Characterization of the interactions between the H3K4 demethylase JARID1A and the SCL-nucleated oncogenic DNA binding complex

Dimple Karia

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Magdalen College

Division of Structural Biology

University of Oxford

Hilary Term 2015

Characterization of the interactions between the H3K4 demethylase JARID1A and the SCL-nucleated oncogenic DNA binding complex

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Dimple Karia

Magdalen College and Division of Structural Biology
University of Oxford, Hilary Term 2015

ABSTRACT

Transcriptional regulation is of key importance to cellular processes such as development and differentiation, and it is dependent on the accessibility of DNA in the chromatin. Methylation and demethylation of histones alter properties of nucleosomes, thereby changing the accessibility of DNA. For many years, histone tail methylation was considered to be irreversible but in the last decade two families of histone demethylases were discovered, lysine specific demethylase 1 (LSD1) and Jumonji C (JMJC) demethylases.

The work presented in this thesis focuses on the chromatin remodeling protein JARID1A, belonging to the JARID1 family of JMJC demethylases that demethylates lysine 4 (K4) on the tail of histone H3. Double and triple methylation of K4 on histone H3 are marks traditionally linked to transcriptionally active regions of DNA. Recently, a translocation fusing the third PHD domain (PHD3) of JARID1A to NUP98 (a common leukaemia fusion partner with transactivation activity) was identified in acute myeloid leukaemia patients (AML). Also, PHD3 was previously shown to mediate interactions between JARID1A and the second LIM domain of LMO2. LMO2 operates within the multiprotein DNA binding complex named "SCL complex" (including proteins SCL/TAL1, LDB1, E47, GATA1) that regulates gene expression at different stages of erythropoiesis.

This thesis presents findings of an investigation into relations between JARID1A and the oncogenic multiprotein complex SCL-E47-LMO2-LDB1-GATA1. The study of the endogenous proteins from erythroid cells using techniques such as Co-immunoprecipitation, GST pull-down and Size-exclusion chromatography showed that JARID1A interacts with SCL and GATA1. These interactions were further characterized biochemically using analytical ultracentrifugation which revealed that the second PHD domain of JARID1A is involved in direct interaction with GATA1. Hence, JARID1A could play a role in regulation of erythropoiesis by modulating activities of transcription factors SCL and GATA1.

This thesis is dedicated to my parents,

Pradip and Priti Karia

DECLARATION OF WORK

The work described in this thesis is my own. For some experiments, certain reagents were obtained from colleagues which has been explicitly stated where appropriate.

Dimple Karia

Magdalen College,
University of Oxford

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Prof. Erika Mancini and Prof. Catherine Porcher for their constant support, patience and supervision throughout my DPhil. I am indebted to Prof. Robert Gilbert for his inputs in this project, help with AUC experiments and for all his support when I needed it the most.

I would like to thank many people who helped me in several facets of this project. Many thanks go to Dr. Sarah Hoosdally, Dr. Hedia Chargroui, Dr. Nicolas Goardon and Dr. Gaetan Juban for their help in the functional experiments and all the support they provided when I was working at WIMM. Thank you to Elisa Hall-Ponselé, Jessica Doondeea, Marina Samitsch and Batchimeg Usukhbayar for their help, enthusiasm and friendship. I am also thankful to all members of Porcher and Vyas groups at the WIMM for making me feel welcome.

I am grateful to everyone in STRUBI for helping me in the lab, for useful discussions and making my time in Oxford very enjoyable. I would like to express my gratitude especially to Dr. Luke Yates for his help with insect cells work, useful discussions in troubleshooting and his support, and Dr. Kamel El Omari for being a great mentor and friend during my summer studentship which inspired me to apply for DPhil. I am also thankful to Tao Ni for his help with AUC experiments and for his friendship. Many thanks go Antonio Biasutto for his help with NMR experiment, Anil Verma for his help with high-throughput protein expression screening, Dr. Mohammad Bahar and Dr. Geoff Sutton for MALS, and Dr. Matthias Zebisch for SPR experiments. I would like to acknowledge Dr. Tom Walter and Dr. Karl Harlos for their help and invaluable guidance in setting up crystallization trials. A special thank you to Margaret Jones for her support in the lab and for her optimism.

I would like to take this opportunity to thank Leukaemia and Lymphoma Research and Nuffield Department of Clinical Medicine for funding my DPhil.

I am truly indebted to my friends and family for their persistent love and support specially my friends Tejas and Hiral for being good listeners.

Finally, a very special thank you to my husband Abhay Kotecha, for his unwavering belief in me and for his constant love and support throughout my DPhil at Oxford.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Å	Angström
AGM	Aorta Gonad Mesonephros
AML	Acute Myeloid Leukaemia
ARID	AT-rich interacting domain
AUC	Analytical ultracentrifugation
BFU-E	Burst Forming Unit-Erythroid
bHLH	basic helix-loop-helix
bHLHZ	basic helix-loop-helix zipper
BSA	Bovine Serum Albumin
C/EBP	CCAAT/Enhancer Binding Protein
CBP	CREB Binding Protein
CenH3	Centromeric H3
CFU-E	Colony Forming Unit-Erythroid
Co-IP	Co-immunoprecipitation
CRC	Chromatin Remodeling Complex
DBP	DNA Binding Protein
DMEM	Dulbeccos Modified Eagles Medium
DMSO	Di-methyl sulfoxide
DTT	Di-thiotritol
EBF	Early B-cell Factor
EDTA	Ethylenediaminetetraacetic acid
ESC	Embryonic Stem Cell
FCS	Fetal Calf Serum
FOG	Friend of GATA1
GATA1	GATA binding protein 1
GFI-1	Growth Factor Independent transcription repressor
HAT	Histone Acetyl Transferase
HMT	Histone Methyl Transferase

HDAC	Histone Deacetylase
HEB	HeLa E-box binding protein
HEK	Human Embryonic Kidney cell line
HIF1	Hypoxia Induced Factor1
HP1	Heterochromatin Protein1
HSC	Haematopoietic Stem Cell
HSP70	Heat Shock Protein 70
hTERT	Telomerase reverse transcriptase
IPTG	Isopropyl-β-D-thiogalactoside
IVTT	*In vitro* transcribed and translated
JARID1	Jumonji AT-rich Interactive Domain1
JmjC	Jumonji C
JmjN	Jumonji N
LB	Luria Bertani Broth
LDB1	LIM domain binding protein 1
LIM-HD	LIM homeodomain
LMO	LIM only protein
LMO1	LIM domain only protein 1
LMO2	LIM domain only protein 2
LSD1	Lysine Specific Demethylase1
MAbs	Monoclonal Antibodies
MEL	Mouse Erythroid Leukaemia cell line
MLL	Mixed Lineage Leukaemia
ml	Millilitre
µl	Microlitre
µM	Micromolar
mm	Millimetre
MS	Mass Spectrometry
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off

NMR	Nuclear Magnetic Resonance
NSCLC	Non-Small Cell Lung Cancer
NUP98	Nucleoporin-98
NuRD	Nucleosome Remodeling and Deacetylase
PBS	Phosphate buffered saline
PCAF	P300/CBP associated factor
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethylenimine
PHD	Plant homeodomain
PHF8	PHD finger protein 8
pRB	Retinoblastoma protein
PRC2	Polycomb Repressive Complex2
RAG1	Recombination Activating Gene1
RBC	Red Blood Cell
RBP2	Retinoblastoma binding protein 2
RPM	Revolutions per minute
RT PCR	Real time Polymerase Chain Reaction
RT	Room temperature
RUNX1	Runt-related transcription factor1
SCL	Stem cell leukaemia protein
SDS-PAGE	Sodium dodecyl sulphate- polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
shRNA	small hairpin RNA
siRNA	small interfering RNA
SirT1	Sirtuin1
T-ALL	T-cell acute lymphoblastic leukaemia
TB	Terrific Broth
TBP	TATA-binding protein
TCA	Tri-chloro acetic acid

TF	Transcription Factor
TSS	Transcription Start Site
Zf	Zinc finger

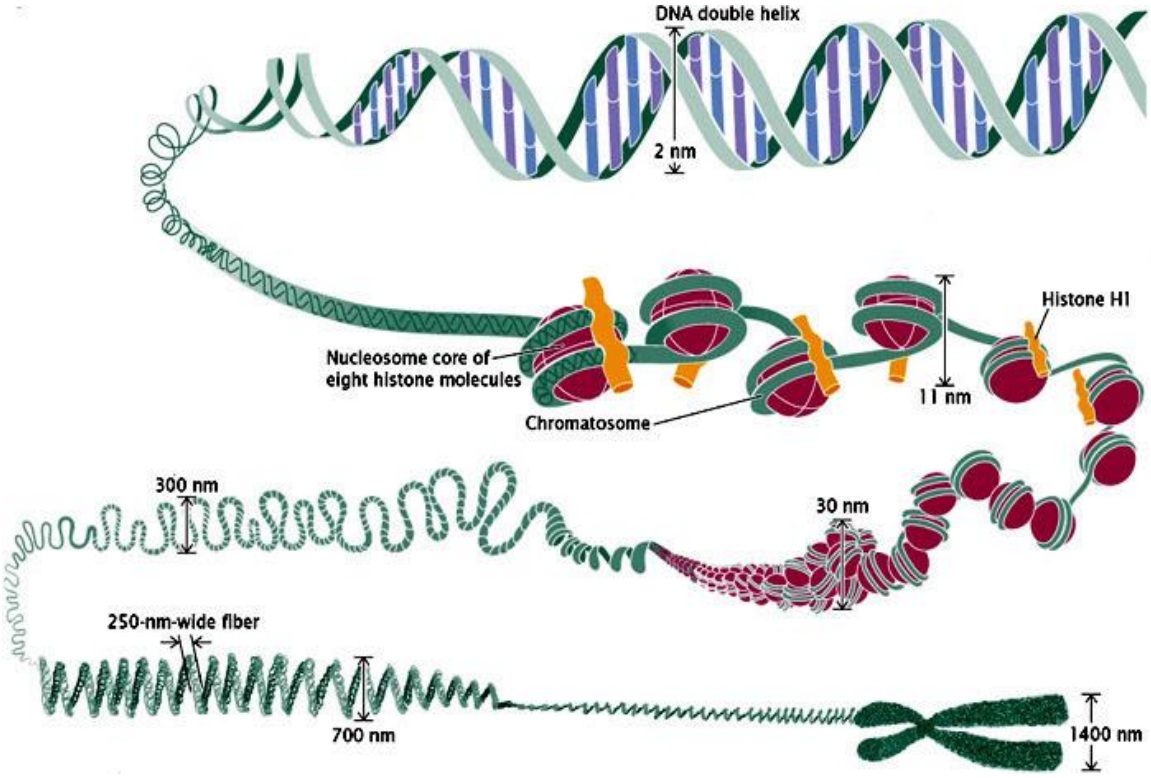
Chapter 1: Introduction

1 Introduction

1.1 Chromatin

In eukaryotes, genomic DNA is hierarchically packed inside the nuclei cells. Several layers of compaction are required to fit the long DNA molecules found in each cell. For instance, human DNA in a single cell measures approximately 1.8m in length and must be compacted several thousand fold to be accommodated in the nucleus of the cell which is only 6µm wide on average. This high order compaction is facilitated by a structure called chromatin. Chromatin is a filamentary nucleoprotein assembly of DNA and proteins that contains twice as much protein as DNA (Ball 2003). Chromatin undergoes several stages of compaction to be packaged within the nucleus of a cell (**Figure 1.1a**). Chromatin organization results in the formation of two states of compaction: heterochromatin (closed conformation) and euchromatin (open conformation). The higher order chromatin structures are highly dynamic and facilitate selective availability of the DNA to be read by proteins that carry out transcription. Hence, the dynamics of higher order compaction of chromatin plays an important role in controlling gene expression and other biological processes inherent to DNA metabolism (Ball 2003).

(a)



(b)

Figure 1.1: Chromatin Organization: (a) DNA is compacted in the form of chromatin, whose fundamental unit is the nucleosome **(b)** The nucleosome is formed by dimers of four histone proteins (two copies each of H2A, H2B, H3, H4) and 146bp of DNA wrapped around it. Higher level compaction might be achieved by folding a string of nucleosomes into a 30nm fibre, which is stabilized by linker histone protein H1. Further compaction leads to higher order structures like chromosomes. [Figure (a) Adapted from (Pierce 2010) and Figure (b) adapted from (Luger et al. 1997)].

1.2 Chromatin Structure and dynamics

1.2.1 Chromatin structure

The fundamental unit of chromatin is the nucleosome. Each nucleosome consists of 146 base pairs (bps) of DNA wrapped in two helical turns around the histone core octamer formed by two copies of each histone H2A, H2B, H3 and H4. These histone proteins have a conserved and highly structured histone core fold and unstructured histone tails protruding outwards from the core as shown in the crystal structure solved by the Richmond group (Luger et al. 1997) (**Figure 1.1b**). The four basic histone proteins associate to form a histone octamer: the H3-H4 heterodimers pair to form a tetramer and the association of this with DNA is the first step in nucleosome assembly; each H2A-H2B heterodimer binds to the H3-H4 tetramer to constitute the octamer. The interactions between these histones are mediated by formation of electrostatic, hydrophobic and hydrogen bonds at the interface. Usually, the core histones interact with DNA via their arginine side-chains. A linker DNA of approximately 10-80 bps connects individual nucleosomes to form a polynucleosome string that is further compacted into a 30nm fibre. The latter is stabilized by binding of histone linker protein H1 (Ball 2003; Felsenfeld and Groudine 2003; Kamakaka and Biggins 2005) (**Figure 1.1a**). Further compaction of chromatin leading into higher order structures is very dynamic and is facilitated by inter-fibre nucleosome-nucleosome interactions and chromatin looping. This plasticity of higher-order chromatin structure plays a key role in regulation of important biological processes such as transcription. Various factors such as core histone post-translational modifications, histone variants and DNA methylation affect the intrinsic properties of nucleosomes thereby regulating nucleosome dynamics (Li and Reinberg 2011). The linker histone H1 has also been proposed to play a crucial role in promoting chromatin high-order structure. Structurally, H1 histones consist of a conserved globular region and long extended amino- and carboxy- termini tails with the latter being involved in determining

their binding properties. H1 histone and its subtypes could be involved in the formation of different chromatin structures that play a role in gene expression.

1.2.2 Histone variants

Although histones are the fundamental packaging components within the chromatin, a fraction of histones arise from their non-allelic variants that have specific expression and localization and can possess significant differences in their primary sequences. These histone variants impart dynamicity to the chromatin structure by changing the stability of the nucleosomes and hence chromatin fibres or by changing the surface residues available for interaction with other histones or with other regulatory proteins. Some of the variants replace the pre-existing histones from nucleosomes during development and differentiation and hence are called replacement histones. The similarity between the basic histones and their non-allelic variants can range from no amino acid difference to highly divergent sequences. Following are different types of histones and their variant forms (Kamakaka and Biggins 2005).

1) **Histone H1**

   The sequence variant forms of H1 histone includes $H1^0$, H5 and sperm and testis-specific variants. These variants have sequence differences in the non-globular N- and C- terminal tail domains. The level of expression of these variants varies in different cell types and during different cellular processes like cell cycle, differentiation and development.

2) Histone H2A

Among all the core histones, H2A is the one which has the most abundant variant forms including H2A.Z, MacroH2A, H2A-Bbd, H2AvD and H2A.X. H2A.Z is the most conserved variant while MacroH2A and H2A-Bbd are restricted only to vertebrates and mammals. These H2A variants differ from the core H2A in terms of both length and sequence of their N- and C- terminal tails. H2A.Z is thought to be involved in both transcriptional activation and repression. It coordinates with Polycomb Repressive Complex2 (PRC2) to regulate gene repression in embryonic stem (ES) cells. It is also localized on the promoters of inducible genes, necessary for the recruitment of the transcription machinery to the promoters of these genes and for the induction of gene expression in mutant cells. H2A.Z also acts as a heterochromatin barrier thereby preventing spreading of the gene silencing factors. MacroH2A variant is localized to the inactive X-chromosome and thought to have a repressive function. It may suppress transcription sterically by blocking access to transcription factors and co-activators. H2A-Bbd lacks the C-terminal tail and destabilizes nucleosomes, thereby assisting in the displacement of nucleosomes during transcription and facilitating gene expression. H2A.X has also been shown to be essential for genome stability. Its phosphorylation at the sites flanking DNA double-stranded breaks is critical for the recruitment of DNA damage response components.

3) Histone H2B

Histone H2B has rarely any variants available. A notable exception is the sperm-specific H2B (SpH2B) in sea urchins, which has a long N-terminal tail that is highly charged.

4) **Histone H3**

Histone H3 possesses three different variant forms: H3.3, H3.4 and centromeric H3 (CenH3). H3.3 and H3.4 variants have the least divergence with only four amino acid differences. Also, H3.3 is expressed throughout the cell-cycle and is localized in transcriptionally active regions of chromatin, unlike its core histone counterparts. CenH3 aids genome stability by acting as an epigenetic mark specifying the site for kinetochore formation. Apart from this, it also plays a major role in directing assembly of the proteinaceous kinetochore structure creating a favourable environment for kinetochore assembly. In this way, CenH3 regulates the conformity of chromosome segregation.

5) **Histone H4**

Histone H4 is a slow evolving protein and hence does not possess any sequence variants.

1.3 Chromatin Remodeling

Regulation of gene expression is essential for cells to carry out various processes. This tight regulation is orchestrated by numerous proteins including cytokines, signaling molecules and transcription factors (TFs). TFs are the essential effectors of the cascade of regulatory mechanisms. They bind specific DNA sequences at the cis-regulatory elements of their target genes. The DNA located at the surface of the nucleosome or in the linker region is accessible to transcription factors, whilst, when the DNA is buried inside the chromatin, it is inaccessible. In order to make DNA accessible, unpacking of chromatin is necessary. The unwinding of chromatin is enabled by protein complexes known as Chromatin Remodeling Complexes (CRCs) (Clapier and Cairns 2009) .

1.3.1 Chromatin Remodeling Complexes

There are two different families of CRCs. One family consists of protein complexes that utilize the energy released by hydrolysis of ATP to move nucleosomes along the length of the DNA molecule to expose DNA binding sites. The second family consists of protein complexes that modify the unstructured histone tail residues by covalent modifications such as methylation, phosphorylation, acetylation and ubiquitination (Zhang and Reinberg 2001).

Often, it is a combination of physical and chemical modifications that take place to modify the chromatin architecture thereby altering the accessibility of DNA to the DNA binding proteins and hence regulating gene expression (Clapier and Cairns 2009) (**Figure 1.2**).

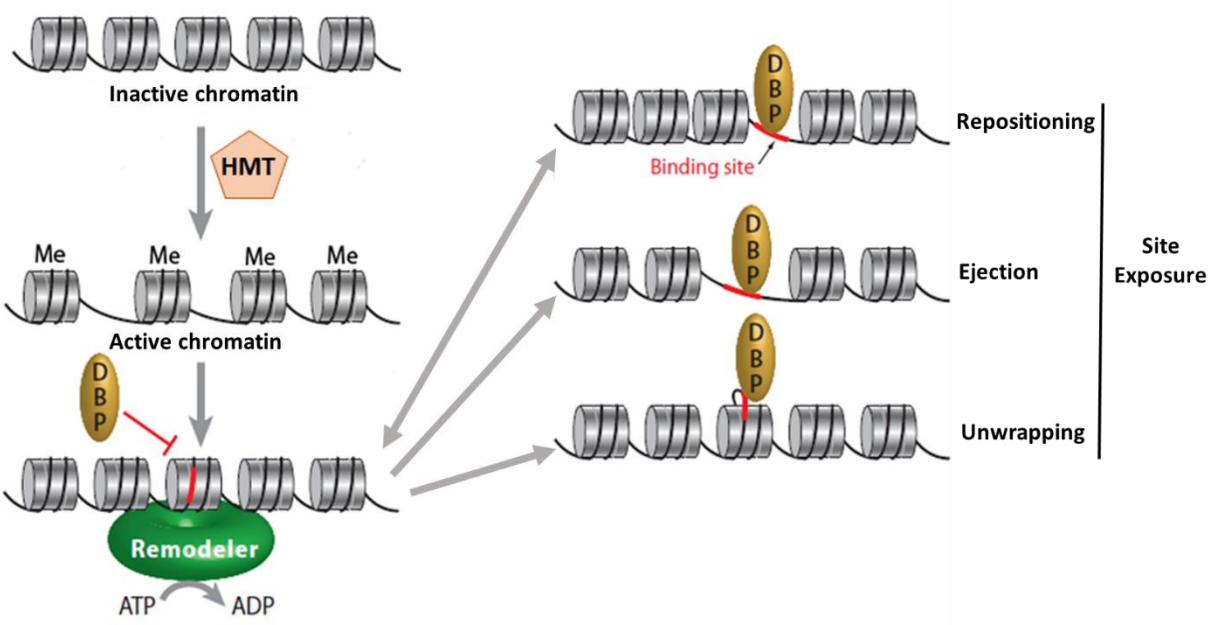


Figure 1.2: Chromatin remodeling. Chromatin remodeling by physical modifications of the nucleosome and post-translation modifications of the unstructured histone tails. HMT=Histone Methyl Transferase, DBP= DNA Binding Protein [Adapted from (Clapier and Cairns 2009)].

1.3.2 Post-translational modifications

Nucleosomes are highly dynamic entities and undergo alterations in their composition, structure and location along the DNA. Histones are subjected to a variety of post-translational modifications by numerous enzymes which change the characteristics of these proteins. These modifications represent a marking system and the presence of a specific or a combination of histone modifications can act as a 'histone code' that generates synergistic or antagonistic interaction affinities for other regulatory proteins that recognise the modified residue. The histone code hypothesis is believed to work in two stages: the first stage includes the establishment of a combinatorial pattern of the histone marks by histone modifying enzymes called the writers, which catalyse the chemical modification of the histone tail residues, and erasers, which eliminate the chemical modifications from the histone tails. The second stage includes the interpretation of the histone code by the effector/reader enzymes that then regulate the biological processes. The charge neutralization resulting from some of the modifications also results in weakening of the histone-DNA interaction thereby facilitating histone release and hence promoting various processes (Chi et al. 2010). These modifications can also occur on histone variants and may enable deposition or eviction of histone variants from the chromatin. For example, H1 histones are phosphorylated on their tails during their deposition on chromatin or before their removal from chromatin. So, the modification could be beneficial in both processes. H2A.Z is reversibly acetylated on its tail lysine residues 4, 7, 10, 13, 16, 21 in *Tetrahymena* and at least one of these lysines is required for cell survival indicating that the charge modulation and dissociation of the tail are important for cell survival. In this way, it could alter higher order chromatin structure thereby modulating chromatin dynamics (Kamakaka and Biggins 2005). The post-translational modifications of histone tails includes methylation, phosphorylation, acetylation and ubiquitination as mentioned earlier (Zhang and Reinberg 2001) (**Figure 1.3**). Often there is

cross-talk between different chemical modifications to affect the downstream processes. For example, deacetylation of lysine 26 (K26) on H1.4 is carried out by SirT1 which physically interacts with PRC2 complex. The latter methylates K26 of H1.4 thereby creating a binding site for heterochromatin protein1 (HP1) to repress transcription (Woodcock and Ghosh 2010; Li and Reinberg 2011).

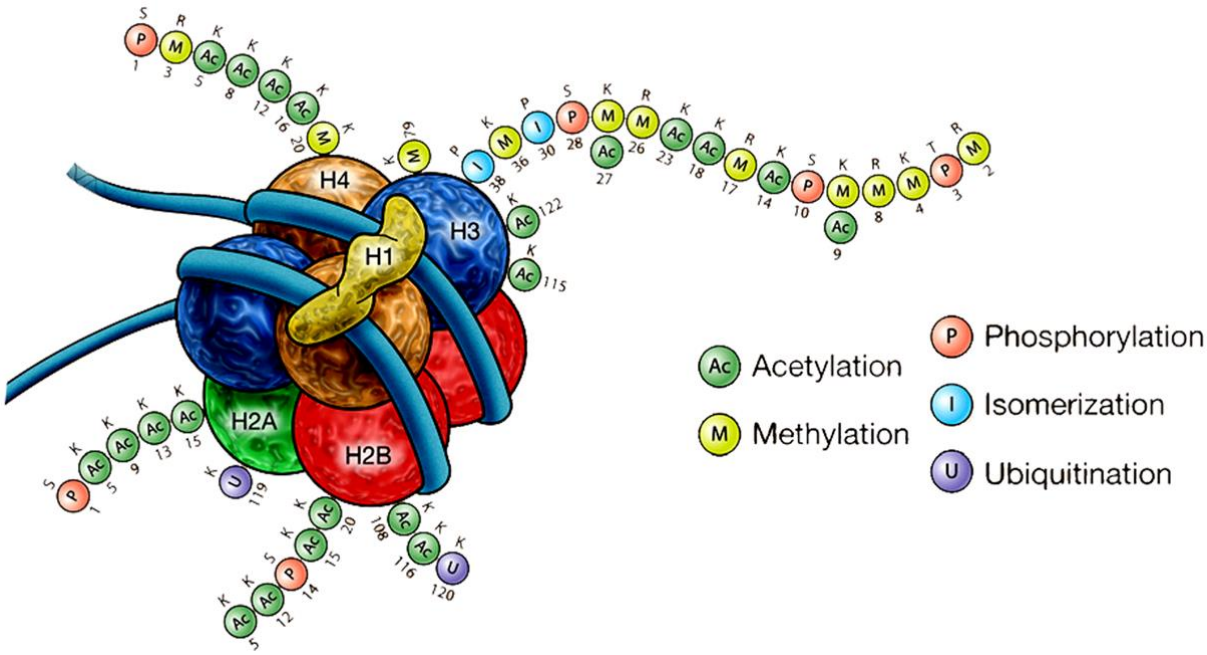


Figure 1.3: Histone tail modifications. The histone tails protruding out of the histone core are subjected to various chemical modifications such as acetylation, methylation and phosphorylation to render chromatin less compact and to expose the DNA binding sites for DNA binding proteins to carry out transcription in a controlled manner [Adapted from (Barber and Rastegar 2010)].

1.3.3 Histone acetylation

Various sites on histone tails are subjected to specific post-translational modifications such as acetylation. Histone acetylation is a highly dynamic process and is regulated by two groups of enzymes: histone acetyl transferases (HATs), which add an acetyl group to the lysine

residues (**Figure 1.3**) and histone deacetylases (HDACs), which remove acetyl group. HATs catalyse transfer of an acetyl group to the Ɛ-amino group of a lysine residue by using acetyl CoA as cofactor thereby neutralizing the positive charge on the lysine residue and thus potentially weakening the electrostatic histone-DNA interactions. There are two classes of HATs: Type A and Type B. The latter are mainly cytoplasmic and acetylate the free histones whereas type A HATs are nuclear and acetylate the lysine residues present on histone tails. Apart from that, some lysine residues present on the histone core are also substrates of Type A HATs. For example, lysine 56 present on the globular histone H3 core is acetylated in humans and the side chain of K56 points towards the major groove of DNA suggesting that acetylation would function in a similar way by disrupting the electrostatic histone-DNA interactions. HDACs act oppositely to HATs and hence they impart positive charge to the lysine residues resulting in strengthening of the electrostatic attraction between histone and DNA thus stabilizing chromatin architecture. Therefore, HDACs are primarily transcription repressors (Bannister and Kouzarides 2011).

1.3.4 Histone Phosphorylation

Histone phosphorylation is also a highly dynamic process and is regulated by kinases, which add the phosphate groups, and phosphatases, which remove the phosphate groups. Phosphorylation occurs predominantly on serines, threonines and tyrosines (**Figure 1.3**). Histone kinases carry out the process by transferring a phosphate group from ATP to the hydroxyl group of the amino acid residue and, in this manner, modify the charge on the target amino acid by accumulating negative charges which in turn influence the chromatin architecture. Most of the substrates of histone kinases are located on the histone tails but there are some on the histone core as well. For instance, tyrosine 41 on histone H3. Less is known about the functions of histone phosphatases (Bannister and Kouzarides 2011).

1.3.5 Histone Ubiquitylation and sumoylation

Ubiquitin is a 76- amino acid polypeptide and hence it leads to bigger changes in the histone it is attached to. It targets lysine residues predominantly on histones H2A and H2B through the sequential action of three enzymes E1- activating, E2- conjugating and E3- ligating (**Figure 1.3**). The substrate specificity and degree of ubiquitylation (mono- or poly-ubiquitylated) are governed by the enzyme complex. Like other histone modifications, ubiquitylation also has a role in regulating gene expression. For example, H2AK119ub1 (mono ubiquitylation of lysine 119 on histone H2A) results in gene silencing whereas H2BK123ub1 is involved in transcriptional initiation and elongation. Ubiquitin marks are removed by the action of isopeptidases called de-ubiquitinating enzymes.

Sumoylation is a very similar modification to ubiquitylation and it involves the addition of a ubiquitin-like modifier to the lysine residues by the action of the same three enzymes that carry out ubiquitylation. It targets lysine residues present on all four histones and acts by opposing acetyl and ubiquitin histone marks, and therefore possesses repressive functions principally (Bannister and Kouzarides 2011).

1.3.6 Histone tail Methylation

Histone tail methylation is a post-translational modification that occurs on lysine and arginine residues. The lysine residues present on Histones H3 (4, 9, 27, 36, 79) and H4 (20) are preferred sites for methylation (**Figure 1.3**). Up to three methyl groups can be added at each lysine residue resulting in four methyl states: unmethylated, monomethylated, dimethylated and trimethylated. Methylation of lysine residues does not affect the overall charge of the residues implying that, unlike acetylation, methylation does not obstruct histone-DNA interaction and it probably plays an indirect regulatory role by creating a favourable

environment for the effector proteins to bind and translate the code into a meaningful biological process (Blackledge and Klose 2010). Histone lysine methylation can mediate both transcription activation and repression depending on the location of the residue which is methylated and the degree of methylation. For example, H3K4 di- or tri-methylation has been linked to transcriptional activation and H3K27 tri-methylation has been linked to transcription repression (**Table 1.2**). H3K4me3 is restricted to the transcription start site (TSS) and marks active promoters and H3K4me2 is often found marking the coding region (Benevolenskaya 2007). Histone methylation is a reversible process and is orchestrated by the action of two groups of enzymes: the writers (Histone methyl transferases) and the erasers (Histone demethylases). The methylation mark is established by SET1 and Mixed Lineage Leukaemia (MLL) family of methyl transferases and is erased by Lysine Specific Demethylases1 (LSD1) and Jumonji AT-rich Interactive Domain1 (JARID1) family of demethylases (Chi et al. 2010). Various protein domains have been identified as the 'methyl readers', which help translate the histone code to elicit its functional meaning. This group comprises the Plant homeodomain (PHD), the Chromodomain, the Tudor domain and the WD40 repeat (Martin and Zhang 2005). For example, heterochromatin protein 1 (HP1) uses its chromodomain to bind to the methyl mark on H3K9 to induce a repressed state (Bannister et al. 2001). Often there are cross-talks between histone demethylases to control the process of transcription.

Mis-regulation of the histone code in terms of reading, writing or interpretation has been linked to cancer initiation or progression, for example Mixed Lineage Leukaemia (MLL) gene arrangement in Leukaemia. The partial tandem duplication of the MLL gene has been found to aberrantly elevate the H3K4 dimethylation and histone acetylation of *Hoxa* genes both of which are associated with transcription activation. Specifically, expression of the

Hoxa9 gene, which is highly expressed in the early haematopoietic stages and silenced upon differentiation, is largely affected in a way that its expression is increased by several folds in terminally differentiated blood cells (Chi et al. 2010). The role of other genes involved in deregulation of the methyl marks on H3K4 and H3K27 and cancer initiation has been listed in **Table 1.1**.

Table 1.1: Genes involved in deregulation of histone methylation mark on H3K4me3 and H3K27me3 and associated with cancer initiation [adapted from (Chi et al. 2010)].

Methyl Mark	Gene name	Deregulation in human cancer
H3K4me		
Writer	*Mll*	MLL gene rearrangement is found in 80% of infant leukaemia and 5-10% of adult acute myeloid and lymphoid leukaemias
	Mll2	MLL2 mutation found in renal cell carcinoma
	Men1	MEN1 is often mutated in endocrine tumours
Reader	*Ing1, Ing2, Ing3, Ing4 and Ing5*	Loss of function mutations of the putative tumour suppressor gene ING5 are associated with solid tumours and somatic mutation of the ING2 gene obstructs binding of H3K4me3 mark.
	Phf3	PHF3 is a PHD finger which reads H3K4me3 mark and its fusion with NUP98 has been linked to myeloid leukaemia initiation.
	Pygo2	PYGO2 is a factor in the βcatenin signaling pathway that is essential for self-renewal of mammary progenitor cells and is found overexpressing in breast cancer cells.
Eraser	*Lsd1*	LSD1 is a part of NuRD-Mi-2 repressive complexes and demethylates H3K4me2/1. It represses metastasis of breast cancer cells and hence is down-regulated in breast cancer.
	Jarid1a	Fusion of PHD finger of JARID1A which recognizes H3K4me3 with NUP98 results in formation of oncoprotein NUP98-PHD. The latter obstructs H3K4me3 binding potential of PHD resulting in enforcing an active state on developmentally critical loci and hence in leukaemia initiation.
	Jarid1b	JARID1B is over-expressed in advanced breast and prostate cancer.
	Jarid1c	Jarid1c inactivating mutations were found in 3% of renal cancers.
	Jhdm1b	Jhdm1b is up-regulated in in retrovirus-induced rat T cell lymphomas.
H3K27me		
Writer	*Ezh2*	EZH2 is over-expressed in various solids tumours in prostate, skin, colon, breast and lung cancers.
Eraser	*Jmjd3*	Down-regulated in lung and liver cancers.
	Utx	Inactivating mutations in Utx is found in some of multiple myeloma, oesophageal squamous cell cancer and renal cell cancers.

General functions of various histone methylation marks have been listed in **Table 1.2** (Martin and Zhang 2005).

Table 1.2: General functions of histone methylation marks.

Methyl marks	Function
H3K4	Gene activation
H3K9	Tri-methylation leads to gene silencing whereas mono- and di-methylation leads to gene activation
H3K27	Gene silencing
H3K36	Gene activation
H3K79	Gene activation
H4K20	Tri-methylation leads to gene silencing whereas mono- and di-methylation leads to gene activation

1.4 Histone demethylases

Common preferred methylation sites are lysine 4, 9, 27, 36, 79 of H3 and lysine 20 of H4. Histone tail methylation is a highly dynamic process like other modifications. However, initially it was believed that methylation of histones is an irreversible process and this mark can only be removed by histone exchange, eviction or maybe by histone tail clipping or dilution during DNA replication. This notion was later proven incorrect upon the discovery of the first histone demethylase named Lysine Specific Demethylase 1 (LSD1) (Shi et al. 2004). Later, another family of enzymes was discovered whose members also possessed demethylase activity. So, to this date, there are two histone demethylase protein families: LSD (Lysine specific demethylases) demethylases and JMJC (Jumonji C) demethylases (Kooistra and Helin 2012).

The **LSD demethylase** family comprises two members, LSD1 and LSD2. LSD1 catalyses demethylation of H3K4me1, H3K4me2 and H3K9me1, H3K9me2. LSD2 only catalyses demethylation of H3K4me1 and H3K4me2. This class of demethylases are flavin amine oxidases and hence they remove the methyl group from its substrates by a flavin adenine dinucleotide (FAD) amine oxidation reaction. The FAD amine oxidation reaction requires a free electron pair at the methylated lysine residue thereby explaining the capability of LSD demethylases only to demethylate mono- and di- methylated lysines (**Figure 1.5a**). LSD1 comprises a SWIRM and amine oxidase domain which are packed together through hydrophobic interactions. The two lobes of the amine oxidase domain form the catalytic cavity, which is relatively large and hence cannot determine the specificity for its substrate that is determined by the catalytic reaction (Kooistra and Helin 2012). The X-ray crystal structure of the catalytic oxidase domain of LSD1 shows that the N-terminal SWIRM and C-terminal oxidase domain pack closely against each other forming a 60Å core from which the tower domain protrudes spanning around 140Å and the catalytic oxidase domain forms the spacious active site cavity (**Figure 1.4**) (Stavropoulos et al. 2006).

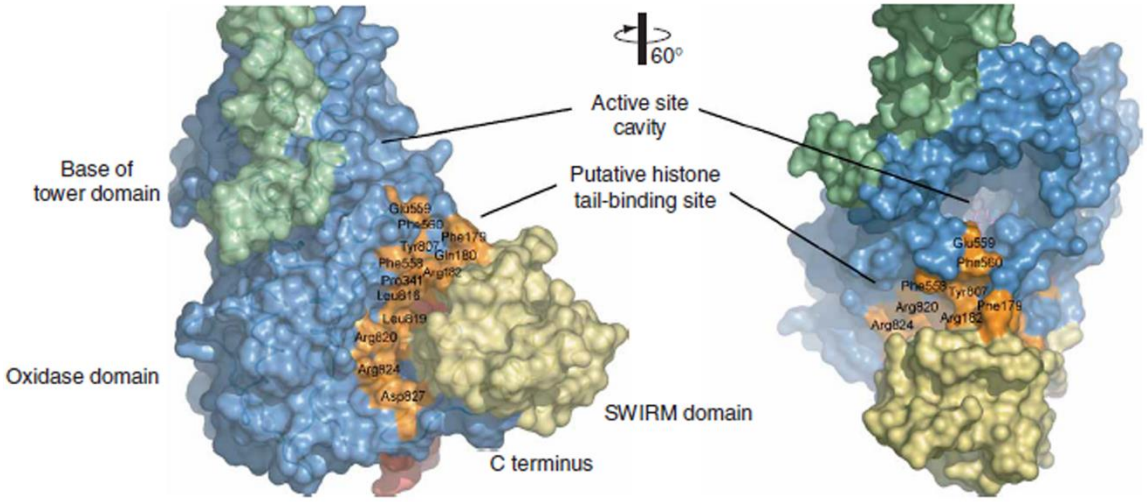


Figure 1.4: X-ray crystallography structure of catalytic oxidase domain of LSD1. X-ray crystal structure of LSD1 shows a spacious active site cavity which disable it for substrate specificity. The formation of the active site cavity is facilitated by the interaction between the C-terminal tail of oxidase domain (blue) and the SWIRM domain (beige). The interface between these two domains (orange) is highly hydrophobic and hence both the domains are held together by a network of van der Waals contacts. [Taken from (Stavropoulos et al. 2006)].

The second family of histone demethylases is comprised of **JMJC demethylases**. All the members of this family contain a JMJC catalytic domain. KDM2A (or JHDM1A) was the first enzyme discovered in this class and it demethylates H3K36me1 and H3K36me2. In total, this family consists of 30 different members, 18 of which have been shown to possess demethylase activity. This family has shown specificity towards H3K4, H3K9, H3K27, H3K36, H4K20 and demethylates di- and tri- methylation (Benevolenskaya 2007). These demethylases are Fe (II) and α-ketoglutarate (α-KG) dependent and carry out demethylation of lysines by a dioxygenase reaction. In this reaction, in presence of Fe (II), oxidative decarboxylation of α-KG takes place resulting with the formation of a hydroxylated methyl-lysine intermediate, which is thermodynamically unstable and hence releasing formaldehyde and the unmethylated lysine substrate (**Figure 1.5b**). The crystal structure of the catalytic

domain of JMJD2A shows that the catalytic cavity has binding pockets for Fe (II) (His188, Glu190 and His276) and α-ketoglutarate (Tyr132, Asn198 and Lys206) and that this determines the specificity for substrate. However, the substrate specificity also depends on the sequence of amino acids around the lysine, so that only peptides of the right conformation will fit into the catalytic cavity (Kooistra and Helin 2012).

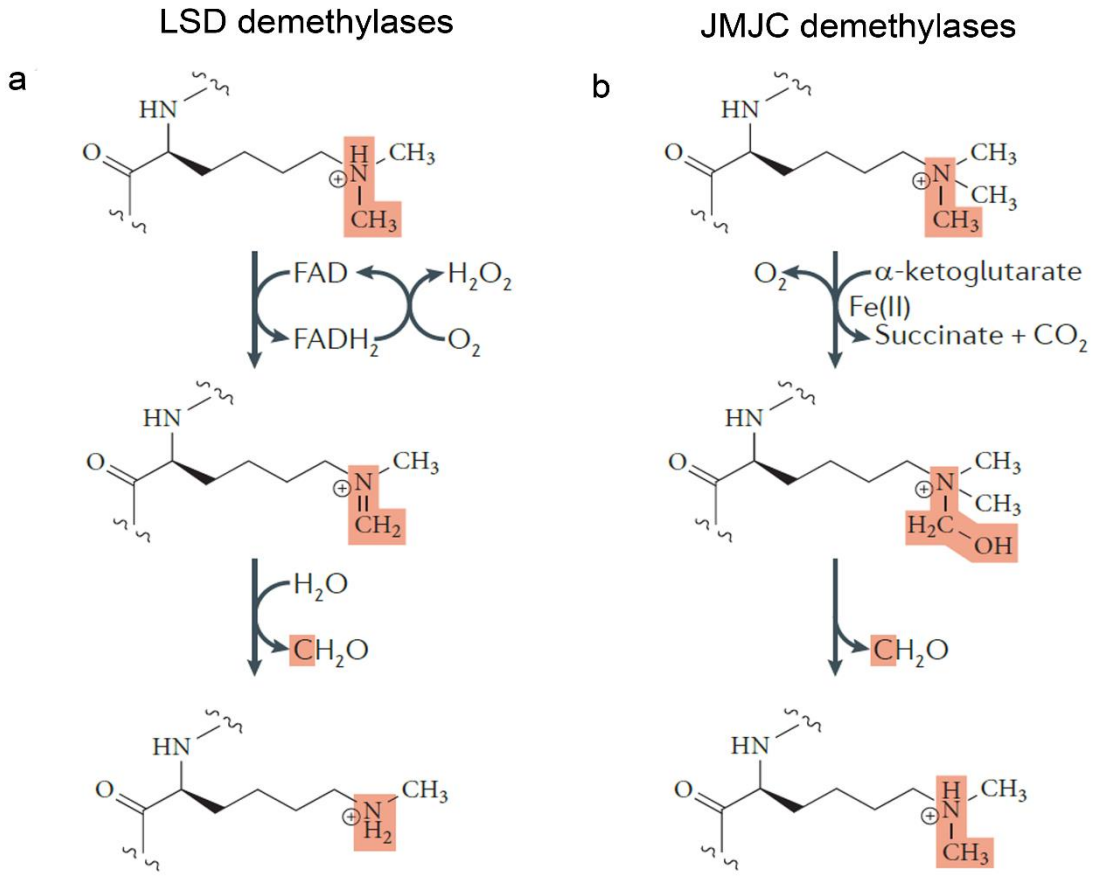


Figure 1.5: The catalytic mechanism of histone demethylation. a) The flavin adenine dinucleotide (FAD)-dependent amine oxidation reaction is used by LSD demethylases to demethylate mono- and dimethylated lysine substrates. **b)** The Fe(II) and α-ketoglutarate dependent dioxygenase reaction is used by JMJC demethylases to demethylate mono-, di- and trimethylated lysine substrates [Taken from (Kooistra and Helin 2012)].

Often there are multiple histone marks deposited on the same residue. Some of the marks are mutually exclusive such as H3K4 and H3K9 methylation whereas some others co-exist on the

same residue, for example H3K4 methylation and histone acetylation. This co-existence or mutual exclusivity of histone marks points towards the fact that there is cross-talk between various histone modifying enzymes and histone marks, and that histone modifications occur in a co-ordinated fashion. The cross-talk can occur between the residues present on the same histone or on different histones. Three different mechanisms for cross-talk involving lysine specific demethylases have been proposed (Verrier et al. 2011):

1) Docking mechanism- In this, a histone demethylase identifies a pre-existing histone modification and binds to it via its modification binding domain and then demethylates the neighboring residue. For example, PHF8 binds to the H3K4me2/3 mark and consecutively demethylates H3K9me1/2. Also, the PRC1 complex binds to H3K27me3 and carries out mono-ubiquitination of H2AK119 (Verrier et al. 2011). This is also exemplified by JARID1C whose one PHD finger binds to the H3K9me3 mark and consecutively demethylates H3K4me3 (Iwase et al. 2007).

2) Co-recruitment mechanism- This involves recruitment of two histone modifying enzymes simultaneously on the same histone leading to establishment of multiple modifications at the same time. For instance, on HOX gene promoters, both H3K4 demethylase JARID1A and H3K27 lysine methyl transferase EZH2 (which is a part of polycomb complex) are recruited at the same time leading to formation of repressive H3K27me3 mark and elimination of active H3K4me3 marks (Verrier et al. 2011).

3) Modification of substrate- In this case modification of a neighboring residue on the histone alters the activity or specificity of a particular histone demethylase. This is

illustrated by the phosphorylation of Thr-6 on H3 which alters the specificity of LSD1 from H3K4me1/2 to H3K9me1/2 (Verrier et al. 2011).

Histone demethylases such as H3K4 demethylases play an important role in development and differentiation. For example, deletion of the JmjC domain of RBR-2 (retinoblastoma binding protein related 2) in *C. elegans* results in defective vulva formation and homozygous mutations of the *lid* (little imaginal discs) homologous protein RBR-2 in *Drosophila* results in larval lethality and developmental defects in surviving flies (Christensen et al. 2007).

