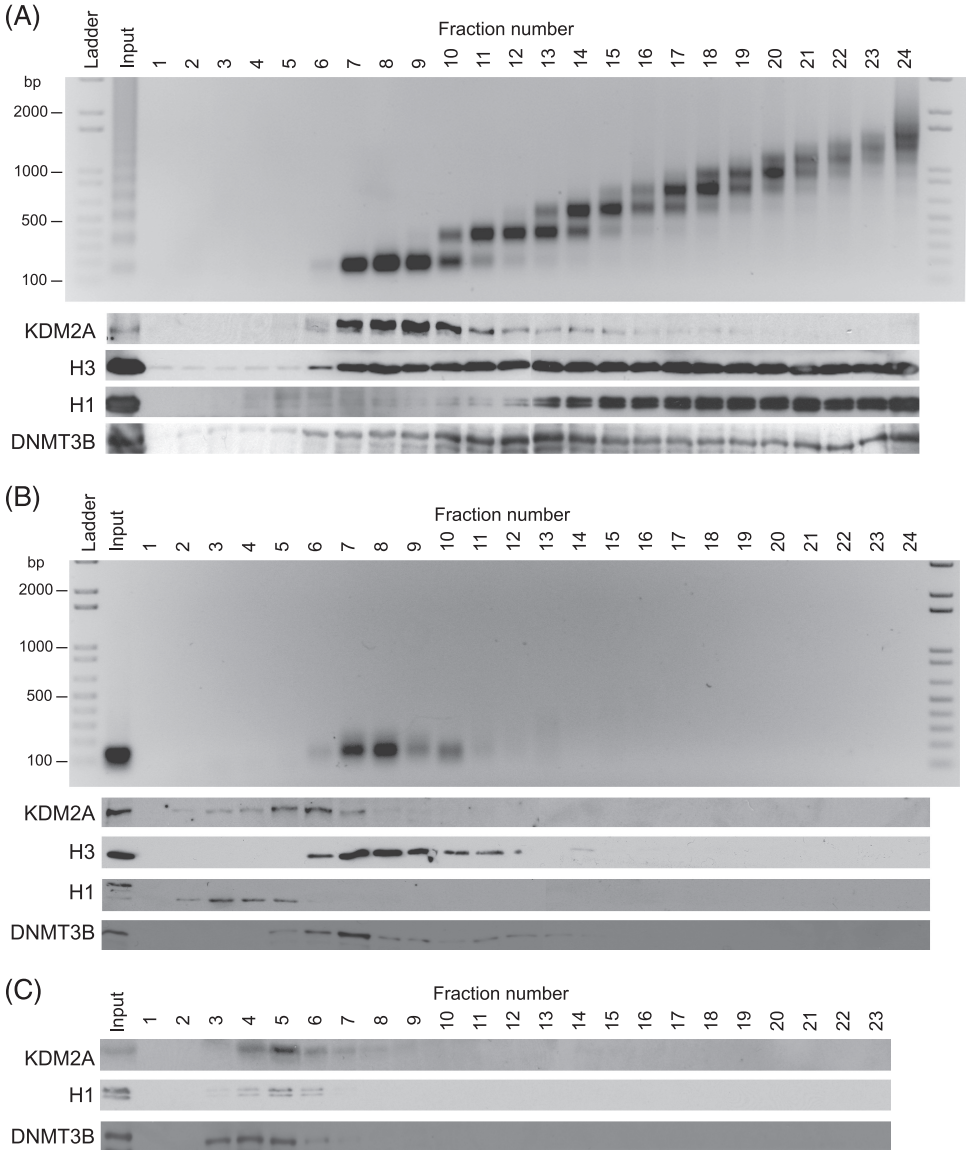


FIG 4 KDM2A interacts with linker DNA *in vivo*. (A) Nuclei were partially digested with micrococcal nuclease to produce polynucleosomes that contain linker DNA. Chromatin was separated by sucrose density gradient centrifugation. The gradient was fractionated from the top into 24 fractions. DNA was purified from each fraction and resolved by agarose gel electrophoresis (top panel), and proteins from each fraction were visualized by Western blotting (bottom panels). DNA and protein analyses include an unfractionated input sample representing 10% of material loaded onto the sucrose gradient. (B) As above, except that native chromatin was digested to completion with micrococcal nuclease to produce mononucleosomes lacking linker DNA. (C) Fractionation of chromatin-free nuclear extract.

mains in the alignment, but the specific aromatic cage that is utilized for methyl-lysine binding and the conserved residues used for lysine binding are completely absent from KDM2A. Together these data suggest that although the PHD domain is conserved across KDM2A orthologues, the function of the PHD domain in KDM2A is not to recognize histone.