1.5 Haematopoiesis

Haematopoiesis is the process of formation of all types of blood cells from Haematopoietic Stem Cells (HSCs) (**Figure 1.6**). HSCs are multipotent cells found in the bone marrow of adult mammals. Because mature blood cells are short-lived, HSCs replenish them throughout the life of an individual. Hence, the two indispensable properties that characterize HSCs are self–renewal and the ability to differentiate into various unilineage progenitor cells, which then commit to a specific lineage. The process of haematopoiesis is dynamic in terms of its sites of occurrence and it proceeds in two waves, primitive and definitive. Primitive haematopoiesis, which occurs in the yolk sac, consists, mainly, of an enhanced production of red blood cells that facilitate oxygenation of tissues as the embryo undergoes rapid growth. This wave of haematopoiesis is temporary and is soon replaced by the definitive haematopoiesis, which occurs in the aorta-gonad mesonephros (AGM), placenta, fetal liver and terminates in the bone marrow which remains the principal site of haematopoiesis throughout post-natal life. The primary source of HSCs is the AGM; they will then colonise

the fetal liver (Orkin and Zon 2008). These totipotent HSCs differentiate into haematopoietic progenitors (MPP) which further form committed precursors that produce specific cell types.

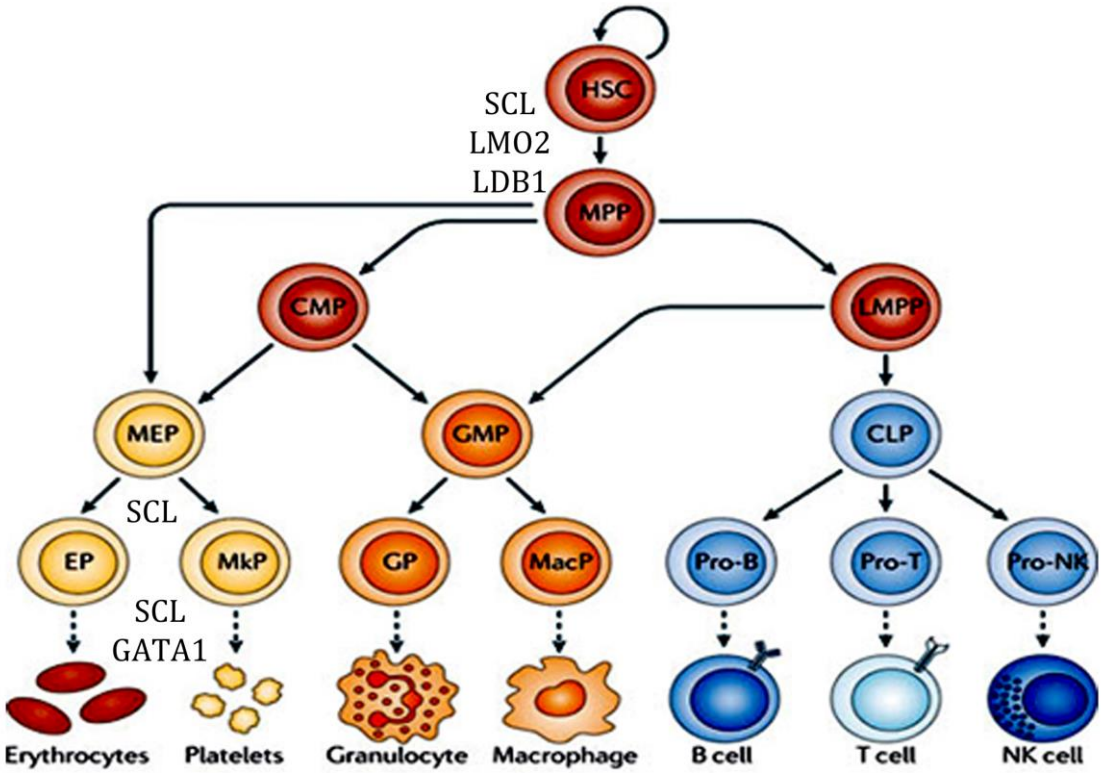


Figure 1.6: Hierarchy of haematopoiesis. Haematopoietic stem cells (HSCs) sit atop of the hierarchy and differentiate giving rise to committed progenitor cells. These progressively lose their multipotency and ultimately give rise to only one lineage. MPP, multi potent progenitor; LMPP, lymphoid primed multipotent progenitor; CMP, common myeloid progenitor; MEP, megakaryocyte-erythrocyte progenitor; GMP, granulocyte-macrophage progenitor; EP, erythrocyte pregenitor; MkP, megakarocyte progenitor; GP, Granulocyte progenitor; MacP, macrophage progenitor; CLP, common lymphoid progenitor; NK, natural killer. SCL and GATA1 are transcription factors, whereas LMO2 and LDB1 are linker and adaptor proteins. [Reproduced from (Cedar and Bergman 2011)].

1.5.1 Erythropoiesis

Erythropoiesis is the process of formation of red blood cells (erythrocytes). During early embryonic development, primitive erythropoiesis occurs in the yolk sac which is then replaced by definitive erythropoiesis that occurs in the fetal liver at first and then is moved to the bone marrow in late fetal life and throughout adulthood. Erythropoiesis is a dynamic multi-step process involving differentiation of pluripotent HSCs into committed erythroid precursors, which then give rise to mature erythrocytes. The first committed erythroid precursor cells are the burst forming unit-erythroid (BFU-E) which give rise to colony forming unit-erythroid (CFU-E). Both these cells are erythroid precursors that are responsive to erythropoietin, a principal regulatory hormone for red cell maturation which is made in the kidneys. CFU-E then differentiates into proerythroblasts, which are committed erythroid progenitors that develop sequentially into erythroblasts through basophilic erythroblasts, polychromatic erythroblasts and orthochromatic erythroblasts stages. The latter then expel nuclei to form reticulocytes. These enucleated cells successively expel all other organelles and are released into blood circulation where they mature into red blood cells (**Figure 1.7**). The cells undergo various morphological changes as they undergo erythropoietic differentiation such as decrease in cell size, nuclear condensation, cytoplasmic acidification and increased haemoglobin concentration (Krantz 1991; Alter 1994; Ingley et al. 2004).

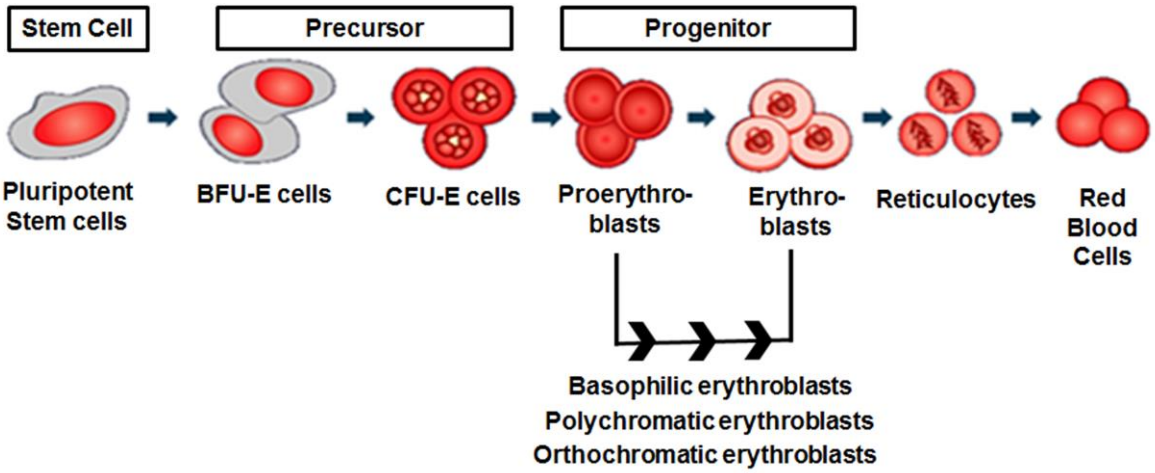


Figure 1.7: Erythropoiesis. Pluripotent Stem cells primed towards erythropoiesis give rise to the first committed erythroid precursors, BFU-E (Burst forming unit-erythroid) which further forms another distinct erythroid precursors, CFU-E (Colony forming unit-erythroid). CFU-Es differentiate into erythroid progenitors, proerythroblasts that sequentially develop into basophilic erythroblasts, polychromatic erythroblasts, orthochromatic erythroblasts (also known as normoblasts) and eventually forming the erythroblasts. The nuclei gets extruded from erythroblasts forming reticulocytes. These enucleated cells successively expel all other organelles before being released into the blood circulation where they mature into Red Blood Cells (RBCs) [Adapted from (Lankhorst and Wish 2010)].

The process of erythropoiesis is widely studied *in vitro* using the Mouse ErythroLeukaemia (MEL) cell line. MEL cells are the erythroid precursor cells that are transformed by the Friend virus complex and are arrested at proerythroblast stage (Friend et al. 1966). MEL cells serve as a valuable model system for studying erythroid differentiation as these cells undergo biochemical and morphological changes upon differentiation with chemical agents like DMSO, which closely recapitulates the normal erythroid differentiation (Harrison 1976). The cells undergo changes such as decreased cell size, nuclear condensation, cytoplasmic acidification, increased transcription of globin genes, increase in haemoglobin synthesis and increase in erythrocyte-specific membrane changes (Harrison 1976; Ingley et al. 2004). Moreover, it has been shown that the majority of proerythroblasts had matured to erythroblasts (also known as normoblasts) which stained positive for benzidine staining

(stains haemoglobin), four days after MEL cells were induced for differentiation by addition of 2% DMSO (Friend et al. 1971).

1.5.2 Transcriptional and epigenetic regulation of Haematopoiesis

Transcription factors are DNA binding proteins that bind to a specific site on DNA and regulate transcription of the downstream gene to RNA. They play a very crucial role in regulation of gene expression and amongst other processes, in cell development, differentiation and growth. These DNA binding proteins can be grouped into classes based on the structural motifs they use for recognition of the specific DNA sequences (Pabo and Sauer 1992):

1) **Helix-turn-helix motif-** This was the very first DNA binding motif discovered. It does not form a stable domain and hence does not work in isolation, but exists as part of larger DNA binding proteins. It consists of highly conserved glycine and several hydrophobic residues which stabilize the organization of the two helices enabling it to be a part of the DNA binding protein. It contacts DNA via its second helix which is also known as the recognition helix (Pabo and Sauer 1992).

2) **Homeodomain-** It contains a helix-turn-helix motif and comprises of an extended N-terminal arm and three α-helices. It makes the majority of contacts with DNA through the residues in helix 3 in which the exposed face fits into the major groove of the DNA while its side chains make extensive contacts with the bases of the DNA. Although most of its contacts are mediated via its third helix, the extended N-terminal arm also makes contacts with the minor groove of the DNA (Pabo and Sauer 1992).

3) **Zinc Finger-** Transcription factors of this family contain a 30 residue repeating unit of a zinc finger with a characteristic sequence pattern of Cys-X_{2or4}-Cys-X_{12}-His-X_{3-5}-His. This domain consists of an anti-parallel β-sheet and an α-helix. The two cysteine residues of the β-sheet and two histidines in the α-helix coordinate a zinc atom thereby stabilizing its structure. Zinc fingers exploit the arginine residues that directly precede the α-helix and the second, third and sixth residues of the latter to make DNA contacts (Pabo and Sauer 1992).

4) **Steroid receptor-** Steroid receptors possess a DNA binding domain with 30 conserved cysteines and a part of this domain folds in the presence of zinc enabling it to recognize the specific DNA sequences (Pabo and Sauer 1992).

5) **Leucine zipper and Helix-loop-Helix-** Proteins containing both these domains play a critical role in development and differentiation. The leucine zipper domain is comprised of 60-80 residues which are divided into two regions, the leucine zipper region that facilitates dimerization and a basic region which contacts the DNA. The leucine zipper region forms two parallel α-helices and the basic region (~30 residues) is rich in arginines and lysines. Helix-loop-helix domains have a structure similar to that of the leucine zippers and contain a basic region for DNA identification, a dimerization region that forms an α-helix, a loop and a second α-helix (Pabo and Sauer 1992).

6) **β-sheet motifs-** These use a β-sheet for the DNA binding in which the β-sheet fits into the major groove of DNA and its side chain residues make contacts with the base pairs; specifically, a lysine contacts a guanine and a threonine binds to an adenine (Pabo and Sauer 1992).

The process of haematopoiesis occurs by coordinating signalling pathways and regulation of gene expression as per the functional requirements of the organism. In haematopoiesis, like in any other differentiation process, lineage commitment and differentiation fall under the control of the relative expression of various transcription factors, which are both ubiquitously expressed and tissue specific (**Figure 1.6**). These transcription factors act in a combinatorial fashion along with chromatin modifying enzymes to activate lineage specific genes (**Figure 1.8**). For example, during haematopoiesis, GATA1, a Zinc finger transcription factor essential for red blood cell maturation and survival, recruits histone acetyl transferase CREB binding protein (CBP)-containing complexes to the β-globin gene locus leading to H3 and H4 acetylation and resulting in high globin gene expression, essential for high haemoglobin synthesis (Rice et al. 2007). In embryonic stem cells, many of the genes possess bivalent chromatin structure; they harbour both activating H3K4me3 marks and repressive H3K27me3 modifications. For instance, the genes coding for master regulators of haematopoietic stem cells (HSC), such as paired box gene 5 (*Pax5*) and *Ebf1* display bivalent status. But as differentiation progresses, these genes lose their bivalency in differentiated cell types (Cedar and Bergman 2011).

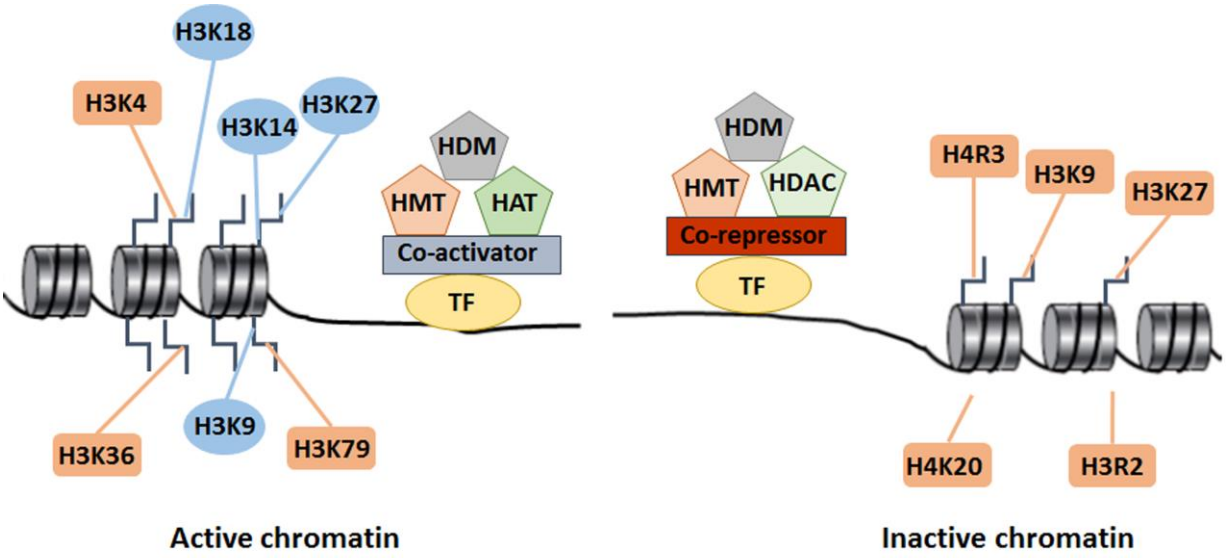


Figure 1.8: Coordination of transcription factors and histone modifying enzymes for regulation of gene expression. Transcription factors (TF) regulate gene expression by recruiting the histone modifying enzymes like histone methyl transferases (HMT), histone acetyl transferases (HAT, histone demethylases (HDM) and histone deacetylases (HDAC) to the target gene promoter. Hyperacetylation and methylation of H3K4, H3K79 and H3K36 have been linked to the active transcription and hypoacetylation, and methylation of H3K9, H3K27 and H4K20 has been associated with the gene repression [Adapted from (Rice et al. 2007)].

Although transcription factor levels are highly variable in the early stages of haematopoiesis, they are stringently regulated by extracellular regulatory signals during later stages of the process. Transcription factors required for haematopoietic development may be distinguished into the factors required for HSC formation like MLL, RUNX1, TEL/ETV6, SCL/TAL1, LMO2 and factors with more lineage-specific roles such as PU.1, GFI-1, C/EBP, GATA1, FOG (Orkin and Zon 2008). The genes encoding these transcription factors are marked specifically based on their expression requirements. For example, the *Gata1* gene is critical for red cell maturation, and is marked by H3K4me3 in the cells where it needs to be expressed, such as red cells, whereas it undergoes demethylation in other cells, such as B cells. Similarly *Rag1*, a gene coding for a lymphoid specific transcription factor, is hyper methylated in B-cells and demethylated in the erythroid cells (Cedar and Bergman 2011).

1.5.3 Pentameric complex nucleated by SCL

Gene expression is regulated by the formation of multiprotein transcription factor complexes and modulation of their composition leads to changes in cell fate. The Stem Cell Leukaemia (SCL/TAL1) oncoprotein is a tissue-specific transcription factor, which nucleates a multiprotein transcription factor complex and acts as an essential regulator at several branch points of the haematopoietic hierarchy (**Figure 1.6**). SCL is also involved in translocational events and participates in the physiopathology of up to 60% of T-cell acute lymphoblastic leukaemias (T-ALL). It belongs to the class II basic helix-loop-helix (bHLH) family of transcription factors, which heterodimerize with ubiquitously expressed class I bHLH proteins such as E47 and bind to DNA through the E-box motif (CANNTG) (Lecuyer and Hoang 2004). SCL also interacts through the conserved interface on its loop and helix2 with LIM domain only 2 (LMO2) which acts as a bridging molecule and interacts with LIM domain-binding protein 1 (LDB1) (Wadman et al. 1997). In red cells, these interactions participate in the nucleation of a pentameric complex containing SCL, E47, LMO2, LDB1 and the haematopoietic TF, GATA1 (**Figure 1.9**). Variation in the composition of this transcription factor complex leads to changes in cell fate (Lecuyer et al. 2007). In the case of SCL, this imparts the flexibility and specificity required for regulation of gene expression in order to control lineage-specific cell fate decisions (Lecuyer and Hoang 2004).

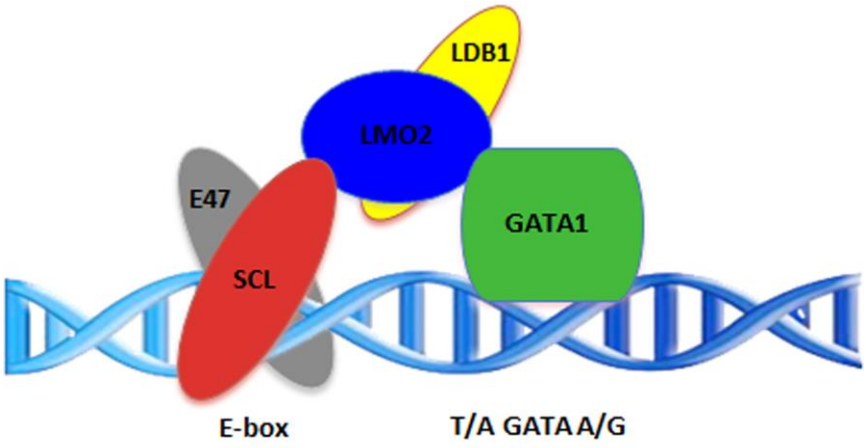


Figure 1.9: Schematic of the Pentameric Complex. SCL belongs to the bHLH family of transcription factors. It heterodimerizes with E47 and binds to the DNA via an E-box motif (CANNTG). SCL also interacts with LMO2, which in turn binds to LDB1. LMO2 can also bind to the transcription factor GATA1 that binds to the GATA DNA binding motif (T/A) GATA (A/G). This results in the formation of a pentameric complex interacting with two DNA binding sites, which are usually separated by 9 base pairs (or one helix turn).

1.5.4 Stem cell Leukaemia Protein (SCL)

SCL is a tissue-specific basic helix-loop-helix transcription factor which was identified because of its persistent susceptibility to reciprocal chromosomal translocation between chromosomes 1 and 14 in T-cell Acute Lymphoblastic Leukaemia. Hence, this nuclear protein is also known as TAL1 (Robb and Begley 1996).

SCL is expressed during normal haematopoiesis in haematopoietic progenitor cells, erythroid, megakaryocytic and mast cell precursors. Apart from this, it is also expressed in endothelial cells. SCL functions at various early and late stages during the establishment of the haematopoietic system. Specifically, it is essential for specification and self-renewal of HSCs and terminal differentiation of erythrocytes. The SCL oncoprotein is essential for the production of embryonic red blood cells in the mouse yolk sac and for the formation of all haematopoietic lineages. SCL−/− embryonic stem (ES) cells fail to differentiate into all

haematopoietic lineages and SCL−/− mice die during embryogenesis between E9-E10.5 due to lack of the red blood cells (Porcher et al. 1996). These observations make it obvious that SCL plays a crucial role both in primitive and definitive haematopoiesis (Orkin 1995; Shivdasani and Orkin 1996; Lecuyer and Hoang 2004). SCL is a direct transcriptional activator in leukaemia involved in chromosomal translocation, interstitial deletions or mutations resulting in its aberrant expression leading to T-ALL which is characterized by accumulation of immature lymphoid progenitor cells (O'Neil et al. 2001). SCL is an obligate factor for cell differentiation towards the erythroid lineage possessing a Helix-loop-Helix (HLH) protein interaction domain (of ~50 residues) preceded by a basic (b) DNA binding region (of ~10 residue). This bHLH domain primarily forms helical homodimers on its own, however, it preferentially forms hetero-dimers with the bHLH domains of E2A proteins (E12 or E47). Normally, E2A proteins act as transcription activators and they bind to the E-box motif present in the regulatory regions of its target genes. However, heterodimerization with SCL represses transactivation by E47 upon DNA binding. Nevertheless, under physiological conditions, where inhibitory HLH Id (inhibitors of DNA binding) proteins are expressed, the SCL-E2A heterodimer is a significantly effective transcriptional activator over E2A homodimers as Inhibitors of differentiation (Id) proteins inhibit DNA binding of the SCL-E2A heterodimers (Voronova and Lee 1994; O'Neil et al. 2001; Ryan et al. 2008). Apart from E2A, SCL also interacts with a cysteine rich protein, LMO2. Like SCL, LMO2 can also be ectopically expressed in T cells. In this case, the two proteins are synergistically responsible for the onset of T-cell leukaemias. They do so by sequestering E2A thereby disrupting normal lymphocyte development resulting in accumulation of immature T-cells leading to T-cell leukaemias (Ryan et al. 2008).

Generally, DNA binding is indispensable for the function of the bHLH family of proteins. However, it has been shown that SCL possesses functions that are independent of its DNA binding activity. Indirect DNA binding activities are sufficient for HSC specification during development whereas direct DNA binding functions with transcriptional auto-regulation govern terminal maturation processes of red blood cells and megakaryotic cells. SCL functions by modulating gene expression in these cells positively and negatively by recruiting cofactors (proteins that assist transcription factors for mediating transcription activation or repression) and chromatin remodeling proteins (Kassouf et al. 2008).

1.5.5 E2A Proteins

E-proteins are transcription factors that belong to the basic helix-loop-helix family of proteins and function in regulatory processes during differentiation, cell growth and survival and tissue- specific gene expression. The E-protein gene family in mammals comprises three genes: E2-2 gene, HeLa E-box binding protein (HEB) and E2A gene. These genes are expressed during the development of haematopoietic, muscle and nervous tissues. Of these genes, the E2A gene encodes two proteins, E12 and E47, as a result of alternative splicing through differential usage of the exon encoding the bHLH domain. E-proteins have a carboxy terminal protein dimerization bHLH domain and a transactivation domain at their amino-terminal regions. Like SCL, E2A proteins form dimers via the HLH domain and recognize specific DNA sequences called E-box motifs through the basic domain. E2A proteins heterodimerize with SCL, and this association plays an important role in haematopoiesis. Id proteins (HLH proteins with no basic DNA-binding domains) inhibit E- proteins' functions by binding to their bHLH domain through the HLH region (Lazorchak et al. 2005).

E2A proteins are expressed in developing lymphocytes and are known to play a fundamental role as transcriptional regulators at various stages of lymphopoiesis. It has been shown by gene targeting studies that the absence of E2A proteins leads to arrest of early B cell development prior to the commencement of immunologlobulin (Ig) heavy chain rearrangements. E2A proteins' expression is prevalent in HSCs, erythroid progenitors and erythroid cells, and they are completely absent once these cells differentiate into myeloid precursors cells (Greenbaum and Zhuang 2002). Their expression is again significantly upregulated during B-lineage commitment in the bone marrow in a dosage dependent manner (Zhuang et al. 2004). Once B-lineage commitment in the bone marrow begins, E2A activates EBF expression. This is followed by synergistic action of E2A and EBF genes, which then upregulates B-cell specific genes. Out of the two E-proteins, E47 is specifically expressed in B- cells. Both E12 and E47 have a high sequence homology and a minor difference in their DNA binding capabilities. However, unlike E47, E12 possesses an inhibitory region at the amino-terminal of the bHLH domain which reduces its dimerizing efficiency rendering it incapable of forming homodimers and making it inefficient in DNA binding activity. Hence, E12 proteins can only form heterodimers with E47 or other bHLH members (Sun and Baltimore 1991) .

1.5.6 LMO proteins

The LIM proteins comprise two families: LMO (LIM only protein) and LIM-HD (LIM-homeodomain). The LIM domains of these proteins are zinc finger domains which bind to two zinc ions facilitating protein-protein interaction. The name LIM comes from the first three proteins in which this motif was discovered: Lin-11, Isl1 and Mec-3. LIM-HD proteins consist of a conserved DNA binding homeodomain and an unstructured C-terminal domain which is diverse in terms of its sequence. LIM-HD proteins act as transcription factors owing

to their DNA binding capability whereas LMO proteins regulate transcription through their association with other transcription factors. This family of proteins plays a critical role in determining cell fate and in developmental processes (Matthews et al. 2008). The LMO proteins are nuclear transcription co-regulators and regulate gene transcription by acting as 'linkers' or 'scaffolding' proteins thereby mediating protein-protein interactions. All LMO proteins possess two tandem LIM domains and no other functional domains. These LIM domains act as a docking site for assembly of multiprotein complexes. LMO proteins bind with high affinity to the nuclear protein LDB1 and this interaction takes place between the N-terminal LIM domain of LMO2 and the 30-residue LID domain of LDB1. So far, four LMO proteins have been identified: LMO1, LMO2, LMO3 and LMO4. In fact, the (LIM-only) family of genes was originally unravelled by the association of LMO1 with the chromosomal translocation t(11;14)(p15;q11). This was followed by the characterization of two more related genes LMO2 and LMO3, of which LMO2 was located at the junction of chromosomal translocation t(11;14)(p13;q11) in T-ALL (Sang et al. 2014).

The LMO proteins are cysteine-rich proteins which consist of two closely spaced LIM domains connected via a small linker with no apparent interaction between the two domains (Dawid et al. 1998). These LIM domains do not show any direct DNA binding activity despite their similarity to GATA1 zinc fingers and their action is restricted to protein-protein interaction (Rabbitts 1998). This is evident from the ability of LMO2 to interact with other proteins; for instance, LMO2 binds to GATA1, LDB1, E47 and SCL in erythroid cells (Valge-Archer et al. 1994; Osada et al. 1995; Ryan et al. 2008; El Omari et al. 2013). LMO2 and SCL are two essential components playing a crucial regulatory role at various stages of haematopoiesis. LMO2 is expressed in erythroid lineages and LMO2−/− mutations also lead to complete failure of yolk sac erythropoiesis causing embryonic lethality at around E10.5

(Warren et al. 1994). The interaction between LMO2 and SCL occurs through a conserved interface in the loop and helix2 of SCL, which in turn stabilizes LMO2 and prevents its degradation by the proteasome. LMO2 levels indeed fall under the control of SCL (Lecuyer et al. 2007). Nevertheless, LMO2 is known to act as a bridging molecule assembling the pentameric DNA-binding complex comprising SCL, E47, LDB1 and GATA1 (Wadman et al. 1997) (**Figure 1.9**). Therefore, over-expression of LMO2 by chromosomal translocation or transgenesis leads to accumulation of aberrant multiprotein complexes responsible for the onset of T-cell leukaemias (Ryan et al. 2008).

However, LMO2 is also essential for erythroid development. It is expressed in the nucleus of the erythroid cells of primitive and definitive erythroid lineages (Warren et al. 1994).

Like LMO2, LMO1 is also an oncogene involved in chromosomal translocations in T-cells leading to T-ALL (Rabbitts 1998). LMO1 is normally not expressed in T-cells but the translocation leads to its ectopic expression. Moreover, it also interacts with SCL (Valge-Archer et al. 1994), LDB1(Valge-Archer et al. 1998) and GATA3 (Ono et al. 1998) in T-cell leukaemia as opposed to LMO2 which interacts with SCL, LDB1 and GATA1. In fact, SCL and LMO1 collaborate to carry out expansion of thymocyte progenitors and inhibit later stages of differentiation resulting in formation of self-renewing leukaemia-initiating cells. This collaboration also favours attainment of activating Notch1 mutations and the hyperactive Notch1 allele accelerates leukaemia initiated by SCL-LMO1 (Tremblay et al. 2010). Moreover, SCL lacking its transactivation domain (domain of transcription factors responsible for mediating gene expression) has been shown to cooperate with the over-expressed LMO1 causing T- cell malignancies in mice. This shows that SCL exerts its effect

through a dominant-negative mechanism (Aplan et al. 1997). The interaction between LMO1 and SCL occurs through the LIM domain and bHLH domain of the former and latter respectively (Wadman et al. 1994). Because of the presence of the bivalent histone marks and the 3' flanking region containing transcriptional enhancer, the LMO1 promoters are primed for ectopic expression in T-cell leukaemias. Following the activation of the promoters, the enhancer is bound by SCL and GATA3 which contribute to the development and maintenance of T-ALL (Oram et al. 2013).

Hence, LMO1 and LMO2 proteins sit at the breakpoints of chromosomal translocation causing their ectopic expression in T-cells, which lead to T-ALL either by facilitating the formation of the SCL DNA binding complex or by sequestering the components of this complex preventing them to carry out their normal function (Valge-Archer et al. 1998).

1.5.7 LIM Domain Binding Protein (LDB1)

LDB1 is a nuclear protein which is ubiquitously expressed and was primarily identified as a cofactor for LIM-homeodomain and LIM-only proteins. It is also known as NLI (Nuclear LIM domain Interactor). It mediates formation of multiprotein complexes involving various transcription factors and their co-regulators. LDB1 acts as an interface for the interaction of other proteins and does not have any enzymatic or DNA binding activity of its own. Based on sequence homology, the LDB family can be classified into two categories: LDB1 and LDB2. All LDB proteins possess a conserved nuclear localization sequence and hence are located mainly in the nucleus. Structurally, LDB family protein members have an amino terminal homodimerization domain and a carboxy terminal LIM Interaction domain (LID), which interacts with LIM domains of LIM-HD and LMO proteins. LDB1 has varied cellular

functions including in neural patterning and in development. *Ldb1* gene aberrations in mice have demonstrated its importance in developmental processes. Indeed, LDB1-null mice show severe anterior-posterior patterning defects resulting in truncated head structures and lack of heart anlage. LDB1 overexpression in immature erythroid cells obstructs their terminal differentiation. Finally, LDB1 also acts as a transcriptional regulator (Matthews and Visvader 2003).

LDB1 interactions with other proteins occur via its structural domains: for instance, homodimerization through its amino terminal self-association domain enhances its interaction with other proteins. The homodimerization domain of LDB1 enables its oligomers to bind to at least two LIM proteins (Matthews et al. 2008). The ability of LDB1 to act synergistically with other proteins is likely to render it capable of determining the development of various cell types in the central nervous system through cell-type specific protein-protein interactions (Matthews and Visvader 2003).

LDB1 interacts with LMO proteins via a 38-residue region called LID (LIM interacting domain) (Matthews et al. 2008). Recombinant forms of the LMO proteins have limited stability and are prone to aggregation thereby preventing their structural characterization. So, in order to increase the stability of the LMO proteins, their LIM domains can be tethered to the $LDB1_{LID}$ by a flexible linker. Using such fusion protein constructs, structures of LMO2 (El Omari et al. 2013) and LMO4 (Deane et al. 2004) proteins were determined to an atomic resolution which gave very valuable insights into the way these LMO proteins interact with LDB1. For example, the prevailing model of T-ALL suggests that ectopic expression of proto-oncogenes SCL and LMO2 in T-cells causes malignancy by sequestration of the E-

proteins (Ryan et al. 2008) with the latter being important regulators of normal T-cell differentiation (Quong et al. 2002). However, the molecular details of this process remained unclear. The crystal structure of the tetrameric complex comprising of SCL/E47 bHLH: LMO2: $LDB1_{LID}$ at 2.9 Å shows that upon binding to the SCL/E47 bHLH heterodimers, LMO2 strengthens the heterodimers by inducing formation of new hydrogen bonds and salt bridges. This results in a rotational movement that weakens the interaction of the heterodimer with DNA. In this way, the crystal structure of the tetrameric complex revealed how the E-protein sequestration model functions (El Omari et al. 2013). Also, the crystal structure of LMO2:$LDB1_{LID}$ at 2.4 Å gave a deep insight into how LDB1 interacts with LMO2. It showed that LDB1-LID forms an extended structure around LMO2 which binds to both of its LIM domains forming hydrogen bonds by backbone-backbone β-strand interactions (**Figure 1.10**). Apart from this, it also forms a network of hydrophobic and electrostatic interactions with LIM1 contributing to all of the electrostatic interactions (El Omari et al. 2011).

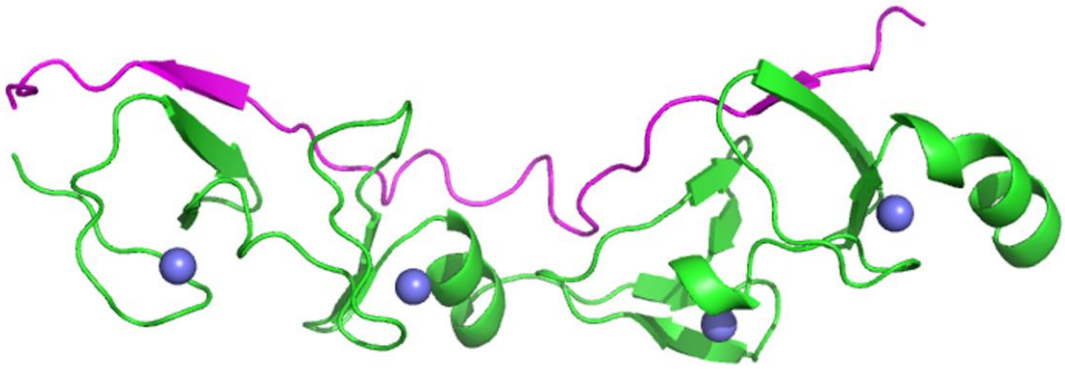


Figure 1.10: X-ray crystallography structure of LMO2:$LDB1_{LID}$ at 2.4 Å resolution. Ribbon diagram displaying the LDB1LID (magenta) forming an extended structure around LMO2 (green) there by binding with both the LIM domains of the latter. The blue balls represents Zn atoms [Taken from (El Omari et al. 2011)].

1.5.8 GATA1

The GATA family contains six different transcription factor proteins, GATA1 to GATA6. These transcription factors have been characterized on the basis of their ability to bind to the consensus DNA sequence (A/T)GATA(A/G) through their two typical C4 (Cys-X_2-Cys-X_{17}-Cys-X_2-Cys) zinc finger motifs. This characteristic DNA binding domain is highly conserved across all the family members. The family has been categorized further into two subfamilies: the haematopoietic group, which consists of GATA1, GATA2, GATA3, which are essential to impart lineage specification for the formation of mature blood cells from HSCs; and the non-haematopoietic group, which contains GATA4, GATA5, GATA6. GATA1 is the founding member of this family as it was first discovered because of its binding specificity to the β-globin 3' enhancer. It favours erythroid lineage and is expressed in both progenitors and differentiated erythroid cells, megakaryocytes, eosinophils and mast cells (**Figure 1.6**) (Ferreira et al. 2005).

GATA1 expression is essential for erythroid differentiation and this is evident from the fact that GATA1 deficient embryonic stem cells form all other types of blood cells but not mature red blood cells. Furthermore, GATA1 null mice display a block in erythropoiesis at the pro-erythroblast stage with mouse embryos dying in mid gestation period between embryonic day E10.5 and E11.5 due to severe anaemia because of the inability of erythroid precursor cells to survive and mature into red blood cells (Pevny et al. 1995; Fujiwara et al. 1996; Ferreira et al. 2005). GATA1 supports viability and maturation of erythroid precursors by averting apoptosis (Weiss and Orkin 1995). This evidence shows that GATA1 is critical for erythropoiesis. Conditional depletion of GATA1 in megakaryocytes resulted in the accumulation of megakaryocyte progenitors and a reduction in platelets, with the ones present showing abnormal morphology (Shivdasani et al. 1997). GATA1 site deletion at the

promoter also led to selective loss of eosinophils (Ferreira et al. 2005). GATA1 possesses three functional domains: N-terminal activation domains and N- and C- Zinc fingers. The C-terminal zinc finger is used for binding to the GATA box DNA sequence whereas the N-finger facilitates stabilization and specificity of this DNA binding along with mediating formation of protein complexes with cofactors. In short, for primitive erythropoiesis, the transactivation domain or N-finger is dispensable; and for definitive erythropoiesis, all the three domains are essential (Ferreira et al. 2005).

1.5.9 Role of Pentameric complex in erythropoiesis

Lineage-specific transcription factors play an important role in red blood cell development. SCL, LMO2 and GATA1 proteins of the pentameric complex play a crucial role in erythropoiesis. This is evident from the fact that homozygous mutations of all three genes result in the embryonic lethality between embryonic days E9-E11.5 due to complete lack of formation of red blood cells (Warren et al. 1994; Porcher et al. 1996; Ferreira et al. 2005). In T-ALL, SCL and LMO2 proto-oncogenes are transformed into oncogenes as a result of translocation. In addition, the pentameric complex SCL: E47: LMO2: LDB1: GATA1, binds to unique, bipartite DNA motif consisting of an E-box (CAGGTG) and GATA site separated by ~9bp (Wadman et al. 1997).

The process of haematopoiesis is partly regulated by the binding of sequence-specific transcription factors to the regulatory regions of target genes in different haematopoietic compartments. These transcription factors associate with various coactivators and corepressors to activate or repress target gene transcription. SCL, which is a core component of the pentameric complex, plays a dual role serving as a transcriptional activator or repressor

in a context-dependent manner. It accomplishes this through recruitment of corepressors such as ETO2, mSin3A, HDAC1/2 or LSD1 and coactivators like P300, PCAF. SCL is associated with LSD1, HDAC1 and HDAC2 in erythroleukaemia and T-cell leukaemia cells. The enzymatic activity of the SCL associated LSD1/HDAC1 complex is down-regulated during early erythroid differentiation; for instance, SCL recruits LSD1 to the silenced p4.2 promoter (regulating gene coding for human erythrocyte protein 4.2 which is an important component of the erythrocyte membrane regulating its stability and flexibility) in undifferentiated MEL cells but not during their differentiation. Upon induction of erythroid differentiation, the SCL-LSD1 interaction is disrupted leading to release of LSD1 from SCL target genes. The promoter then becomes enriched in H3K4me2 marks and becomes active. This shows that SCL precludes differentiation of erythroid progenitors by recruitment of LSD1 and switches to a transcriptional activator during differentiation of erythroid cells by reducing its interaction with LSD1. Therefore, dynamic regulation of the association of SCL and LSD1 may determine the onset of the erythroid differentiation programme (Hu et al. 2009a). The SCL-LSD1 interaction is modulated by phosphorylation of serine 172 in SCL. This phosphorylation event destabilizes the interaction and results in H3K4 hypermethylation of target promoters and consequent activation of the target genes (Li et al. 2012).

SCL has also been shown to be associated with the corepressors mSin3A and HDAC1 in MEL and human T-ALL cells, however, this association was perturbed upon induction of differentiation of MEL cells by DMSO. In fact, the bHLH domain of SCL was determined to be both necessary and sufficient for this association. This describes how SCL carries out transcriptional regulation by association with corepressors (Huang and Brandt 2000). SCL also recruits another corepressor, ETO2, through interaction with E2A in the erythroid cells. ETO2 represses SCL target genes leading to expansion of erythroid progenitors. However, at

the onset of erythroid differentiation, the stoichiometry of ETO2 in the SCL complex changes and the interaction is lost in the late stages of erythroid differentiation resulting in expression of genes vital for terminal erythroid maturation. This may occur by dissociation of ETO2 from SCL or down-regulation of the ETO2 expression. ETO2 may be carrying out this function by recruiting other corepressors or preventing the DNA binding (Schuh et al. 2005; Goardon et al. 2006).

In order to serve its role of transcriptional activator, SCL also associates with coactivators P300 and P300/CBP-associated factor (P/CAF), both of which are histone acetyltransferases (HATs). Moreover, the interaction between SCL and P300 is direct (Huang et al. 1999). Upon erythroid differentiation, these interactions increase thereby increasing the acetylation of SCL which positively affects the transcriptional activity of SCL in two ways: by significantly increasing its DNA binding activity and destabilizing the association of SCL with corepressor mSin3A, specifically by P/CAF. Besides, the latter enhances SCL mediated transcription by direct acetylation of local promoter chromatin (Huang et al. 1999; Huang et al. 2000). In a nutshell, SCL recruits corepressors during early stages of erythropoiesis which are then replaced by coactivators during the differentiation process for the transcription of SCL target genes.

GATA1 associates with a transcriptional coactivator CREB-binding protein (CBP) which is a HAT. CBP regulates the transcriptional activity of GATA1 by directly acetylating the latter at two highly conserved lysine-rich motifs near the zinc finger domains. However, the acetylation has no effect on the DNA-binding activity of GATA1. Expression of an adenovirus oncoprotein E1A interferes with the CBP activity which in turn fails to acetylate

GATA1. The failure of acetylation of GATA1 leads to inhibition of erythroid differentiation thereby abrogating the expression of GATA1 target genes like α- and β- globin genes. Hence, CBP stimulates GATA1 activity in an E1A sensitive manner. It could be doing this by (i) inducing conformational change in GATA1, resulting in exposure of its activation domain, or (ii) by serving as a link between GATA1 and the transcriptional machinery, or (iii) by acetylating histones at the target genes promoters thereby maintaining an open chromatin configuration poised for transcription activation (Blobel et al. 1998; Hung et al. 1999). The physical interaction between GATA1 and FOG1 (Friend of GATA1) is essential for erythroid cell maturation. It has been shown that FOG1 modulates repressive functions of GATA1 by recruiting NuRD repressor complex (Nucleosome Remodeling and Deacetylase) leading to histone deacetylation (Hong et al. 2005).

In this way, the complex interacts with various cofactors including chromatin remodelling proteins such as histone acetyl transferases P300 and PCAF and histone methyl demethylase LSD1 to regulate gene expression in erythropoiesis (Hu et al. 2009b) (**Figure 1.11**).

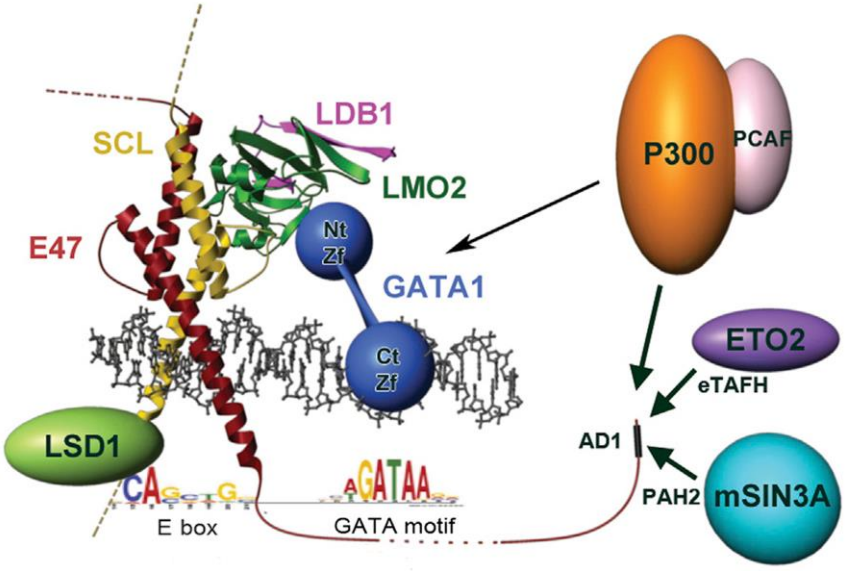


Figure 1.11: Model representing the Pentameric Complex. The pentameric complex interacts with the cofactors including histone modifying enzymes such as histone methyl transferases P300, PCAF and histone demethylase LSD1 [Taken from (El Omari et al. 2013)].

1.6 JARID1 family of proteins

JmjC proteins derived their name from the presence in their sequence of a jumonji domain. The name jumonji (which means cruciform in Japanese) was originally derived from a mouse mutation that affected neural tube development and produced a cruciform-shaped neural tube. The JARID1 family of proteins is one such JmjC domain containing class of lysine specific demethylase enzymes that catalyse the removal of methyl groups from lysine 4 of histone H3. Drosophila, worms and yeast possess only one ortholog of this protein, Lid, rbr-2 and Jhd2 respectively whereas in mammals, this superfamily of proteins comprises of JARID1A (RBP2/KDM5A), JARID1B (Plu-1), JARID1C (SMCX) and JARID1D (SMCY). These four proteins contain the catalytic JmjC domain responsible for histone demethylation and the JmjN domain, which is important for maintaining the structural integrity of the JmjC domain. JARID1 members also contain an AT-rich-interacting domain (ARID) responsible for their

DNA binding activity and a C5H2C zinc finger supposedly involved in protein-protein interactions. Finally, all four JARID1 proteins have plant homeodomain-associated protein domains (PHDs), which are protein-protein interaction domains and histone-methyl-lysine binding motifs. The JARID1 family differs in the number of PHD domains; JARID1A and 1B contain three, whilst JARID1C and 1D contain two PHDs (**Figure 1.12**) (Secombe and Eisenman 2007).

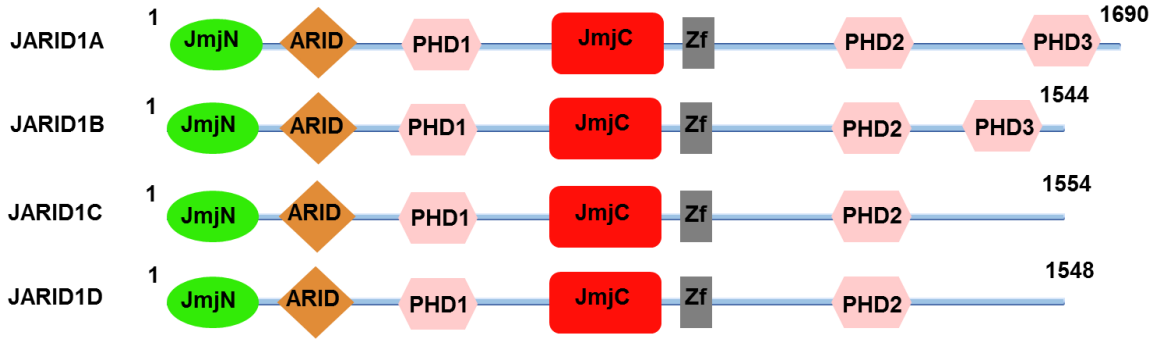


Figure 1.12: Structure of the JARID1 proteins. JARID1 proteins contain several conserved domains: The JmjC domain (red) has demethylase activity, JmjN (green) regulates the JmjC activity, ARID domain (brown) has DNA binding activity, Zinc finger (gray) and PHD domains (pink) are predicted to be involved in protein-protein interactions.

JARID1A is a ubiquitously expressed nuclear protein comprised of 1690 amino acids and a molecular weight of ≈200kDa. It specifically demethylates di- and tri- methylated H3K4. Human JARID1A is evolutionarily conserved, with homologs in mouse (JARID1A), *Drosophila* (LID), *C. elegans* (RBR-2) and *S. pombe* (LID2). Its catalytic domain, JmjC, is highly conserved throughout different species and across the JARID1 proteins. Overall, there is a 52, 50 and 49% sequence identity between JARID1A and JARID1B, JARID1C; and JARID1D respectively (Secombe and Eisenman 2007).

JARID1B also known as PLU-1/KDM5B, is a H3K4 demethylase that demethylates H3K4me2/3 but not H3K4me1. It was first isolated because of its overexpression in breast cancer cell lines. It was also found to be expressed in 90% of invasive ductal carcinomas. JARID1B has been identified as a crucial enzyme affecting the proliferative capacity of breast cancer cells by repression of tumour suppressor genes like BRCA1. Hence, down-regulation of JARID1B in breast cancer cells leads to a decline in the tumour forming potential of these cells in xenograft mouse models (Yamane et al. 2007).

JARID1C/ KDM5C/ SMCX is an H3K4me2/3 demethylase and its deregulation has mostly been implicated in X-linked mental retardation. It is a part of the REST co-repressor complex and is required for REST mediated repression during neurogenesis (Tahiliani et al. 2007). However, it could also be linked to renal cell carcinoma as it was found mutated in 3% of clinical tumour samples (Dalgliesh et al. 2010).

JARID1D/ KDM5D/ SMCY is homologous to JARID1C and is located on the Y-chromosome. It has not been specifically linked to any human diseases yet. However, deletion analysis of Y-chromosome specific genes from prostate cancer specimens showed that 52% of the patients had JARID1D deleted (Perinchery et al. 2000).

1.7 JARID1A

JARID1A was originally described as a nuclear protein that binds to the retinoblastoma protein (pRB) leading to promotion of cellular differentiation, hence it is also known as retinoblastoma binding protein 2 (RBP2). It is also called KDM5A (lysine specific demethylase 5A, here after called JARID1A) as more recently its function as a histone lysine

demethylating enzyme has become evident. It demethylates H3K4me2/3/1 processively, but fails to initiate demethylation of H3K4me1 (Klose et al. 2007). The JARID1A protein possesses three PHD domains. An interesting observation was that JARID1A-PHD3 binds to H3K4me3, which is the substrate for JmjC domain probably acting as a 'boundary factor' preventing demethylation whereas JARID1A-PHD1 binds to H3K4me0 which is the product of JmjC domain possibly acting as a 'seed' to propagate demethylation (Upadhyay et al. 2011). Specifically, the PHD3 domain of JARID1A was found to be responsible for binding to H3K4me3 marks. The X-ray crystal structure of the PHD3 domain of JARID1A in complex with an H3K4me3 peptide shows that tryptophan residues at position 1625 and 1635 facilitate this interaction (mutations in these residues resulted in disruption of the interaction). Both of these tryptophan residues are evolutionary conserved across JARID1A and JARID1B homologues in human, mouse, chicken and fly (**Figure 1.13**). The X-ray crystallographic structure of JARID1A-PHD3 with H3K4me3 clearly shows that the binding is facilitated by three events: 1) pairing between the two anti-parallel β-sheets of the H3 backbone and JARID1A-PHD3; 2) formation of a hydrophobic cleft by the two orthogonally aligned tryptophan residues, W1625 and W1635, to anchor the H3K4me3 side chain which is stacked between the two indole rings of the tryptophans with intermolecular contacts; and 3) positioning of the H3R2 residue in an acidic pocket formed by Glu1627/Asp1629/Asp1633 (**Figure 1.14**) (Wang et al. 2009). Since H3K4me3 is generally found in active gene regions, JARID1A has been suggested to function as a general transcriptional repressor.

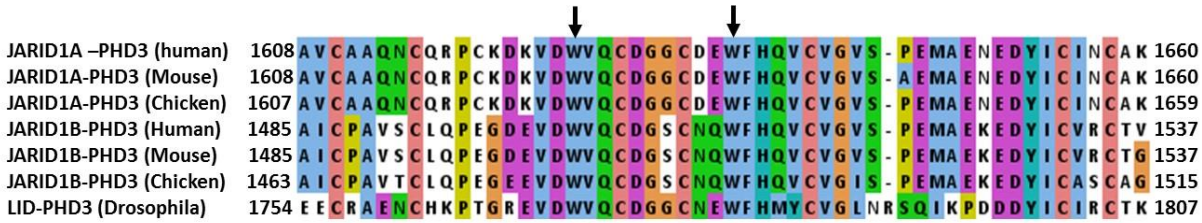


Figure 1.13: Sequence alignment of the third PHD domains of JARID1 proteins. Sequence alignment of the third PHD domains of JARID1 homologues across human, mouse, chicken and fly was carried out using ClustalW and represented using Jalview 2.8.1 with Clustalx colour scheme. The two conserved tryptophan residues which are critical for H3K4me3 mark identification are indicated with the black arrows on the top.

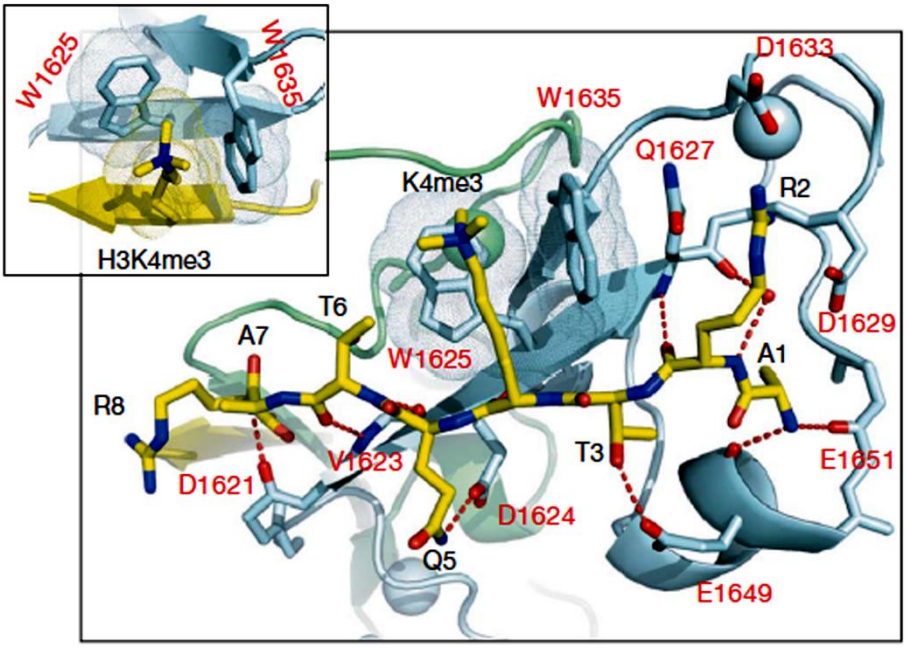


Figure 1.14: X-ray crystallography structure of JARID1A-PHD3 in complex with H3K4me3 at 1.9 A resolution. JARID1A-PHD3 is shown in cyan and H3K4me3 in yellow with the numeration for each represented in red and black respectively. The blown up view shows the magnified picture of positioning of H3K4me3 in the hydrophobic cleft formed by two tryptophan residues of JARID1A-PHD3, W1625 and W1635 [Taken from (Wang et al. 2009)].

The JmjC domain of JARID1A is responsible for its demethylase activity and mutation in the iron binding site of this domain (H483A) leads to abrogation of the activity (Klose et al.

2007). In a similar study, it was shown, when two of the three residues believed to form the Fe (II) binding pocket (histidine 483 and glutamic acid 485) were mutated to glycine and glutamine respectively, the demethylase activity of JARID1A was lost. These residues are conserved across all four JARID1 proteins as shown in **Figure 1.15** (Christensen et al. 2007).

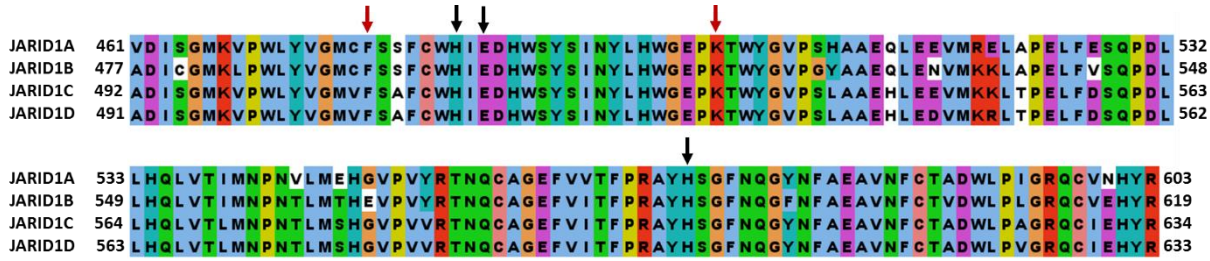


Figure 1.15: Sequence alignment of JmjC domains of all four JARID proteins. Sequence alignment of JmjC domains of mouse JARID1A, JARID1B, JARID1C and JARID1D was carried out using ClustalW and represented using Jalview 2.8.1 with Clustalx colour scheme. The conserved histidine and glutamic acid residues which are essential for Fe (II) binding are indicated with black arrows on the top and the conserved phenylalanine and lysine residues which are involved in α-KG binding are indicated with red arrows on the top. All these residues are necessary for their demethylase activity.

The ARID domain of JARID1A has been shown to bind to a CCGCCC DNA motif. Deletion of the ARID domain led to the failure of the enzyme in reducing H3K4me3 levels indicating that intact ARID domain is critical for JARID1A's demethylase activity. An NMR solution structure of the ARID domain showed that it consists of six helices (H1-H6) and two loop regions (L1 and L2) with H2, H3 and H4 forming a prominent U shape. Chemical shift perturbation upon DNA binding suggested that the binding interface of ARID is possibly located on the L1 and the helix-loop-helix motif formed by H4, L2 and H5 (**Figure 1.16**). Mutational analysis using EMSA showed that L2 and H5 is majorly involved in the sequence-specific DNA binding, specifically Lysine 152. Moreover, mutational analysis using the transactivation assay employing a reporter gene system based on the BRD2

promoter (a target gene of JARID1A) showed that the ARID domain residues necessary for identifying the CCGCCC motif are essential for JARID1A to regulate transcription of its target genes (Tu et al. 2008).

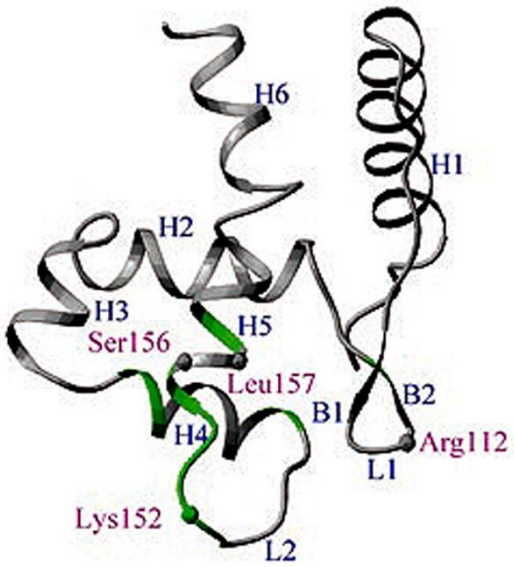


Figure 1.16: NMR solution structure of ARID domain of JARID1A. A ribbon diagram of the JARID1A ARID domain is shown with the residues at the DNA binding interface labelled and Lys 152 on loop 2 shown in green. The latter contributes majorly in the DNA binding activity [Taken from (Tu et al. 2008)].

Another good example of how atomic structures can provide important information for understanding mechanisms of action is the crystal structure of KDM7A, which demethylates H3K9me2/1, H3K27me2/1 and that of KDM7B which demethylates H3K9me2/1. Both these proteins possess N-terminal PHD domains which bind to H3K4me3 and C-terminal JmjC domains that demethylate H3K9me3 or H3K27me3. KDM7B assumes a bent conformation, which enables both of its domains to bind their targets thereby increasing substrate interaction, whereas KDM7A forms an extended conformation which prevents access of JmjC to the H3K9me2, if the PHD domain is bound to H3K4me3 (Upadhyay et al. 2011).

The following table shows the list of histone lysine demethylases whose domain structures are solved (**Table 1.3)**.

Table 1.3: Lysine demethylase with known structures with their PDB IDs [List constructed based on the structures deposited in Protein Data Bank (PDB) and adapted from (Verrier et al. 2011)].

Protein family	Systematic name	Substrate specificity	Structures solved
LSD	KDM1A/LSD1	H3K4me2/1 H3K9me2/1	LSD1 Full length (2H94) LSD1-Co-REST (2IW5)
	KDM1B/LSD2	H3K4me2/1	With H3 (4GU0)
JmjC	KDM2A/FBXL11A/JHDM1A	H3K36me2/1	FL with & w/o α-KG-Fe^{2+} (2YU1, 2YU2)
	KDM2B/FBXL10B/JHM1B	H3K36me2/1	Zn fingers (4o64)
	KDM3A/JMJD1A/JHDM2A	H3K9me2/1	none
	KDM3B/JMJD1B/JHDM2B	H3K9me2/1	JmjC domain (4C8D)
	KDM4A/JMJD2A/JHDM3A	H3K9me3/2 H3K36me3/2	With inhibitors (3RVH, 4GD4) With H3K36 peptides (2P5B)
	KDM4B/JMJD2B	H3K9me3/2 H3K36me3/2	With H3K9me3/pyridine 2,4 dicarboxylic acid (4LXL)
	KDM4C/JMJD2C/GASC1	H3K9me3/2 H3K36me3/2	JmjC domain (2XML) Tudor domain (2XDP)
	KDM4D/JMJD2D	H3K9me3	JmjC with H3K9me3 and α-KG (4Hoo)
	KDM4E/JMJ2E	H3K9me3/2	None
	KDM5A/ JARID1A/ RBP2	H3K4me3/2	ARID domain (2JXJ) PHD3-H3K4me3 (3GL6)
	KDM5B/JARID1B/PLU1	H3K4me3/2	ARID domain (2EQY) PHD1 (2MNY)
	KDM5C/JARID1C/SMCX	H3K4me3/2	ARID domain (2JRZ)
	KDM5D/JARID1D/SMCY	H3K4me3/2	ARID domain (2YQE) PHD1 (2E6R)
	KDM6A/UTX	H3K27me3/2	JmjC-NOG-Ni(II)-H3K27me3(3AVR)
	KDM6B/JMJD3/KIA0346	H3K27me3/2	With H3K27me3 (4EZH) Free (4EZ4)
	KDM7A/KIAA1718	H3K9me2/1 H3K27me2/1	JmjC domain (3KV9)
	KDM7B/PHF8/KIAA1111	H3K9me2/1 H3K27me2/1	JmjC domain (3K3N) PHD domain (1WEP)
	KDM8/JMJD5	H3K36me2	JmjC- α-KG (3UYJ)

1.7.1 Role of JARID1A in gene regulation, development and differentiation

JARID1A contributes to the process of transcription by interacting with pRB (Benevolenskaya et al. 2005), TBP (TATA-binding protein), p107 (Kim et al. 1994), nuclear receptors (Chan and Hong 2001), Myc (Secombe et al. 2007), Sin3B/HDAC (van Oevelen et al. 2008), Mad1 (Ge et al. 2010) and RBP-J (Liefke et al. 2010).

The tumour suppressor gene pRB carries out its function via the retinoblastoma pathway, which promotes cellular senescence by repressing transcription mediated by E2F that acts as a transcription suppressor in G1/S phase transitions of the cell cycle. pRB has been shown to co-operate with JARID1A and JARID1B for the demethylation of H3K4me3 on the E2F responsive target genes. This results in the establishment of a repressive state on the latter thereby aiding in cell cycle arrest, which is crucial for cell senescence (Chicas et al. 2012). Moreover, both JARID1A and pRB (Retinoblastoma protein) co-localize in the nucleoli and both these proteins interact and promote cellular differentiation. This is done by pRB, which neutralizes JARID1A thereby activating its target genes involved in maintenance of active chromatin state for differentiation (Benevolenskaya et al. 2005).

For cellular differentiation to proceed, cell cycle genes need to be permanently silenced. H3K4 demethylation by JARID1A is critical for repression of cell cycle genes and it also coordinates with E2F4 to carry out transcriptional repression during differentiation (Beshiri et al. 2012). In terminally differentiated skeleton muscle cells, E2F4 is shown to recruit Sin3B-JARID1A-HDAC to a subset of its target genes and the latter results in permanent gene silencing by removal of active acetylation and methylation histone marks. This event

coincides with the establishment of the repressive H3K27 methylation thereby varying the nucleosome structure and formation of heterochromatin leading to gene silencing (van Oevelen et al. 2008). Furthermore, Genome-wide analysis of JARID1A revealed its inhibitory effects on cell differentiation dependent cell cycle genes like PCNA (Lopez-Bigas et al. 2008).

Nuclear receptors are ligand-inducible transcription factors, which play a role in diverse cellular processes like differentiation, development and homeostasis. Interaction of nuclear receptors with JARID1A has been shown to be critical for optimal transcriptional activation and both N- and C- terminal regions of JARID1A are essential for this purpose (Chan and Hong 2001).

Myc is a proto-oncogene which regulates several of its target genes involved in cell growth by binding to the E-box DNA sequence via its bHLHZ domain (basic helix-loop-helix zipper). Binding of the JARID1A homologue LID has been identified to be essential for dMyc induced expression of Nop60B gene, which is a growth regulatory gene (Secombe et al. 2007). The Myc/Max/Mad proteins are bHLHZ transcription factors and whilst Myc/Max heterodimers bind to E-box to activate gene expression, Mad/Max heterodimers repress gene expression. Mad has been shown to recruit JARID1A to a Myc target, hTERT (telomerase reverse transcriptase) that codes for telomerase enzyme required for telomere elongation. JARID1A represses hTERT and could prevent malignant transformation of leukemic cells (Ge et al. 2010).

Also, JARID1A forms an integral part of the Notch-RBP-J complex. RBP-J is a transcriptional regulator important in the Notch signalling pathway, which controls various cellular processes such as proliferation, differentiation, and stem cell maintenance during embryogenesis. RBP-J acts as a repressor when not bound to Notch proteins and an activator when bound to Notch proteins. In absence of Notch signalling, RBP-J recruits JARID1A, which demethylates its target genes resulting in their repression (Liefke et al. 2010).

JARID1A has also been found to be physically and functionally associated with NuRD (Nucleosome Remodeling and Deacetylase complex) with its C-terminal region (JmjC-PHD3) and also sharing a number of target genes with that of CHD4, the catalytic subunit of NuRD (Nishibuchi et al. 2014).

JARID1A plays a non-redundant role in circadian oscillator function by forming a complex with CLOCK and BMAL1 transcription factors enhancing their transcriptional activity on Per2 promoter thereby influencing the circadian clock (DiTacchio et al. 2011).

The JARID1A homolog in *C. elegans* rbr-2 plays a role in vulva development and mutation resulting in the elimination of its entire JmjC demethylase domain leads to undeveloped vulva formation (Christensen et al. 2007). The intrinsic H3K27me3 activity of PRC2 (polycomb repressive complex2) exhibited by its EZH2 component is crucial for PcG-mediated gene silencing in mouse ES (embryonic stem) cells. The Kristian group showed by genome wide location analysis that JARID1A is linked with a large number of PRC2 target genes in mouse ES cells and that PRC2 recruits JARID1A to its target genes for

transcriptional repression during ES cell differentiation by reducing H3K4me3 levels and increasing H3K27me3 levels (Pasini et al. 2008).

1.7.2 Role of JARID1A in haematopoiesis and leukaemia

Evidence suggests that JARID1A might play a role in normal and malignant haematopoiesis. As mentioned earlier, JARID1A was discovered due to its association with pRb, a tumour suppressor protein, suggesting a possible connection of JARID1A with tumourigenesis. Peripheral blood analysis of JARID1A knock-out mice showed neutrophilia and analysis of their haematopoietic stem cell and myeloid precursor compartments revealed a significant decrease in the rate of apoptosis and enhanced survival and increased cycling of cells in these compartments (Klose et al. 2007).

Acute myeloid leukaemia (AML) is believed to emerge as a result of two types of mutations: Type-I mutations which impart proliferative/survival advantage and Type-II mutations which are usually chromosomal translocations involving formation of fusion proteins and causing loss of haematopoietic differentiation (de Rooij et al. 2013). JARID1A has been identified as a fusion partner gene of Nucleoporin-98 (NUP98) (which is a common fusion partner found in many leukaemias) in a subset of AML patients in which chromosomal translocation of NUP98 resulted in fusion of its amino terminus with the C-terminus of the PHD motif of JARID1A (van Zutven et al. 2006). NUP98/JARID1A fusion was also found in about 11% of the pediatric acute megakaryoblastic leukaemia cases in which overexpression of the Hox A and Hox B gene clusters was observed (de Rooij et al. 2013). Moreover, this fusion protein arrested haematopoietic differentiation in murine models and induced AML. This was carried out by abrogating the H3K4me3 binding potential of PHD3 finger, thereby enforcing an

active state on developmentally critical loci of haematopoiesis such as the genes coding for the transcription factors Hox, Meis1a, Gata3 and interfering with the transition cellular states which is critical for leukaemia initiation (Wang et al. 2009).

Several lines of evidence suggest an interaction between JARID1A and the SCL-containing multiprotein complexes. JARID1A has been shown to interact with the second LIM domain of LMO2 via its PHD3 domain (Mao et al. 1997). In T-cells, LMO2 act as a proto-oncogene whose expression results in the accumulation of immature thymocytes to overt leukaemia. Published data also show interactions of JARID1A with pRB (Defeo-Jones et al. 1991), Sin3 (van Oevelen et al. 2008) and histone deacetylase HDAC1 (Lee et al. 2009), all known partners of SCL [pRB (Vitelli et al. 2000), Sin3 (Huang and Brandt 2000) and HDAC1 (Huang and Brandt 2000)]. Altogether, this suggested a potential physical interaction between JARID1A and the SCL protein complexes. Also, there are shared activities at a functional level: JARID1A regulates cycling activity in the HSC and myeloid compartments increasing entry into G1 (Klose et al. 2007) and SCL regulates G0 to G1 progression in HSCs specifically by maintaining LT-HSCs in quiescence (Lacombe et al. 2010).

1.7.3 Role of JARID1A in other diseases

JARID1A has also been linked to other forms of diseases. In general, as H3K4me3 is an activation mark, JARID1A is considered to possess repressive function owing to its ability to demethylate H3K4me3/me2. However, in the presence of retinoblastoma protein (pRB), it seems to be functioning as a co-activator of transcription (Benevolenskaya et al. 2005). In line with this finding, it was also shown that the loss of demethylase activity of JARID1A

does inhibit tumourigenesis induced by deletion of tumour suppressor genes Rb1 and Men1 (Lin et al. 2011).

In several forms of lung cancers, recurrence of the disease is observed after the patients respond well to the anti-cancer drugs. This is attributed to the existence of a small population of cancer cells, which engage a dynamic survival strategy acquiring a reversible drug-tolerant state to prevent their complete extermination. Upon examination, elevated levels of expression of JARID1A were observed in these drug tolerant persisters obtained from exposure of an NSCLC (Non-Small Cell Lung Cancer)-derived cell line (PC9) with high concentration of drugs (Sharma et al. 2010). JARID1A was also found to be over-expressed in human lung cancer tissues obtained from NSCLC patients (Teng et al. 2013) (Qi et al. 2014). Moreover, JARID1A over-expression significantly increased tumour angiogenesis (Qi et al. 2014) and it was shown to be critical for proliferation, motility, migration, invasion and metastasis of lung cancer cells. All these roles of JARID1A in lung cancer were attributed to its DNA-binding and demethylase activities (Teng et al. 2013).

Like in lung cancer, JARID1A is over-expressed in breast cancer cells thereby conferring drug resistance to these cells. It was also shown that overexpression of JARID1A altered H3K4 methylation levels in these cells significantly, thereby repressing a set of genes like CDKI (cyclin dependent kinase inhibitors) that play a crucial role in cell cycle transition and mediating apoptotic cell death. Hence, knock down of JARID1A in these cells using shRNA caused growth inhibition and increased activity of EGFR inhibitors against them (Hou et al. 2012). Very recently, the Yan group has shown that JARID1A plays a crucial role in breast cancer metastasis and tumour progression. JARID1A upregulates metastasis-related genes

such as tenacin C (TNC), which is an essential component of the metastatic niche and is required for metastatic colonization of breast cancer cells in lungs by enhancing tumour cell migration and invasion, important stages of metastasis (Cao et al. 2014).

In summary, JARID1A has been implicated in different aspects of tumourigenesis and indifferent types of cancer. However, the molecular mechanisms involved require further investigation.

1.8 Aims and overview of this thesis

As described above, mounting evidence shows that chromatin remodeling proteins play a crucial role in regulation of gene expression via their interactions with various transcription factors. Earlier studies suggested that JARID1A could have a function in haematopoiesis and in the onset of leukaemia. This led us to investigate the role and molecular mechanisms of JARID1A in this context. We began by asking whether JARID1A is part of, or interacts with, the haematopoietic specific multiprotein complex nucleated by SCL. To unravel the interaction between the endogenous proteins, we employed several techniques such as co-immunoprecipitation and size-exclusion chromatography. As an interaction between JARID1A and the SCL complex was discovered, we then continued dissecting which domain(s) of JARID1A could be responsible for mediating this interaction. Furthermore, we attempted to determine several structures of JARID1A domains and complexes to shed light on the function of JARID1A. These included; 1) determining the structure of a complex of JARID1A and SCL complex interacting proteins; 2) investigating the structure of individual domains of JARID1A to unravel the structural motifs that enable this protein to carry out specific functions; 3) determining the structure of an alternative form of the SCL complex,

containing LMO1, which is also involved in causing T-cell leukaemia; 4) crystallization of LMO1 on its own.

Chapter 2 details the general methods and materials used for all the experiments described in this thesis and also provides a brief description of the theory of analytical ultracentrifugation and principles behind crystallization. More specific conditions used for each experiment are described later in their respective results chapters.

Chapter 3 describes the interaction studies performed predominantly in the mouse erythroid leukaemia (MEL) cell line, which closely recapitulates erythropoiesis upon induction of their differentiation. Different techniques which were used involved co-immunoprecipitation (Co-IP) and reverse Co-IP, size-exclusion chromatography and mass spectrometry. Real time PCR studies displayed correlation between expression patterns of all three genes: *Jarid1a, Scl* and *Gata1* at different stages of haematopoiesis in mouse bone marrow and fetal liver cells indicating the existence of possible interaction between the three. Initially, the interactions between endogenous JARID1A and SCL multiprotein complex were studied using Co-IP which revealed the interaction of JARID1A with SCL and GATA1. This was further confirmed by reverse Co-IP.

Chapter 4 highlights the more targeted approach used in the determination of specific domain(s) of JARID1A involved in the interactions with the SCL complex. The techniques used for this purpose were a mammalian two-hybrid assay, GST pull-down assay and analytical ultracentrifugation (AUC). It was determined that the PHD2 domain of JARID1A is responsible for its association with GATA1.

This was followed by attempts to carry out structural characterization of the interaction between the PHD2 domain and GATA1, and individual domains of JARID1A. Cloning, expression, purification and high-throughput crystallization trials of PHD1, PHD2 and PHD2+GATA1 are described in Chapter 5. In parallel, expression, purification and crystallization of the LMO1 protein and LMO1-E47/SCL bHLH complex was also carried out which are shown in Chapter 6. Of all the proteins tested, only LMO1 proteins yielded crystals which were subsequently examined for diffraction at Diamond Light Source. These crystallization conditions were further subjected to optimization to improve upon the diffraction of these crystals. The optimization screens performed are also explained.

Chapter 7 presents the conclusions and future work. In this chapter, all the results of the work carried out in this thesis are discussed in the context of the known functions of the SCL complex. Some future experiments which can be carried out to further understand the relevance of the described interactions between JARID1A and SCL/GATA1 have also been highlighted.

Chapter 2: Methods and Materials

2 Methods and Materials

2.1 Cell Culture

MEL (mouse erythroid leukaemia cell line) and HEK (human embryonic kidney cell line) cells were grown and maintained in Dulbecco's modified Eagle medium (DMEM) (Invitrogen) containing 10% fetal calf serum (FCS) (PAA), 1.2% L-glutamine (Gibco, Life Technologies) and 1.2% penicillin/streptomycin (Gibco, Life Technologies).

2.2 Nuclear Extract preparation

Nuclear extracts were isolated from HEK/MEL cells by washing the cells in 1×PBS and resuspending them in 10× bed volume of ice cold Buffer A (10mM HEPES pH 7.9, 1.5mM $MgCl_2$, 10mM KCl). The cells were then enabled to swell and burst by incubation on ice for 10 minutes, mixed occasionally and vortexed for 10 seconds. The nuclei were pelleted by centrifugation for 1 minute at 16,300×g and resuspended in 5× bed volume of ice cold Buffer C (20mM HEPES pH 7.9, 1.5mM $MgCl_2$, 420mM NaCl, 25% glycerol) and incubated for 1 hour at 4°C on a rotary shaker. Nuclear extracts were obtained by centrifugation for 5 minutes at 16,300×g on a table top centrifuge.

2.3 Co-immunoprecipitation

MEL cells were used for all the co-immunoprecipitation (Co-IP) assays. MEL cells are undifferentiated erythroid cells which are arrested at the proerythroblast stage and they recapitulate erythropoiesis upon induction of differentiation by DMSO. Hence, MEL cells can be used as a model system to study events occurring in erythropoiesis. For the Co-IP assay, 100µl Protein G Dynabeads (Invitrogen) (1:3-beads: protein) were equilibrated in dilution buffer (50mM Tris HCl pH 7.5, 0.3%NP40, 1× Protease Inhibitor tablet, 1mM DTT)

and blocked by resuspending in blocking buffer (0.2mg/ml BSA, 0.4mg/ml sonicated salmon sperm DNA, dilution buffer) by incubation at 4°C overnight on a rotary shaker.

On the second day, 1mg of nuclear extracts were diluted in dilution buffer and pre-cleared by incubation with corresponding normal IgG (depending upon the IP antibody) and 50µl of blocked beads for 1-3 hours at 4°C on a rotary shaker. For the IP, 5µg of the corresponding IP antibodies were added to the pre-cleared nuclear extracts and incubated at 4°C for 1 hour on a rotary shaker. Then, the blocked beads were washed three times with dilution buffer and added to the nuclear extract and antibody mixture and incubated at 4°C overnight on a rotary shaker. IgG was used as a negative control.

On the third day, the supernatant was removed and the beads were washed 5 times with 4× bed volume of wash buffer (50mM Tris HCl pH 7.5, 0.3%NP40, 1× protease inhibitor tablet, 1mM DTT, 150mM NaCl). The bound material was eluted by boiling the beads with 1× NuPAGE LDS sample buffer for 10 minutes and analysed by SDS-PAGE and Western blotting.

2.4 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples from Co-IP assays and from protein purification were analysed by SDS-PAGE, which separates denatured proteins on the basis of their molecular weight. Two different types of gels were used: self-poured Bio-Rad Mini-Protean gels were used for checking protein expression and purification, whereas 4-12% Bis-tris and 3-8% Tris-acetate gradient gels (Invitrogen) were used for Co-IP and mass spectrometry experiments. 1× Laemmli running buffer (0.025M Tris Base, 0.192M Glycine, 1% SDS) for running self-

poured gels using a BioRad Mini-Protean Tetra cell gel system. 1× MOPS SDS running buffer (NuPAGE, Invitrogen) for 4-12% Bis-tris gels and 1× Tris-acetate SDS running buffer (NuPAGE, Invitrogen) for 3-8% Tris-acetate gels were used for running precast gels using NuPage Novex gel system (Invitrogen). The recipe for casting self-poured gels is given in **Table 2.1**.

Table 2.1: Recipe for SDS-PAGE Gels; Gels were cast and run on a BioRad Mini-Protean gel system at 150-200 V in 1× Laemmli running buffer.

Resolving gel						
Gel (%)	**Acrylamide * (ml)**	**Gel Buffer** (ml)**	**H_2O (ml)**	**TEMED (ml)**	**10% APS (ml)**	**10% SDS (ml)**
15	2.5	1.3	1.1	0.01	0.05	0.05
18	3	1.3	0.59	0.01	0.05	0.05
* 30% w/v acrylamide, 0.8% bis acrylamide (37:5:1) **1.5M Tris-HCL, pH 8.8						

Stacking gel						
Gel (%)	**Acrylamide * (ml)**	**Gel Buffer** (ml)**	**H_2O (ml)**	**TEMED (ml)**	**10% APS (ml)**	**10% SDS (ml)**
4.5	0.5	0.38	2.1	0.08	0.03	0.03
* 1M Tris-HCl, pH 6.8						

Protein samples were mixed with 1× Laemmli loading dye (250mM Tris HCl, pH 6.8, 20%w/v glycerol, 10% SDS, 5% β-mercaptoethanol, 5% bromophenol blue), heated at 100°C for 10 minutes and spun down before loading onto the gel. 15µl of the protein samples were loaded onto polyacrylamide gels alongside appropriate molecular weight markers [Spectra high range multi-colour protein ladder (300-40kDa) and PageRuler prestained protein ladder (250-10kDa) from Thermo Fisher Scientific].

Gels were stained using Coomassie staining solution (0.05% Coomassie brilliant blue G250, 50% isopropanol, 10% acetic acid, MQ water) and destained using destaining solution (40% isopropanol, 10% acetic acid, MQ water).

2.5 Western blotting

The samples including Input (nuclear extract before pull down (PD)), IP (immunoprecipitated material), unbound (after pull down) and IgG (negative control) were divided into two and loaded onto 3-8% Tris-acetate gel (Invitrogen) (to be able to resolve JARID1A) and 4-12% Bis-tris gel (Invitrogen) (to be able to resolve GATA1, SCL, LDB1, LMO2). After electrophoresis, the proteins were transferred onto the Hybond PVDF membrane (GE Healthcare) using standard procedures and blocked in 5% dried skimmed milk (Marvel) in 1×PBST (1×PBS with 0.1% Tween20) overnight at 4°C on a shaker. The blocked membrane was probed with the primary antibody (refer to **Table 2.2**) and incubated for 2 hours at room temperature (RT) on a shaker. This was followed by 3 washes of the membrane in 1× PBST for 10 minutes. The membrane was then probed with the secondary antibody conjugated with HRP, incubated for 1 hour at RT, washed as before and developed using ECL Prime Reagent (GE Healthcare).

Table 2.2. List of antibodies with their dilutions and Source.

Antibody	Dilution	Source
SCL/TAL1	1:500 (for IP) 1:200 (for PD)	Santacruz (sc-12984)
LDB1/Clim2	1:500 (for IP) 1:200 (for PD)	Santacruz(sc-11198)
Gata-1 (M20)	1:500 (for IP) 1:200 (for PD)	Santacruz (sc-1234)
LMO2	1:10,000	ABD serotec (MCA2744GA)
JARID1A	1:500 (for IP) 1:200 (for PD)	Abcam (ab78322)
Streptavidin-HRP	1:20,000	Thermo scientific (N100)

2.6 Mass Spectrometry Analysis

3mg of HEK nuclear extracts were immunoprecipitated with JARID1A antibody as described above and run on 3-8% Tris-acetate gel. The gel was stained by Silver staining using Silver Quest staining kit (Invitrogen), bands were excised from the IP lane and sent for analysis by peptide mass fingerprinting to Prof. Benedikt Kessler's lab at the Target Discovery Institute (NDM, Oxford).

2.7 Gel Filtration Chromatography

Gel filtration chromatography, also known as size-exclusion chromatography, separates proteins on the basis of their size and shape. To determine if JARID1A fractionates and co-elutes at the same elution volumes as that of the other components of the pentameric complex, 2.5mg MEL nuclear extracts were loaded onto a Superose-6 column (GE Healthcare) equilibrated in the running buffer (20mM HEPES pH 7.6, 200mM NaCl, 1mM DTT, 0.2mM PMSF, 10% glycerol). Running buffer was pumped through the column at a flow rate of 0.5ml/min and 0.5ml fractions were collected. The eluted fractions were precipitated by trichloroacetic acid (TCA) and analysed by SDS-PAGE and Western Blotting.

2.8 Trichloroacetic acid (TCA) Precipitation

1/10 sample volume of 0.15% Sodium deoxycholate was added to the sample, vortexed and incubated at RT for 10 minutes. 1/10 sample volume of 72% TCA was then added to each sample, vortexed and incubated at RT for 5 minutes. Precipitated proteins were pelleted by centrifugation at 16,300×g for 10 minutes and re-suspended in resuspension buffer (4%SDS, 0.2M Tris-HCL pH 7.5, 0.15M NaOH, 1× loading buffer, 1mM DTT).

2.9 JARID1A knockdown by shRNA

The small interfering RNA (siRNA) sequences published in (Benevolenskaya et al. 2005) and new siRNA sequences designed using the Target Finder application from Ambion were cloned into the pSuperamp vector to produce small hairpin RNAs (shRNAs). In order to check the efficiency of knockdown by these shRNAs, these vectors were transfected into HEK cells using Lipofectamine 2000 (Invitrogen). Post 48 hours, nuclear extraction was carried out from these transfected cells and Western blot analysis was performed. The sequences of the shRNAs are listed in **Table 2.3**.

Table 2.3: Sequences of the siRNAs along with the restriction sites (underlined) and loop region (bold).

si RNA	Source	Sequence
siRNA No.2	Eurofins, MWG	Sense- 5'GGAAGATCTCCGCCAAGAACATTCCAGCCT**TTCAAG AGA**3' Antisense- 5'CCCAAGCTTGAATTCAAAAAGCCAAGAACATTCCAGCCT**TC TCTTGAA**3'
siRNA No.2SC	Eurofins, MWG	Sense- 5'GGAAGATCTCCGCCTTCACAAGACTACCAG**TTCAAGAGA**3' Antisense- 5'CCCAAGCTTGAATTCAAAAAGCCTTCACAAGACTACCAG**TC TCTTGAA3**'
siRNA No.3	Eurofins, MWG	Sense- 5'GGAAGATCTCCCTTCTGTACTGCTGACTGG**TTCAAGAGA**3' Antisense- 5'CCCAAGCTTGAATTCAAAAACTTCTGTACTGCTGACTGG**TCT CTTGAA**3'
RBP2si1	Eurofins, MWG	Sense- 5'GGAAGATCTCCGGATTGGCAACCTCCATTT**TTCAAGAGA**3' Antisense- 5'CCCAAGCTTGAATTCAAAAAGGATTGGCAACCTCCATTT**TC TCTTGAA**3'
RBP2si2	Eurofins, MWG	Sense- 5'GGAAGATCTCCGCAAGACCTCGGCACAGTATT**TTCAAGAGA** 3' Antisense- 5'CCCAAGCTTGAATTCAAAAAGCAAGACCTCGGCACAGTATT **TCTCTTGAA**3'

2.10 Transfection

Transfection of the plasmid and siRNA was carried out using Lipofectamine 2000 (Invitrogen). 8μg of plasmid DNA and 20μl of lipofectamine reagent were diluted in 500μl of OptiMEM (Invitrogen), incubated at RT for 5 minutes and mixed together. The mixture was incubated at RT for 20 minutes and then added on the HEK cells grown to 90% confluency in a T-25 flask (25cm^2 surface area). This was followed by addition of 4.5ml of DMEM (without serum, L-glutamine and Penstrep) and incubated at 37°C for 48 hours.

2.11 Real time PCR by Taqman assay

The cDNA samples from different stages of haematopoiesis were a kind gift from Julia Drapper (former PhD student in lab). The cDNA samples were diluted to 20ng/μl and 10 μl of each cDNA sample and water were added to the 96-well plate and mixed with 12.5μl of a mix containing 200μM of primers, and 100μM of probe in 1× universal PCR mastermix (Applied Biosystems). Every sample was analysed in duplicate and run on a 7500 Fast Real Time PCR system. To estimate the range of linearity, MEL cell cDNA was used. The mean threshold cycle (ct) (number of cycles required for the fluorescent signal to cross the background levels) value for each sample was calculated and relative gene expression levels for each gene listed below were calculated relative to the expression of the house keeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Primers and probes for the Taqman assay were designed using the Primer Express software from Applied Biosystems or ordered predesigned. Their sequences are listed in the **Table 2.4**.

Table 2.4: Primer and probe sequences of *Scl*, *Gata1* and *Gapdh* used in Real Time Assay.

Gene	Source	Sequence/ Reference
Jarid1a	Eurogentec	F=5'CTGTAATCCGGAGCGACTTGTAT3' R=5'GGATAGCGGTACCTAAGACATTTGTT3' P=5'TCATCCAACTGATTTGTGCTCTTGCCC3'
Scl	Applied Biosystems	Mm01187033_m1
Gata1	Eurogentec	F=5'AGAGAAGCTGAGGCCTACAGA3' R=5'CAGGAATTCCCTCCATACTGTTGAG3' P=CACTCCCCAGTCTTTC3'
Gapdh	Eurogentec	F=5'CATCCATGACAACTTTGGTATCGT3' R=5'CAGTCTTCTGGGTGGCAGTGA3' P=AAGGACTCATGACCACAGTCCATGC3'

2.12 MEL cell differentiation

1×10^6 cells/ml were seeded in T-175 ($175cm^2$ surface area) flask one day before differentiation. Next day, differentiation was induced by adding 2% DMSO to the final concentration, mixed gently by swirling the flask and incubated at 37°C for 4 days. 5×10^4 cells were taken at Day0 (D0), Day2 (D2) and Day4 (D4) by cytospin for staining and the rest was used for nuclear extraction.

MGG Staining

Slides were immersed in undiluted May-Grunwald solution for 3 minutes and rinsed with 1×PBS. These were then immersed in 5% Giemsa/PBS for 15 minutes, rinsed with water, air-dried and mounted with Entellan mounting medium and fixed with a coverslip.

2.13 Protein expression and Purification

2.13.1 Growth Media

LB Medium

Luria Bertani (LB) medium was prepared by resuspending 25g of the LB broth (Miller) in 1L dH_2O and autoclaved at 121°C for 15-20 minutes at 15psi pressure.

LB agar Plates

LB agar was prepared as LB medium followed by addition of 15g/L of agar before autoclaving. The media were melted in a microwave and when the temperature reached 40°C, ampicillin (100mg/ml stock) was added to get the final concentration of 100µg/ml. Approximately, 20ml of media was used for pouring each plate.