KDM2A binds linker DNA *in vivo*. Our *in vitro* nucleosome binding assays indicate that KDM2A interacts with chromatin through recognition of linker DNA via the ZF-CXXC DNA binding domain. To understand if this binding mechanism is utilized *in vivo*, nuclei were isolated from mouse embryonic stem cells and digested with micrococcal nuclease. Digestion was done either under limiting reaction conditions that produced a polynucleosomal ladder of approximately 1 to 10 nucleosomes or under complete conditions that resulted in exclusively mononucleosome fragments (Fig. 4A and B) (16). If endogenous KDM2A associates with linker DNA *in vivo* it would remain associated with polynucleosomal chromatin that has intact linker regions but not with more completely digested mononucleosomal chromatin that lacks linker DNA. To examine whether this was the case, chromatin was isolated using limited digestion and fractionated by density sucrose gradient centrifugation. A nucleosome repeat length DNA ladder was observed in gradient fractions that have histone H3, indicating intact chromatin (Fig. 4A). When the same fractions were analyzed by Western blotting, KDM2A was found exclusively in the chromatin-containing fractions. This suggests

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

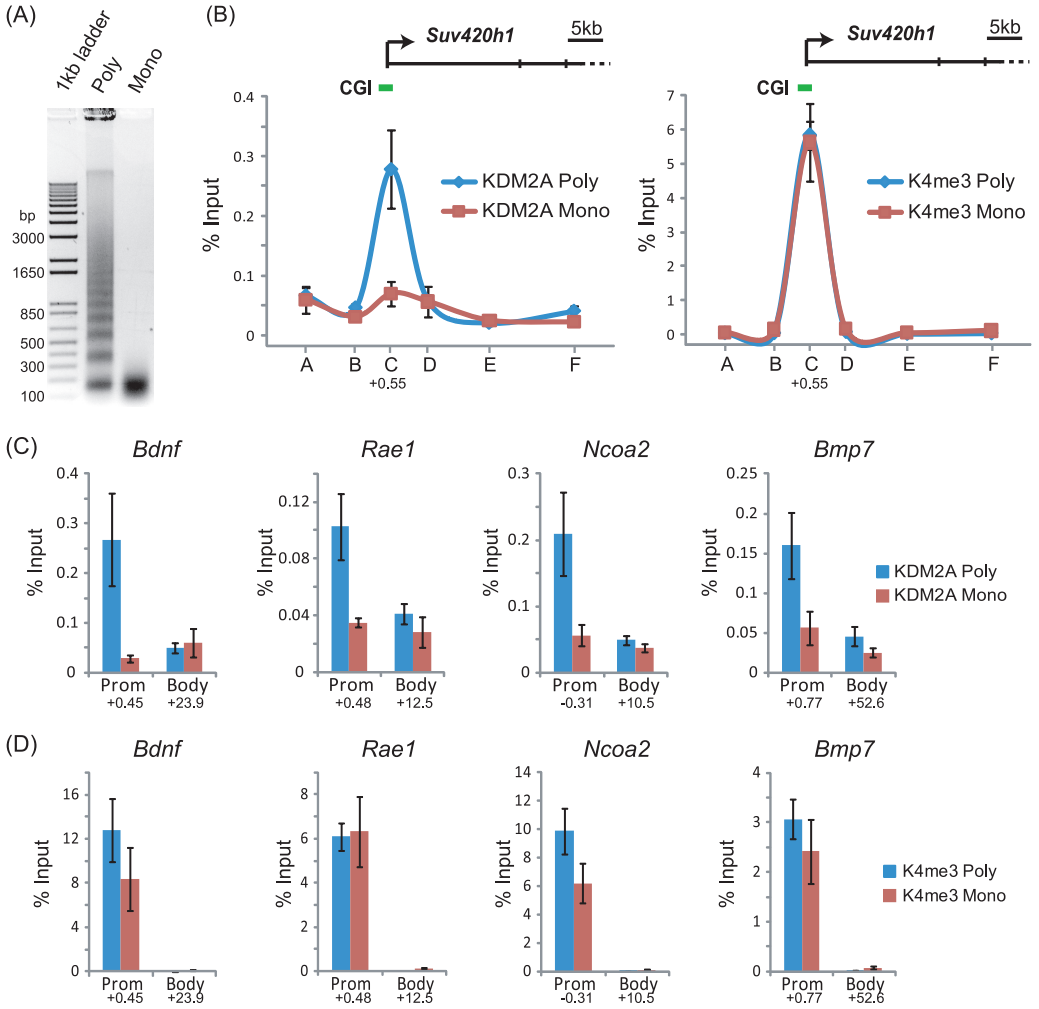


FIG 5 KDM2A is targeted to CpG islands through recognition of linker DNA. (A) Formaldehyde-fixed chromatin was digested with micrococcal nuclease to produce polynucleosomes that contain linker DNA and mononucleosomes lacking linker DNA. (B) KDM2A binding to the Suv420h1 gene was analyzed by ChIP at tiled positions across the gene (A to F on the *x* axis) by quantitative PCR on both chromatin preparations. KDM2A binding (left panel) in the polynucleosomal fraction (blue line) corresponded to the CpG island region, and this signal was lost in the mononucleosomal preparation (red line). In contrast, the signal for H3K4me3 over the same regions remained constant for both samples (right panel). (C and D) The same analysis was extended to a series of four more CpG island genes, analyzing both the CpG island region and body region. In all instances, digestion to mononucleosomes released KDM2A from CpG island chromatin (C), while the H3K4me3 signal remained relatively constant (D). Together, these