LB auto-induction media

The recipe for LB auto-induction media is taken from (Studier 2005) and is given in **Table 2.5**. Trace elements consisted of 50mM $FeCl_3$, 20mM $CaCl_2$ and 10mM $ZnSO_4$.

Table 2.5: LB auto-induction media recipe for 1L.			
Volume of each reagent to be added	**Reagent**	**Components of the reagents**	**Final concentration**
958 ml	LB media	Tryptone Yeast extract NaCl	
20 ml	50×M	Na_2HPO_4 KH_2PO_4 NH_4Cl Na_2SO_4	25 mM 25 mM 50 mM 5 mM
20 ml	50×5052	Glycerol Glucose α-lactose	0.5% 0.05% 0.2%
2 ml	1 M $MgSO_4$	$MgSO_4$	2 mM
0.2 ml	1000×trace elements	Trace elements	0.2 ×

TB auto-induction media

Terrific broth (TB) was made by dissolving 48g of TB powder (Melford) per litre of dH_2O followed by addition of 4ml of 100% glycerol. For TB auto-induction media, 20ml of 50 ×5052 solution (glycerol, glucose and α-lactose) was added to autoclaved TB media.

Sf-900II media for Sf9 cell growth

Sf9 cell culture, Sf9-900I media (Invitrogen) supplemented with 100µg/ml penicillin and 100µg/ml streptomycin was used. For protein expression, 1% FCS was also added to the media.

2.13.2 Protein expression and purification in *E.coli*

Purified plasmid DNA containing the gene of interest was transformed in chemically competent *E.coli* expression strains (Rossetta and BL21) as per the protocol described in section 2.14.4. Single colonies were picked up to make glycerol stocks (section 2.14.7). 35ml of LB media supplemented with ampicillin (100µg/ml) was inoculated from the glycerol stock and incubated at 37°C overnight in a shaker at 180 RPM. Next day, the preculture was inoculated in 1L of LB media (the volume of media was scaled up depending on the level of expression of each protein) supplemented with ampicillin (50µg/ml) and incubated at 37°C in a shaker at 210 RPM until OD_{600} reached 0.6. Protein expression was induced by adding 1mM IPTG and incubated at 20°C overnight at 210 RPM in a shaking incubator. For protein expression in LB/TB auto induction media, overnight grown precultures were inoculated in the media and incubated at 37°C for 8 hours in a shaking incubator at 210 RPM. After 8 hours, the temperature was decreased to 20°C and cultures were further incubated for 18-20 hours in a shaking incubator at 210 RPM.

Next day, the cells were harvested by centrifugation at 5,053×g for 15 minutes in a Beckman Coulter Avanti J20 using JLA 8.1 rotor. This was followed by resuspension of the cells in 20ml of lysis buffer (50mM HEPES,pH 8, 400mM NaCl, 10mM Imidazole, 0.5% tween 20, 34µl of sigma protease inhibitor, 12.5 units of DNaseI from Sigma). Cells were disrupted by sonication (Sonics Vibracell) for 2 minutes on ice with alternating bursts for 10 seconds and resting period of 15 seconds. The cell lysate was then clarified by centrifugation at 48,384×g for 30 minutes in a Beckman Avanti J-30I using JA 30.50 Ti rotor.

For His tagged proteins, the supernatant was then loaded onto Talon or Ni-NTA (Clontech) affinity resin (in the case of Ni-NTA resin, 20mM Imidazole was added in the lysis buffer)

pre-equilibrated with the lysis buffer and incubated for 1 hour at 4°C on a rotary shaker. The beads were then washed with the washing buffer (50mM HEPES,pH 8, 400mM NaCl, 10mM Imidazole) to eliminate non-specifically bound proteins. The over-expressed protein was eluted by the elution buffer (50mM HEPES,pH 8, 200mM NaCl, 500mM imidazole). The pellet after lysis, beads, flow- through, wash and elution fractions were then analysed by SDS polyacrylamide gel electrophoresis.

For GST tagged proteins, cell lysis was carried out in lysis buffer (50mM HEPES,pH 8, 400mM NaCl, 0.5% tween 20, 34µl of sigma protease inhibitor, 12.5 units of DNaseI from Sigma) and the clarified lysate was mixed with glutathione beads. This was followed by overnight incubation at 4°C on a rotary shaker. Next day, beads were washed with the wash buffer (50mM HEPES,pH 8, 400mM NaCl). The over-expressed protein was eluted using the elution buffer (50mM HEPES,pH 8, 200mM NaCl and 15mM reduced glutathione).

The eluted protein fractions were concentrated using a concentrator (Millipore) of Molecular weight cut-off (MWCO) appropriately chosen on the basis of the molecular weight of the protein. The concentrated protein was further purified by size-exclusion chromatography, by loading it on a HiLoad 16/60 Superdex 200 (or 75) prep grade column (GE Healthcare).

Note that the procedure described here is for cell pellets from 1L culture. The volumes of buffers and sonication times were increased depending on the volume of the culture. Also, this is a generic protocol and specific conditions used for each protein will be detailed in the respective sections.

GST Tag cleavage

For the proteins that were to be used for analytical ultracentrifugation (AUC) and crystallization, protein was not eluted off glutathione beads and on-column GST tag cleavage was performed. For this, at the end of the affinity purification, beads were washed and resuspended in the low salt buffer. Glutathione beads containing the protein were re-suspended in the tag cleavage buffer (50mM HEPES, pH 8, 200mM NaCl, 1mM DTT) in 1:1 ratio. 200µl of 3C protease (Pierce) was added to beads, which were then were incubated at 4°C on a rotary shaker for 24 hours. Following incubation, the buffer (now containing the protein without the tag) was collected and along with beads (after cleavage) was analysed by SDS-PAGE gel to check the success of the tag cleavage reaction and purity of the protein without the tag.

2.13.3 Protein expression in Sf9 cells

Baculovirus expression systems offer several advantages over *E.coli* expression system, importantly it allows for posttranslational modifications to proteins of eukaryotic origin. Baculovirus infects insect cells and the baculovirus used in this study is *Autographa californica* nuclear polyhedrosis virus (AcNPV). This virus needs a protein named polyhedrin to encapsulate virions upon infection. So, for the expression of protein using this system, the gene coding for polyhedrin is replaced with the gene coding for the recombinant protein. The insect cells used were from worm *Spodoptera frugiperda.*

2.13.3.1 SF9 cell culture

SF9 cells were grown and maintained in GIBCO™ Sf 900II media (Invitrogen) supplemented with 100µg/ml penicillin and 100µg/ml streptomycin. Cells were cultured in conical flasks by incubation at 27.5°C at 100 RPM in a shaking incubator. Cells were maintained at a density

of 0.8-1×10^6 cells per ml of culture. An approximate cell density was determined by counting the number of cells in the sample volume using haemocytometer.

2.13.3.2 Recombinant baculovirus generation

Sf9 cells were cultured to 1×10^6 cells/ ml density as described in the previous section. Freshly prepared plasmid containing the gene of interest was clarified with chloroform by mixing 40µl of the plasmid with 20µl chloroform, mixing by vortex, centrifuging at 16,300×g for 10 minutes and pipetting the top layer of DNA into a fresh microfuge tube. Approximately, 1×10^6 Sf9 cells were seeded in 35mm well of a 6-well plate with 2ml of Sf 900II media. Cells were incubated at room temperature for 15-20 minutes to enable them to adhere to the surface to form a monolayer.

For recombinant baculovirus generation, these cells were co-transfected with bacmid DNA and plasmid DNA coding for the recombinant protein using the Fugene transfection kit. For transfection of each well, 100µl of Sf 900II media (without antibiotics) was mixed with 1.5µl of the Bacmid DNA (0.5µg) and 5µl of the purified plasmid DNA (1-2µg) that encodes for the ORF1629 baculovirus elements and protein of interest respectively [Solution1]. In a separate tube, 100µl of Sf 900II media (without antibiotics) was mixed with 8µl of the Fugene transfection reagent [Solution2]. Solutions 1 and 2 were mixed together gently and incubated at RT for 20 minutes. The fugene transfection reagent/DNA mixture was diluted with 800µl of Sf 900II media (without antibiotics). Media were removed from top of the Sf9 monolayer and the diluted fugene/DNA mixture was gently applied in a drop wise fashion and mixed by swirling the plate to ensure uniform spread of the mixture over the cells. Cells were incubated in an incubator at 27.5°C in a box containing wet blotting paper to maintain humidity. Next day, 1ml of Sf 900II media (without antibiotics) was added onto cells

followed by incubation for 5days. The recombinant baculovirus was harvested from the culture media (2ml) and clarified by centrifugation at 500×g for 5 minutes at room temperature. The virus was stored at 4°C until used.

2.13.3.3 Baculovirus Amplification

50ml of suspension culture of Sf9 cells was made using Sf 900II media supplemented with 100µg/ml penicillin and 100µg/ml streptomycin in a 1L conical flask (to ensure proper aeration) with a cell density of 2×10^6 cells/ ml for virus amplification. Cells were infected with 1ml recombinant virus from the first infection and incubated at 27.5°C, 100RPM in a shaking incubator for 3 days. The supernatant containing amplified virus was harvested 3 days post infection by centrifugation at 1000×g for 10 minutes.

2.13.3.4 Protein expression and purification

For protein expression, the recombinant baculovirus was generated and amplified as described in the previous section. Sf9 cells were cultured in suspension in Sf 900II media supplemented with 100µg/ml penicillin and 100µg/ml streptomycin to a density of 2×10^6 cells/ ml. 500ml of cell culture was then infected with 5ml of the amplified virus stock in a 2L conical flask and incubated at 27.5°C, 120RPM. The media was also supplemented with 1% FCS. 3 days (72 hours) post infection, cells were harvested by centrifugation at 1000×g at 4°C.

Protein purification was carried out as described in section 2.13.2 except lysis buffer (50mM HEPES,pH 8, 200mM NaCl, 20mM Imidazole, 1% NP40, 34µl of sigma protease inhibitor, 25 units of DNaseI) was used. Cells were disrupted by sonication only for 2 minutes with alternating bursts (10 seconds) and resting period (15 seconds) at 40% amplitude on ice as

insect cells do not possess cell wall. The clarified cell lysate was incubated with 3ml bed volume of Ni-NTA affinity resin pre-equilibrated with the lysis buffer, for 30 minutes at 4°C. Beads were washed with wash buffers with an increasing concentration of imidazole (50mM HEPES,pH 8, 200mM NaCl, 20/50/250 mM Imidazole) and eluted with the elution buffer (50mM HEPES,pH 8, 200mM NaCl, 500mM Imidazole).

2.14 Mammalian two hybrid assays

2.14.1 Construct design

Four constructs were designed including full length (FL), JmjN-JmjC, Zf- PHD3 and PHD3 (Refer to Chapter 4, section 4.2.1 for details of construct design).

2.14.2 Polymerase chain reaction

Polymerase chain reaction was carried out in PTC-225 Peltier Thermal Cycler (MJ Research) using the Expand Long Template PCR System (Roche). Thermal cycling was carried out using conditions described in **Table 2.6**.

Table 2.6: Thermal cycling conditions used in PCR.			
Steps	**Temperature**	**Time**	**Cycles**
Initial Denaturation	94°C	2 min	1
Denaturation Annealing Elongation	94°C 65°C 68°C	10 sec 30 sec 4 min	9
Denaturation Annealing Elongation	94°C 65°C 68°C	15 sec 30 sec 4 min+20 sec cycle elongation for each successive cycle	19
Final Elongation	68°C	7 min	1
Cooling	4°C	Unlimited time	

The final 50µl reaction mixture was made in nuclease free water and composed of 5µl of 10× PCR buffer system1, 1.75µl of 10mM dNTP (Invitrogen), 1.5µl of 10µM forward and reverse primers (Invitrogen), 5µl of 100ng/µl template DNA (Mouse *Jarid1a* cDNA optimized for mammalian expression from GeneArt) and 0.75µl expand enzyme mix. The reaction mixture was made on ice. pVP16 LMO2, pVP16 SCL, pM LMO2, pVP16 SCL and pM E47 plasmids were gifted by Dr. Kamel El Omari. Primers for all other constructs are listed in **Table 2.7**.

Table 2.7: *Jarid1a* and *Gata1* constructs' primer sequences. Primers used in cloning of *Jarid1a* constructs and *Gata1* in pVP16 and pM vectors.

Gene	Source	Sequence/ Reference
JmjN-JmjC	Invitrogen	F=5'GGGGAATTCGCCTCTGTGGGCCCTGGCGG3' R=5'GGGTCTAGACTAGGCGGCCATCTTGAAGATCAGTTCCTC3'
Zf-Phd3	Invitrogen	F=5'GGGGAATTCGACCCCGAGTGCCTGGATGTGGG3' R=5'GGGTCTAGACTAGCTTGTCTCTTTCAGATCCTCCATGGGC3'
Phd3	Invitrogen	F=5'GGGGAATTCAGCGGAGCCGAGGAATCCGACGATG3' R=Same as Zf-Phd3 construct
Gata1	Invitrogen	F=5'GGGGGATCCTCGATTTTCCTGGTCTAGGGGCCCTGG3' R=5'GGGTCTAGACTAAGAACTGAGTGGGGCGATCACGCTG3'

2.14.3 Cloning

Mouse *Jarid1a* cDNA optimized for mammalian expression (from GeneArt) was used as a template to carry out PCR as described in section 2.14.2. PCR products and vectors (pVP16 and pM) were digested with restriction enzymes EcoRI and XbaI (NEB) using the standard protocol recommended by manufacturers. Specifically, double digest using EcoRI and XbaI restriction enzymes was carried out in a volume of 20µl containing NEB buffer4, 1× BSA and 1U enzyme/ µg of DNA. The mixture was incubated at 37°C for 1.5 to 2 hours.

The digested PCR products were further purified using the PCR purification kit (Qiagen) and the digested vectors were run on 1% agarose gel and gel extracted using the gel extraction kit (Qiagen). The purified products were quantified by running on 1% agarose gel by densitometry using the Image Lab software (Bio-Rad).

Prior to ligation, the linearized vectors were dephosphorylated with phosphatases to prevent re-circularization of the empty vector. For this, the linearized vector DNA was treated with Shrimp alkaline phosphatases (Roche) in presence of the buffer (0.5M Tris-HCl, pH 8.5, 50mM $MgCl_2$) and incubated for 30 minutes at 37°C.

The insert DNA was mixed with the linearized, dephosphorylated vector DNA in the molar ratio of 1 vector: 3 insert in a typical 10 µl reaction and ligated using T4 DNA Ligase (NEB) by incubation at 16°C overnight. 2µl of the ligated products were transformed into *E.coli* DH5α XL-10 gold competent cells (Agilent Technologies). Constructs were tested by restriction digestion and sequencing.

2.14.4 Transformation

1µl (~100ng) of plasmid DNA was mixed with 50-70µl of the chemically competent *E.coli* cells and incubated on ice for 20 minutes. The choice of *E.coli* strain was dependent on whether the transformation was carried out for protein expression (*E.coli* rossetta (DE3) pLysS and BL21 strains from Invitrogen) or for plasmid amplification (*E.coli* DH5α from Invitrogen). Cells were then heat shocked at 42°C for 45 seconds and immediately transferred back on ice for 5 minutes. This was followed by addition of 250µl of SOC media (Invitrogen) and incubation at 37°C in a shaker for 45 minutes to 1 hour for plasmid recovery. Cells were then plated on LB agar plates containing 100µg/ml ampicillin or carbenicillin and incubated overnight at 37°C.

2.14.5 Plasmid isolation and quantification

A single colony from the transformed plates was picked up and inoculated in 5ml LB media with appropriate antibiotics and incubated at 37°C for ~16 hours in a shaker at 200RPM.

Next day, cells were harvested from the overnight grown cultures by centrifugation at 3,724×g at 4°C for 15 minutes in a Beckman Coulter Allegra X-15R centrifuge. Plasmid DNA was isolated from cells by using QIAprep® Spin Miniprep Kit (Qiagen). Precisely, this involved lysing cells in alkaline conditions followed by precipitation of genomic DNA, proteins and lipids under neutral conditions. Precipitates were then pelleted down by centrifugation at 16,300×g on a bench top centrifuge (Eppendorf) for 10 minutes during which the plasmid DNA got adsorbed on the silica gel membrane of spin columns under high salt conditions. Salt was then removed by washing the membrane with 70% ethanol and the plasmid DNA was eluted in Milli-Q (MQ) water and quantified by Nanodrop 1000 Spectrophotometer (Thermo Scientific).

The spectrophotometer carried out DNA quantification on the basis of absorption at 260nm and 280nm based on the Beer-Lambert law: $A = \varepsilon cl$ (where ε- extinction coefficient, A- absorbance, c- concentration, l- pathlength). DNA absorbs at 260nm and protein absorbs at 280nm, hence, 260:280 ratio higher than 1.85 was considered pure.

2.14.6 DNA sequencing

DNA sequences of plasmids were confirmed either at the internal sequencing service at WIMM or by Source Biosciences. Sequences were analysed using the MacVector and the Chromas softwares.

2.14.7 Glycerol stocks

Clones were stored in the form of glycerol stocks for long-term storage. For this, a single colony was picked up from LB agar plates and grown overnight at 37°C in LB media

containing appropriate antibiotics. 1ml of the overnight grown culture was mixed with 1ml of autoclaved 50% glycerol in a cryovial and stored at -80°C.

2.14.8 Agarose Gel Electrophoresis

1% (w/v) agarose gels were prepared in 1×TAE buffer (0.04M Tris-Acetate and 0.001M EDTA). Sybr green (SafeImager™, Invitrogen) was added to the gels at a final concentration of 1×. 5µl samples were mixed with 1µl of 6× loading buffer. Samples were loaded onto the gel alongside the appropriate molecular weight markers (Thermo Scientific). Gels were run at 100 volts in 1× TAE buffer. DNA bands were visualized and/or quantified using Gel doc (Bio-Rad).

2.14.9 Transfection

HEK cells were transfected using Polyethylenimine (PEI). 1µl of each plasmid (pVP16 and pM containing the inserts); pVP16 (200ng/µl), pM (200ng/µl), Firefly Luciferase (200ng/µl) and pRenilla (100ng/µl) were diluted in 100µl of OPTI-MEM+GlutaMAX (Gibco) media, mixed with 6µl PEI, vortexed, incubated for 10minutes at RT and added drop wise to cells containing 600µl of DMEM+FCS+L-glutamine and gently mixed. Transfected cells were then incubated at 37°C for 48 hours.

Cells were then washed with 1×PBS to remove dead cells and medium and then resuspended in 100µl of 1×Passive Lysis Buffer (Promega Kit). 20µl of cell lysates were dispensed in a black 96 well plate and the assay was performed as per the recommended protocol of the dual luciferase reporter assay system (Promega).

2.15 GST Pull down assay of JARID1A domains and *In vitro* transcribed and translated (IVTT) SCL and LDB1

2.15.1 *In vitro* transcription and translation

Plasmids pcDNA SCL, pcDNA LDB1, pcDNA LMO2, pcDNA GATA1 and pcDNA ETO2 were a generous gift from Dr. Nicolas Gordon (former post doc in the lab). The *in vitro* transcription and translation reactions were carried out using the TNT® Quick Coupled Transcription/Translation kit (Promega) which involved mixing of 1µg of the plasmid with TNT master mix and 1mM Methionine as shown in **Table 2.8**.

Table 2.8: In-vitro transcription and translation reaction.

TNT Master mix	40µl
Methionine (1mM)	1 µl
Plasmid DNA (1 µg)	2 µl
MilliQ water	7 µl
Total	**50 µl**

The reaction mixture was incubated at 30°C for 90 minutes and then stored at -20°C until used.

2.15.2 GST pull down assay

GST tagged proteins were purified as described in section (2.13.2). For the GST pull down assay, the protein was not eluted off glutathione beads after affinity purification. The bound protein was quantified by the Bradford's assay. Approximately 500µg of protein was used. Beads were equilibrated with the binding buffer (40mM HEPES, pH 7.8, 50mM KCl, 0.2mM EDTA, 5mM $MgCl_2$, 0.1% Triton X-100, 10% glycerol, 1.5mM DTT, 0.5mg/ml BSA) for 1 hour at 4°C on rotary shaker. After 1 hour, IVTT products were mixed with beads followed by incubation at 4°C overnight on a rotary shaker.

Next day, beads were washed 5 times with 4× bed volume of the wash buffer (binding buffer without BSA). All the washing buffer was removed and the bound material was eluted by boiling beads with 1×NuPAGE LDS Sample buffer for 10 minutes and analysed by SDS-PAGE and Western blotting.

2.16 Protein expression and purification for crystallization

2.16.1 Construct Design

In order to make proteins more amenable to crystallization, flexible loops and inter-domain regions are removed by molecular cloning. Construct design for protein crystallization was carried out using the high through-put infusion cloning system. Boundaries for all constructs were designed based on disorder prediction by the RONN software (www.oppf.ox.ac.uk/RONN) and secondary structure prediction by the Phyre2 software (www.sbg.bio.ic.ac.uk/phyre2).

2.16.2 High through-put cloning of JARID1A domains for crystallization

All constructs were cloned into pOPINF (for N-terminal His-tag), pOPINE (for C-terminal His-tag) and pOPINJ (for GST tag) vectors from the pOPIN vector suite developed by Oxford Protein Production Facility (OPPF). The fusion tag, parent vector and forward and reverse primer extensions for vectors used are shown in **Table 2.9**. The fusion tags can be cleaved by using 3C protease. The advantage of using pOPIN vectors is that they can be used for all three protein expression systems: *E.coli*, Insect cells and Mammalian cells as these vectors contain T7lacO promoter for protein expression in *E.coli*, P10 promoter/lef-2 and 1629 baculo elements for protein expression in insect cells and CMV enhancer/β-actin promoter for protein expression in mammalian cells.

Table 2.9: List of pOPIN vectors used and their specifications.				
Vector	**Fusion Tag**	**Parent vector/ Antibiotic resistance**	**Forward Primer Extension**	**Reverse Primer Extension**
pOPIN E	C-terminal His tag --KHHHHHH	pTriEx2/Ampicillin	aggagatataccatg	gtgatggtgatgttt
pOPIN F	N-terminal His tag -- MAHHHHHHSSGL EVLFQ*GP	pTriEx2/Ampicillin	aagttctgtttcaggg cccg	atggtctagaaagc ttta
pOPIN J	N-terminal His and GST tags MAHHHHHHSSG-GST-LEVLFQ*GP	pTriEx2/Ampicillin	aagttctgtttcaggg cccg	atggtctagaaagc ttta
* represents 3C cleavage site				

The In-Fusion® HD cloning kit from Clontech was used for cloning which employs the In-Fusion enzyme that fuses DNA fragments precisely by recognizing a 15 bp overlap at their ends by homologous recombination thereby facilitating ligation independent cloning (LIC). The basic molecular biology techniques used were the same as described in section 2.14. Forward and reverse primers for the PCR were designed using the OPINE (www.oppf.ox.ac.uk), a web based tool, which adds a 15bp extension to primers homologous to the cloning site in the target vector. Primers were designed so that the melting temperature was around 65°C-68°C with a GC content of 40%-60%. While designing primers with the OPINE suite, a 15bp oligonucleotide sequence homologous to the specific cloning site in the target vector is automatically appended to primers.

All primers were ordered from Invitrogen. Lyophilized primers were resuspended in MQ water to obtain a stock of 100µM that was then diluted to 1µM for use in the PCR reactions. Sequences for forward and reverse primers for all constructs are listed in **Table 2.10**.

Table 2.10: List of all the JARID1A constructs with their primer sequences.				
Sr. No	**Protein**	**Tag**	**Forward primer**	**Reverse Primer**
1	JmjN	N-His & GST	aagttctgtttcagggcccgCCAGAGTGCCCCGTGTTCGAG	atggtctagaaagctttaCTCGTTCAGCCGCTGCACCC
		C-His	aggagatataccatgCCAGAGTGCCCCGTGTTCGAG	gtgatggtgatgtttCTCGTTCAGCCGCTGCACCC
2	ARID	N-His & GST	aagttctgtttcagggcccgGAAGCCATGACCAGAGTGCGGC	atggtctagaaagctttaGCCGCTCTGGAACAGCTCGTAG
		C-His	aggagatataccatgGAAGCCATGACCAGAGTGCGGC	gtgatggtgatgtttGCCGCTCTGGAACAGCTCGTAG
3	JmjN+ARID	N-His & GST	aagttctgtttcagggcccgCCAGAGTGCCCCGTGTTCGAG	atggtctagaaagctttaGCCGCTCTGGAACAGCTCG
		C-His	aggagatataccatgCCAGAGTGCCCCGTGTTCGAG	gtgatggtgatgtttGCCGCTCTGGAACAGCTCG
4	PHD1	N-His & GST	aagttctgtttcagggcccgCGGAAGGGCACCCTGAGCG	atggtctagaaagctttaCTCGCGCACGGCCTGCTC
		C-His	aggagatataccatgCGGAAGGGCACCCTGAGCG	gtgatggtgatgtttCTCGCGCACGGCCTGCTC
5	JmjC	N-His & GST	aagttctgtttcagggcccgGAACAGAGCGTGCTGGCCCAC	atggtctagaaagctttaGCTGAACACGCAGTGCCGC
		C-His	aggagatataccatgGAACAGAGCGTGCTGGCCCAC	gtgatggtgatgtttGCTGAACACGCAGTGCCGC
6	PHD1+JmjC	N-His & GST	aagttctgtttcagggcccgCAGCGGAAGGGCACCCTGAG	atggtctagaaagctttaGCTGAACACGCAGTGCCGC
		C-His	aggagatataccatgCAGCGGAAGGGCACCCTGAG	gtgatggtgatgtttGCTGAACACGCAGTGCCGC
7	Zf(short)	N-His & GST	aagttctgtttcagggcccgGACGACGAGAGACAGTGCAGCG	atggtctagaaagctttaCCGCACTTTCACGCCGTACAG
		C-His	aggagatataccatgGACGACGAGAGACAGTGCAGCG	gtgatggtgatgtttCCGCACTTTCACGCCGTACAG
8	Zf(long)	N-His & GST	aagttctgtttcagggcccgGACGACGAGAGACAGTGCAGCG	atggtctagaaagctttaCTTCTTGTGGTTGAAGCTGGCGC
		C-His	aggagatataccatgGACGACG	gtgatggtgatgtttCTTCTTGTG

			AGAGACAGTGCAGCG	GTTGAAGCTGGCGC
9	JmjC+Zf	N-His & GST	aagttctgtttcagggcccgGAACAGAGCGTGCTGGCCCAC	atggtctagaaagctttaCTTCTTGTGGTTGAAGCTGGCGC
		C-His	aggagatataccatgGAACAGAGCGTGCTGGCCCAC	gtgatggtgatgtttCTTCTTGTGGTTGAAGCTGGCGC
10	PHD2	N-His & GST	aagttctgtttcagggcccgGAACGGATCGAGGAAGTGAAGTTCTGC	atggtctagaaagctttaTCTGGGCCGTCTGCTCCG
		C-His	aggagatataccatgGAACGGATCGAGGAAGTGAAGTTCTGC	gtgatggtgatgtttTCTGGGCCGTCTGCTCCG
11	PHD3	N-His & GST	aagttctgtttcagggcccgGCCGTGTGTGCCGCCCAG	atggtctagaaagctttaGGCGCAGTTGATGCAGATGTAGTCC
		C-His	aggagatataccatgGCCGTGTGTGCCGCCCAG	gtgatggtgatgtttGGCGCAGTTGATGCAGATGTAGTCC
*Lowercase nucleotides highlight homologous sequence				

The amplified products were purified using the PCR purification kit (Qiagen) and the target vectors were linearized by restriction digestion using HindIII and KpnI restriction sites for pOPINF and pOPINJ. For pOPINE, NcoI and PmeI restriction sites were used. The infusion reaction was carried out by mixing 30-50ng of the purified PCR product, 100 ng of the linearized vector, 2µl of 5X In-Fusion HD enzyme premix and nuclease free water to a final volume of 10µl. The reaction mixture was incubated for 15 minutes at 37°C followed by 15 minutes at 50°C in a PCR thermocycler (Bio-rad). 2µl of the In-fusion reaction product was transformed into library efficiency *E. coli* DH5α competent cells (Life technologies) using the standard transformation protocol as described in section 2.14.4. The positive clones were selected utilizing blue-white selection by spreading the transformed cells on LB agar plates containing 100µg/ml ampicillin, 1:1000 dilution of 20% (w/v) X-Gal in DMF and 1mM IPTG (final concentration). White colonies were picked, inoculated in LB media

supplemented with 100µg/ml ampicillin and incubated overnight at 37°C in a shaking incubator at 180 RPM. Next day, cells were spun down and the plasmid DNA isolated using the Qiaprep Spin Miniprep kit (Qiagen). Plasmids were sequenced by using the Source Biosciences sequencing services.

2.16.2.1 Polymerase Chain Reaction

Polymerase chain reaction was carried out in a T100™ Thermal Cycler (Bio-Rad) using the Phusion Flash High-Fidelity Master Mix (Thermo Scientific). A typical PCR reaction mixture composition is shown in **Table 2.11** and thermal cycling conditions are shown in **Table 2.12**.

Table 2.11: PCR reaction mixture composition.

MQ water	19.1µl
2× Phusion Flash PCR Master mix	25µl
Forward Primer (1µM)	1µl
Reverse Primer (1µM)	1µl
Template (5ng)	2.9µl
DMSO (100%)	1µl

Table 2.12: Thermal cycling condition used in PCR.

Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	10 sec	1
Denaturation Annealing Elongation	98°C 68°C 72°C	1 sec 40 sec 20 sec	30
Final Elongation	72°C	1 min	1
Cooling	4°C	Unlimited time	

2.16.3 Protein expression and purification

Expression tests for all JARID1A constructs were performed in two different *E. coli* strains: *Rosetta and BL21 Lemo,* using the Power Prime broth (Molecular Dimensions) and the TB auto-induction media to attain high cell density. Single colonies from transformed bacteria were picked, inoculated in 0.7ml of the Power Prime broth supplemented with 100µg/ml ampicillin in a 96 deep-well block, sealed with gas-permeable adhesive seals and incubated overnight at 37°C at 225 RPM. 125µl of overnight cultures (precultures) were then inoculated into 2.5ml of the Power Prime broth for IPTG induction and the Overnight Express™ Instant TB medium for auto-induction, supplemented with 100µg/ml ampicillin in a 24-deep well plates. Cultures were then incubated at 37°C at 225 RPM until OD at 600nm reached 0.5 (~5hours). IPTG induction cultures were then induced by addition of 1mM IPTG and incubated overnight (~18 hours) at 20°C at 225RPM. Auto-induction cultures were incubated for ~7hours at 37°C at 225RPM and temperature was then reduced to 20°C for a further incubation of 20-24 hours. Cells were then harvested by centrifugation at 6000×g for 10 minutes and media was decanted from cell pellets by inverting plates and allowing them to drain on paper towels. Plates were sealed and stored at -80°C.

Plates were then thawed at room temperature. From this point onwards the purification was automated and carried out by the Qiagen BioRobot 8000, which was equipped with all the required buffers. Cells were resuspended in 210µl of lysis buffer (50mM NaH_2PO_4 (pH 8), 300mM NaCl, 10mM Imidazole, 1% v/v Tween 20) supplemented with 1mg/ml lysozyme 400units/ml of Dnase-I followed by incubation at room temperature for 30 minutes to allow the action of Lysozyme and Dnase-I. The cell lysate was then clarified by centrifugation at 5000×g for 30 minutes at 4°C, mixed with 20µl of equilibrated Ni-NTA magnetic beads and incubated for 30 minutes at room temperature on a plate shaker. Beads were washed 3 times

with 200µl of the wash buffer (50mM NaH_2PO_4 (pH 8), 300mM NaCl, 20mM Imidazole, 0.05% v/v Tween 20) by mixing for 5 minutes and eluted with 50µl of the elution buffer (50mM NaH_2PO_4 (pH 8), 300mM NaCl, 250mM Imidazole, 0.05% v/v Tween 20) by mixing for 1 minute. Eluted proteins were analysed by SDS-PAGE using 26 wells 10% Bis-Tris gel (Invitrogen) and gels were stained by the Instant*Blue* stain (Expedeon).

The large-scale protein expression using the classical IPTG induction system in the Power Prime broth was carried out as described in section 2.13.2. However, in case of the TB auto-induction medium, overnight grown pre-cultures were inoculated in the TB auto-induction media, which was constructed as shown in section 2.13.1. Following inoculation, cells were incubated at 37°C for 8 hours in a shaking incubator at 210RPM. After 8 hours, the temperature was decreased to 20°C and cultures were further incubated for 18 hours.

Protein purification was carried out as described in section 2.13.2 except lysis buffer (50mM HEPES, pH 8, 400mM NaCl, 10mM Imidazole, 0.5% tween 20, 34µl of sigma protease inhibitor), wash buffer (50mM HEPES, pH 8, 400mM NaCl, 20mM Imidazole) and elution buffer (50mM HEPES, pH 8, 200mM NaCl, 500mM Imidazole) were used. Specific conditions for the purification of each protein will be discussed in the results chapters.

2.17 Protein crystallization

2.17.1 Basic principles of protein crystallization

The determination of the three- dimensional structure of a protein at an atomic level necessitates several steps; to begin with purification of proteins to homogeneity and then formation of periodically ordered crystals. Protein crystals are very fragile and are prone to

dehydration and hence are always kept in their crystallization solution, also known as mother liquor. This is also required while freezing crystals for diffraction data collection.

The process of crystallization of biological macromolecules occurs gradually by altering the protein solubility in order to attain a state of super saturation. Crystallization is a thermodynamic process that strives for the minimization of the free energy of the system. When a protein is in solution, it is in equilibrium with the solvent. Protein crystallization is stimulated by addition of precipitants which compete with the protein molecules for solvent. As a result, the system undergoes internal changes and is thermodynamically driven to a new equilibrium state there by attaining a supersaturated state. At this stage, protein molecules will come together and form many new non-covalent interactions, thus minimising the free energy of the system and resulting in phase separation where protein molecules separate from the solution leading to formation of crystal nuclei. The formation of the crystal-nucleating site is the first phase transition where molecules form an ordered state and in highly favourable conditions, molecules self-assemble into highly ordered crystals. The process of crystallization can be schematically illustrated by the protein crystallization phase diagram (**Figure 2.1**) (Chayen 2004).

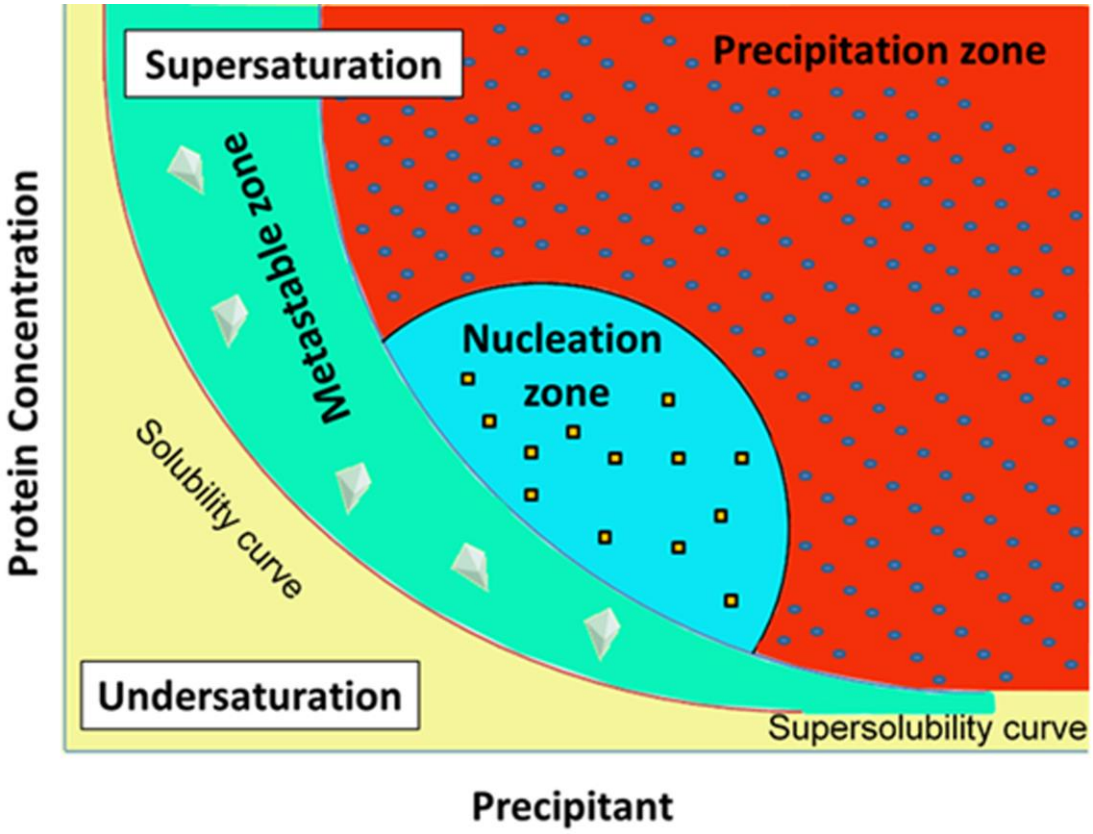


Figure 2.1: Illustration of crystallization phase diagram: The crystallization process can be schematically represented by a phase diagram which consists of four zones representing different degrees of supersaturation: Precipitation zone (high supersaturation), where the protein molecules will precipitate, Nucleation zone (moderate supersaturation), where the formation of nucleation sites are intiated, Metastable Zone (Lower supersaturation, where crystals are formed and may grow but no more nucleation events occur, Undersaturation zone, where proteins are in solution and will never crystallize [Adapted from (Chayen 2004)].

2.17.2 Protein Crystallization

Obtaining highly ordered diffracting crystals can be difficult and is often referred to as the bottleneck in crystallography. There are various processes by which protein crystals can be obtained including the vapour diffusion batch method, crystallization by dialysis, free interface diffusion and Microfluidics (Chayen and Saridakis 2008). Vapour diffusion is the most commonly used technique for crystallogenesis because of its extensive sampling of a wide range of crystallization parameters. This method can be executed in two ways: sitting

drop or hanging drop technique. In both techniques, the purified protein solution is mixed with the precipitant solution and is equilibrated in a closed system against the reservoir, also known as mother liquor containing the same precipitant at higher concentration. As water from the protein+precipitant droplet evaporates and diffuses into the reservoir, the precipitant concentration in the droplet goes up thereby pushing the protein molecule in the supersaturated state driving crystallogenesis.

2.17.3 Crystallization screening

Even if pure, highly soluble protein is successfully obtained, producing high-quality diffracting crystals is often a challenge on the way to determining 3-D structure as crystallization is a complex, multi-parametric process in which all parameters involved need to create a highly favourable environment to attain the state of supersaturation. The latter is dependent on and can be achieved by varying several parameters like concentration of precipitants, protein, additives, pH, temperature, buffer conditions that affect the surface charge on the protein etc.

There are numerous crystal screens available commercially which are widely used and are referred to as sparse matrices. These screens are comprised of a large number of combinations of solutions that enable us to explore various conditions such as precipitant concentration, pH, ionic strength and specific additives. Most of these screens contain precipitants that have been previously successfully used in crystallizing other proteins and allow proficient screening of commonly used precipitants such as salts (eg. Ammonium sulphate, ammonium acetate, lithium sulphate, NaCl), polymers (eg. high molecular weight PEGs) and organic solvents (eg. 2-propanethanol, 2, 4-methylpentanediol) at a range of pHs.

The first step for finding crystallization conditions for a new protein is to screen sparse matrices to obtain promising initial hits, followed by a 'fine screening' to improve crystal growth and quality. High-throughput crystallization protocols developed by the OPPF were used to perform the initial screening which were based on a nanolitre scale sitting drop vapour-diffusion method using standard crystallization screening kits in 96-well crystallization plates. For this, commercially available sparse matrix crystallization screens were reformatted into 96-well format as listed in **Table 2.13** by Dr. Tom Walter who manages the crystallization facility.

Table 2.13: 96-well crystallization screens.		
In-House Screen	**Commercial Screens**	**Source**
Block 1	Crystal Screen I Crystal Screen II	Hampton Research
Block 2	Wizard Screen I Wizard Screen II	Hampton Research
Block 3	PEG/Ion Grid Screen 6000 Grid Screen Ammonium Sulphate	Hampton Research
Block 4	Natrix Crystal Screen Cryo	Hampton Research
Block 5	Grid Screen PEG/LiCl Grid Screen Sodium Chloride Grid Screen MPD Quick Screen	Hampton Research
Index	Index Screen (Index HT)	Hampton Research
SaltRX	SaltRX	Hampton Research
PACT	PACT premier™	Hampton Research
KersFast	Protein-Nucleic acid complex crystal screen	KeraFast
Nuc-Pro	JBScreen Nuc-Pro	MiTeGen
Morpheus	Morpheus®	Molecular Dimensions

Crystallization screens were set up by dispensing 100µl of the precipitant solutions into the Greiner plate reservoir chambers of the CrystalQuick 96 well crystallization plate (Greiner) using the Robbins Hydra 96 micro-dispenser (Matrix Technologies Ltd, UK). 100nl of protein was then dispensed as sitting drops on to the central well of the raised platform by the Cartesian robot (Cartesian Technologies Microsys MIC4000) followed by addition of 100nl

of reservoir to each crystallization platform from their corresponding reservoir wells. The crystallization plate was then immediately sealed manually with transparent self-adhesive foils (Viewseal, Greiner).

The crystallization plates were then stored at 20°C in the Homebase automated storage vault (The Automation Partnership, Ltd, UK) or at 4°C in a Rock Imager1000 (Formulatrix) depending upon the desired temperature. Plates at 20°C were imaged using an Oasis 1700 imaging system (Vecco Instruments) and plates at 4°C were imaged using the in-built imaging system of the Rock Imager 1000 (Formulatrix) (Walter et al. 2005).

2.17.4 Crystallization Optimization

Once the potential crystallization conditions are established by coarse screening, the second step is to perform 'fine screening' of components within the crystallization condition. This is most commonly achieved by executing a finer grid screen of the major component of the hit condition (eg. Salt, precipitant or solvent) along with a pH screen around the potential condition obtained.

2.17.5 Additive Screening

Sometimes addition of small molecules to the crystallization conditions could alter protein solubility, interaction between protein-protein molecules and aid crystal growth. The 72 reagents of Hampton Additive screen were reformatted into 96-well format by Dr. Tom Walter. The standard crystallization plate was set up using a single crystallization condition (the potential condition obtained in coarse screen) in which 100µl of reservoir was dispensed employing the Robogo (MWG Biotech, UK). 100nl of protein was dispensed onto the crystallization platform as sitting drops followed by addition of 100nl of the reservoir by the

Cartesian robot (Cartesian Technologies Microsys MIC4000). 100nl of the non-volatile reagents are dispensed onto the protein-reservoir droplet whereas for volatile reagents, 20µl is added to the reservoir after the reservoir is added to the droplet. The non-volatile reagents obtained from the Hampton Research were diluted 5:1 with water before using them (Walter et al. 2005).

2.17.6 Three-row optimization

In three- row optimization, only F, G and H rows (36 well) are used. In this method, the original screen is dispensed on the crystallization plate reservoir wells and simultaneously also diluted with MQ water in 1%-3% steps by the Robogo (MWG Biotech, UK) leading to the formation of mother liquor which contains 100% to 67% of the screen components. Crystallization droplets are generated by the Cartesian robot (Cartesian Technologies Microsys MIC4000) such that it optimizes protein: mother liquor ratio by dispensing protein such that the final protein: mother liquor ratios are 1:1, 1:2 and 2:1 in rows F, G and H respectively of the crystallization plate. The plate was then immediately sealed manually with transparent self-adhesive foils (Viewseal, Grenier) and stored in the storage vault at the required temperature.

2.17.7 Micro seeding

Sometimes spontaneous nucleation during crystallization leads to the formation of several tiny, low quality crystals (micro crystals). One way to overcome this issue is by bypassing the nucleation zone altogether by inserting ready-made nuclei into the metastable zone. This leads to the epitaxial growth of the protein crystal on the provided crystalline surface and may lead to growth of better diffracting crystals (Chayen and Saridakis 2008).

Micro seeding was performed by harvesting microcrystals in their mother liquor and crushing them with Teflon beads ('Seed Bead', Hampton Research) by vortexing for 90 seconds. 50nl of the seed stock was added to each well after setting up the crystallization screen of the protein in the same or different potential crystal yielding condition (Walter et al. 2008).

2.17.8 Crystallization of PHD2 and PHD2-GATA1 complex

Crystallization of PHD2 domain of JARID1A (6.67mg/ml in 50mM HEPES, pH 8, 200mM NaCl, 1mM DTT) purified from *E. coli* cells was carried out by mixing 100nl of protein and 100nl of the reservoir solution as described earlier in section 2.17.3. Crystallization screens Block1, Block3, PACT premier, Index, SaltRx and PEGRx were set up (**Table 2.13**).

Crystallization of PHD2-GATA1 complex was carried out by mixing PHD2 (3.3mg/ml in 50mM HEPES, pH 8, 200mM NaCl, 1mM DTT) and GATA1 (7.7mg/ml in 50mM Tris, pH 8), 200mM NaCl, 1mM DTT) in equimolar ratio (1:1). Crystallization screens which were set up are Block1, Block2, Block3, Block4, Index, PACT Premier, PEGR$_X$, KeraFast, NucPro (**Table 2.13**).

2.17.9 Crystallization of PHD1 domain of JARID1A

Crystallization of PHD1 domain of JARID1A (5.26mg/ml in 50mM HEPES, pH 8, 200mM NaCl, 1mM DTT) purified from Sf9 cells was carried out by mixing 100nl of protein and 100nl of the reservoir solution as described earlier in section 2.17.3. Crystallization screens which were set up are Block2, Block4, NucPro and KeraFast (**Table 2.13**).

2.17.10 Crystallization of LMO1

Full length LMO1 protein is unstable and is prone to aggregation which makes it very difficult to purify. Hence, FLINC1 construct was designed by the fusion of N- and C-terminal LIM domains of LMO1 with LID domain of LDB1 proteins. These two proteins, LMO1 and LDB1 were connected by a flexible linker that is 11 amino acid residues long (GGSGGHMGSGG).

Crystallization of LMO1 (50mM HEPES, pH 8, 500mM NaCl, 1mM DTT) was carried out by mixing 100nl of protein mixture with 100nl of reservoir as described in section 2.17.3. The following crystallization screens were set up: Block1, Block2, Block3, Block4, Index, $PEGR_X$, KeraFast, NucPro, and Morpheus (**Table 2.13**). Crystals were obtained in the $PEGR_X$ screen that were then optimized using additive screen (section 2.17.5.), three-row optimization (section 2.17.6.) and micro seeding (section 2.17.7.).

2.18 Multi-angle light scaterring (MALS)

The complex formation between FLINC1 and E47/SCL bHLH was confirmed using MALS which was carried out in flow mode meaning that MALS detector was coupled downstream to SEC. MALS was performed using a Superdex-75 10/300 column (GE Healthcare) with an injection volume of 90µl. First the column was equilibrated in buffer [50mM HEPES (pH 8), 200mM NaCl, 1mM DTT] followed by running of Ovalbumin with a molecular weight of 44kDa through the column as a standard to determine the elution volume at which the complex could be expected to elute. Two different concentrations of the complex were run on MALS- 1mg/ml and 2mg/ml to ensure that the complex formation is not concentration dependant. The protein fractions eluted from the column were detected using a Dawn Heleos II (Wyatt Technology) at 20°C.

2.19 Analytical Ultracentrifugation

Analytical Ultracentrifugation (AUC) is a versatile biophysical technique used for the characterization of proteins and protein complexes. In AUC, sedimentation of macromolecules in the centrifugal field is monitored in solution without interaction with any matrix or surface and without any chemical modification of the protein, which allows the determination of their thermodynamic and hydrodynamic properties. One of the established techniques used for the measurement of the solute gradients in AUC experiments is the real-time measurement of optical signal (absorbance or interference) as a function of radial distance. Two types of experiments exist, sedimentation velocity, which is a hydrodynamic method most commonly used for measurement of static interactions, and sedimentation equilibrium, which is a thermodynamic method usually used for measurement of dynamic interactions (Howlett et al. 2006) In this thesis, only sedimentation velocity experiments were performed and hence only those will be described in detail.

2.19.1 Basic Principles of sedimentation velocity and hydrodynamics

When centrifugal force is applied on to the macromolecules during sedimentation, a depletion of macromolecules occurs at the meniscus thereby forming a concentration boundary that moves towards the bottom of the centrifuge cell as a function of time.

When a solute suspended in a solvent is subjected to gravitational field, three different types of forces are exerted on it as shown in (**Figure 2.2**) (Lebowitz et al. 2002).

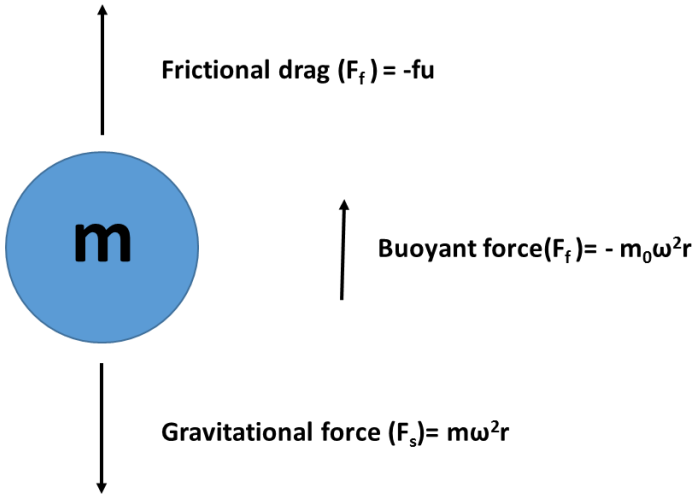


Figure 2.2: Forces acting on a molecule in AUC

First is the **gravitational force (F_s)**, which is proportional to the mass of the particle and its acceleration. The latter in case of a spinning rotor is determined by the distance of the particle from the axis of rotation (r) and square of angular velocity (ω).

$$F_s = m\omega^2 r = \frac{M}{N}\omega^2 r$$

(Equation 1)

m is the mass of a single solute molecule in grams, M is the molar weight of the solute (in grams/mole) and N is Avogadro's number.

Second is the **buoyant force (F_b)**, which is proportional to the weight of fluid displaced by the solute. The negative sign shows that the force is imposed in the opposite direction to that of the gravitational force.

$$F_b = -m_0\omega^2 r$$

(Equation 2)

m_0 is the mass of fluid displaced which is proportional to the density of the solvent (ρ).

$$m_0 = m\bar{v}\rho = \frac{M}{N}\bar{v}\rho$$

(Equation 3)

$\bar{v}$ is the volume in ml occupied by each gram of the solute in solution (partial specific volume).

If the density of the solute is greater than that of the solvent, the solute will start sedimenting radially towards the bottom of the cell. So, the third type of force which is exerted on the solute particle is the **Frictional drag (F_f)** which is proportional to the velocity with which the particle sediments towards the bottom of the cell (u). Frictional drag is in action in the opposite direction to the gravitational force.

$$F_f = -fu$$

(Equation 4)

f is the frictional coefficient which relies on the size and shape of the solute.

All the three forces, gravitational, buoyant and frictional described above balance rapidly within very short time (less than 1μs). Therefore,

$$F_s + F_b + F_f = 0$$

(Equation 5)

Combining equations 1, 3 and 4,

$$\frac{M}{N}\omega^2 r - \frac{M}{N}\bar{v}\rho\omega^2 r - fu = 0$$

(Equation 6)

Rearranging,

$$\frac{M}{N}\omega^2 r(1 - \bar{v}\rho) - fu = 0$$

(Equation 7)

$$\frac{M(1 - \bar{v}\rho)}{Nf} = \frac{u}{\omega^2 r} = s$$

(Equation 8)

Where s is the sedimentation coefficient, which is the velocity of the solute per gravitational acceleration. Sedimentation coefficient is directly proportional to the buoyant effective molecular weight of the solute (molecular weight which is corrected for the effects of buoyancy) and inversely proportional to the frictional coefficient as shown in equation 8 (Ralston).

2.19.2 Experimental set-up for AUC

Analytical Ultracentrifugation experiments were carried out using a Beckman Optima XL-I analytical ultracentrifuge operating in velocity mode at 20°C. Experiments were performed utilizing 280 nm absorbance optics. For sample loading, cell components were assembled such that the 3mm double sector shaped-centrepiece, which is surrounded by two thick optical-grade quartz lenses, forms a cavity for the sample. The centerpiece forms two sectors:

one for the reference and another for the sample. The rotor (An 60Ti) was spun at 1,29,024×g (40,000 RPM) under vacuum and double-sector cells enable the optical detector system to measure the absorbance of the sample taking into consideration absorbance by any of the buffer components. The optical detection system collects data every time the cell comes in alignment with the optical light path and measures the difference in the absorbance between the sample and the reference sectors. A total of 150 scans were collected with an interval of 6 minutes as each scan takes 2 minutes and 3 scans were taken, one for each cell.

2.19.3 Sample preparation for AUC

AUC was carried out to analyse the interactions between PHD domains of JARID1A and N and C zinc fingers of GATA1. Proteins were diluted to 1mg/ml concentration for their individual runs. For the mixture, equimolar concentrations of both the purified proteins were mixed together and loaded onto the cell. Three samples were run in a single experiment, two individual protein samples and third was a mixture of both. Both proteins were in the buffer 50mM HEPES, pH 8, 200mM NaCl and 1mM DTT.

2.19.4 Data analysis for AUC

Data analysis was carried out using the SedFit software (by Schuck, http://www.analyticalultracentrifugation.com). Data from scans 10-120 were used for the analysis aimed at eliminating artefacts arising from irregularities at the meniscus. Analysis was carried out using continuous c(s) model to calculate the sedimentation coefficients of protein species in the sample at a resolution of 100 and also in $c(s,f/f_0)$ mode to compare the different types of molecules present in the sample with varied hydrodynamic radii at a resolution of 10. The frictional coefficient is a unitless entity and is the ratio between the experimental frictional coefficient and the minimum frictional coefficient if the molecules in

the sample were compact spheres (f/f_0). Analysed data were then plotted using the pro Fit software (QuantumSoft) and fitted with a Gaussian function to determine peak centres, which represent the sedimentation coefficients of the species. Sedimentation coefficient (s) was plotted on X-axis, continuous sedimentation coefficient distribution (c(s)) on Y1-axis, molecular weight on Y2-axis and $c(s,f/f_0)$ was plotted over a third dimension which was represented by a heat plot.

Chapter 3: Interaction of endogenous JARID1A with the SCL multiprotein complex

3 Interaction of endogenous JARID1A with the SCL multiprotein complex

3.1 Introduction

In this chapter, I will discuss the techniques and experiments performed to study endogenous interactions between JARID1A and the SCL multiprotein transcription factor complex. Klose and colleagues showed that the knock down of JARID1A in mouse models leads to a significantly decreased rate of apoptosis and enhanced cycling of cells in the haematopoietic stem cell and myeloid compartments (Klose et al. 2007). This suggests that JARID1A could play a role in haematopoiesis. Additionally, it was previously shown that JARID1A interacts with LMO2 (Mao et al. 1997). As LMO2 is a part of the SCL multi-protein complex, it is possible that JARID1A interacts with this complex and that this interaction accounts for its reported role in haematopoiesis.

To investigate the role of JARID1A in haematopoiesis, we began by performing size-exclusion chromatography of nuclear extracts derived from Mouse ErythroLeukaemia (MEL) cells, which closely recapitulate erythropoiesis upon differentiation. MEL cells are erythroid precursor cells that are transformed by the Friend virus complex and are arrested at the proerythroblast stage (Friend et al. 1966). MEL cells serve as a valuable model system for studying erythroid differentiation as these cells undergo biochemical and morphological changes upon differentiation, which closely resemble normal erythroid differentiation. These events include chromatin condensation, accumulation of globin mRNAs, increased haemoglobin synthesis and erythroid specific membrane changes (Harrison 1976). MEL cell differentiation can be induced using various chemical agents such as nicotinamide and its

analogues, butyric acid, DMSO, N-, N- dimethylformamide, and other highly polar compounds like N, N-dimethylacetamide, acetamide, N-methyl formamide etc (Friend et al. 1971; Leder and Leder 1975; Tanaka et al. 1975; Terada et al. 1979).

3.2 Analysis of multiprotein complexes by size-exclusion chromatography

To probe interactions between the endogenous JARID1A and its potential binding partners and to analyse the composition and size of complexes, protein fractionation experiments were carried out by Size-exclusion chromatography (SEC). SEC separates proteins based on their size and shape. Proteins with higher molecular weight are eluted first and those with lower molecular weight are eluted later. So, if proteins are in a complex, these are eluted earlier compared to when they exist in isolation. Proteins from the MEL cell nuclear extracts were fractionated using an analytical Superose-6 10/300 gel-filtration column calibrated with low and high molecular weight (MW) markers (Appendix-IV, Figure A.14). The void volume of the Superose-6 10/300 column is 8ml and its separation range is 5 to 5000 kDa. All the elution peak fractions were probed with antibodies against JARID1A (using monoclonal antibodies from Abcam), GATA1, SCL, LDB1 and LMO2 using western blot analysis. The elution profile showed that JARID1A elutes in a number of different fractions of high MW with other proteins of the complex (**Figure 3.1**). The first peak in the UV absorbance spectra was observed in correspondence to the void volume that generally corresponds to aggregates of proteins/ complexes too large to be resolved by a Superose-6 column. On the right shoulder of this peak (fractions that eluted between 8-12ml, the molecular weight marker thyroglobulin with size of 669kDa eluted at 12 ml under same buffer conditions and flowrate), JARID1A (200kDa) was detected together with SCL (45kDa), GATA1 (45kDa), LMO2 (15kDa) and LDB1 (46kDa). This shoulder is therefore likely to contain a big multiprotein complex that includes JARID1A and the SCL complex. In a separate second

experiment this shoulder could be resolved as a distinct peak at 11ml elution volume (Appendix-IV, Figure A.13).

In a second elution peak (fractions that eluted at 13-14ml, the molecular weight marker ferritin with size of 440kDa eluted at 14.5ml under same buffer conditions and flowrate), JARID1A was also detected together with SCL, LMO2, LDB1 and GATA1 corresponding to a total molecular weight of 410kDa. In a third elution peak (fractions that eluted at 15-16ml, the molecular weight marker aldolase with size of 158kDa eluted at 16ml under same buffer conditions and flowrate) SCL, LMO2, LDB1 and GATA1 were detected corresponding to a total molecular weight of 200kDa probably forming the "pentameric" complex (Wadman et al. 1997). Elution peak fractions were not probed for E47 as working E47 antibodies were not available at that time, however, it is likely to be present in the complex as SCL forms heterodimers with E47 (Hsu et al. 1991). JARID1A was not detected in this peak. SCL, LDB1 and GATA1 are also detected in fractions that eluted at 17ml at which the molecular weight marker ovalbumin with size of 43kDa eluted under same buffer conditions and flowrate, which is compatible with the three proteins present individually. The fractions eluted between 19-21ml showed the presence of LMO2 in isolation as the molecular weight marker Ribonuclease A with size of 13.5kDa eluted at 19ml under same buffer conditions and flowrate (**Figure 3.1** and Appendix-IV, Figure A.14).

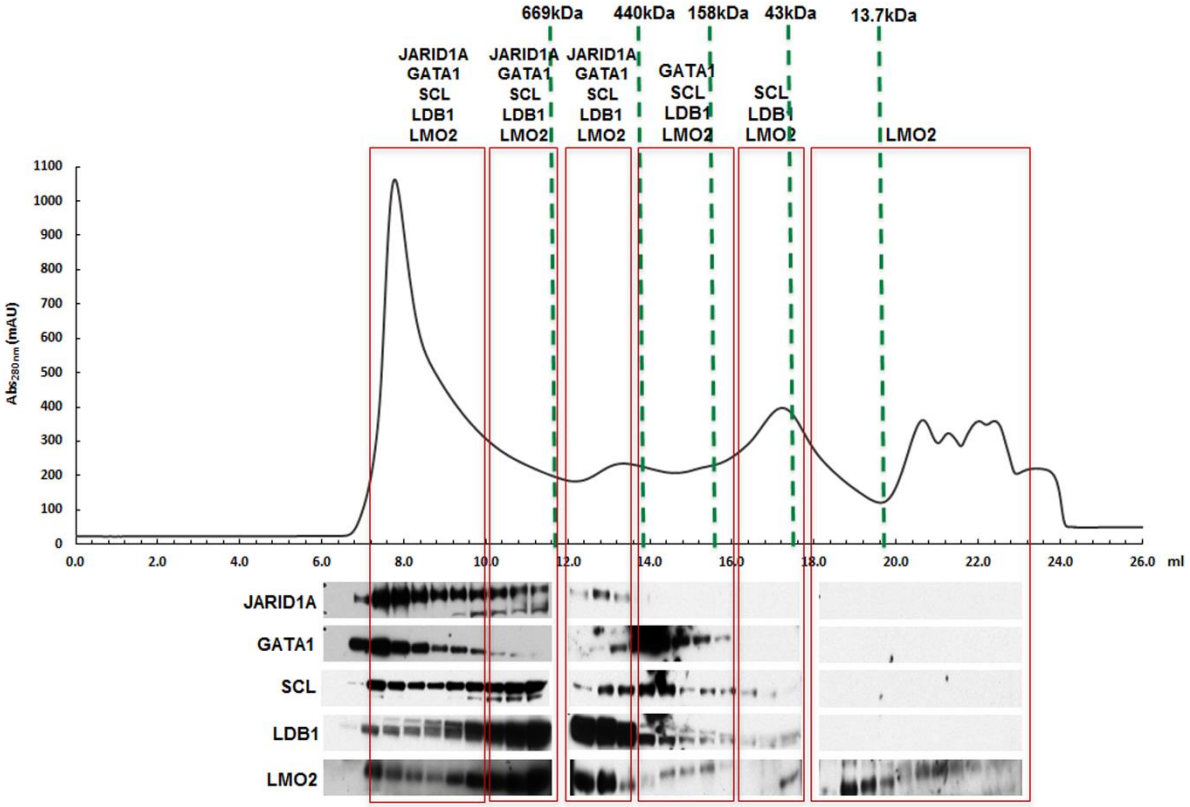


Figure 3.1: **Size-exclusion chromatography.** MEL cell nuclear extracts were fractionated by size-exclusion chromatography and the eluted fractions were Tri-chloro acetic acid precipitated and analysed by western blotting. The X-axis represents the elution volumes in ml and the Y-axis represents the UV absorption of proteins at 280nm wavelength. The proteins detected at different elution volumes are indicated at the top of the blue boxes. The green vertical dotted lines highlight the elution volume at which the standard molecular weight markers corresponding to the molecular weights indicated at the top eluted under same experimental conditions.

In conclusion, JARID1A co-elutes in two separate elution volumes with components of the SCL complex. The first one contains SCL/LMO2/LDB1/GATA1 and has an apparent MW of ~669kDa or higher, whilst the second one is smaller, with an apparent MW of around 440kDa which also contains SCL, LMO2, LDB1 and GATA1.

3.3 Analysis of interaction between JARID1A and the SCL complex by Co-immunoprecipitation assays

To further characterise the endogenous interactions between JARID1A and the SCL complex, we performed Co-immunoprecipitation (Co-IP) assay using MEL cell nuclear extracts.

3.3.1 Optimization of immunoprecipitation of JARID1A

Immunoprecipitation relies on the use of an antibody to precipitate a protein out of solution followed by detection of the protein by western blot. This technique requires optimisation of the protocol conditions in order to get a good signal from the protein and a low background which assures the identity of the endogenous protein detected. First, we tested three different commercial antibodies to find the antibody that would bind JARID1A with higher affinity (this translates into higher amounts of protein being precipitated out of solution) (i) Monoclonal (mouse monoclonal raised against residues 1416-1434 of human JARID1A), (ii) Purified polyclonal (rabbit polyclonal raised against 1675 and 1722 residues of human JARID1A) and (iii) Unpurified polyclonal (rabbit polyclonal raised against residues 1440-1687 of human JARID1A); all antibodies were purchased from Abcam®. Although the monoclonal antibodies were raised against human peptide, they have specificity towards mouse JARID1A owing to the presence of exactly the same peptide in mouse.

Antibodies were initially tested for binding by performing western blotting analysis on the nuclear extracts generated from Human embryonic kidney cells (HEK). All three antibodies could identify endogenous JARID1A, however only the mouse monoclonal antibodies showed a strong intensity for JARID1A (**Figure 3.2a**). As a result, all immunoprecipitation

(IP) assays were performed using the monoclonal antibodies. Interestingly, two bands (~250kDa and ~180kDa) could be detected using any of the three antibodies.

All IP assays were performed using nuclear extracts from MEL cells and Protein G beads that have recombinant Protein G covalently coupled to their surface. Protein G has affinity with the Fc region of antibodies. As a negative control, nuclear extracts mixed with immunoglobin G (IgG) were used. To optimise the IP protocols, different ratios of Protein G beads to nuclear extract (1:2, 1:3, 1:5 and 1:7) were used. As shown in **Figure 3.2b,** the ratio 1:3 was found to be the optimal since the Immunoglobulin G (IgG) lane (lane 5) was clear and the background was minimal. JARID1A monoclonal antibodies and a 1:3 Protein G beads to nuclear extract ratio were therefore used for all subsequent experiments unless specified.

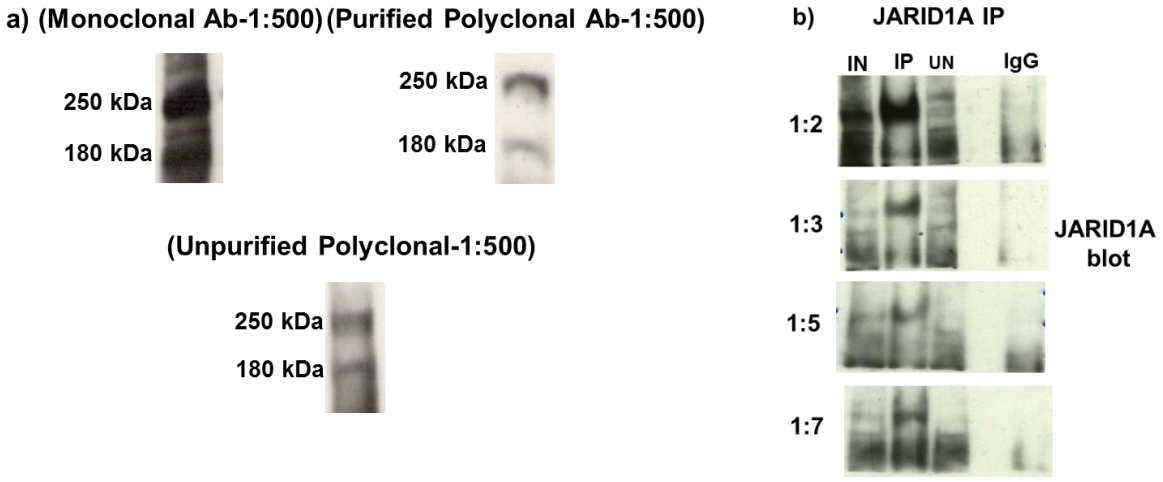
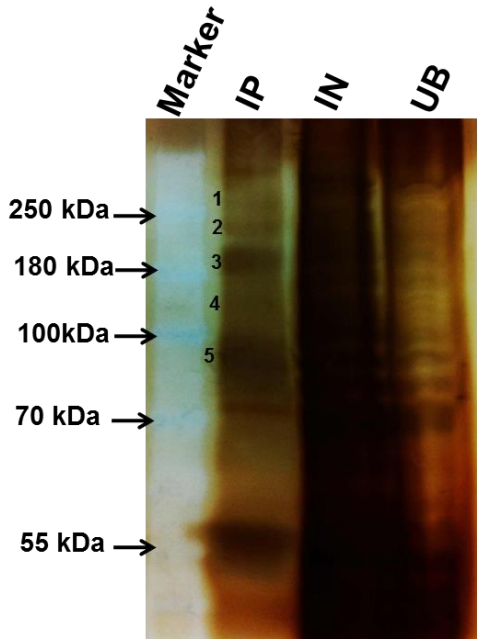


Figure 3.2: JARID1A IP. a) Different JARID1A antibodies, monoclonal, purified polyclonal and unpurified polyclonal were screened by western blotting of HEK cell nuclear extracts. **b)** JARID1A IP was carried out on MEL nuclear extracts using JARID1A monoclonal antibodies and 1:2, 1:3, 1:5 and 1:7 beads: protein ratios. Shown in the figure are input (IN), immunoprecipitation (IP), unbound (UN) and IgG control fractions.

As seen in **Figure 3.2a**, the antibodies are highlighting two bands at 250kDa and 180kDa. The band at ~250kDa is heavier than expected for JARID1A (MW 196kDa), but it is likely to be the full-length protein as it was characterized by mass spectrometry described below. This could be due to the post-translational modifications of JARID1A. The second band could be a degradation product.

In humans, two isoforms of JARID1A have been reported [Uniprot P29375.1 (192kDa) and P29375.2 (186kDa)]. So, an alternative explanation is that the second band at ~180kDa is the smaller isoform of JARID1A. The second isoform lacks the PHD3 domain. Since JARID1A antibodies used were raised against a region between PHD2 and PHD3 of human JARID1A, they could be recognising both isoforms.

To verify the identity of bands, we used the Mass Spectrometry analysis (MS). For MS, JARID1A IP (we used 4mg of nuclear extract instead of the usual 1mg, whilst the ratio of protein G beads to nuclear extract was kept the same at 1:3) was performed and the precipitate resolved on a SDS-PAGE gel and stained by silver staining. Silver staining was used instead of Coomassie due to its higher sensitivity. IP (immunoprecipitated), IN (input to show that JARID1A is expressed in HEK), and UB (unbound) fractions were analysed. The two bands from the immunoprecipitated sample as shown in **Figure 3.3**, were excised and analysed by MS. The results confirm that both bands correspond to JARID1A.



Band no.	Matches	Sequences	Hits
Band 1	5	3	JARID1A
	1	1	Nucleolar transcription factor1
Band2	16	12	JARID1A
	9	6	Microtubule associated protein 4
	1	1	Nitrogen permease regulator3
Band 3	29	17	Microtubule associated protein 4
	5	5	JARID1A
	2	2	Prosalusin
	1	1	Platlet glycoprotein Ib alpha chain
Band 4	6	4	Poly[ADP-ribose] polymerase1
	7	5	Nucleolin
	2	2	Roudabout homolog 1
	2	2	ATP-citrate synthase
	5	5	ATP-binding cassette subfamily
	2	2	Latrophilin-1
	2	2	WD repeat-containing protein 7
Band 5	24	15	Heat shock protein HSP 90-beta
	19	12	Heat shock protein HSP 90-alpha
	12	8	Putative heat shock protein HSP 90
	2	2	Elongation factor 2
	3	2	Olfactory receptor 6K3
	3	2	Microtubule associated protein 4
	1	1	Tubulin alpha-1C chain

Figure 3.3: Silver stained gel showing JARID1A IP. JARID1A IP was performed using the monoclonal antibodies and the resulting proteins were analyzed by SDS-PAGE followed by silver staining. The dominant bands were excised and used for the MS analysis. The bands selected for the MS analysis are numbered. Shown in the gel (left) are input (IN), immunoprecipitation (IP), unbound (UN) fractions. Hits obtained from MS for the five bands are listed in the table (right).

3.3.2 Optimization of JARID1A Co-immunoprecipitation

Next, to probe potential interactions of JARID1A with components of the SCL multiprotein complex, we performed co-immunoprecipitation assays where JARID1A antibodies were used for IP and the co-immunoprecipitated proteins probed by western blot analysis using antibodies against SCL, LMO2, LDB1 and GATA1.

Since the molecular weight of the IgG heavy chain (50kDa) is similar to that of SCL (45kDa) and GATA1 (45kDa), the protein G beads and JARID1A antibodies were cross-linked to try and reduce the contamination from the IgG. Although the intensity of the heavy chain band was reduced inexplicably, there was also a significant reduction in the efficiency of the IP. Furthermore, a non-specific band appeared in the negative control (shown by an arrow in

Figure 3.4). As we wanted to maintain optimal efficiency of the JARID1A IP, we decided not to use crosslinking in subsequent experiments.

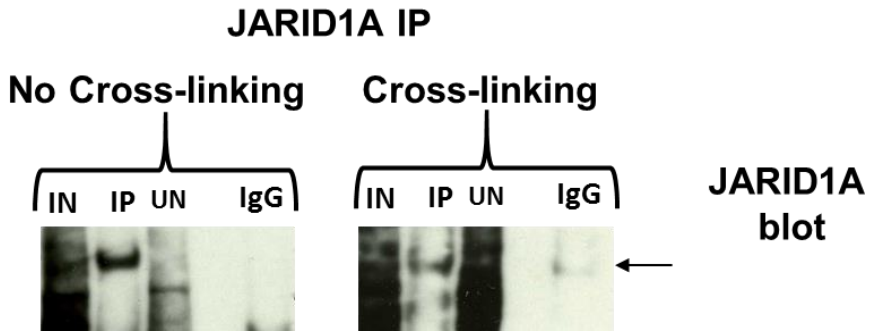


Figure 3.4: JARID1A IP with and without cross-linking of the Protein G beads. JARID1A IP was performed using MEL nuclear extracts with and without chemically cross-linked Protein G dynabeads to the JARID1A antibodies and analyzed by western blotting. Shown in the figure are input (IN), immunoprecipitation (IP), unbound (UN) and IgG control fractions.

All Co-IP were therefore performed without crosslinking and after blocking beads with salmon sperm DNA and BSA, and pre-clearing the nuclear extracts as described in Chapter 2, Section 2.3. Despite working efficiently in immunoprecipitating JARID1A, the monoclonal antibodies failed to co-immunoprecipitate SCL, GATA1, LDB1 or LMO2 (not shown) possibly because monoclonal antibodies (MAbs) are blocking the binding site on JARID1A. We therefore decided to use the JARID1A purified polyclonal antibodies we had initially discarded (**Figure 3.2a**). The polyclonal antibodies are known to bind the C-terminal of JARID1A. The antibodies successfully pulled down JARID1A, SCL and GATA1 (**Figure 3.5**). However, LDB1 and LMO2 were not pulled down.

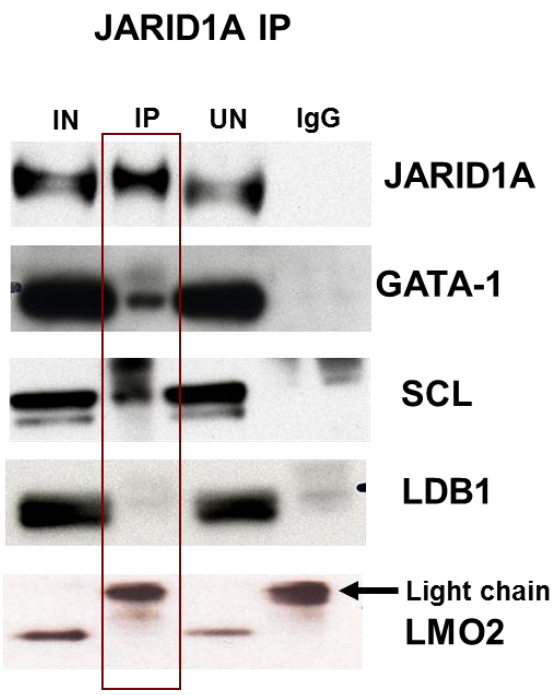


Figure 3.5: JARID1A IP using polyclonal antibody. JARID1A IP was perfomed using the MEL nuclear extracts employing polyclonal JARID1A antibodies and western blotting analysis was carried out for the detection of JARID1A, GATA1, SCL, LDB1 and LMO2. Shown in the figure are input (IN), immunoprecipitation (IP), unbound (UN) and IgG control fractions.

These results confirm that the endogenous JARID1A interacts with the endogenous SCL and GATA1 in MEL cells. As both SCL and GATA1 play a crucial role in erythropoiesis, it is possible that JARID1A, a chromatin remodeling protein, is also involved in the regulation of erythropoiesis. Since JARID1A does not interact with LMO2 or LDB1, therefore, it is possible that it does not need LMO2 or LDB1 for mediating functions of SCL or GATA1.

3.3.3 Optimization of Reverse Co-immunoprecipitation

To further detail interactions between SCL/GATA1 and JARID1A, we performed reverse Co-IPs. SCL, LDB1 and LMO2 IPs were performed in a way similar to what is described above (using TAL1 C-21, LDB1 Clim2 and LMO2 antibodies respectively. For antibody

dilutions, refer to table 2.2 in Chapter 2) and the pulled down proteins were probed for JARID1A. JARID1A was reproducibly detected in the SCL IP **(Figure 3.6a** top panel**)**.

JARID1A was not detected in the LDB1 IP **(Figure 3.6b** top panel**)**. A weak signal for JARID1A was detected in LMO2 IP **(Figure 3.6c** top panel**)**, however, LMO2 was not detected in the JARID1A Co-IP **(Figure 3.5** bottom most panel**)**. Therefore, it cannot be deducted if JARID1A interacts with LMO2 or not.

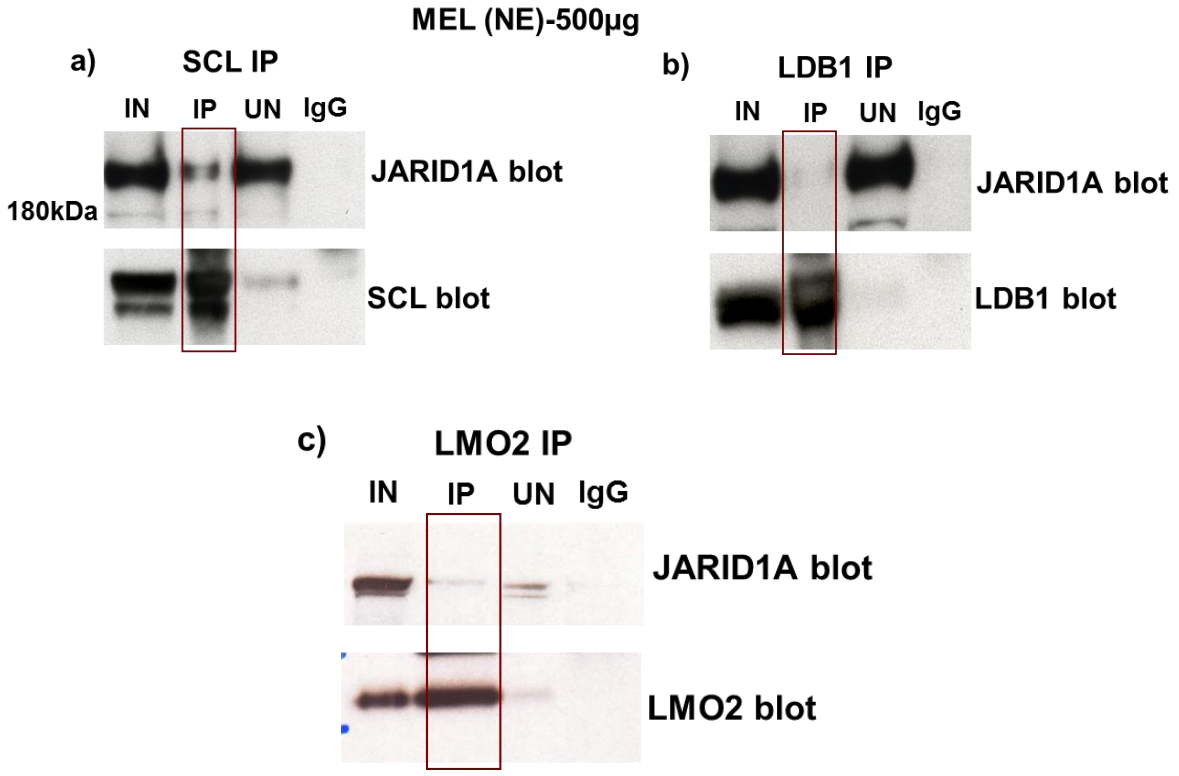


Figure 3.6: SCL, LDB1 and LMO2 IP. SCL, LDB1 and LMO2 IPs (a, b and c respectively) were performed from 500 µg of MEL cell nuclear extracts (NE) and the western blotting analysis was carried out for the detection of JARID1A. Shown in the figure are input (IN), immunoprecipitation (IP), unbound (UN) and IgG control fractions.

The protocol for GATA1 IP required optimisation. First, it was performed using the GATA1 N6 antibody (monoclonal). However, no GATA1 could be pulled down. To improve the

efficiency of the IP, a rabbit anti-rat bridging antibody was used to increase the number of epitopes for the monoclonal antibody on the Protein G beads. N6 antibodies were also cross-linked on the protein G beads and compared with the bridging method (**Figure 3.7**). GATA1 IP with bridging antibodies also failed to detect GATA1 (bottom left panel) or JARID1A (top left panel). A band was observed at a slightly higher position in the GATA1 blot compared to that of the input (IN) and unbound (UN) bands; a similar (although much weaker) band was also observed in the IgG lane implying that the band corresponds to the IgG heavy chain. In cross-linking experiment, a very weak band was detected in the GATA1 blot (IP lane, bottom right panel). When probed for JARID1A, a band was detected (IP lane, Top right panel) however, the same band was also detected in the IgG lane. Hence, none of these methods could be used unequivocally to detail interactions between the endogenous GATA1 and JARID1A when using GATA1 N6 antibody.

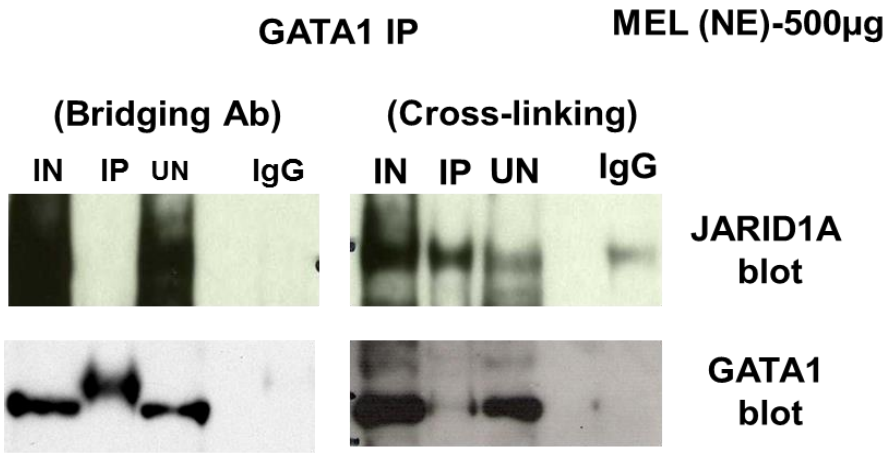


Figure 3.7: Comparison of GATA1 IP using bridging antibodies and cross-linked Protein G beads. GATA1 IP was performed with the N6 antibodies using both techniques on 500 µg MEL nuclear extracts (NE) and western blotting was performed to detect JARID1A. Shown in the figure are input (IN), immunoprecipitation (IP), unbound (UN) and IgG control fractions.

Different GATA1 antibodies were therefore tested. In parallel, two different Protein G beads: nuclear extract ratios were also tested (1:10 and 1:3) (**Figure 3.8**). The IP of GATA1 was most efficient with GATA1 M20 antibodies (notice the very weak signal in the UN lane, **Figure 3.8a**, left bottom panel). These antibodies were also able to co-immunoprecipitate JARID1A. The detection of JARID1A was less efficient when a ratio of 1:3 beads to nuclear extract ratio was used, compared to a 1:10 ratio (**Figure 3.8b)**. All subsequent GATA1 IPs therefore use 1:10 ratio. For antibodies dilutions refer to table 2.2 in chapter 2.

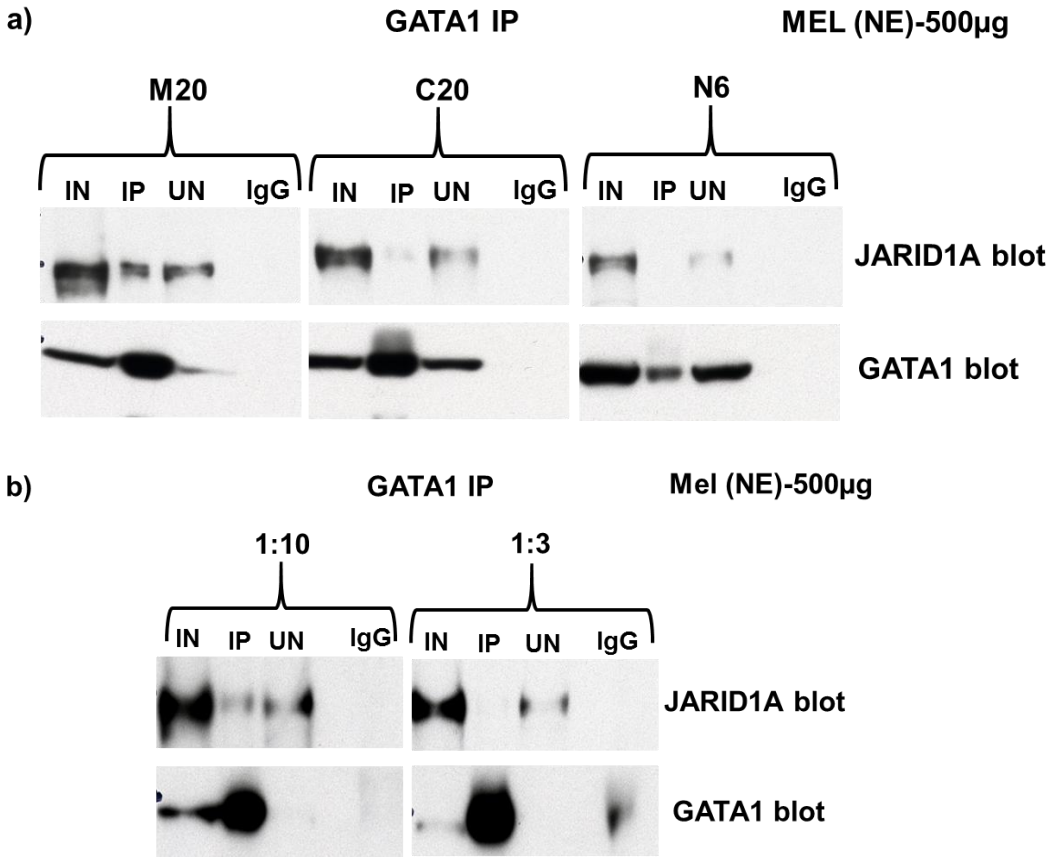


Figure 3.8: Optimizing GATA1 IP using different antibodies and beads: protein ratio. **a)** GATA1 IP was optimized on 500 µg of MEL cell extracts using different antibodies (M20, C20 and N6). **b)** IP was performed with the GATA1 M20 antibodies and 1: 10 and 1:3 Protein G beads: protein ratios and analysed by western blotting. Shown in the figure are input (IN), immunoprecipitation (IP), unbound (UN) and IgG control fractions.

The results indicate that JARID1A interacts with transcription factors SCL and GATA1 in MEL cells, but doesn't interact with LDB1. A faint JARID1A band was observed in LMO2 IP, but the reverse IP did not show any LMO2, hence its interaction with LMO2 needs to be further investigated. As both SCL and GATA1 play a crucial role in erythropoiesis, it is possible that JARID1A, a chromatin remodeling protein, is also involved in the regulation of erythropoiesis.

3.4 Interaction of JARID1A with SCL and GATA1 at D0, D2 and D4 of MEL cell differentiation

To study the interaction of JARID1A with SCL and GATA1 during erythropoiesis, we first analysed JARID1A expression levels at day0 (D0), day2 (D2) and day4 (D4) of MEL cell differentiation.

3.4.1 JARID1A levels at different stages of MEL cell differentiation

MEL cells are blocked at an immature, proerythroblast stage. Terminal erythroid differentiation can be induced with chemical agents, such as DMSO. Once cells undergo differentiation, morphological and molecular changes occur including decrease in the nuclear:cytoplasm ratio, establishment of the gene expression programme necessary for the activation of erythroid-specific genes and abundant production of haemoglobin resulting from increased transcription of globin genes (Fraser and Curtis 1987). The differentiation of MEL cells was induced by using 2% DMSO (Friend et al. 1971). The differentiated cells were identified by staining with May-Grünwald Giemsa (MGG) stain, which stains the nucleus purple and cytoplasm light blue. This highlights the changing nuclear: cytoplasm ratio in differentiating MEL cells. Some of the morphological changes observed over 4 days of differentiation are shown in **Figure 3.9**. D0 represents proerythroblast stage with prominent

nuclei. D2 represents polychromatic erythroblasts with small, dense nuclei with more prominent cytoplasm and D4 represents orthochromatic erythroblasts (also known as normoblasts), which also have small, dense nuclei and more prominent cytoplasm. Although benzidine staining which stains haemoglobin was not performed, at D4 benzidine positive cells are expected as orthochromatic erythroblasts do produce haemoglobin (Friend et al. 1971; Harrison 1976) Also, when cells at D4 were centrifuged, the cell pellets were red in colour further confirming the presence of haemoglobin. Hence, these cells recapitulate normal erythroid maturation.