observations indicate that KDM2A interacts with linker DNA at CpG islands. In all cases, error bars correspond to standard errors of the means from biological triplicates. Numbers below the data for promoter (Prom) and body primer sets indicate the position of the primer set (kb from center of amplicon) with respect to the major transcription start site of the gene in question.

KDM2A is intimately associated with chromatin that contains linker DNA. Interestingly, KDM2A was enriched in the mono-trinucleosome fractions, suggesting that it may predominantly associate with nuclease-sensitive chromatin, a feature that is characteristic of CpG island elements (35). In agreement with previous observations, histone H1 is preferentially enriched in the oligonucleosomal fractions, suggesting that these regions of the gradient contain the rapidly sedimenting, compact and more nuclease-resistant chromatin (23). In contrast, when chromatin was digested to a homogenous mononucleosome population that lacks linker DNA and resolved on the same density gradient, a single mononucleosome DNA band was observed that cofractionated with histone H3, again indicating intact chromatin (Fig. 4B). Importantly, at this point, when KDM2A was analyzed by Western blotting, it was no longer associated with the chromatin-containing fraction but instead migrated in the same position on the gradient as is observed when chromatin-free nuclear extracts are fractionated in a similar manner (Fig. 4C). This indicates that KDM2A chromatin binding is abrogated when linker DNA is removed. In support of the contention that this binding is dependent on linker DNA, these observations mirrored the behavior of histone H1, which also interacts with linker DNA (Fig. 4A to C). The loss of KDM2A from the mononucleosomal fraction appears to be due to the loss of linker DNA as DNMT3B, a protein previously shown to interact with nucleosomes (16), remains associated with the mononucleosomal fraction (Fig. 4C). Together these observations demonstrate that KDM2A in bulk chromatin relies on linker DNA to bind chromatin.

KDM2A binds linker DNA in CpG islands. KDM2A binding to bulk chromatin relies on intact linker DNA. Our recent work demonstrated that KDM2A binds to CpG islands *in vivo* (4). Therefore, we wanted to know whether targeted interaction of