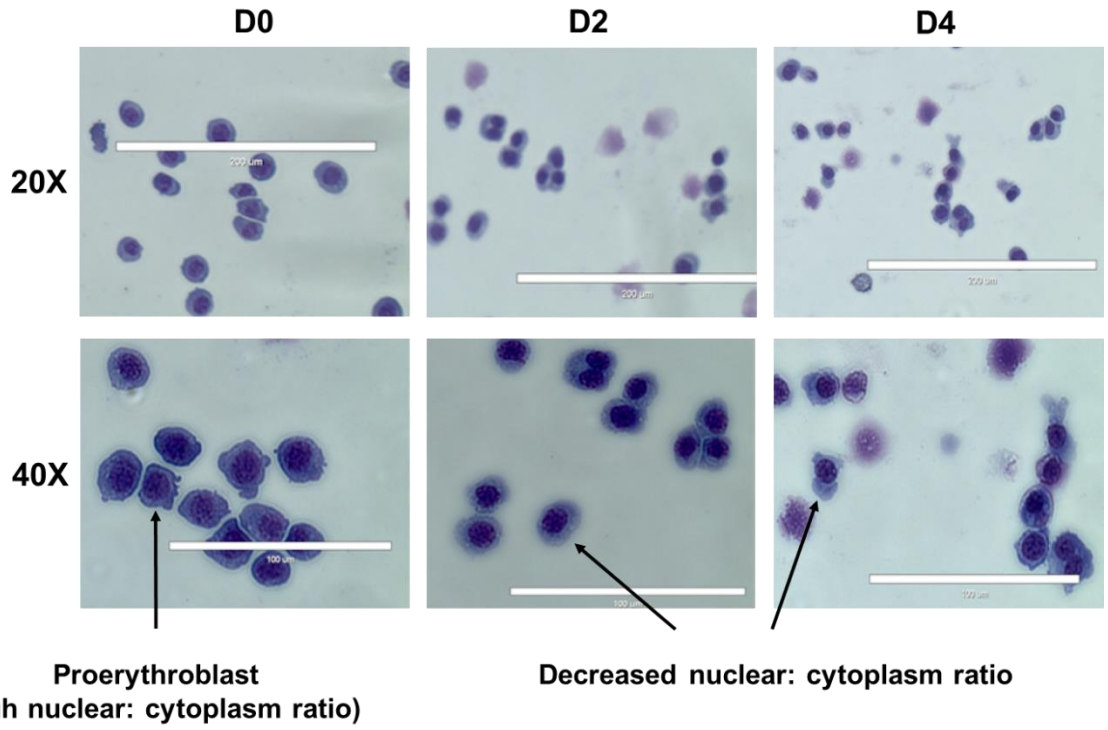


Figure 3.9: MGG staining of MEL cells at day 0 (D0), day 2 (D2) and day 4 (D4) of differentiation. MEL cell differentiation was induced by addition of 2% DMSO. Cells from D0, D2 and D4 were cytospun on glass slides and stained with MGG. Cells were imaged at two different magnifications, 20X and 40X and the white bars represent the scales.

We followed the expression pattern of JARID1A using western blot analysis and found that JARID1A protein levels seemed to slightly decrease as cells undergo differentiation (**Figure 3.10**).

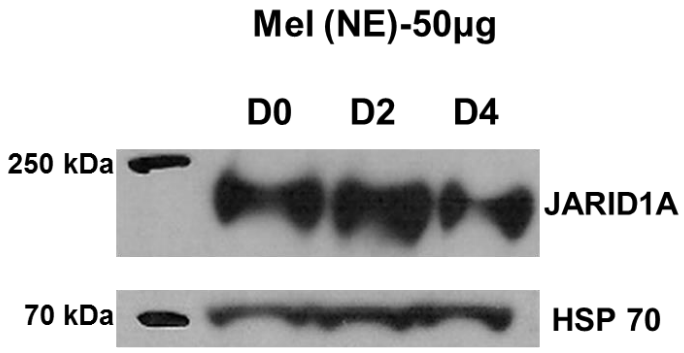


Figure 3.10: Expression levels of JARID1A at D0, D2 and D4 of MEL cells differentiation. JARID1A expression was checked at D0, D2 and D4 of MEL cell differentiation by western blotting. Heat shock protein 70 (HSP70) was used as loading control.

3.4.2 SCL and GATA1 Co-immunoprecipitation assay at D0, D2 and D4 of MEL cell differentiation

To determine if the interaction of JARID1A with SCL and GATA1 varies during MEL cell differentiation, SCL and GATA1 co-immunoprecipitation assays were performed using nuclear extracts from D0, D2 and D4 of MEL cells differentiation and analysed by Western blot (**Figure 3.11**). SCL expression levels do not change throughout differentiation, which is evident from the similar exposure times of the SCL blot at D0, D2 and D4. However, the interaction of SCL with JARID1A vary through differentiation. The interaction is stronger at D2 compared to that at D0 and D4. Specifically, the smaller isoform of JARID1A is preferentially pulled down with SCL at D0 and D4. Unfortunately, the western blot for LMO2 is not clear enough to analyse. Upon repeating this experiment once more, these results were reproducible.

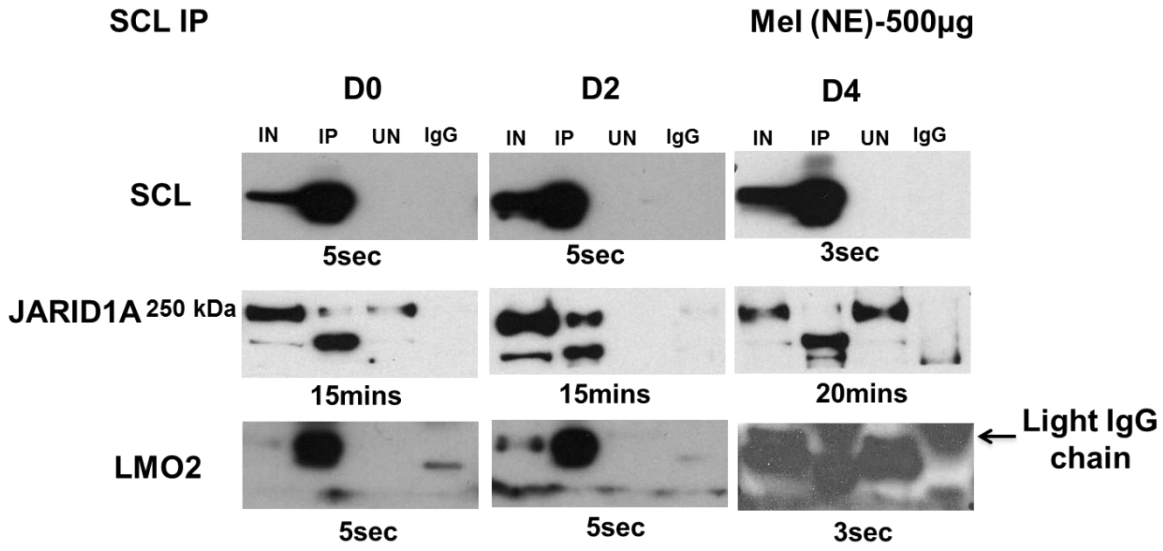


Figure 3.11: SCL co-immunoprecipitation at D0, D2 and D4 of MEL cell differentiation. SCL IP was performed on nuclear extracts prepared from MEL cells at D0, D2 and D4 of differentiation and the western blotting analysis was carried out to check its interaction with JARID1A and LMO2. Shown in the figure are input (IN), immunoprecipitation (IP), unbound (UN) and IgG control fractions. The exposure time for each blot is mentioned at the bottom.

GATA1 expression decreases from D0 to D4 (**Figure 3.12**, notice the increased exposure time at D4). This is reflected in the strength of the JARID1A signal, which also appears to be weaker. Thus, the interaction between JARID1A and GATA1 probably remains constant throughout differentiation. Unfortunately, the western blot for LMO2 is not clear enough to analyse. Upon repeating this experiment once more, these results were reproducible.

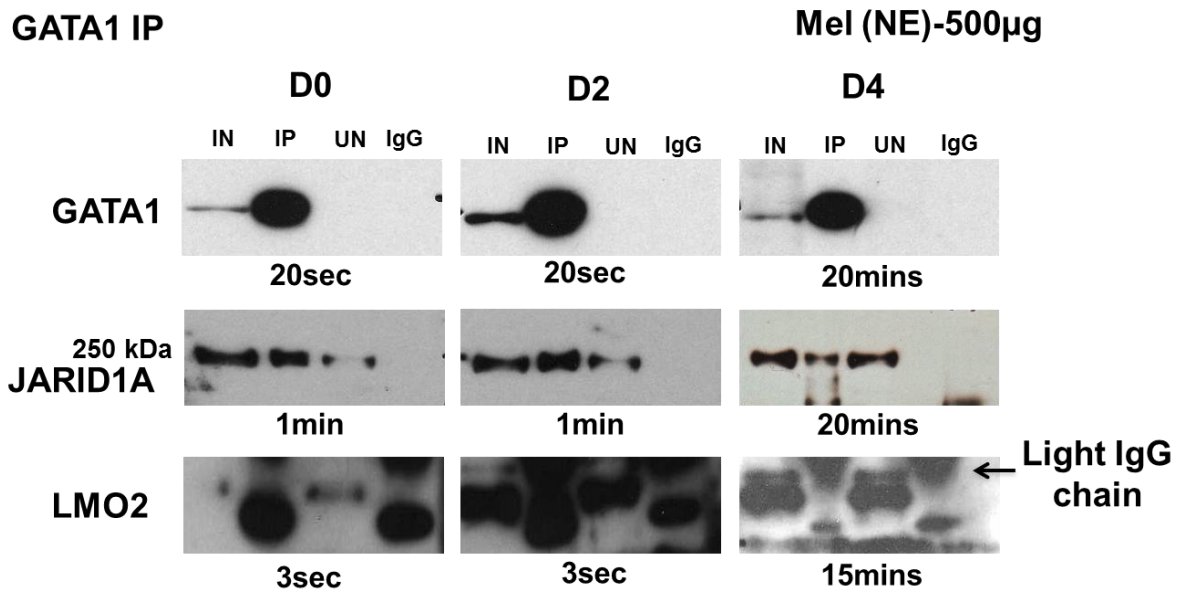


Figure 3.12: GATA1 co-immunoprecipitation at D0, D2 and D4 of MEL cell differentiation. GATA-1 IP was performed at D0, D2 and D4 of MEL cell differentiation and the western blotting analysis was carried out to check its interaction with JARID1A and LMO2. Shown in the figure are input (IN), immunoprecipitation (IP), unbound (UN) and IgG control fractions. The exposure time for each blot is mentioned at the bottom.

In summary, we can detect interactions between JARID1A, SCL and GATA1 throughout differentiation of MEL cells. However, the pattern of interactions of JARID1A with SCL and GATA1 differ with JARID1A and SCL interaction being stronger at D2 compared to D0 and D4, and JARID1A and GATA1 interaction remaining the same.

3.5 Comparison of mRNA levels of *Jarid1a, Scl* and *Gata1* at different stages of haematopoiesis

Since fetal liver and bone marrow are the primary sites for haematopoiesis in fetus and adults respectively, the mRNA levels of *Jarid1a, Scl* and *Gata1* at different stages of haematopoiesis were compared in cells obtained from these locations. Real Time PCR (RT-PCR) was performed on the cDNA samples obtained from different haematopoietic

compartments of mouse bone marrow and fetal liver. It was observed that, although levels of expression of these genes vary in these compartments, they follow a similar pattern of expression in bone marrow and fetal liver (**Figure 3.13** and **Figure 3.14** respectively). All three genes are expressed at high levels in the early stages of haematopoietic differentiation, ie the haematopoietic stem cell (HSC) compartments [Long term HSCs (LT HSC), short term HSC (ST HSC)], at lower levels in lymphoid and myeloid progenitor compartments [lymphoid primed multipotent progenitor (LMPP), Pre granulocyte-macrophage (Pre GM), granulocyte-macrophage progenitor (GMP)] and at higher levels in the megakaryocyte/erythrocyte precursor and more mature compartments [Pre megakaryocyte-erythrocyte (Pre MegE), Pre colony forming unit-erythroid cells (Pre CFUe), colony forming unit-erythroid cells (CFUe), megakaryocyte progenitor (MkP). Pre CFUe and CFUe are erythroid precursors]. Figures presented here are representative experiments of 3 biological replicates.

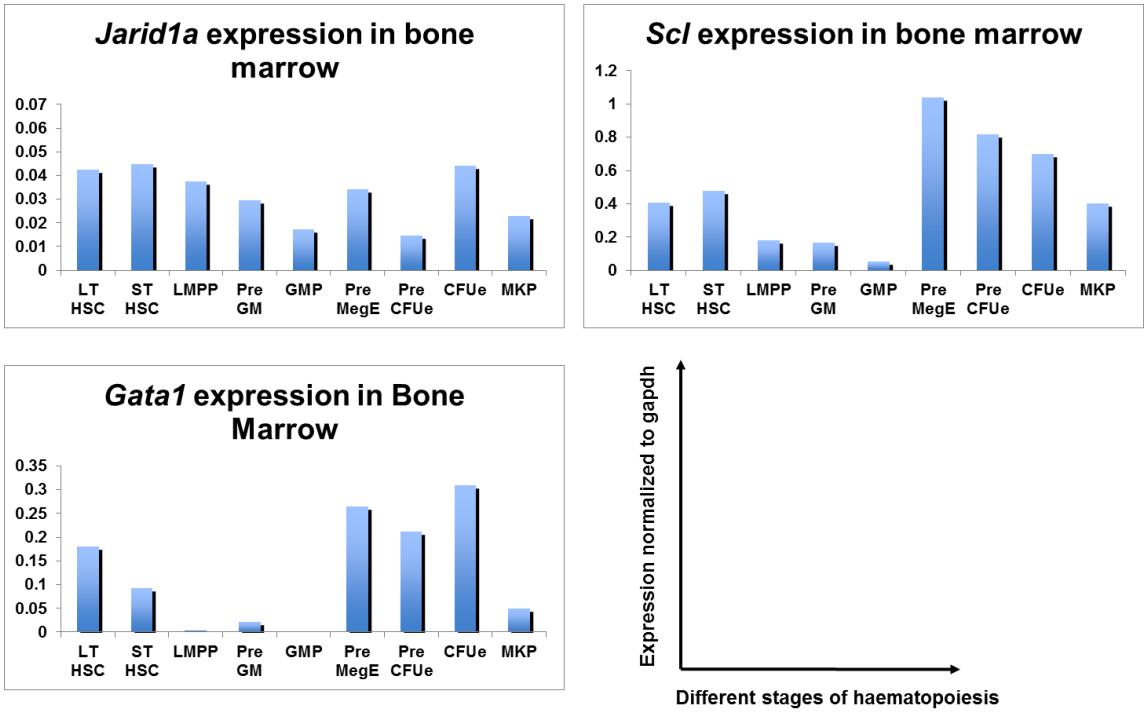


Figure 3.13: Determination of mRNA levels of *Jarid1a*, *Scl* and *Gata1* in mouse bone marrow by RT-PCR. Expression levels of *Jarid1a*, *Scl* and *Gata1* mRNAs were checked by performing RT-PCR on cDNA samples obtained from fractionated mouse bone marrow. Long term HSCs (LT HSC), short term HSC (ST HSC), lymphoid primed multipotent progenitor (LMPP), Pre granulocyte-macrophage (Pre GM), granulocyte-macrophage progenitor (GMP), Pre megakaryocyte-erythrocyte (Pre MegE), Pre colony forming unit (Pre CFUe), colony forming unit (CFUe), megakaryocyte progenitor (MkP).

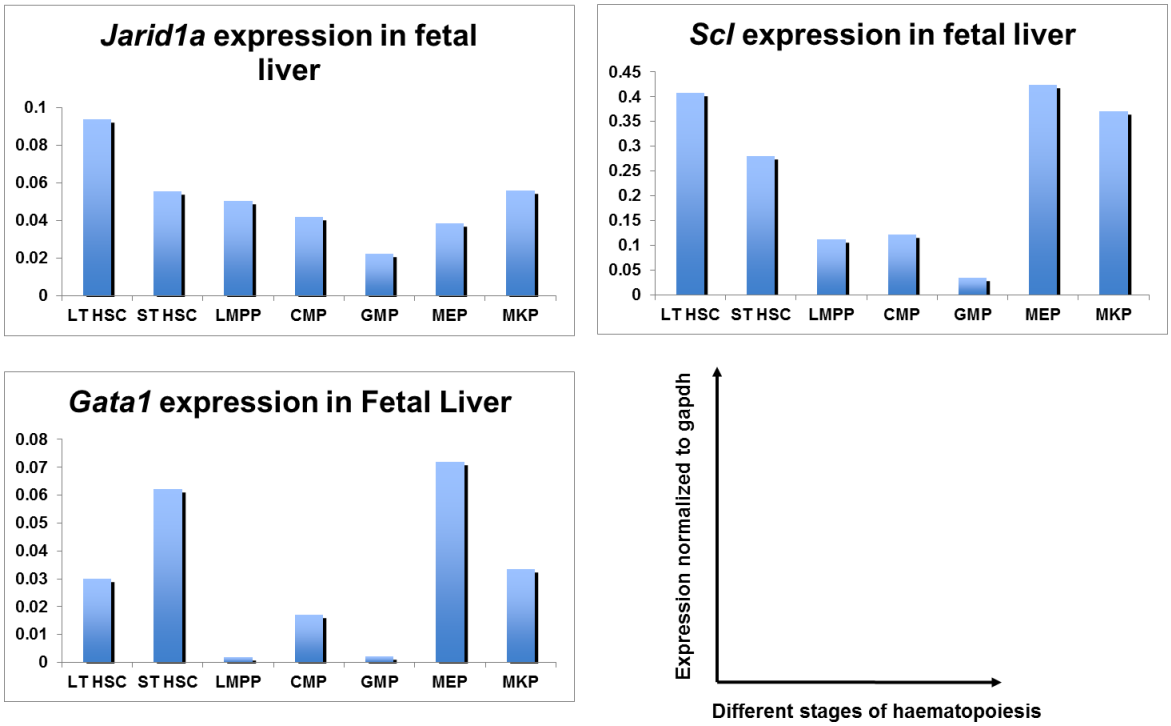


Figure 3.14: Determination of mRNA levels of *Jarid1a*, *Scl* and *Gata1* in mouse fetal liver by RT-PCR. Expression levels of *Jarid1a*, *Scl* and *Gata1* mRNAs were checked by performing RT-PCR on the cDNA samples obtained from fractionated mouse Fetal liver. Long term HSCs (LT HSC), short term HSC (ST HSC), Lymphoid primed multipotent progenitor (LMPP), common myeloid progenitor (CMP), granulocyte-macrophage progenitor (GMP), megakaryocyte-erythrocyte progenitor (MEP), megakaryocyte progenitor (MkP).

Scl is a critical transcription factor expressed in the early stages of haematopoiesis (Lecuyer and Hoang 2004). It is interesting to observe that, like *Scl*, *Jarid1a* is also expressed at relatively high levels in the stem cell compartment and in the more mature compartments. Hence, it is possible that the two proteins might interact with each other *in vivo*. Although it is to be noted that not all proteins possessing the same pattern of expression necessarily interact. The expression pattern of *Gata1* was also similar to that of *Jarid1a* and *Scl*.

Taken together with the similar phenotypes of *Jarid1a* -/- and *Scl* +/- HSCs (as discussed in the Introduction Chapter), these results suggest a possible functional interaction of these two proteins in stem cells.

3.6 JARID1A knock down by shRNA

To examine the function of JARID1A specifically in erythroid cells, we used a shRNA knockdown approach. We started with a shRNA approach instead of siRNA as the former can be persistently expressed whereas the latter gets degraded. Two *Jarid1a* shRNAs were designed using the published siRNA sequences No.2 and No.3 (Benevolenskaya et al. 2005). A scrambled sequence (No.2 SC) was also used as a negative control. To ascertain if JARID1A had been successfully knocked down, shRNAs were transfected in Human embryonic kidney cells (HEK). The final aim was to knock down JARID1A in MEL cells but initially HEK cells were used to check the efficiency of knockdown as these cells are easier to transfect. The protein and mRNA levels of JARID1A were analysed by western blotting and RT PCR respectively, 48 hours post transfection (**Figure 3.15 a and b**). No significant decrease in JARID1A levels could be detected. To check the transfection efficiency, a GFP plasmid was transfected in parallel showing nearly 100% transfection efficiency.

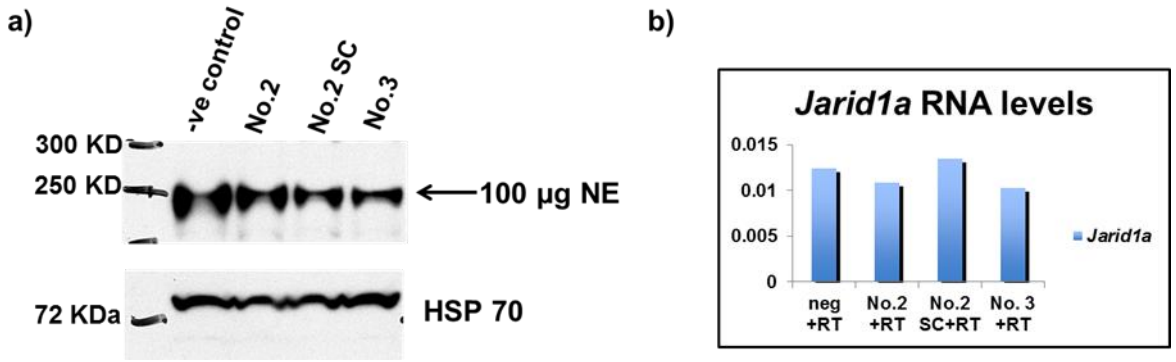


Figure 3.15: Knock down of JARID1A by shRNA. (a) Western blotting analysis was carried out using 100µg HEK nuclear extracts obtained from HEK cells transfected with *Jarid1a* shRNAs, where No.2 and No.3 were test and No.2 SC is a control. The top lane shows the JARID1A blot and the bottom lane represents the HSP70 blot which was used as loading control. **(b)** Relative expression levels of *Jarid1a* mRNA in HEK cells transfected with shRNAs normalized to *gapdh*.

Next, we tested siRNA sequences for knocking down *Jarid1a* instead of shRNA as used in (Benevolenskaya et al. 2005). HEK cells were transfected using Lipofectamine 2000 (Invitrogen) with sequences No.2, No.2SC, No.3 and also with No.2 and No.3 together as sometimes more than one siRNA is required to successfully knockdown a gene. The western blot analysis carried out on nuclear extracts obtained from transfected cells suggested that the combination of No.2 and No.3 sequences was able to knock down JARID1A levels to a certain extent as shown in **Figure 3.16a**. However, when the experiment was repeated, the knockdown was not reproducible as shown in **Figure 3.16b**.

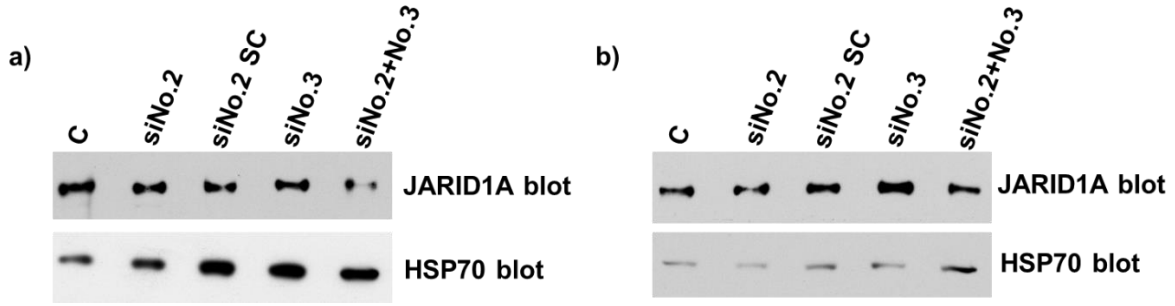


Figure 3.16: Knock down of JARID1A by siRNA. a) Western blot analysis of nuclear extracts from HEK cells transfected with Jarid1a siRNAs; siNo.2, siNo.2SC which is control, siNo.3 and siNo2+No.3. **b)** Replicate experiment with all the parameters exactly same as shown in a).

To probe for the efficiency of the transfection reagent, we also tested Lipofectamine RNAiMax (Invitrogen) and Oligofectamine (Invitrogen), which are reported to be valuable transfection reagents for siRNA. However, no JARID1A knockdown could be observed for any of the siRNAs sequences using these reagents. Next, we purchased ON-TARGETplus SMARTpool siRNA, which consists of four oligos against JARID1A from Dharmacon. Two different transfection reagents, Lipofectamine 2000 and Oligofectamine were tested in MEL cells and six different time points were taken at 6, 12, 18, 24, 48 and 72 hours post transfection. Nuclear extraction was carried out for control (C) and siRNA transfected (si) cells. Nuclear extracts were subject to western blotting using JARID1A and HSP70 antibodies. There is a reduction in the band intensities for si (6h), si (12h), si (18h) for oligofectamine transfected cells and si (6h) for lipofectamine transfected cell but their respective controls also show a decrease in the band intensity (**Figure 3.17**). Hence, the siRNA pool failed to knockdown JARID1A. Also in these blots, the full length JARID1A protein is barely visible. This could be due to poor western transfer.

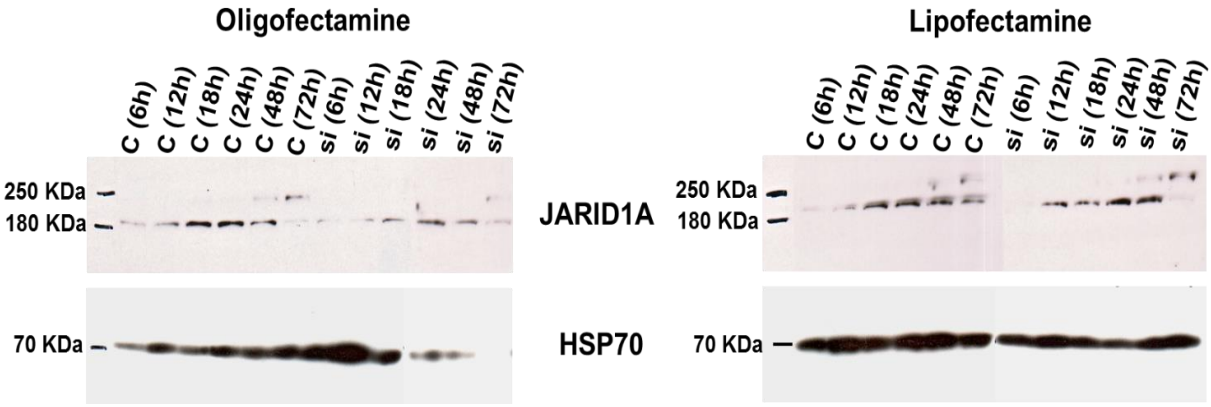


Figure 3.17: Knock down of JARID1A by ON-TARGETplus SMARTpool siRNA. Western blot analysis of nuclear extracts from MEL cells transfected with ON-TARGETplus SMARTpool siRNA from Dharmacon using oligofectamine (left) and lipofectamine (right). Six different time points were taken post transfection, 6h, 12h, 18h, 24h, 48h and 72h. 'C' and 'si' represents control and siRNA transfected samples respectively.

In conclusion, we were not able to knockdown JARID1A to study its role in erythropoiesis. Therefore, more siRNA sequences need to be tested for successful JARID1A knockdown in MEL cells which could potentially throw some light on the possible function of JARID1A in erythropoiesis if any.

3.7 Conclusion

The co-immunoprecipitation experiments we performed show that the H3K4 demethylase JARID1A interacts with the oncogenic and key haematopoietic transcription factors SCL and GATA1 in MEL cells, a transformed cell line and a model of erythroid differentiation. The co-elution of JARID1A with components of the DNA binding complex of SCL-LMO2-LDB1-GATA1 shows that it may be a part of the pentameric complex (Wadman et al. 1997). As explained earlier, elution peak fractions were not probed for E47 as working E47 antibodies were not available at that time, however, it is likely to be present in the complex as SCL forms heterodimers with E47 (Hsu et al. 1991). Furthermore, JARID1A, SCL and GATA1 also follow similar expression patterns across various stages of mouse

haematopoietic differentiation. Taken together, these results suggest a possible role of JARID1A in modulating activities of the transcription factors SCL and GATA1, and in turn, in regulating erythroid differentiation.

Chapter 4: Analysis of the direct interactions between JARID1A and the SCL complex

4 Analysis of the direct interactions between JARID1A and the SCL complex

4.1 Introduction

It is important to determine which domain(s) of JARID1A mediate direct interactions with components of the SCL complex to better understand how this multiprotein complex may form and as a prelude to performing crystallography studies. To this purpose, we employed different techniques such as mammalian two-hybrid assay, GST pull-down assay and Analytical Ultra-centrifugation (AUC).

4.2 Mammalian two-hybrid assay of JARID1A with SCL complex components

To start analysing the nature and extent of the direct interactions between JARID1A and components of the SCL complex, a mammalian two-hybrid assay was employed. As an initial crude approach, JARID1A was divided into two halves and cloned into pGal4 and pVP16 vectors as described below.

4.2.1 Construct Design

Two constructs were designed to include the N-terminal (named JmjN-JmjC) and the C-terminal (named Zf-PHD3) halves of JARID1A. Secondary structure prediction as well as disorder prediction software (RONN) were used to define the boundaries for these constructs (**Figure 4.1a**). A third construct containing only the PHD3 domain was also designed as it had been previously reported that LMO2 interacts with the PHD3 domain of JARID1A (Mao et al. 1997) (**Figure 4.1b**).

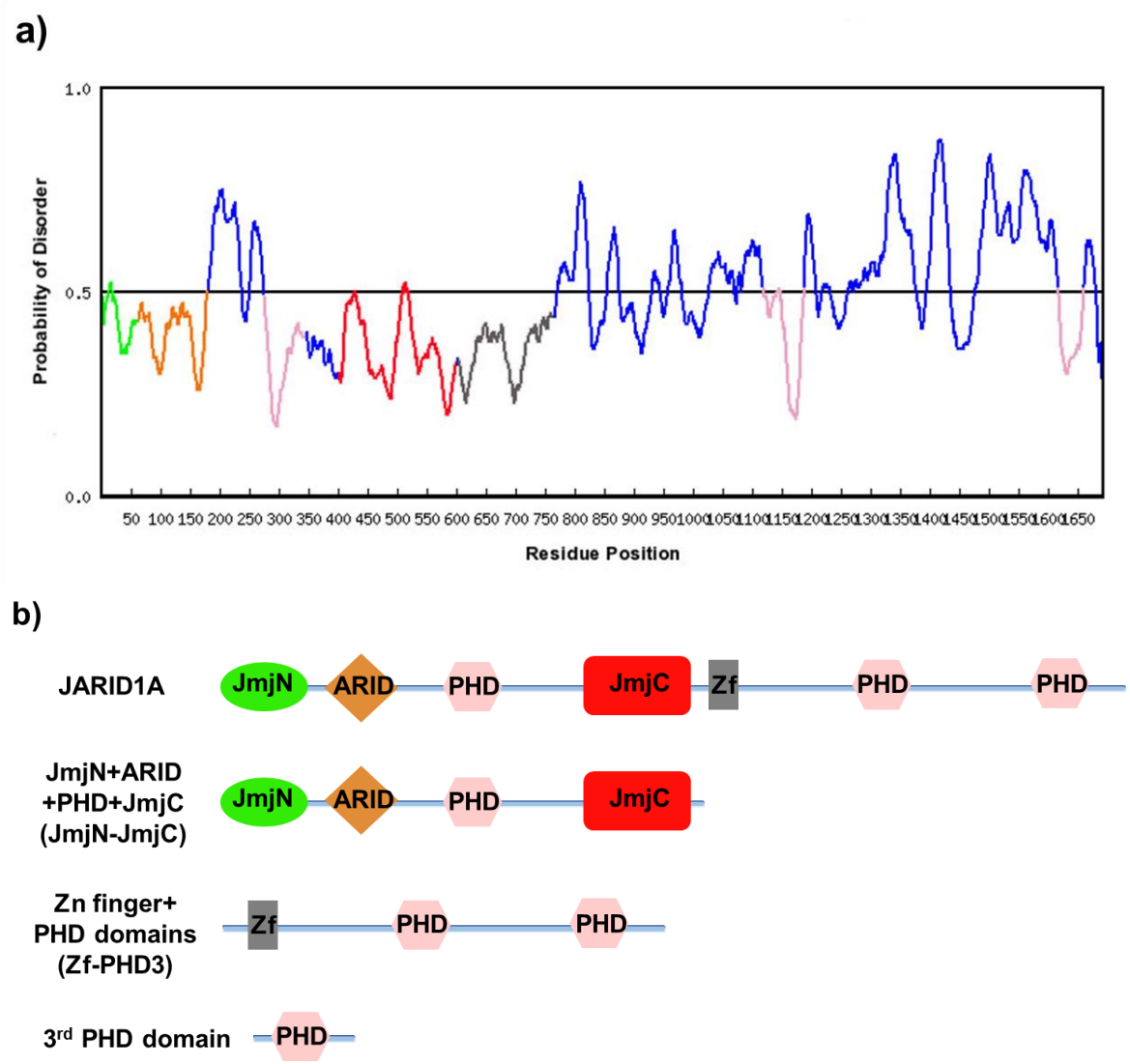


Figure 4.1: Construct Design for JARID1A. a) Probability of disorder plot and secondary structure prediction of JARID1A was calculated using RONN software, where everything above 0.5 is predicted to be disordered. **b)** Based on disorder and secondary structure prediction, different constructs were designed to investigate the interactions of JARID1A with the SCL complex components.

The full length (FL) and truncated versions of JARID1A were cloned in vectors pVP16 (transcription activation domain, V) and pM (Gal4 DNA binding domain, G) for the Mammalian two-hybrid assay. The assay was performed by using a dual luciferase reporter assay system, which involves simultaneous expression and measurement of two individual reporter enzymes: firefly luciferase, acting as a control reporter whose activity serves as a baseline response and Renilla luciferase, acting as an experimental reporter. Normalizing the

activity of the experimental reporter to the activity of the internal control minimizes the experimental variability caused by factors such as cell viability and transfection efficiencies. To perform the assay, one of the two potential protein partners was fused to the DNA binding domain while the other was fused to the transcription activation domain. Both vectors along with pRenilla and pLuciferase were transfected into HEK cells using Polyethyleneimine (PEI). If the two proteins interact, they will bring the Gal4 DNA binding domain and the VP16 activation domain in close proximity thereby forming a transcription activation complex that binds and activates the promoter of the reporter gene (**Figure 4.2a**). The luminescent signals from both luciferases (first firefly and then Renilla) were then recorded. The assay was carried out using different combinations of JARID1A constructs with SCL, E47, LMO2 and GATA1. Plasmids G SCL (Gal4 SCL), V-SCL (VP16 SCL), G-LMO2 (Gal4 LMO2), V-LMO2 (VP16 LMO2), and G-E47 (Gal4 E47) were gifts from Dr. Kamel El Omari. Individual plasmids were used as negative control and for the determination of the background activity. The western blot analysis showed that all constructs were expressing (**Figure 4.2b**).

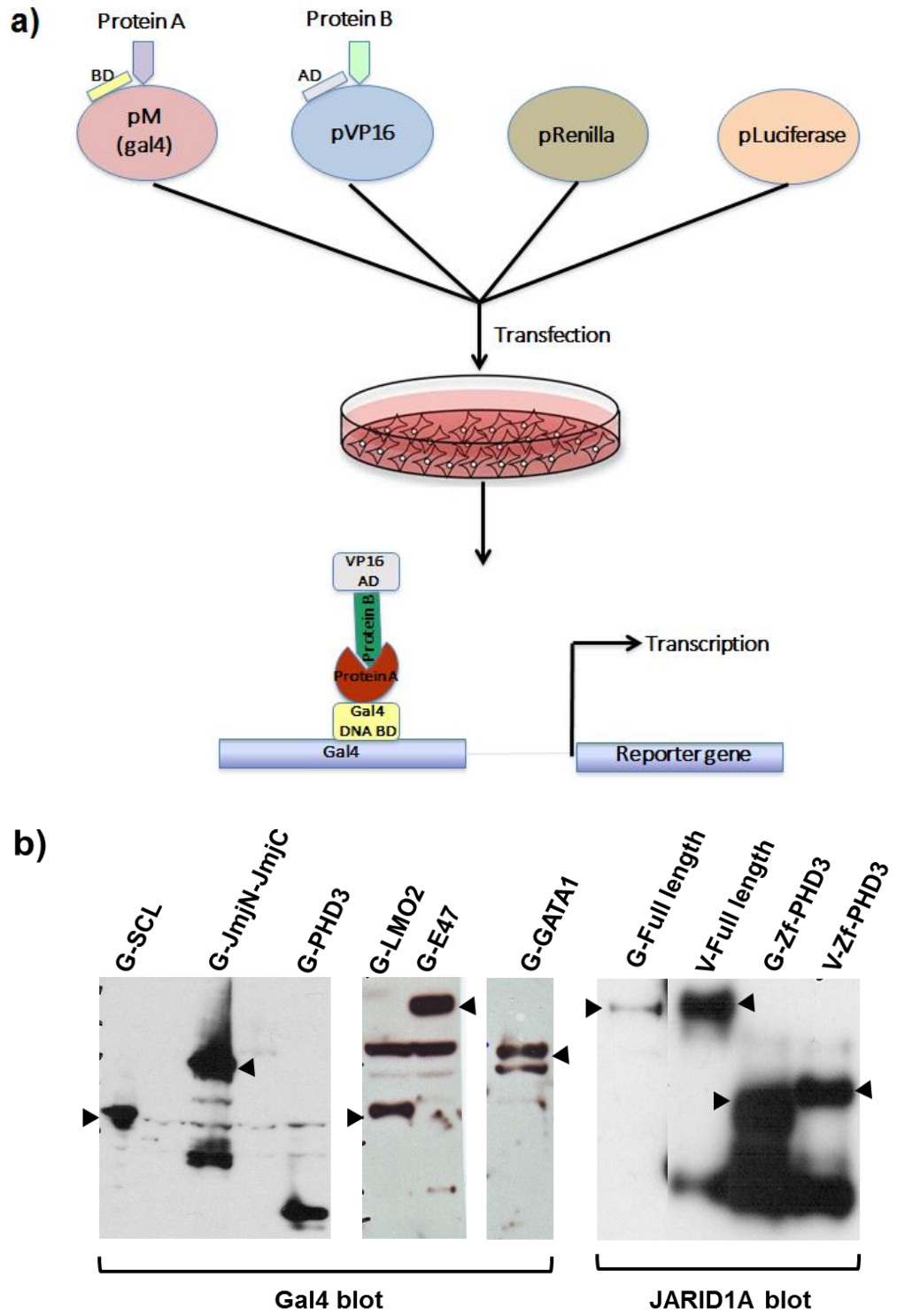


Figure 4.2: Mammalian two hybrid assay. a) Schematic representation of the Mammalian two-hybrid assay. **b)** Western blotting analysis of nuclear extracts obtained from cells transfected with JARID1A constructs. JARID1A full length and C-terminal (Zf-PHD3) constructs were probed with JARID1A antibodies whereas N-terminal (JmjN-JmjC) and PHD3 constructs were probed with GAL4 antibodies as JARID1A antibodies recognize the region between PHD2 and PHD3 domains. SCL, E47, LMO2 and GATA1 constructs were probed with GAL4 antibodies in western blotting. Black arrows indicate proteins of correct size.

First, the mammalian two-hybrid assay was performed to analyse the potential interactions between SCL and JARID1A full length, JmjN-JmjC (N-terminal domains), Zf-PHD3 (C-terminal domains), and PHD3 domain constructs. Both combinations were used: G-SCL and V-JARID1A constructs; and V-SCL and G-JARID1A constructs. None of the JARID1A constructs showed any interaction with SCL in this assay while the positive controls (G-SCL+V-LMO2, V-SCL+G-E47 and V-SCL+G-LMO2) worked well (**Figure 4.3**)

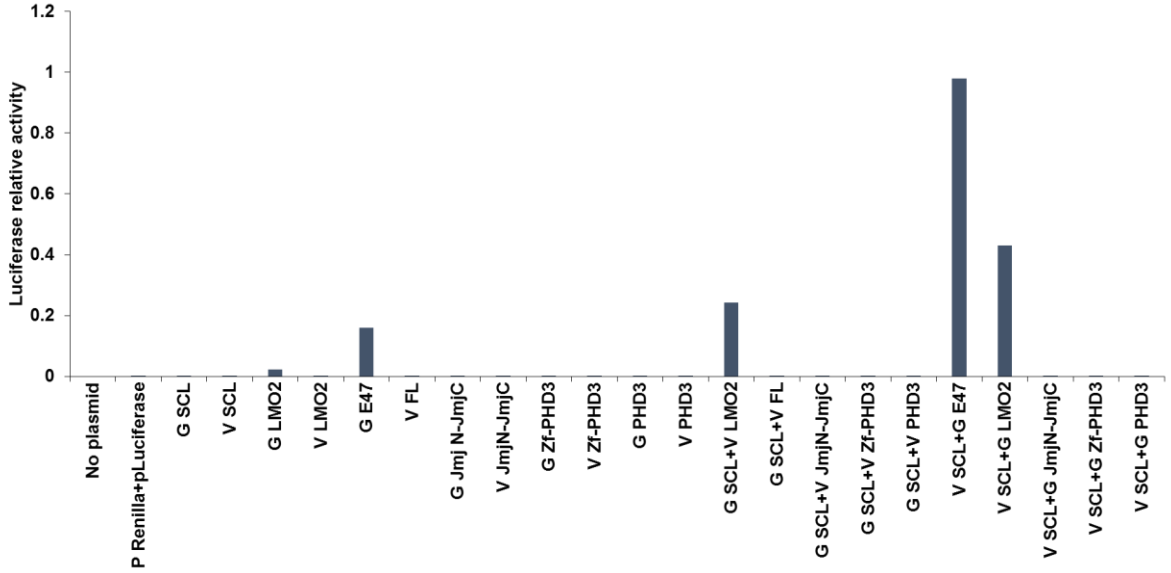


Figure 4.3: Mammalian two-hybrid assay in HEK293 cells to investigate the potential interactions between JARID1A and SCL. Mammalian two-hybrid assay was performed with G-SCL and V-JARID1A full length, V-JmjN-JmjC, V-Zf-PHD3 and V-PHD3; and also by inverting the tags, using a dual luciferase reporter assay system.

Next, interactions between JARID1A and LMO2 were investigated. Combinations of G-LMO2 and V-FL, V-PHD3 were used. The assay was also performed using reverse combinations. Along with this, the interaction between G-E47 and V-JARID1A full length was also checked to determine if JARID1A full length showed any interaction with E47. V-SCL+G-E47 and V-SCL+ G-LMO2 were used as positive controls, which showed good luciferase activity. However, none of the JARID1A constructs showed any interaction with

LMO2 or E47. The same controls were used as in previous assay for the determination of background luciferase activity. (**Figure 4.4**).

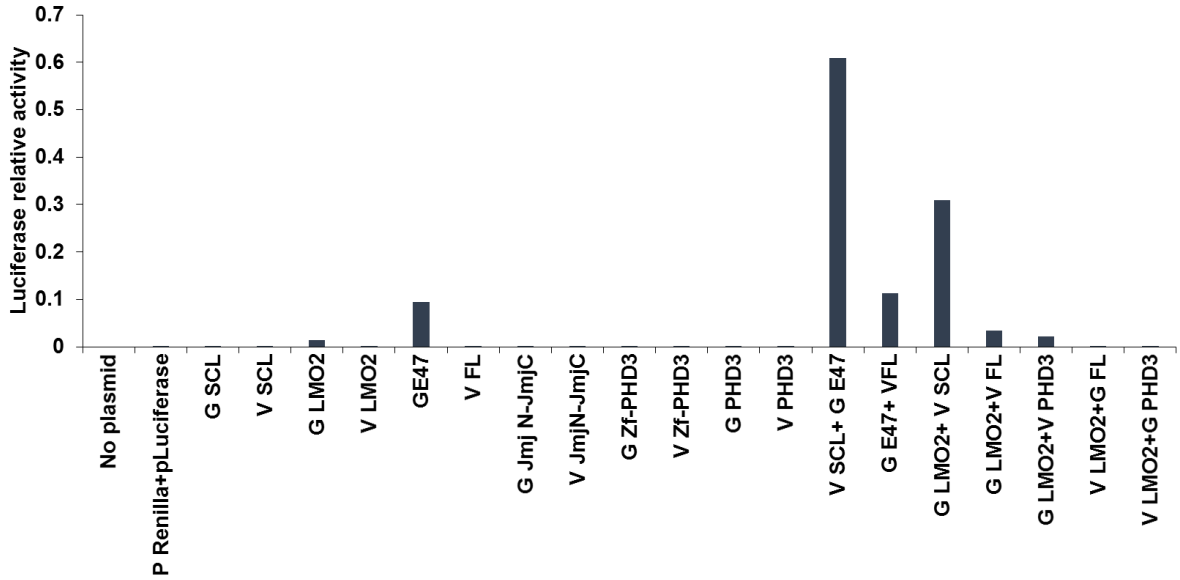


Figure 4.4: Mammalian two-hybrid assay in HEK293 cells to investigate the interaction between JARID1A and LMO2. Mammalian two-hybrid assay was performed with G-LMO2 and V-JARID1A full length, V-PHD3; and also by inverting tags, using dual luciferase reporter assay system.

The potential interactions between JARID1A and GATA1 were also examined. Combinations of G GATA1 and V JARID1A full length, V JmjN- JmjC, V Zf-PHD3, V PHD3 were used. V SCL+G E47 and V SCL+ G LMO2 were used as positive controls and individual plasmids for the determination of background luciferase activity. Once again, no interaction could be observed between the JARID1A constructs and GATA1 (**Figure 4.5**).