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

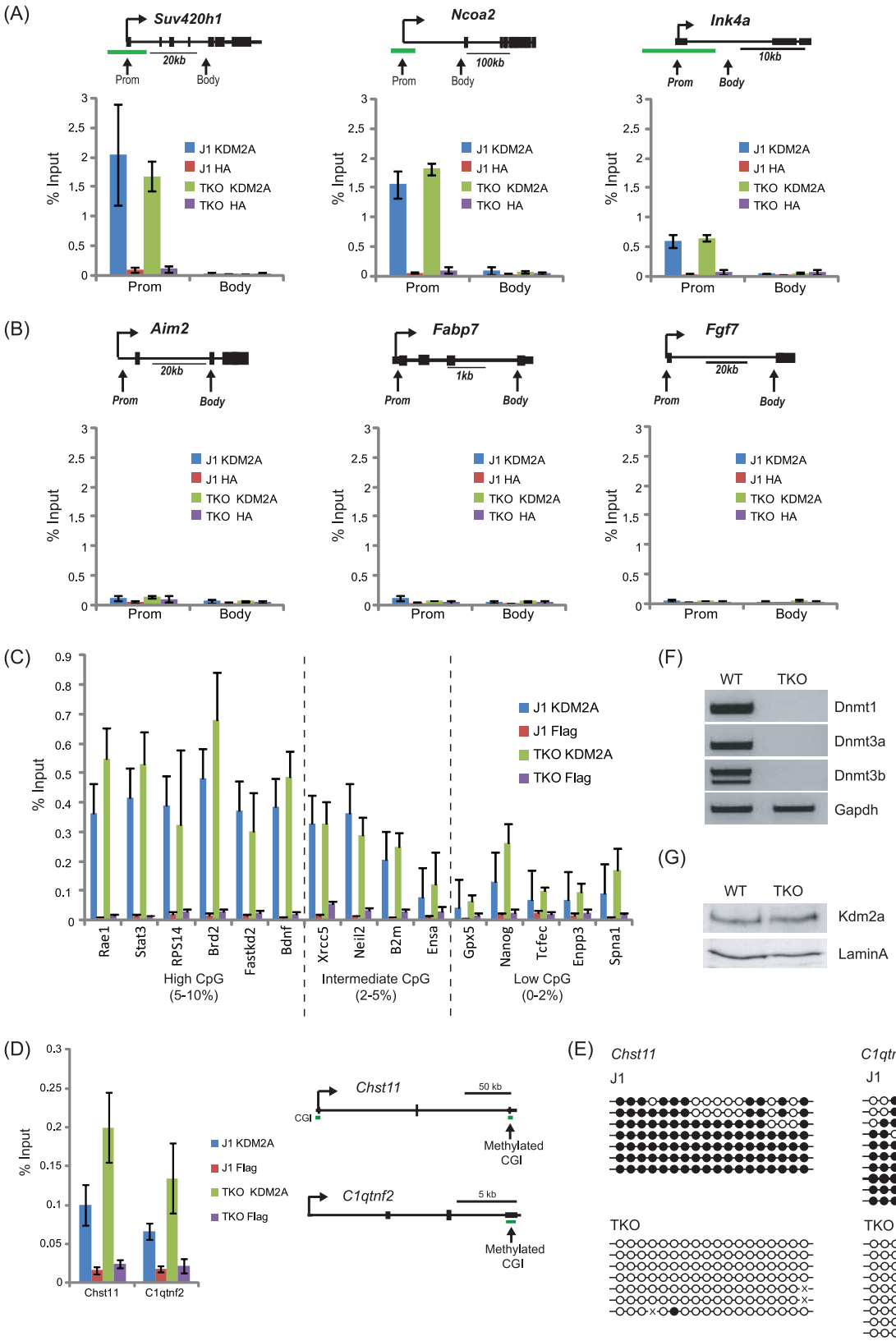


FIG 6 KDM2A remains bound specifically to CpG islands in cells lacking DNA methylation. TKO mouse embryonic stem cells that lack all three known DNA methyltransferases (F) and have no detectable DNA methylation were obtained. ChIP was used to analyze KDM2A binding to three CpG island genes (A) and three non-CpG island genes (B) in both the WT parent line and the TKO cell line. In the three CpG island regions, KDM2A binding remained the same, indicating that loss of methylation elsewhere in the genome did not cause a loss of binding. In addition, the body of the CpG island genes (A) and non-CpG island genes (B)