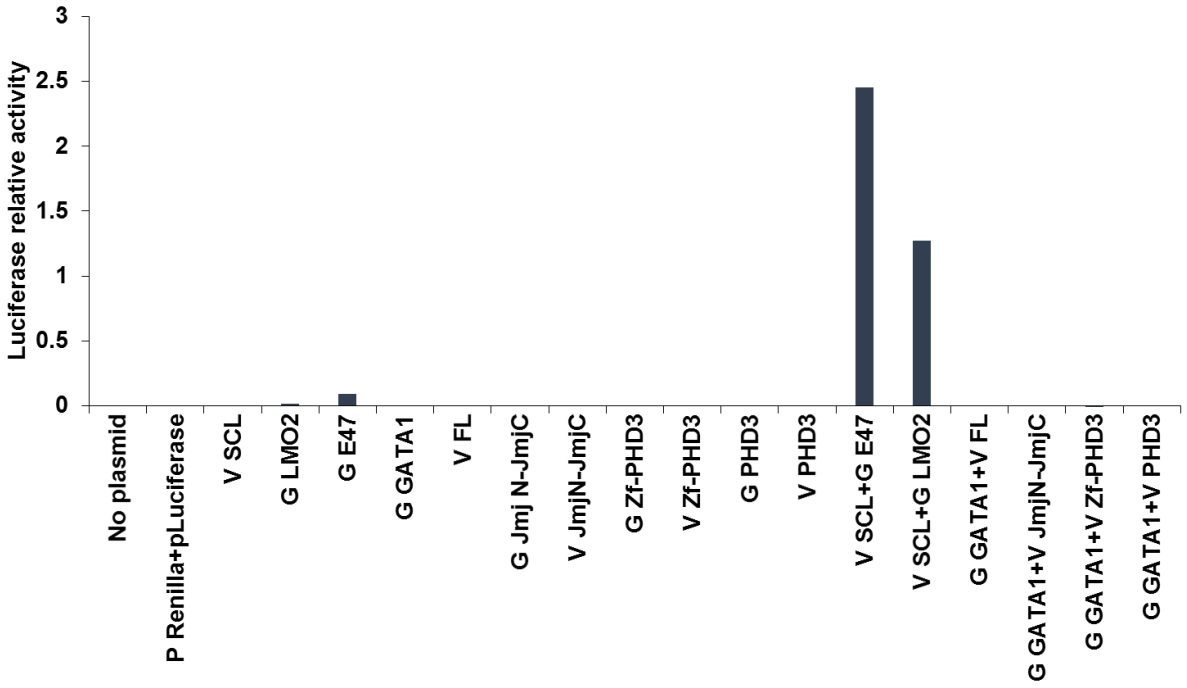


Figure 4.5: Mammalian two-hybrid assay to investigate the interactions between JARID1A and GATA1. Mammalian two-hybrid assay was performed with G-GATA1 and V-JARID1A full length, V-JmjN-JmjC, V-Zf-PHD3 and V-PHD3; and also by inverting the tags, using dual luciferase reporter assay system.

In conclusion, no interaction could be observed between JARID1A constructs and SCL, E47, LMO2 or GATA1 in any of combinations. This could be because more than one component is required for interactions to occur. As an alternative assay, GST pull-down assays were performed using GST-tagged JARID1A constructs and *in vitro* transcribed and translated SCL complex components.

4.3 GST Pull-down assay of JARID1A with *in-vitro* transcribed and translated SCL and LDB1

The N- and C-terminal halves of constructs described above were tagged with GST and GST pull-downs were performed with *in vitro* transcribed and translated (IVTT) SCL, E47, LMO2, LDB1 and ETO2. Although ETO2, a co-repressor, is not a part of the core SCL complex, it has been shown to associate with SCL in erythroid cells (Schuh et al. 2005) and,

therefore, could potentially be part of a multiprotein SCL/JARID1A complex. Plasmids for IVTT were a generous gift from Dr. Nicolas Goardon. Unfortunately, *in vitro* translation of GATA1 did not work (when checked by western blotting using antibodies against GATA1) and hence it could not be tested for the interaction with JARID1A in this assay.

GST, GST LMO2, GST JmjN-JmjC and GST Zf-PHD3 were expressed in *E. coli BL21* cells by IPTG induction and purified by affinity chromatography as described in Chapter 2, section 2.13.2 (**Figure 4.6**). The protein bound to beads was quantified by Bradford's assay prior to using it for GST pull-down assay.

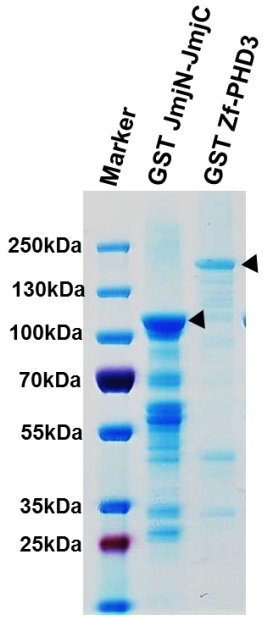


Figure 4.6: Purification of GST JmjN-JmjC and GST Zf-PHD3: SDS-PAGE gel stained with coomassie staining showing expression of GST JmjN-JmjC and GST Zf-PHD3. The correct protein bands are indicated by black arrows.

In vitro transcription and translation was carried out using TNT Quick coupled Transcription/Translation system from Promega, which employs rabbit reticulocyte lysates.

GST pull-down assay was carried out by mixing GST, GST LMO2, GST JmjN-JmjC and GST Zf-PHD3 bound to glutathione beads which were resuspended in binding buffer, with 50µl of IVTT SCL/LDB1 and subjected to western blot analysis using antibodies raised against SCL, E47, LMO2, LDB1 and ETO2 following the incubation period. Both N- and C-terminal parts of JARID1A showed interaction with SCL and LDB1 as represented by red and blue boxes respectively (**Figure 4.7** and **Figure 4.8** respectively).

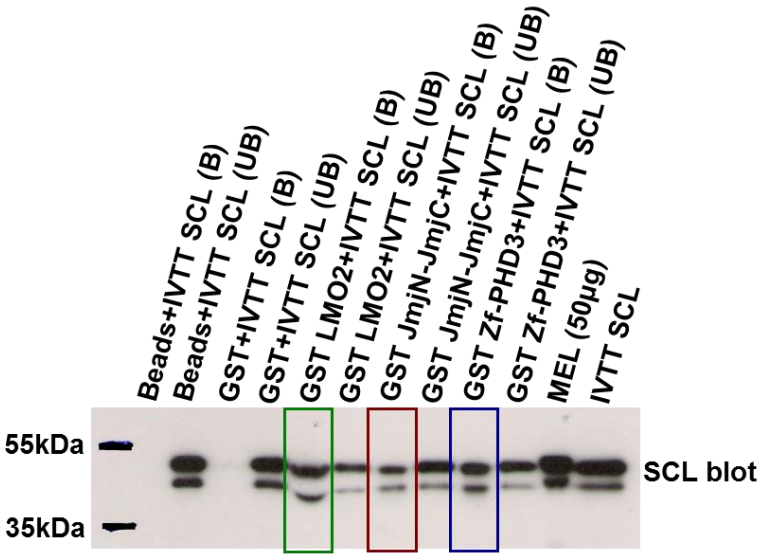


Figure 4.7: GST Pull-down assay to check the interaction between N- and C- terminal parts of JARID1A and SCL. GST pull-down assay was performed using GST JmjN-JmjC (N-terminal, highlighted with red box) and GST Zf-PHD3 (C-terminal, highlighted with blue box) with IVTT SCL. Green box represents the positive control GST LMO2/ IVTT SCL. 'B' represents beads/bound fraction and 'UB' unbound fraction.

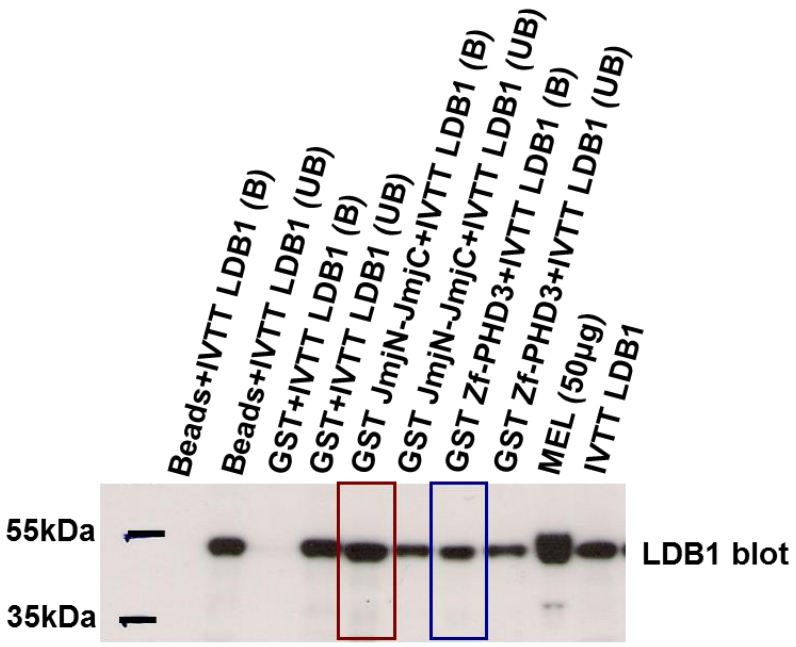


Figure 4.8: GST Pull-down assay to check the interaction between N- and C- terminal parts of JARID1A and LDB1. GST pull-down assay was performed using GST JmjN-JmjC (N-terminal, represented with red box) and GST Zf-PHD3 (C-terminal, represented with blue box) with IVTT LDB1. 'B' represents beads/bound fraction and 'UB' unbound fraction.

GST LMO2/ IVTT SCL was used as positive control (represented by a green box) and glutathione beads, GST protein were used as negative controls showing that the interaction was not due to non-specific binding with beads or GST since there was no band in beads fraction. GST pull-down with E47 and ETO2 did not show any interaction (**Figure 4.9**). In the E47 blot, the correct size E47 band is indicated by a black arrow. The LMO2 blot did not show any bands corresponding to the size of LMO2 protein and hence it could not be analysed. In case of pull-down with E47, GST LMO2 and IVTT E47 was also included as a negative control. Beads/ bound (B) and unbound (UB) fractions were loaded onto the gel for each pull down. MEL cell nuclear extracts were also loaded on the gel along with beads/bound and unbound fractions of each pull-down to compare bands position with the endogenous SCL, LDB1, and ETO2. A sample of IVTT products was loaded to ascertain that IVTT reaction had worked.

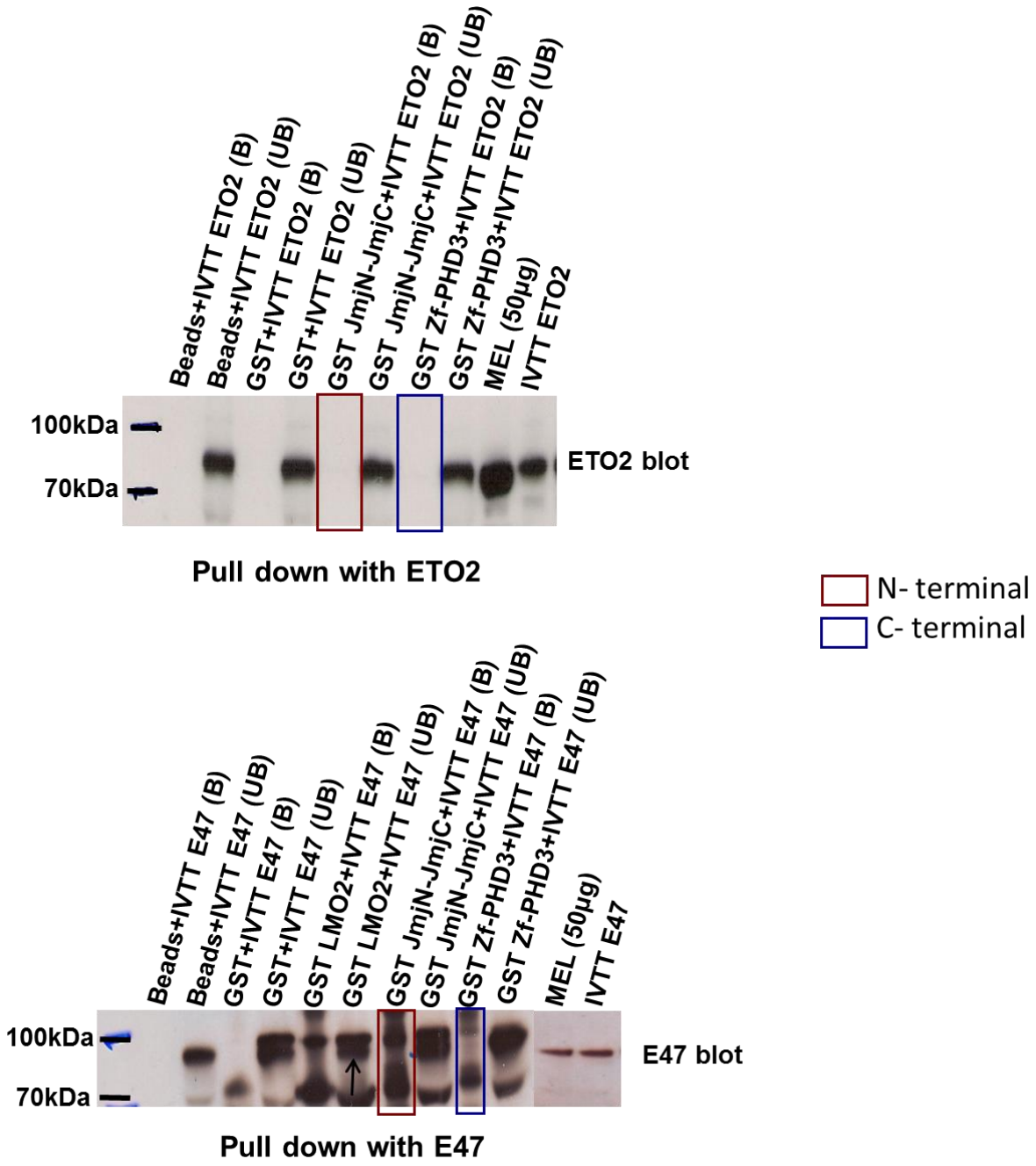


Figure 4.9: GST Pull-down assay to check the interaction between N- and C- terminal parts of JARID1A and ETO2/E47. GST pull-down assay was performed using GST JmjN-JmjC (N-terminal, represented with red box) and GST Zf-PHD3 (C-terminal, represented with blue box) with IVTT ETO2 and IVTT E47. The black arrow indicates the correct E47 band. 'B' represents beads/bound fraction and 'UB' unbound fraction. The top panel is the ETO2 blot and the bottom panel is the E47 blot.

As SCL and LDB1 showed interaction with both the N- and C- terminal domains of JARID1A (See the pulled-down lanes highlighted by red and blue boxes in **Figure 4.7** and **Figure 4.8**), a battery of constructs comprising individual domains of JARID1A were designed with GST tag. Three double domain constructs were also designed for domains that did not have any unstructured regions between them. The schematic of JARID1A and the list of constructs are shown in **Figure 4.11**. These constructs were also designed with His tag on the N- and C-terminals which were used for the crystallization trials as described in Chapter 5.

4.3.1 Cloning of JARID1A constructs

Construct design and cloning were carried out as described in Chapter 2, sections 2.15.1 and 2.15.2. All sequences coding for the JARID1A domains to be expressed were obtained by PCR. The PCR products were analysed by agarose gel electrophoresis using 1% agarose gel. All the PCR reactions worked successfully (**Figure 4.10**).

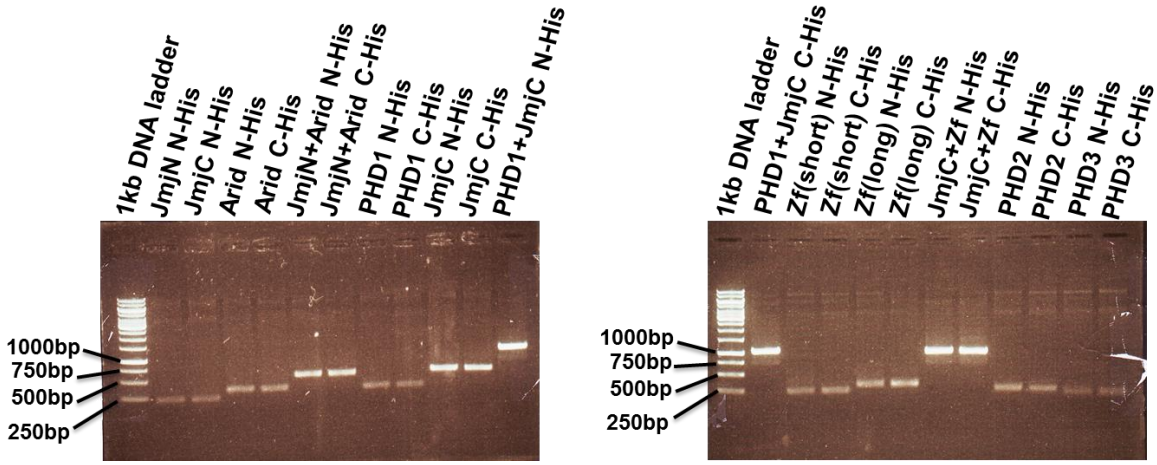


Figure 4.10: PCR of the JARID1A clones. Agarose gel electrophoresis analysis of the PCR products for cloning of the JARID1A constructs.

The amplified products were then cloned into pOPINE, pOPINEF and pOPINEJ vectors using In-Fusion enzyme, which incorporates the amplified PCR product into the linearized vector by homologous recombination. Plasmids were then sequenced (Source Biosciences sequencing services). Sequences matched the desired regions of JARID1A domains.

SDS-PAGE analysis confirmed expression of all constructs (**Figure 4.11**). Specifically Arid, JmjN+Arid, PHD1, Zf, PHD2 and PHD3 constructs expressed at high levels. Expression and purification of all the 11 constructs were carried out using the same protocol used for GST JmjN-JmjC and GST Zf- PHD3 purification as described earlier.

Sr. no.	Clones
1	JmjN
2	Arid
3	JmjN+Arid
4	PHD1
5	JmjC
6	PHD1+JmjC
7	Zf (short)
8	Zf (long)
9	JmjC+Zf
10	PHD2
11	PHD3

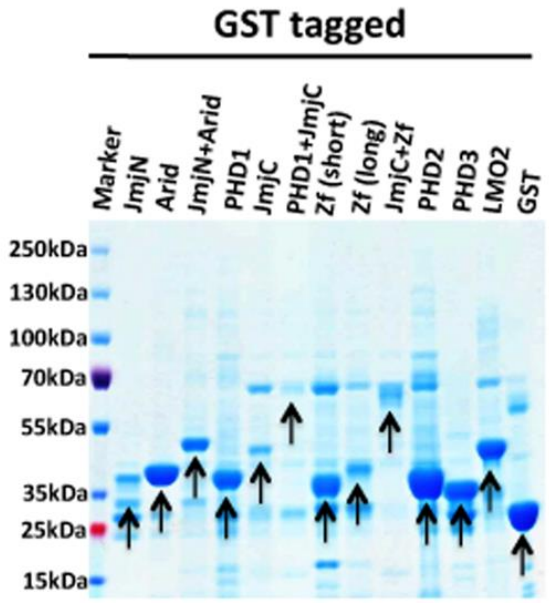


Figure 4.11: JARID1A constructs designed for GST Pull-down assay. Top panel: schematic representation of JARID1A full length protein and its domains. Bottom panel: list of GST-tagged JARID1A constructs (left) and SDS-PAGE gel showing their expression (right).

The GST pull-down assay was performed using each of the JARID1A constructs with IVTT SCL and LDB1.

The western blot analysis of the pull-down between JmjN, ARID, JmjN+ARID constructs and SCL (top) and LDB1 (bottom) is shown in **Figure 4.12**. The bound fractions for JmjN, ARID and JmjN+ARID are represented by red, blue and orange boxes respectively and the green box indicates the positive control (GST LMO2/IVTT SCL). None of the domains showed interaction with SCL or LDB1 individually, however, the double domain construct JmjN+ARID showed interaction with LDB1.

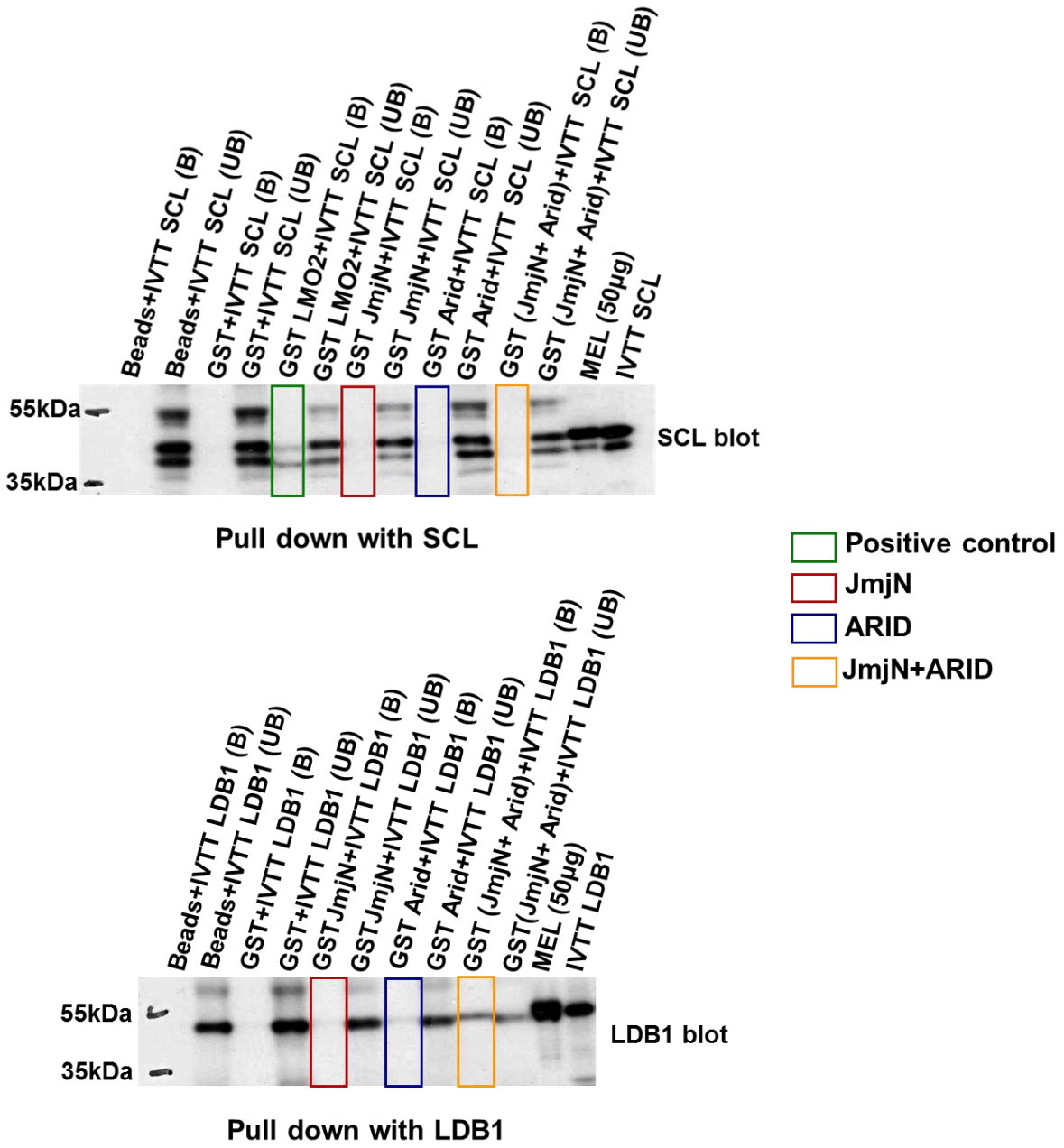


Figure 4.12: GST pull-down assay to check the interaction between JARID1A constructs and SCL/LDB1. GST pull down assay of GST JmjN (indicated in red), GST ARID (indicated in blue) and GST JmjN+ARID (indicated in orange) with IVTT SCL and LDB1. Green box represents the positive control GST LMO2/ IVTT SCL. 'B' represents beads/bound fraction and 'UB' unbound fraction. The top panel is the SCL blot and the bottom panel is the LDB1 blot.

The western blot analysis of the pull-down between PHD1, JmjC, PHD1+JmjC constructs and SCL (top) and LDB1 (bottom) is represented in **Figure 4.13** in which the bound fractions for PHD1, JmjC and PHD1+JmjC are represented by red, blue and orange boxes respectively in each blot. All three constructs showed interaction with SCL and LDB1. But the expression profile (**Figure 4.11**) shows that only the PHD1 domain is expressed at high levels in this set and hence we can only be confident about the interaction between PHD1 and SCL/LDB1.

The western blot analysis of the pull-down between Zf (short), Zf (long), JmjC+Zf constructs and SCL (top) and LDB1 (bottom) is shown in **Figure 4.14**. Two zinc finger constructs (Zf long and Zf short) were generated as there were two possible C-terminal boundaries for this construct ending with hydrophilic residues. The Zf (long) possess an extra 23 amino acid residues at the C-terminus. The bound fractions for Zf (short), Zf (long) and JmjC+Zf are represented by red, blue and orange boxes respectively in each blot. Zf did not interact with SCL or LDB1 but JmjC+Zf interacted with both SCL and LDB1.

The western blot analysis of the pull-down between PHD2, PHD3 domains and SCL (top) and LDB1 (bottom) is shown in **Figure 4.15**. Red and blue boxes in each blot represent the bound fractions for PHD2 and PHD3 respectively. PHD2 interacts with LDB1 but it is difficult to say if the weak bands observed in the bound fraction of PHD2 in the SCL blot and PHD3 in the SCL and LDB1 blots are resulting from non-specific interactions. This was further investigated by using higher ionic strength in binding and wash buffers.

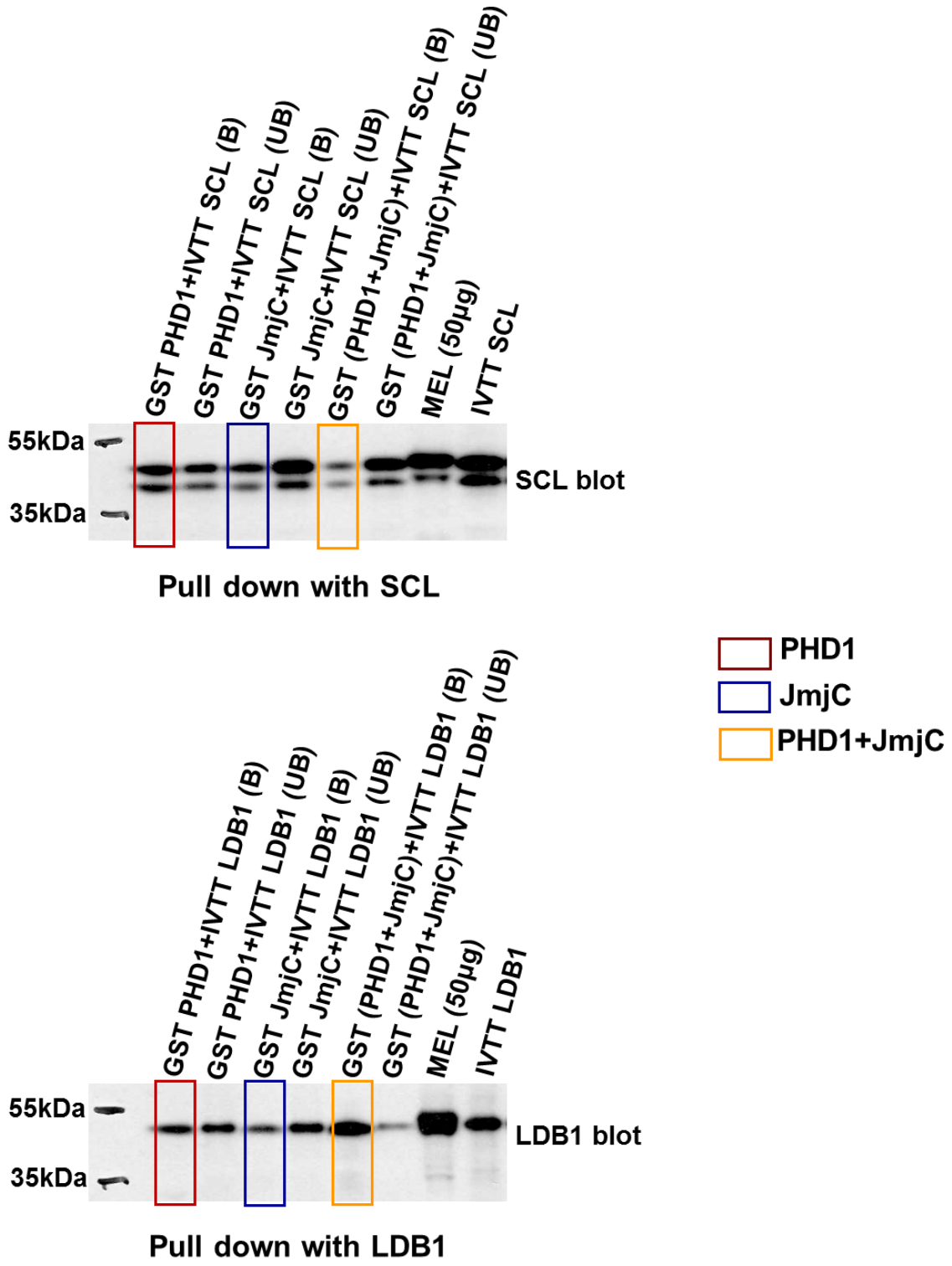


Figure 4.13: GST pull-down assay to check the interaction between JARID1A constructs and SCL/LDB1. GST pull down assay of GST PHD1 (indicated in red), GST JmjC (indicated in blue) and GST PHD1+JmjC (indicated in orange) with IVTT SCL and LDB1. 'B' represents beads/bound fraction and 'UB' unbound fraction. The top panel is the SCL blot and the bottom panel is the LDB1 blot.

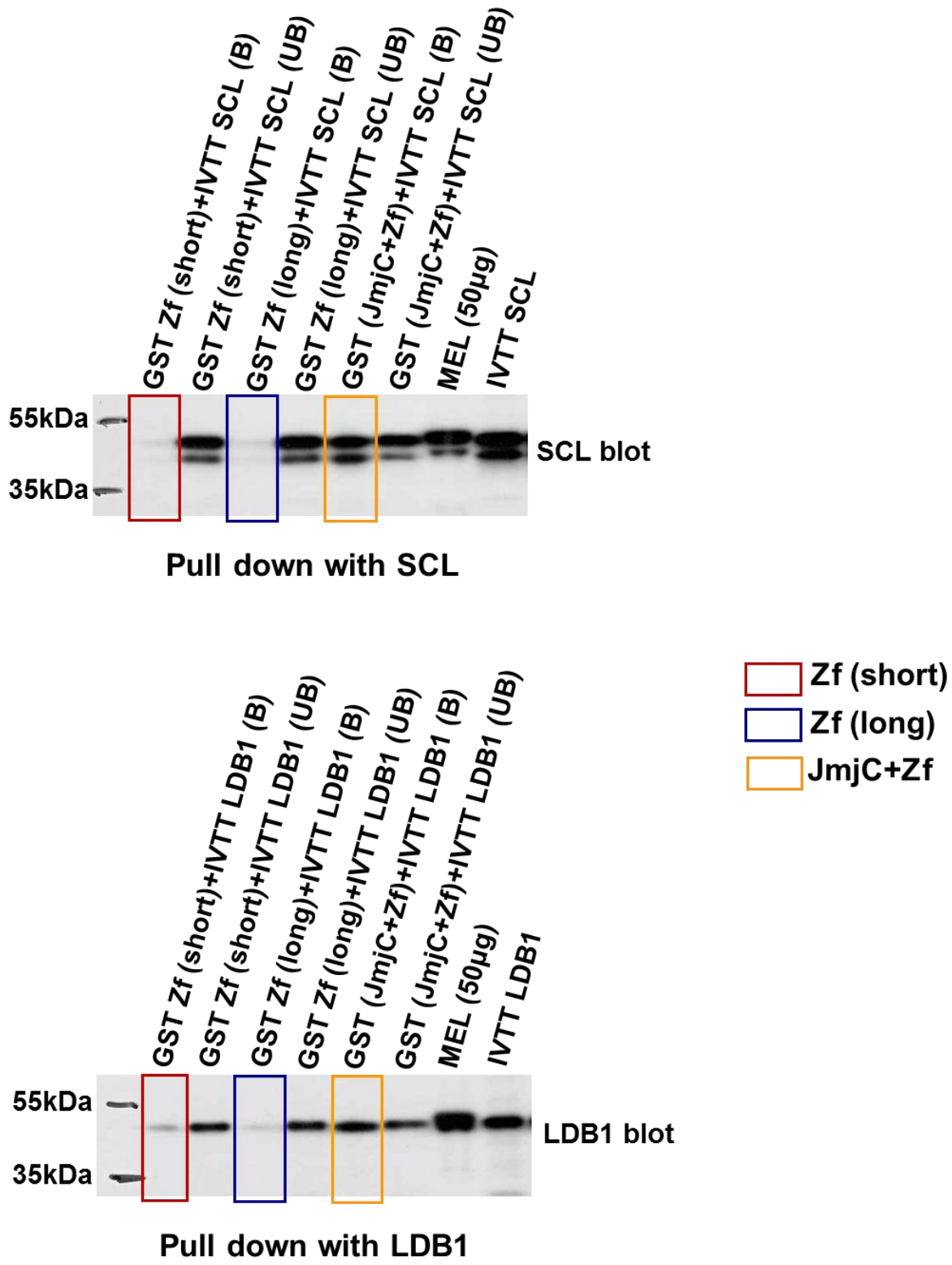


Figure 4.14: GST pull-down assay to check the interaction between JARID1A constructs and SCL/LDB1. GST pull down assay of GST Zf (short) (indicated in red), GST Zf (long) (indicated in blue) and GST JmjC+Zf (indicated in orange) with IVTT SCL and LDB1. 'B' represents beads/bound fraction and 'UB' unbound fraction. The top panel is the SCL blot and the bottom panel is the LDB1 blot.

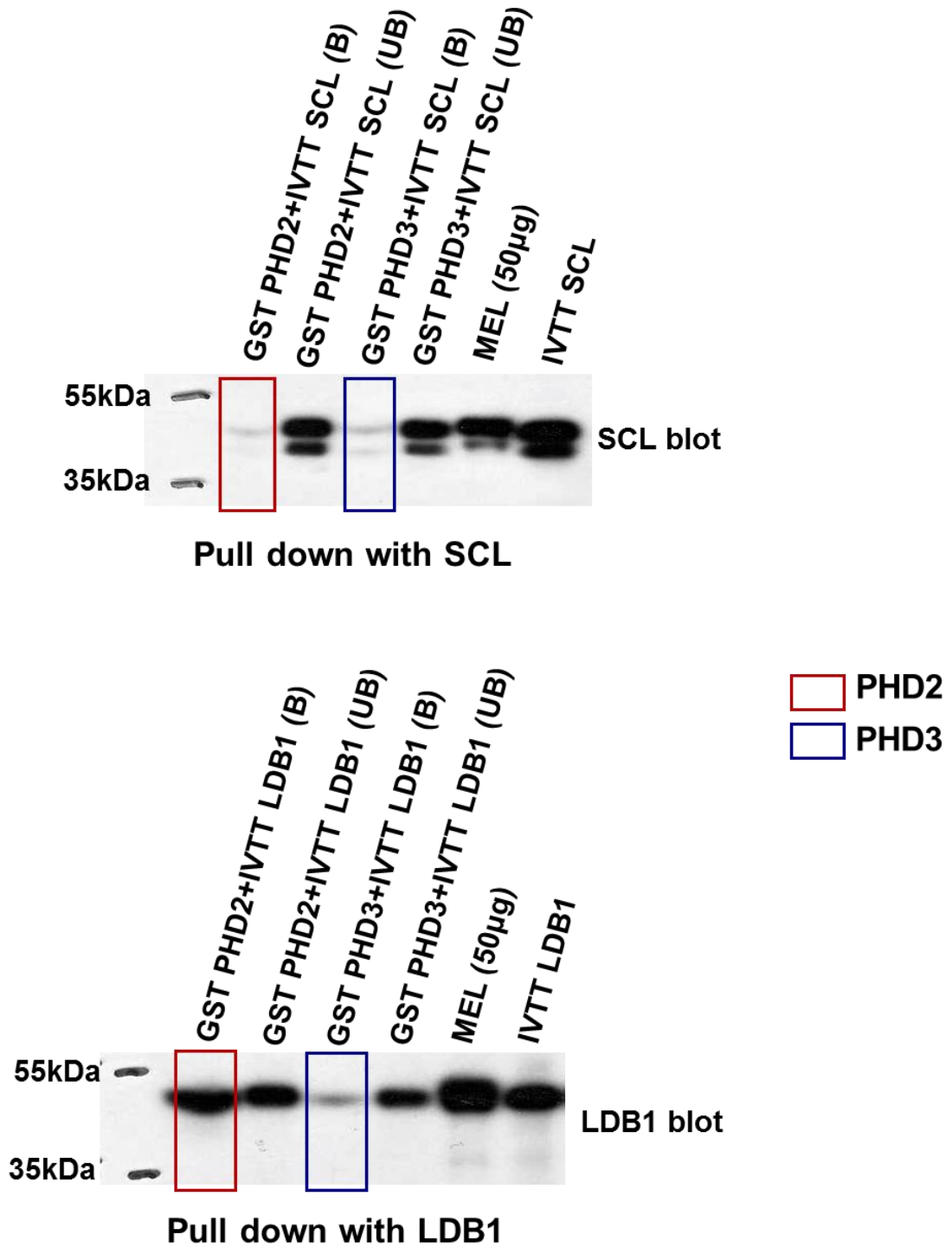


Figure 4.15: GST pull-down assay to check the interaction between JARID1A constructs and SCL/LDB1. GST pull down assay of GST PHD2 (indicated in red) and GST PHD3 (indicated in blue with IVTT SCL and LDB1. 'B' represents beads/bound fraction and 'UB' unbound fraction. The top panel is the SCL blot and the bottom panel is the LDB1 blot.

To determine if the weak bands observed in the bound fractions of Zf (short) and Zf (long) in the SCL and LDB1 blots (the red and blue boxes in **Figure 4.14**); PHD2 in SCL blot (the red box in SCL blot of **Figure 4.15**); and PHD3 in both SCL and LDB1 blots (the blue boxes in **Figure 4.15**) were resulting from non-specific interactions, pull-downs were repeated under more stringent binding conditions using higher ionic strength buffers (by including 400mM NaCl in both binding and wash buffers). The western blot analysis of the pull-down between Zf (short), Zf (long), PHD2, PHD3 domains and SCL (right) and LDB1 (left) in the new condition is shown in **Figure 4.16**. The bound fractions of Zf (short), Zf (long), PHD2 and PHD3 are represented by red, blue, orange and yellow boxes respectively in each blot and the green box indicates the positive control GST LMO2/IVTT SCL. High salt concentration in the binding and wash buffers increased the intensities of bands in the SCL blot. However, the ones in the LDB1 blot remained the same.

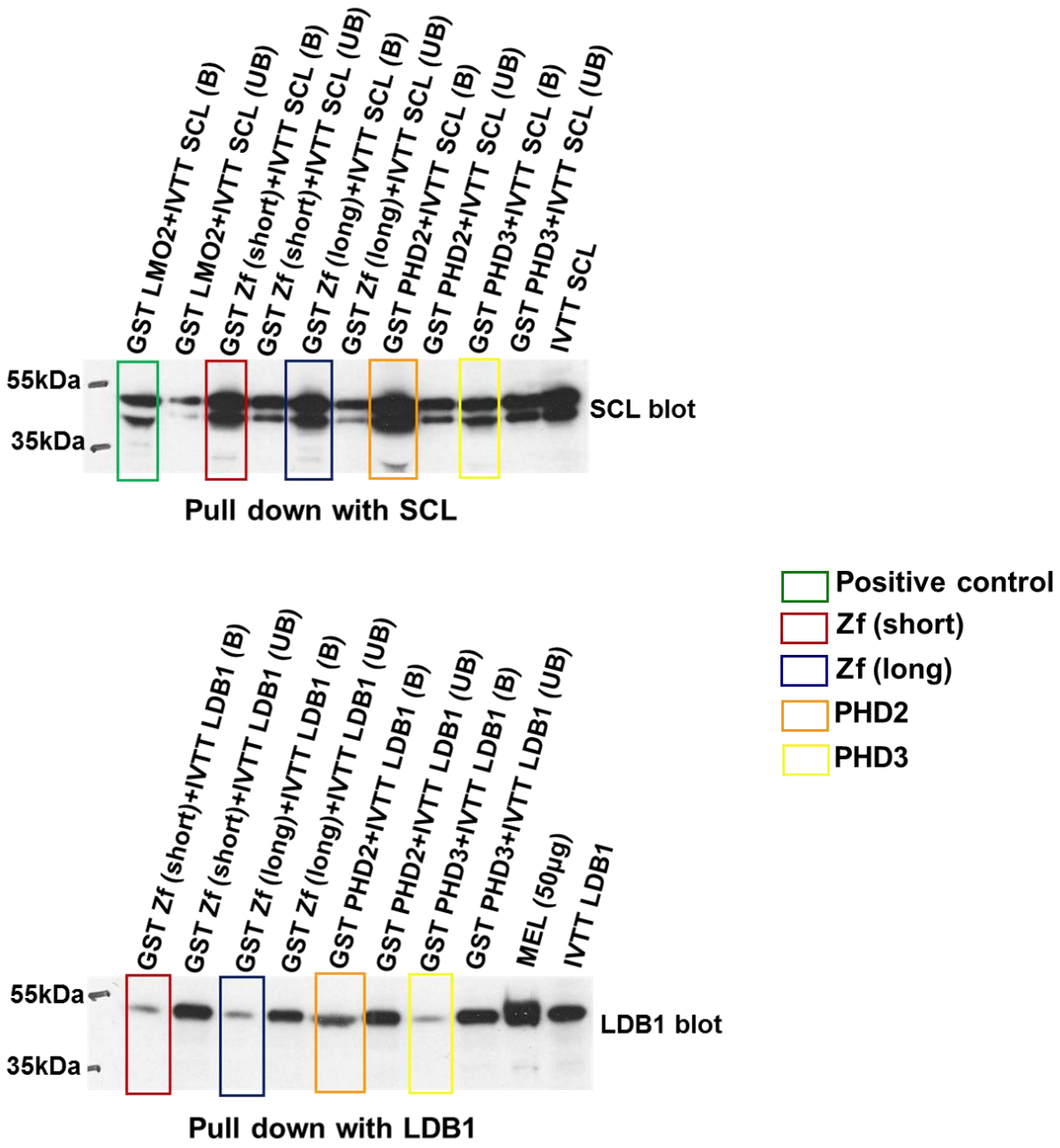


Figure 4.16: GST pull-down assay to check the interaction between some JARID1A constructs and SCL/LDB1 in high salt (400mM NaCl). GST pull down assay of GST Zf (short) (indicated in red), GST Zf (long) (indicated in blue), GST PHD2 (indicated in orange) and GST PHD3 (indicated in yellow) with IVTT SCL and LDB1. Green box represents the positive control GST LMO2/ IVTT SCL. 'B' represents beads/bound fraction and 'UB' unbound fraction. The top panel is the SCL blot and the bottom panel is the LDB1 blot.

The above results are hard to interpret as high salt resulted in an increase in the band intensities, contrary to what was expected. The strength of interaction, as monitored by the band intensity, was therefore investigated as a function of the individual components of the buffer. The composition of all buffers is shown in **Figure 4.17a**. First, all salts were removed from the buffer (buffer-1) but this increased intensities of bands (bands highlighted with red, blue, green and yellow boxes for buffer-1 in **Figure 4.17b**). Next, in order to check the effect of KCl and NaCl on the interaction, KCl (buffer-2, which was the standard buffer used for all pull-downs) was replaced with NaCl (buffer-3). Neither KCl nor NaCl showed any changes (compare the weak bands highlighted in with red, blue, green and yellow boxes for buffer-2 and buffer-3 in **Figure 4.17b** with SCL blots of **Figure 4.14** and **Figure 4.15**). Next, $MgCl_2$ was eliminated from the buffer (buffer-4) to determine its effect on pull-downs. However, $MgCl_2$ did not show any effect (the weak bands highlighted with red, blue, green and yellow boxes for buffer-4 in **Figure 4.17c**). The effect of $MgCl_2$ was also investigated on pull-downs under physiological conditions (using PBS as buffer). The interaction of SCL with all four JARID1A constructs increased under this condition in absence and presence of $MgCl_2$ (represented by the strong bands highlighted in with red, blue, green and yellow boxes for buffer-5 and buffer-6 in **Figure 4.17c**) showing that $MgCl_2$ does not alter any of the interactions.

In conclusion, there were some discrepancies in the GST pull-down assay for some of the constructs; however, it became obvious that the PHD domains, which are in general believed to be responsible for protein-protein interactions, are likely involved in interaction of JARID1A with SCL and LDB1.