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

KDM2A with promoter-associated CpG islands relies on linker DNA. To address this possibility, a modified chromatin immunoprecipitation (ChIP) approach was designed whereby nonhistone factors were mildly fixed to chromatin using formaldehyde and then nuclei were isolated and digested under limiting or complete conditions using micrococcal nuclease (Fig. 5A). This resulted in chromatin that either contains linker DNA or lacks linker DNA. KDM2A was then immunoprecipitated, and the isolated DNA was analyzed by quantitative real-time PCR. Initially amplicons spanning the CpG island containing the Suv420H1 gene were analyzed in either the limited or complete digestions (Fig. 5B). The KDM2A signal in the limited digestion revealed a normal KDM2A ChIP signal that peaked over the CpG island and was absent from the rest of the gene. However, when the same regions were analyzed in the complete digestions, the KDM2A ChIP signal was completely lost (Fig. 5B, left panel). This loss of signal is not due to preferential overdigestion of CpG island chromatin in these complete digestions, as the amount of H3K4me3 ChIP signal at the CpG island regions was the same in both chromatin preparations (Fig. 5B, right panel). A further four CpG island genes were analyzed at both the CpG island and body for both KDM2A and H3K4me3 (Fig. 4C and D). In all instances, removal of linker DNA abrogated KDM2A binding but left CpG island chromatin intact, as verified by H3K4me3 ChIP. Together these observations are in keeping with our *in vitro* observations and demonstrate that KDM2A binds to CpG island chromatin through recognition of DNA between nucleosomes that encompasses linker DNA or more extended nucleosome-free regions.

KDM2A is retained at CpG islands in the absence of DNA methylation. We have previously shown that KDM2A binds to nonmethylated CpG islands and that the acquisition of DNA methylation in a small fraction of CpG islands during differentiation is sufficient to block DNA binding (4). CpG islands contain a much higher density of CpGs than the rest of the genome, and outside CpG islands, the remaining CpG dinucleotides are heavily methylated. Presumably these two features of the mammalian genome lead to efficient nucleation of KDM2A to CpG islands. To understand if DNA methylation in bulk chromatin is a determining factor for excluding KDM2A from these regions of the genome, mouse triple knockout (TKO) embryonic stem cells that lack both *de novo* methyltransferases DNMT3A and -B and the maintenance methyltransferase DNMT1 were obtained (Fig. 6F) (48). These cells do not have any detectable CpG methylation. To examine whether the presence of nonmethylated DNA outside CpG islands in TKO cells alters KDM2A binding profiles, ChIP analysis was carried out on a panel of genes with CpG island promoters and genes with non-CpG island promoters, analyzing both promoter and body regions in wild-type (WT) and TKO cell lines (Fig. 6A and B). Interestingly, a loss of DNA methylation had no impact on the normal nucleation of KDM2A at the CpG island promoters examined (Fig. 6A) and did not create new binding sites at several non-CpG island promoters (Fig. 6B). Furthermore, the presence of nonmethylated DNA in the body of both classes of genes did not result in *de novo* binding of KDM2A over these regions of the genes tested. Importantly, KDM2A profiles were also maintained in the TKO cells when promoter regions of various CpG content were analyzed (Fig. 6C). However, loss of DNA methylation at some weak CpG islands that are subject to DNA hypermethylation in normal ES cells did lead to a modest renucleation of KDM2A in the TKO cells (Fig. 6D and E). This is in agreement with previous observations that CpG island methylation is a barrier to KDM2A binding (4). Importantly, the maintenance of normal KDM2A binding profiles at the genes tested in the TKO cells was not simply due to increased KDM2A expression, as the levels of KDM2A were the same in the parental WT and TKO cell lines (Fig. 6G). From the regions analyzed here, it appears that in situations where a CpG island acquires methylation, KDM2A binding is inhibited (4) (Fig. 6D), but loss of DNA methylation elsewhere in the genome does not significantly impact KDM2A binding (Fig. 6A, B, and C). Based on the observation that KMD2A is incapable of binding to nucleosome-bound DNA, it appears that low CpG dinucleotide frequency and occlusion of CpG binding sites by nucleosomes is a major barrier to KDM2A binding outside CpG islands. Therefore, KDM2A recognition of CpG island chromatin appears to rely on both nonmethylated CpG and accessibility of this binding site in extranucleosomal DNA.

DISCUSSION

Despite over 25 years of active research on CpG islands, the mechanisms underpinning the function of these interesting elements remain poorly understood. The recent discovery that ZF-CXXC domain-containing proteins act as interpreters that translate the CpG island signal into a unique chromatin architecture has revealed that CpG islands are not simply passive genomic elements but that they function proactively in marking regulatory element chromatin (4, 45). Based on these observations, there appears to be an intimate relationship between CpG island chromatin structure and function.

To better understand this relationship, we have taken advantage of a chromatin reconstitution system to define how KDM2A binding to CpG islands is achieved. Importantly, this work reveals that nucleosomal DNA is a major barrier to chromatin binding *in vitro* and that the ZF-CXXC domain specifically recognizes DNA between nucleosomes, including linker DNA or nucleosome-free regions. Despite the fact that KDM2A encodes a PHD domain, it does not appear to use this domain in recognition of chromatin. *In vivo*, recognition of CpG island promoters by KDM2A is achieved through interaction with linker DNA, revealing that a combination of nonmethylated CpG dinucleotide content and linker DNA accessibility may be utilized by ZF-CXXC proteins to recognize the CpG island signal and translate this into chromatin modification. As most ZF-CXXC domain-containing proteins function to modify chromatin, this would suggest that there is an intimate and

(both promoter and body) did not acquire KDM2A binding. (C) This analysis was then expanded to analyze a series of gene promoters that have high, intermediate, and low CpG content to demonstrate that KDM2A nucleation is largely unaffected by a loss of DNA methylation. (D) A similar ChIP analysis was performed at two CpG islands that are hypermethylated in wild-type J1 cells. Loss of DNA methylation in the TKO cells caused a modest increase in KDM2A nucleation at these weak CpG islands. (E) Bisulfite sequencing demonstrated that the Chst11 and C1qtnf2 CpG islands are hypermethylated in J1 but not TKO mouse ES cells. (G) Importantly, KDM2A protein levels remained the same in the WT and TKO cell lines. Error bars in all ChIP experiment data correspond to standard errors of the means from biological triplicates.

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

potentially self-reinforcing cycle that defines CpG island functionality through chromatin modification and accessibility. For example, genome-wide studies analyzing nucleosome occupancy of transcription start sites suggest that CpG island-containing genes have high occupancy and more regularly phased nucleosomes (47) than non-CpG island genes and other regions of the genome. It has been suggested that this could arise in part through the underlying GC-rich sequence (46, 47). Together, elevated nonmethylated CpG content and regular nucleosome arrangement may create a situation at CpG islands where linker DNA is largely accessible and occupied by ZF-CXXC proteins. Indeed, both ZF-CXXC-containing proteins KDM2A and CFP1 occupy over 90% of CpG islands genome-wide and appear to be excluded from only a very small subset of CpG island genes that are highly repressed (4, 40, 45). Nucleation of KDM2A at CpG islands removes H3K36me2, a modification that would normally recruit repressive histone deacetylase complexes and is inhibitory to transcriptional initiation (6, 42). In contrast, CFP1 leads to deposition of H3K4me3, a modification that correlates positively with transcriptional potential (12). Interestingly, H3K4me3 can act as a nucleation site for the NURF chromatin remodelling complex through the effector protein BPTF and the SAGA histone acetyltransferase complex through the effector protein SGF29 (49, 50). Therefore, nucleation of these activities at CpG islands following CFP1-mediated deposition of H3K4me3 could result in a reinforcement of nucleosome phasing and histone acetylation, in keeping with genome-wide studies indicating that CpG island chromatin is often more accessible and acetylated than surrounding and bulk chromatin (34, 35, 44, 51). Based on these properties, one can envisage that CpG islands may therefore function in a self-reinforcing manner whereby organized nucleosome structure and accessibility may permit access by ZF-CXXC DNA binding factors to nucleosome-free DNA, as described here, which then initiates a chromatin architecture through histone modification that allows effector protein-based reinforcement of accessibility. Importantly, this cascade could easily be reinitiated following cell division through binding of ZF-CXXC proteins to nonmethylated CpG islands in daughter cells and provide a simple mechanism by which CpG island chromatin architecture could be copied, based on a DNA template, to cellular progeny. Clearly, future studies are required to address these possibilities.

CpG islands are associated with the majority of genes, and ZF-CXXC proteins bind almost all CpG islands. This includes genes that cover the entire gene expression spectrum, including repressed, poised, and active genes. Together, these observations suggest that the presence of a CpG island does not directly impact the defined expression state of associated genes. Therefore, how do CpG islands impact gene regulatory element function and why are they so broadly associated with genes? Our work and the work of others indicate that chromatin accessibility (35, 51) and defined chromatin modification architecture (4, 45) are common features of CpG islands. Therefore, one possibility is that CpG islands use ZF-CXXC-dependent chromatin modification to set up a constitutively fluid and accessible chromatin structure that permits efficient recognition of gene regulatory elements by DNA binding transcription factors and perhaps eventually the core transcriptional machinery during gene activation. The fact that chromatin accessibility is a central feature of transcription factor nucleation is supported by recent work demonstrating that the binding of glucocorticoid receptor (GR) following hormone stimulation required not only the presence of the GR binding motif but also an accessible chromatin environment prior to stimulation for efficient GR nucleation (17). This is in agreement with other genome-wide studies analyzing transcription factor binding to chromatin *in vivo* that indicate not all theoretical binding sites are recognized equally (11, 13, 18, 24, 32). Therefore, the capacity of CpG islands to create permissive and accessible chromatin may be particularly important for highlighting gene regulatory elements in large and complex mammalian genomes. Given their association with both silent and active genes, this model places CpG islands as facilitators of gene regulatory processes as opposed to defining absolute transcriptional output. In keeping with this model, a recent study demonstrated that induction of CpG island-containing immediate early genes occurs rapidly without a requirement for the chromatin remodelling machinery (33). In contrast, induction of non-CpG island-containing immediate early genes relies on the function of the ATP-dependent chromatin remodelling factor BRG-1 (33). Therefore, it is tempting to speculate that chromatin modifications and accessibility that are intrinsic to CpG islands through the function of ZF-CXXC proteins contribute to this transcriptionally permissive and responsive state.

As we begin to uncover the factors that interpret CpG islands, the more apparent it becomes that chromatin structure and modification are important in defining CpG island function. Our observations that the CpG island signal is interpreted through recognition of linker DNA by the ZF-CXXC domain and an emerging trend from ChIP-seq studies analyzing transcription factor nucleation imply that nucleosome structure, occupancy, and modification may be important factors in gene regulatory element recognition and function *in vivo*. Therefore, a detailed mechanistic understanding of how DNA binding proteins interface with defined chromatin templates *in vitro*, and not simply their underlying naked DNA sequence motifs, in addition to genome-wide profiling endeavors will be central to understanding how selection and nucleation at target sites are achieved *in vivo*.

ACKNOWLEDGMENTS

We thank the late Jonathan Widom and Tom Owen-Hughes for the 601 nucleosome-positioning sequences; Yi Zhang, Timothy Richmond, and Matthew Simon for histone expression constructs; Sigrun Rumpel and Cheryl Arrowsmith for the histone peptide array analysis; James Parker for help with structural images; Masaki Okano for the TKO ES cells; and Neil Brockdorff for antibody reagents and valuable discussion during the course of this work.

This work is funded in the Klose laboratory by the Wellcome Trust, Cancer Research UK, EMBO, The Lister Institute of Preventive Medicine, and the Medical Research Council.

REFERENCES

1. **Bannister AJ, et al.** 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature **410**:120–124.
2. **Bird A, Taggart M, Frommer M, Miller OJ, Macleod D.** 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell **40**:91–99.
3. **Blackledge NP, Klose R.** 2011. CpG island chromatin: a platform for gene regulation. Epigenetics **6**:147–152.
4. **Blackledge NP, et al.** 2010. CpG islands recruit a histone H3 lysine 36 demethylase. Mol. Cell **38**:179–190.
5. **Capili AD, Schultz DC, Rauscher IF, Borden KL.** 2001. Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. EMBO J. **20**: 165–177.
6. **Carrozza MJ, et al.** 2005. Histone H3 methylation by Set2 directs deacety-

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford

lation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell **123**:581–592.
7. **Cierpicki T, et al.** 2010. Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat. Struct. Mol. Biol. **17**:62–68.
8. **Dhalluin C, et al.** 1999. Structure and ligand of a histone acetyltransferase bromodomain. Nature **399**:491–496.
9. **Dyer PN, et al.** 2004. Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol. **375**:23–44.
10. **Engeholm M, et al.** 2009. Nucleosomes can invade DNA territories occupied by their neighbors. Nat. Struct. Mol. Biol. **16**:151–158.
11. **Farnham PJ.** 2009. Insights from genomic profiling of transcription factors. Nat. Rev. Genet. **10**:605–616.
12. **Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA.** 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell **130**:77–88.
13. **Guertin MJ, Lis JT.** 2010. Chromatin landscape dictates HSF binding to target DNA elements. PLoS Genet. **6**:e1001114.
14. **Hodges C, Bintu L, Lubkowska L, Kashlev M, Bustamante C.** 2009. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. Science **325**:626–628.
15. **Izban MG, Luse DS.** 1991. Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. Genes Dev. **5**:683–696.
16. **Jeong S, et al.** 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Mol. Cell. Biol. **29**: 5366–5376.
17. **John S, et al.** 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat. Genet. **43**:264–268.
18. **Johnson DS, Mortazavi A, Myers RM, Wold B.** 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science **316**:1497–1502.
19. **Kalkhoven E, Teunissen H, Houweling A, Verrijzer CP, Zantema A.** 2002. The PHD type zinc finger is an integral part of the CBP acetyltransferase domain. Mol. Cell. Biol. **22**:1961–1970.
20. **Klose RJ, Bird AP.** 2006. Genomic DNA methylation: the mark and its mediators. Trends Biochem. Sci. **31**:89–97.
21. **Kouzarides T.** 2007. Chromatin modifications and their function. Cell **128**:693–705.
22. **Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T.** 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature **410**:116–120.
23. **Li G, et al.** 2010. Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. Mol. Cell **38**:41–53.
24. **Lidor Nili E, et al.** 2010. p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy. Genome Res. **20**:1361–1368.
25. **Lomvardas S, Thanos D.** 2002. Modifying gene expression programs by altering core promoter chromatin architecture. Cell **110**:261–271.
26. **Lomvardas S, Thanos D.** 2001. Nucleosome sliding via TBP DNA binding in vivo. Cell **106**:685–696.
27. **Lorch Y, LaPointe JW, Kornberg RD.** 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell **49**:203–210.
28. **Lowary PT, Widom J.** 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. **276**:19–42.
29. **Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ.** 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389**:251–260.
30. **Nady N, Min J, Kareta MS, Chedin F, Arrowsmith CH.** 2008. A SPOT on the chromatin landscape? Histone peptide arrays as a tool for epigenetic research. Trends Biochem. Sci. **33**:305–313.
31. **Orphanides G, Wu WH, Lane WS, Hampsey M, Reinberg D.** 1999. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature **400**:284–288.
32. **Pique-Regi R, et al.** 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. **21**:447–455.
33. **Ramirez-Carrozzi VR, et al.** 2009. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell **138**:114–128.
34. **Roh TY, Cuddapah S, Zhao K.** 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev. **19**:542–552.
35. **Sabo PJ, et al.** 2004. Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. Proc. Natl. Acad. Sci. U. S. A. **101**:4537–4542.
36. **Saunders A, et al.** 2003. Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. Science **301**:1094–1096.
37. **Saxonov S, Berg P, Brutlag DL.** 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl. Acad. Sci. U. S. A. **103**:1412–1417.
38. **Shogren-Knaak M, et al.** 2006. Histone H4–K16 acetylation controls chromatin structure and protein interactions. Science **311**:844–847.
39. **Simon MD, et al.** 2007. The site-specific installation of methyl-lysine analogs into recombinant histones. Cell **128**:1003–1012.
40. **Smith E, Shilatifard A.** 2010. The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. Mol. Cell **40**:689–701.
41. **Song J, Rechkoblit O, Bestor TH, Patel DJ.** 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science **331**:1036–1040.
42. **Strahl BD, et al.** 2002. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol. Cell. Biol. **22**: 1298–1306.
43. **Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ.** 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. **14**:1025–1040.
44. **Tazi J, Bird A.** 1990. Alternative chromatin structure at CpG islands. Cell **60**:909–920.
45. **Thomson JP, et al.** 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature **464**:1082–1086.
46. **Tillo D, Hughes TR.** 2009. G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics **10**:442.
47. **Tillo D, et al.** 2010. High nucleosome occupancy is encoded at human regulatory sequences. PLoS One **5**:e9129.
48. **Tsumura A, et al.** 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells **11**:805–814.
49. **Vermeulen M, et al.** 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell **142**:967–980.
50. **Wysocka J, et al.** 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature **442**:86–90.
51. **Xi H, et al.** 2007. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. PLoS Genet. **3**:e136.

Downloaded from http://mcb.asm.org/ on January 27, 2012 by University of Oxford