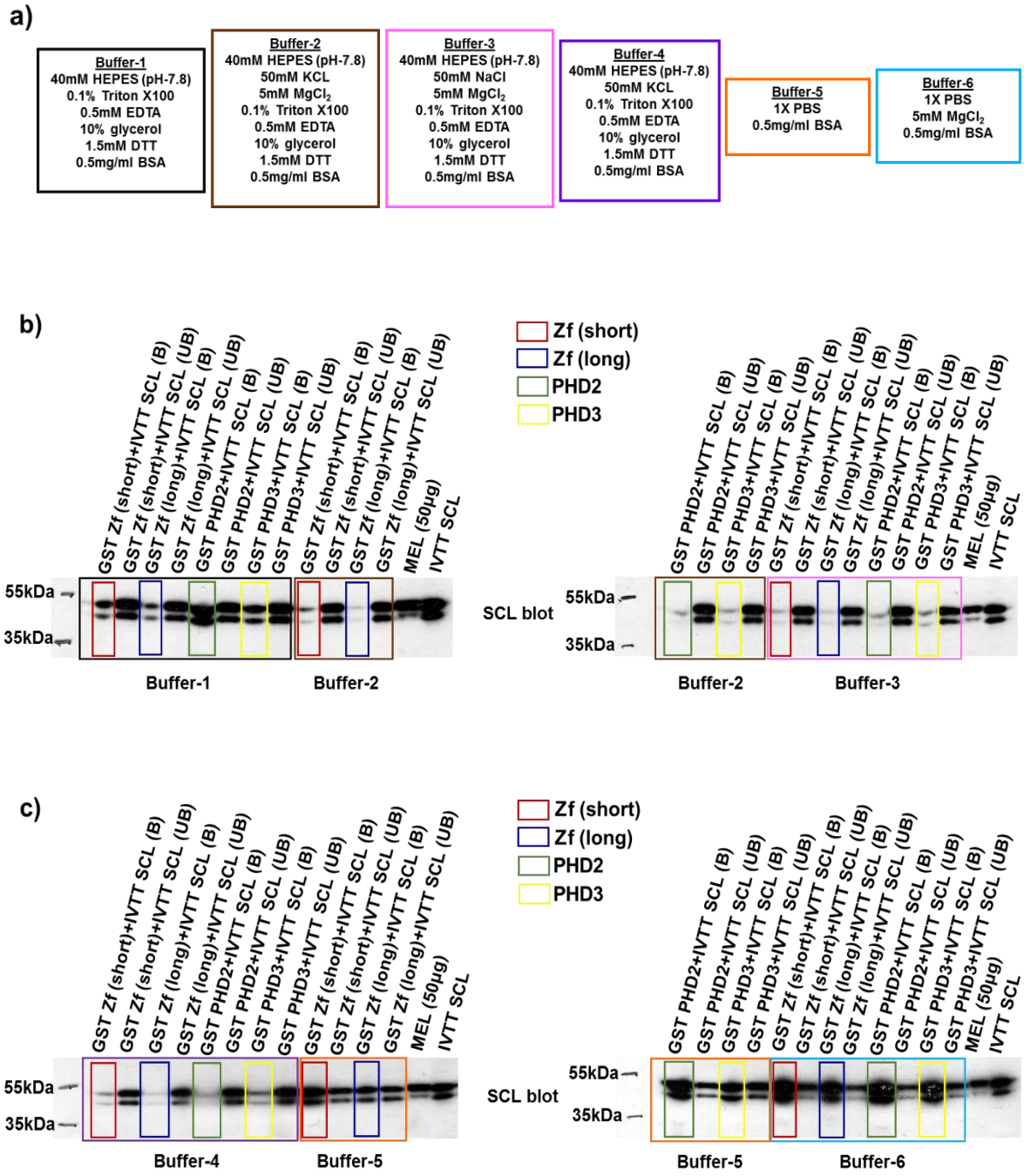


Figure 4.17: GST pull-down assay to check the interaction between some JARID1A constructs and SCL/LDB1 with different buffers. a) List of different buffers (1-6) used for the pull-downs. **b)** and **c)** GST pull down assay of GST Zf (short) (indicated in red), GST Zf (long) (indicated in blue), GST PHD2 (indicated in green) and GST PHD3 (indicated in yellow) with IVTT SCL. The outer boxes are colour coded depending upon the buffers used as shown in (a) black for buffer-1, brown for buffer-2, magenta for buffer-3, violet for buffer-4, orange for buffer-5 and cyan for buffer-6. 'B' represents beads/bound fraction and 'UB' unbound fraction

A summary of interactions obtained between JARID1A constructs and SCL/ LDB1 in GST pull-down assays is presented in **Table 4.1**.

Table 4.1: Summary of interactions obtained in GST pull-down assay.

Constructs	Binding with SCL	Binding with LDB1
JmjN	-	-
Arid	-	-
JmjN+Arid	-	+++
PHD1	+++	+++
JmjC	+++	+++
PHD1+JmjC	++	++++
Zf (short)	-	-
Zf (long)	-	-
JmjC+Zf	+++	+++
PHD2	+	++++
PHD3	+	+

++++ (strong)→high signal in bound lane compared to unbound
+++ (binding)→equal signal in bound and unbound lanes
++ (weak) →weak signal in bound lane compared to unbound
+(very weak) →very weak signal in bound lane obtained only after longer exposure and very strong signal in unbound lane
- → No interaction

To further dissect these findings another more sensitive technique called Analytical ultra-centrifugation (AUC) was used to study interactions of the three PHD domains of JARID1A with the SCL complex components.

4.4 Analytical Ultracentrifugation of JARID1A PHD domains with the complex components

Analytical ultracentrifugation (AUC) is a powerful technique for the quantitative analysis of macromolecules in solution, in which the sedimentation of molecules in a centrifugal field is monitored in real time without the need of binding them to matrices or surfaces.

E47/SCL bHLH (consisting of the bHLH domains of SCL and E47), FLINC2 (containing LIM domains of LMO2 fused to LID domain of LDB1) and GATA1 NFCF (containing N- and C-terminal zinc fingers) were used (**Figure 4.18**), as full-length proteins are either very unstable or cannot be expressed and purified. These constructs contain the structured domains of the corresponding proteins. The generation of LDB1 construct for expression of stable, soluble protein was planned, however, it could not be followed up due to shortage of time. Hence, the interaction of these constructs with JARID1A were examined. E47/SCL bHLH and LMO2 constructs were gifted by Dr. Kamel El Omari and the GATA1 NFCF construct was provided by Dr. Olga Platonova. AUC of all three PHD domains of JARID1A with E47/SCL bHLH, LMO2 and GATA1 NFCF was performed. Two types of methods can be used for AUC experiments: sedimentation velocity and sedimentation equilibrium. The sedimentation velocity method was selected as it is a hydrodynamic method used to define size and interactions of macromolecules whereas the sedimentation equilibrium method is a thermodynamic method used to define assembly stoichiometry and association constants.

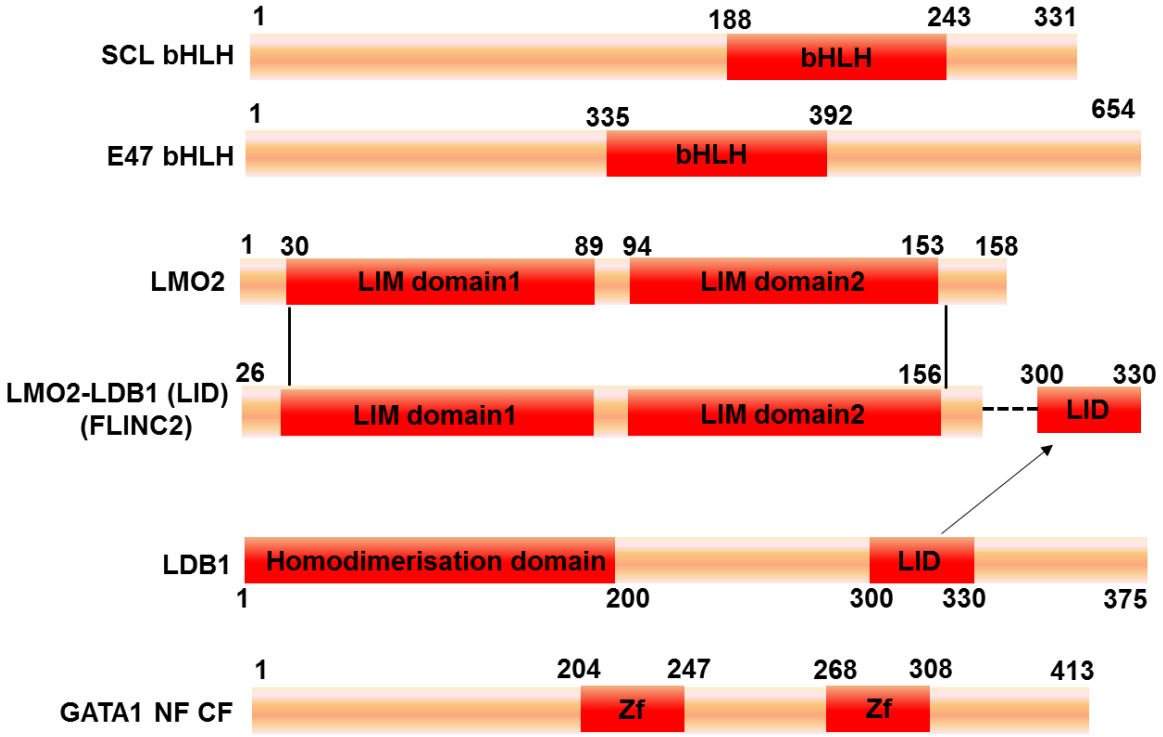


Figure 4.18: Construct Design for SCL, E47, LMO2 and GATA1. E47/SCL bHLH (containing bHLH domains of SCL and E47), FLINC2 (containing LIM domains of LMO2 fused to LID domain of LDB1) and GATA1 NFCF (containing both the N- and C- terminal zinc fingers) constructs were used for AUC experiments.

4.4.1 Protein purifications for AUC

The PHD2 protein was purified as described in chapter 5, section 5.3. LMO2, E47/SCL bHLH and GATA1 NFCF were purified as described in the sections below. PHD1 and PHD3 proteins were also used for AUC which will be described later.

LMO2 Purification

The recombinant forms of full length LMO proteins have inadequate solubility and stability and are prone to aggregation, which makes it difficult to biochemically or structurally analyse

them. Nevertheless, their stability is enhanced when they interact with LDB1, which acts as a scaffolding protein. Therefore, a fusion protein was previously designed by combining the LID domain of LDB1 and the first LIM domain of LMO2, which is called FLIN2 (Fusion of the LIM interacting domain of LDB1 and the N-terminal LIM domain of LMO2) (Deane et al. 2001). In this fusion protein, both proteins were connected by an 11 amino acid long linker. Glycine residues were included to confer flexibility and serine residues were included to confer solubility to the fusion protein and the sequence of the linker was GGSGGHMGSGG

We have generated a similar fusion protein but now including both N and C- termini domains of LMO2. Hence, the fusion protein is named as FLINC2 (Fusion of the LIM interacting domain of LDB1 and the N and C- termini domains of LMO2).

FLINC2 was expressed and purified from *E. coli Rosetta* cells as described in Chapter 2, section 2.13.2. Elution fractions were concentrated using a 10kDa MWCO concentrator and further purified by Size exclusion chromatography (SEC) using Superdex 75 16/60 column (GE Healthcare). The protein peak fractions were pooled together and concentrated using a 10kDa MWCO concentrator. The purification of FLINC2 is shown in **Figure 4.19**, in which the SEC chromatogram profile of FLINC2 displays two major peaks (**Figure 4.19a**). The first peak is just after the void volume indicating the presence of aggregates. The second peak, which is indicated by a black arrow, is the FLINC2 protein peak. The SDS-PAGE analysis of samples from the affinity chromatography, wherein pellets (after sonication), beads post elution, flow-through post binding, wash, elution1 (first concentrated fraction) and elution2 (rest of the eluted fractions) fractions were analysed (**Figure 4.19b**). Both the elution

fractions show presence of FLINC2 protein represented by a black arrow. The SDS-PAGE analysis of peak fractions from SEC is shown in **Figure 4.19c**. The purified FLINC2 was obtained which was used for AUC experiments.

E47/SCL bHLH Purification

E47/SCL bHLH was expressed in *E. coli Rosetta* cells and purified using the same protocol used for LMO2 except that 3kDa MWCO was used to concentrate proteins obtained after affinity chromatography and SEC. The purification of E47/SCL bHLH is shown in **Figure 4.20**. The SEC chromatogram profile of E47/SCL bHLH showed a single major peak as represented in **Figure 4.20a**. This is the E47/SCL bHLH peak and is indicated by a black arrow. The SDS-PAGE analysis of samples from the affinity chromatography, wherein pellets (after sonication), beads post elution, flow-through post binding, wash, elution1 (first concentrated fraction) and elution2 (rest of the eluted fractions) fractions were analysed is displayed in **Figure 4.20b**. Both the elution fractions show the over-expressed E47/SCL bHLH protein indicated by a black arrow. The SDS-PAGE analysis of peak fractions from SEC, where peak fractions are indicated in **Figure 4.20c**. Partially purified E47/SCL bHLH was obtained which was used for AUC experiments.

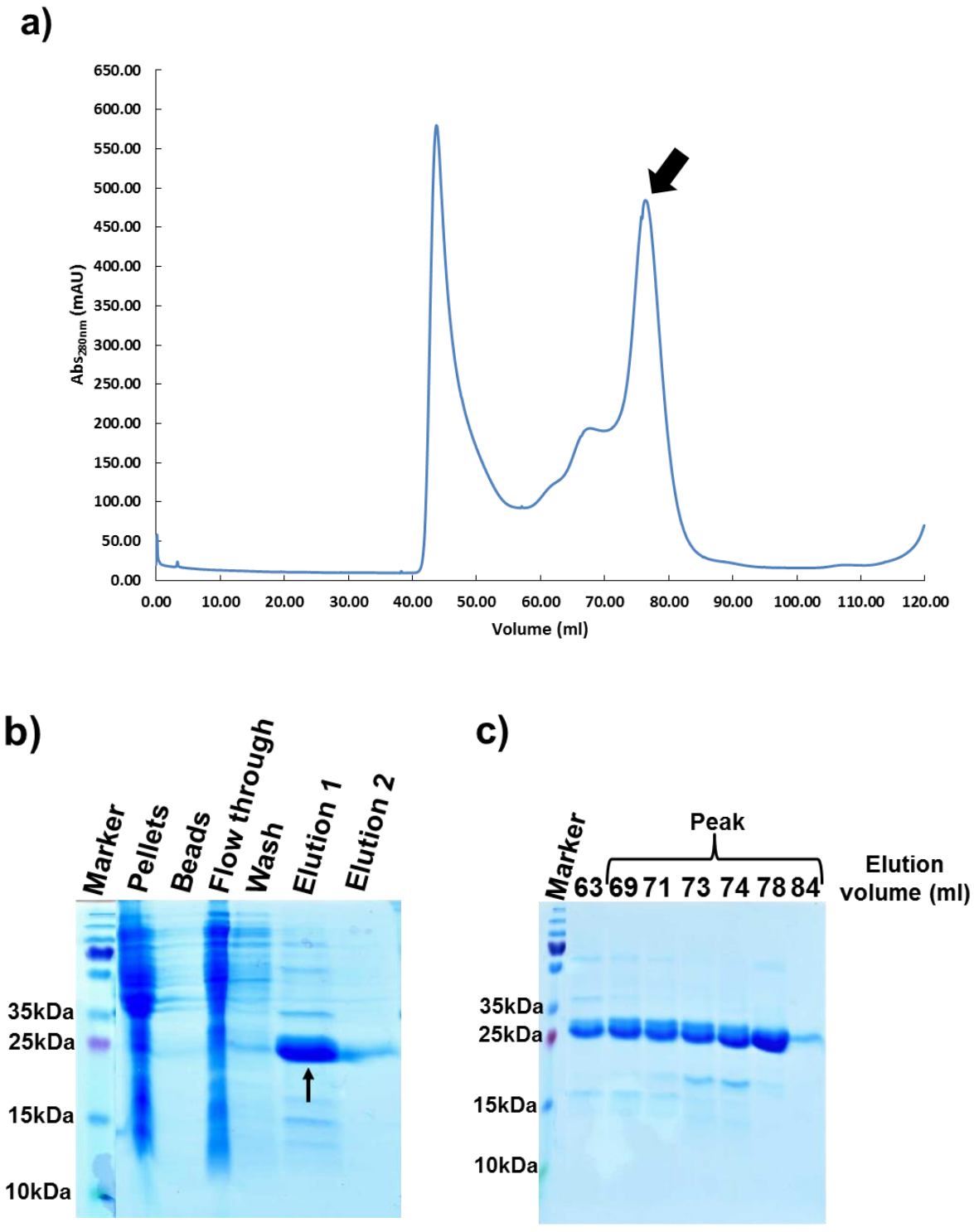


Figure 4.19: Purification of FLINC2. a) Size- exclusion chromatography (SEC) of the FLINC2 protein with a molecular weight of 22kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of FLINC2 (protein band indicated by a black arrow). **c)** SDS-PAGE analysis of peak fractions from SEC of FLINC2.

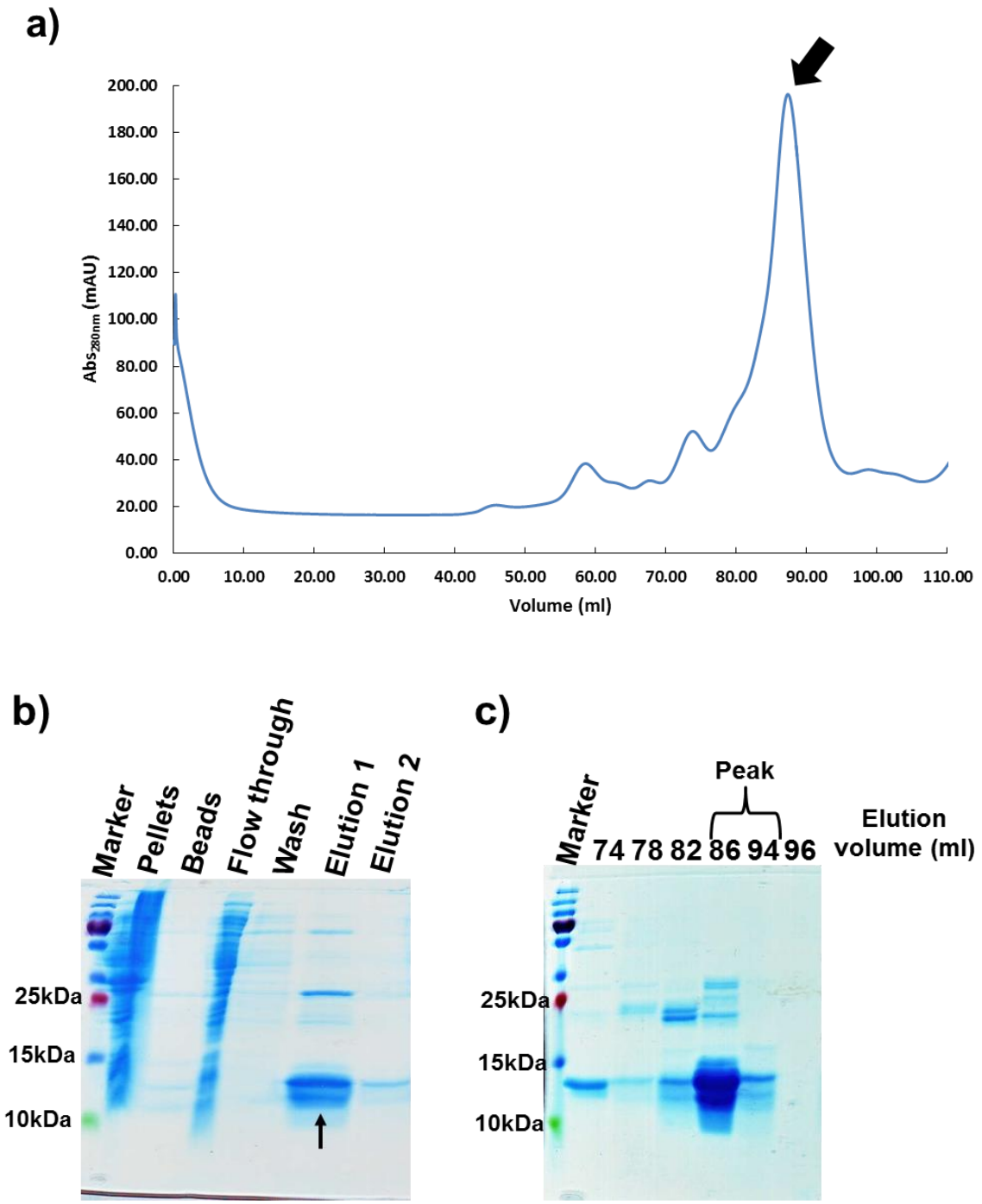


Figure 4.20: Purification of E47/SCL bHLH. a) Size- exclusion chromatography (SEC) of the E47/SCL bHLH protein with a molecular weight of 19kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of E47/SCL bHLH (the protein band indicated by a black arrow). **c)** SDS-PAGE analysis of peak fractions from SEC of E47/SCL bHLH.

GATA1 NFCF purification

GATA1 NFCF was expressed in *E. coli Rosetta* cells, which were cultured in LB auto-induction media. The overall purification protocol used for GATA1 NFCF purification was the same as that for FLINC2 except that a different lysis buffer (50mM HEPES,pH 8, 500mM NaCl, 10mM Imidazole, 1mM β-mercapto ethanol (β-ME), 5mM $MgCl_2$, lysozyme, 34µl of sigma protease inhibitor, 12.5 units/L of DNaseI), wash buffer (50mM HEPES,pH 8, 500mM NaCl, 30mM Imidazole, 1mM β-ME, 5mM $MgCl_2$) and elution buffer (50mM HEPES,pH 8, 500mM NaCl, 500mM Imidazole, 1mM β-ME, 5mM $MgCl_2$) were used. Also, cells were disrupted in a cell disruptor at 30 kpsi at 4°C instead of sonication. The purification of GATA1 NFCF is shown in **Figure 4.21**. The SEC chromatogram profile, in which the major protein peak indicated by a black arrow represents purified GATA1 NFCF is shown in **Figure 4.21a**. The SDS-PAGE analysis of samples from the affinity chromatography, where pellets (after sonication), beads post elution, flow-through post binding, wash, elution1 (first concentrated fraction) and elution2 (rest of the eluted fractions) fractions were analysed is represented by **Figure 4.21b.** Both the elution fractions show the presence of GATA1 NFCF protein indicated by a black arrow. The SDS-PAGE analysis of peak fractions from SEC of GATA1 NFCF is shown in **Figure 4.21c**. The peak fractions which were pooled together and concentrated are highlighted. The purified GATA1 NFCF protein was used for AUC experiments and also for co-crystallization as described in chapter 5, section 5.7.

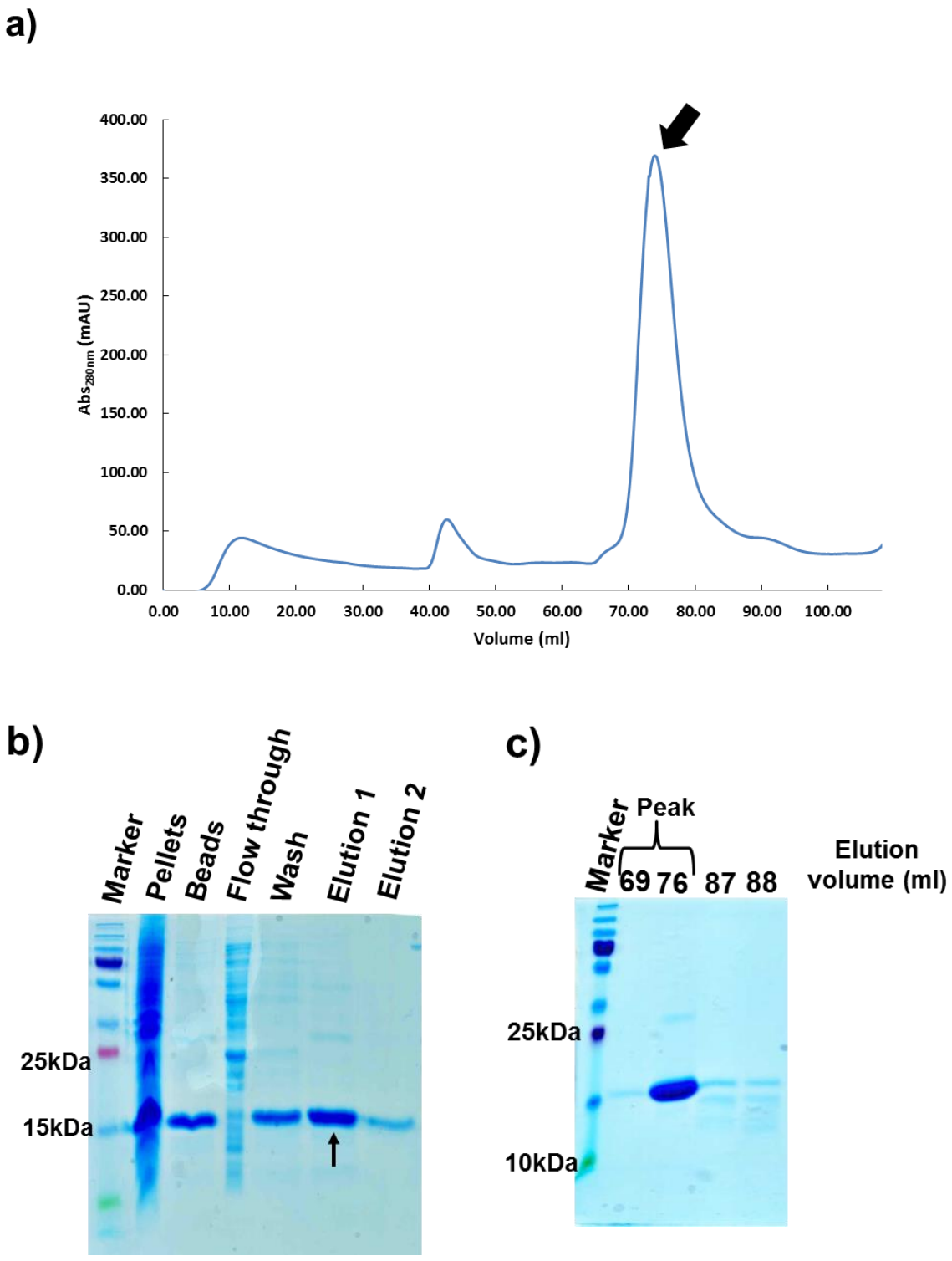


Figure 4.21: Purification of GATA1 NFCF. a) Size- exclusion chromatography (SEC) of the GATA1 NFCF protein with a molecular weight of 15kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of GATA1 NFCF (protein band indicated by a black arrow). **c)** SDS-PAGE analysis of peak fractions from SEC of GATA1 NFCF.

4.4.2 Analytical Ultracentrifugation

For AUC, samples were loaded on to the 3mm path length AUC cells and centrifuged at 1,29,024xg at 20°C. To begin with, His-tagged PHD2 domain was used and AUC was performed in the following combinations: PHD2: FLINC2, PHD2: GATA1 NFCF and PHD2: E47/SCL bHLH.

Of these combinations, one potential interaction was observed; PHD2 and GATA1 NFCF (**Figure 4.22**). The model independent c(s,ff_0) analysis of PHD2 absorbance data displayed two peaks (**Figure 4.22a**). The first broad peak could be that of unfolded PHD2 as unfolded proteins tend to have broad elongated shape with varied diffusion rates that has a boundary effect that is indicated by the broad shape of the curve. Also, they tend to sediment slowly, hence the lower s-value than that of the mono-dispersed protein. The second sharp, major peak represents properly folded PHD2. The species has an estimated molecular weight of about 10kDa, which is in line with the expected theoretical molecular weight of PHD2.

The model independent c(s,ff_0) analysis of GATA1 NFCF absorbance data displayed a single peak. The species has a sedimentation coefficient of 1.5S and an estimated molecular weight 15kDa, which is in line with the expected theoretical molecular weight of GATA1 NFCF (**Figure 4.22b**).

When the apparent sedimentation coefficient distribution (c(s)) profiles of PHD2 (red), GATA1 NFCF (green) and PHD2+GATA1 NFCF(blue) mixture were superimposed, PHD2 and GATA1 NFCF on their own formed a sharp, well defined peaks at 1.4S and 1.5S

respectively. After both proteins were mixed in equimolar ratios, a single sedimentation coefficient peak was observed at 1.67S suggesting that these proteins interact (**Figure 4.22c**).

The model independent c(s,ff_0) analysis of absorbance data obtained from equimolar mixture of PHD2 and GATA1 NFCF resulted in two peaks. The first peak has a molecular weight of 15kDa indicating unbound GATA1 and the second peak has a molecular weight of 20kDa which is slightly lower than the theoretical molecular weight of PHD2+GATA1 NFCF (25kDa). However, a clear shift in the sedimentation coefficient to 1.67S, which is higher than that of PHD2 (1.4S) and GATA1 NFCF (1.5S) individually, is indicative that both proteins do form a complex (**Figure 4.22d**). Interestingly, the unbound GATA1 NFCF tends to sediment slowly which is indicated by the lower sedimentation coefficient (1S) than expected (1.5S) (**Figure 4.22b**). This could be due to the complex affecting the sedimentation rate of the unbound GATA1 NFCF.

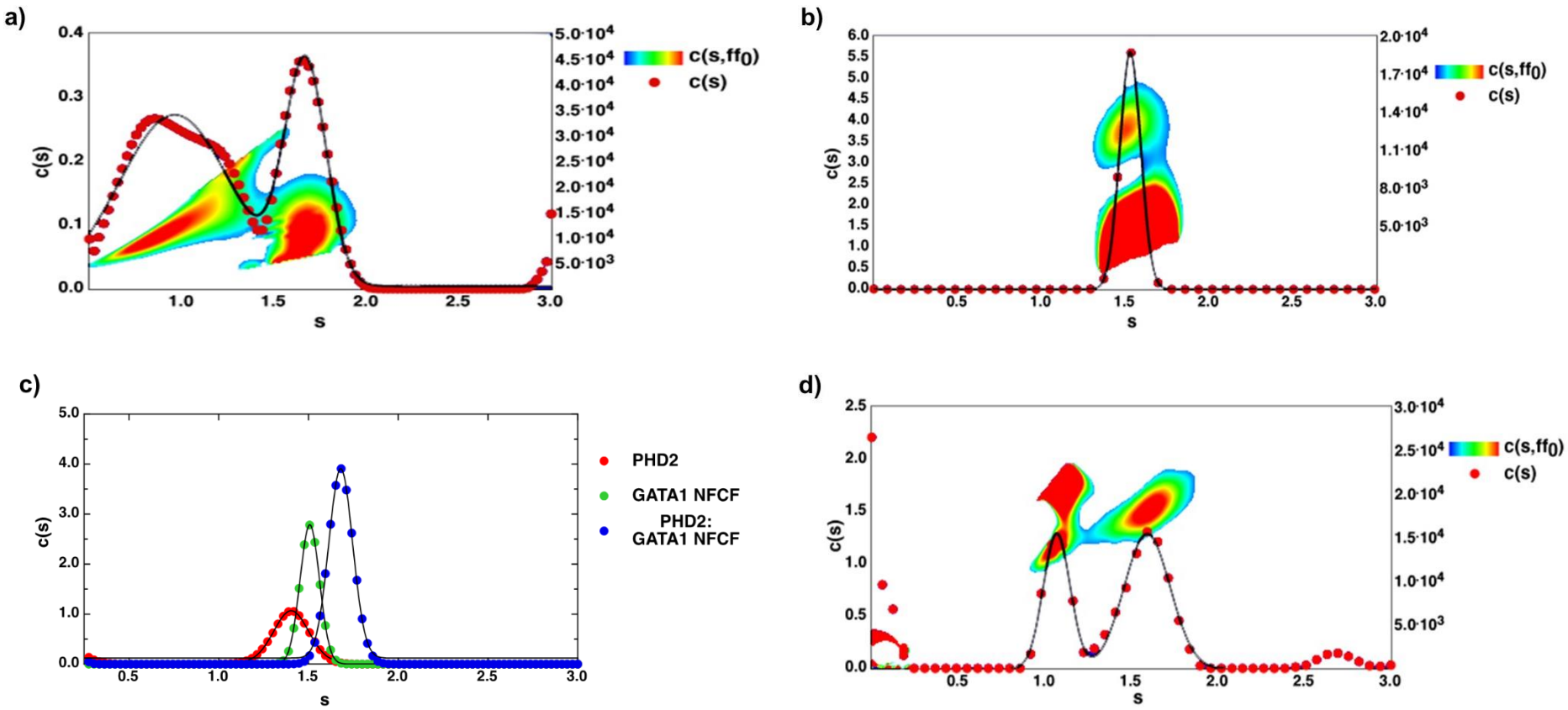


Figure 4.22: AUC sedimentation velocity experiment of PHD2 and GATA1 NFCF. a), b) and d) represent c(s,ff$_0$) analysis for PHD2, GATA1 NFCF and PHD2+GATA1 NFCF absorbance data respectively. The c(s) value has predominant species fitted with a Gaussian function, s=1.4±0.1S for PHD2, s= 1.5±0.05S for GATA1 NFCF and s=1.67±0.06S for PHD2+GATA1 NFCF. The Y2 axis shows the molecular weight and the temperature plot represents the distribution of the apparent frictional coefficients in the sample which is plotted over a third dimension (see key for contour plots). **c)** apparent sedimentation coefficient distribution (c(s)) profiles of PHD2 (red), GATA1 NFCF (green) and PHD2+GATA1 NFCF (blue) mixture superimposed. All the graphs were generated using the pro Fit software.

FLINC2-E47/SCL bHLH combination was used as a positive control (**Figure 4.23**). The model independent c(s,ff$_0$) analysis of FLINC2 absorbance data displayed a single, major peak. The species has a sedimentation coefficient of 2.1S and the estimated molecular weight 22kDa, which is in line with the expected theoretical molecular weight of FLINC2 (**Figure 4.23a**).

The model independent c(s,ff$_0$) analysis of E47/SCL bHLH absorbance data displayed a major peak at 2S and estimated molecular weight 20kDa which is in line with the expected theoretical molecular weight of E47/SCL bHLH (**Figure 4.23b**).

The superimposition of the apparent sedimentation coefficient distribution (c(s)) profiles of FLINC2 (red), E47/SCL bHLH (green) and FLINC2+E47/SCL bHLH (blue) showed that FLINC2 and E47/SCL bHLH on their own formed well defined peaks at 2.1S and 2S respectively. When both proteins were mixed in equimolar ratios, a single broad sedimentation coefficient peak was observed at 2.8S, clearly demonstrating an interaction between these proteins (**Figure 4.23c**).

The model independent c(s,ff$_0$) analysis of the absorbance data obtained from equimolar mixture of FLINC2 and E47/SCL bHLH resulted in three peaks. The first peak has a molecular weight of about 22kDa indicating unbound FLINC2. The broad complex peak from the sedimentation coefficient distribution analysis resolved into two peaks in c(s,ff$_0$) analysis each with molecular weight of 35kDa which is close to the molecular weight of FLINC2+E47/SCL bHLH (40kDa) confirming the complex formation between these proteins. The formation of two peaks with different c(s,ff$_0$) with same molecular weights is

indicative of the fact that two different conformations of the complex are present (**Figure 4.23d**). Similar to that of PHD2+GATA1, the sedimentation coefficient of unbound FLINC2 was lower (1.5S) than expected (2.1S) (**Figure 4.23b**). Again this could be due to the complex affecting the sedimentation rate of the unbound FLINC2.

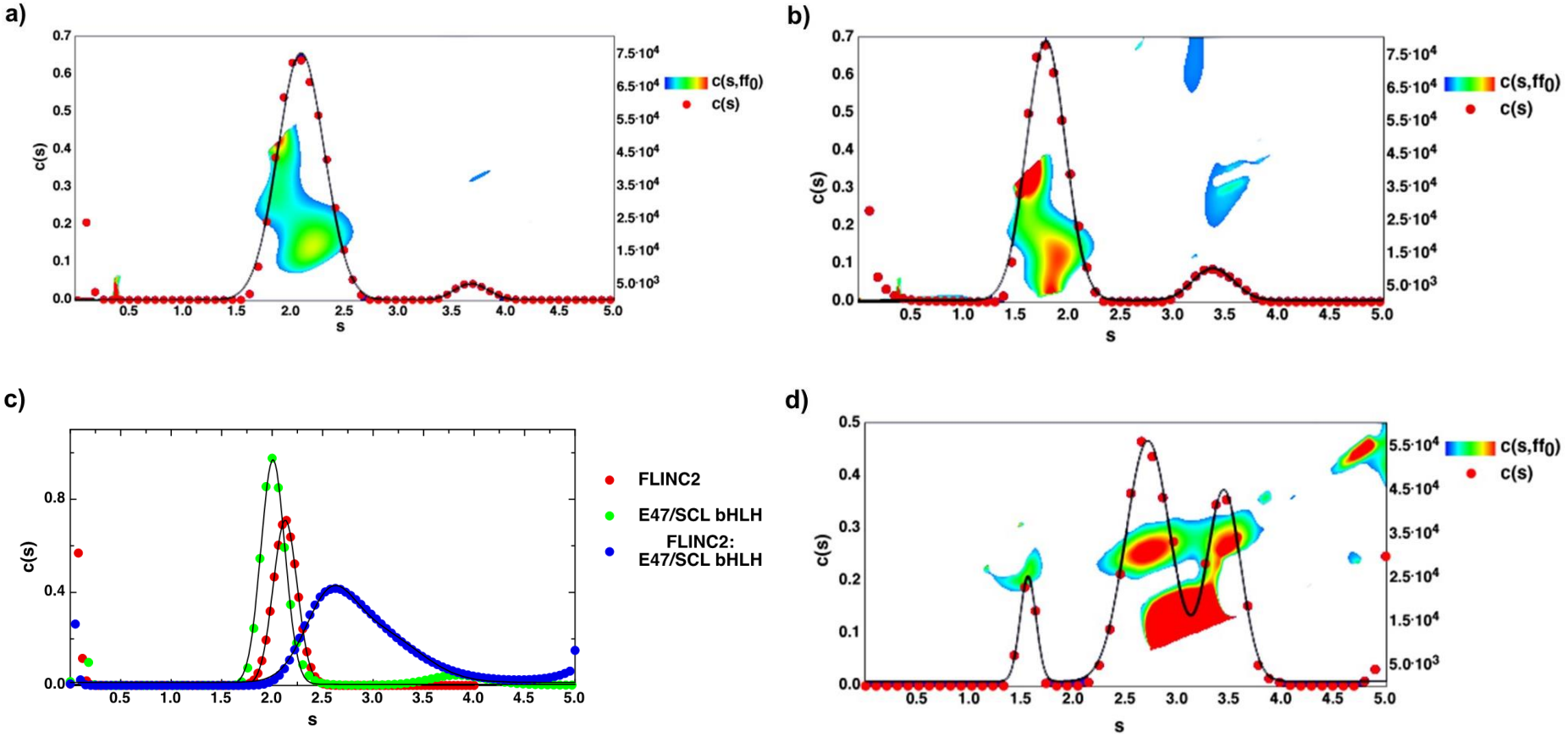


Figure 4.23: AUC sedimentation velocity experiment of FLINC2 and E47/SCL bHLH. a), b) and d) represent c(s,ff$_0$) analysis for FLINC2, E47/SCL bHLH and FLINC2+E47/SCL bHLH absorbance data respectively. The c(s) value has predominant species fitted with a Gaussian function, s=2.1±0.1S for FLINC2, s=2±0.1S for E47/SCL bHLH and s=2.8±0.4S for FLINC2+E47/SCL bHLH. The Y2 axis shows the molecular weight and the temperature plot represents the distribution of the apparent frictional coefficients in the sample which is plotted over a third dimension (see key for contour plots). **c)** apparent sedimentation coefficient distribution (c(s)) profiles of FLINC2 (red), E47/SCL bHLH (green) and FLINC2+E47/SCL bHLH (blue) mixture superimposed. All the graphs were generated using the pro Fit software.

The expression of His-tagged PHD2 in *E.coli* cells was good enough to provide suitable amounts of purified protein to perform AUC experiments. However, the expression levels of His-tagged PHD1 and PHD3 constructs were very low and not enough for AUC. Hence, these were expressed as GST-tagged proteins (which expressed very well yielding mg quantities of protein) and then the tag was cleaved using HRV 3C protease. However, upon cleaving the GST tag, proteins precipitated.

It was therefore decided to use the uncleaved GST-tagged versions of the three PHD domains, although clearly the presence of a tag can complicate the analysis of the data as GST can dimerize.

All three proteins were over-expressed in *E. coli* rossetta cells by IPTG induction and purified using affinity chromatography followed by size-exclusion chromatography (SEC) using a Superdex 200 16/60 column. Fractions from SEC were analysed by SDS-PAGE and the relevant fractions were concentrated using 10kDa MWCO (molecular weight cut-off) concentrator tubes.

The purifications of GST PHD1, GST PHD2 and GST PHD3 proteins are shown in **Figure 4.24**, **Figure 4.25** and **Figure 4.26** respectively. In all the figures, panel ‘a’ represents the SEC chromatography profile in which the major peak representing the correct protein peak is indicated by a black arrow. Panel ‘b’ shows the SDS-PAGE analysis of samples from affinity chromatography, where pellets (after sonication), beads post elution, flow-through post binding, wash, elution1 (first concentrated fraction) and elution2 (rest of the eluted fractions) fractions were analysed. The elution1 fractions showed two bands, one at 35kDa and a

smaller band at 25kDa which is indicated by a black arrow. The expected size for all three proteins is 36kDa. Mass spectrometry performed on the second band observed at around 25kDa indicated that this band too is that of the corresponding protein being purified. These bands are therefore likely to be degradation products. Panel 'c' shows SDS-PAGE analysis of peak fractions from SEC. The protein peaks representing the mono-dispersed proteins are highlighted.

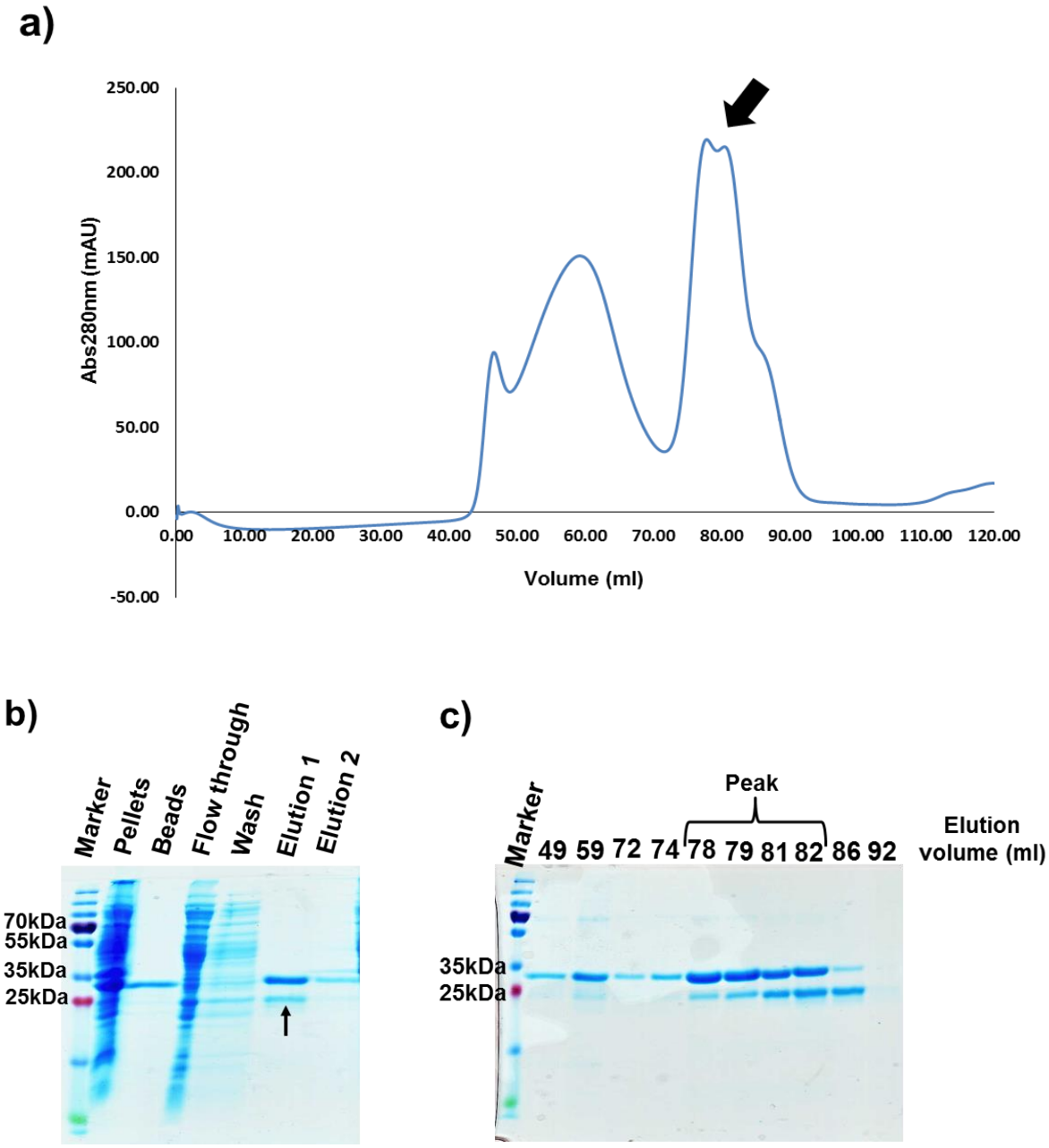


Figure 4.24: Purification of GST PHD1. a) Size- exclusion chromatography (SEC) of the GST PHD1 protein with a molecular weight of 36kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of GST PHD1. **c)** SDS-PAGE analysis of peak fractions from SEC of GST PHD1.

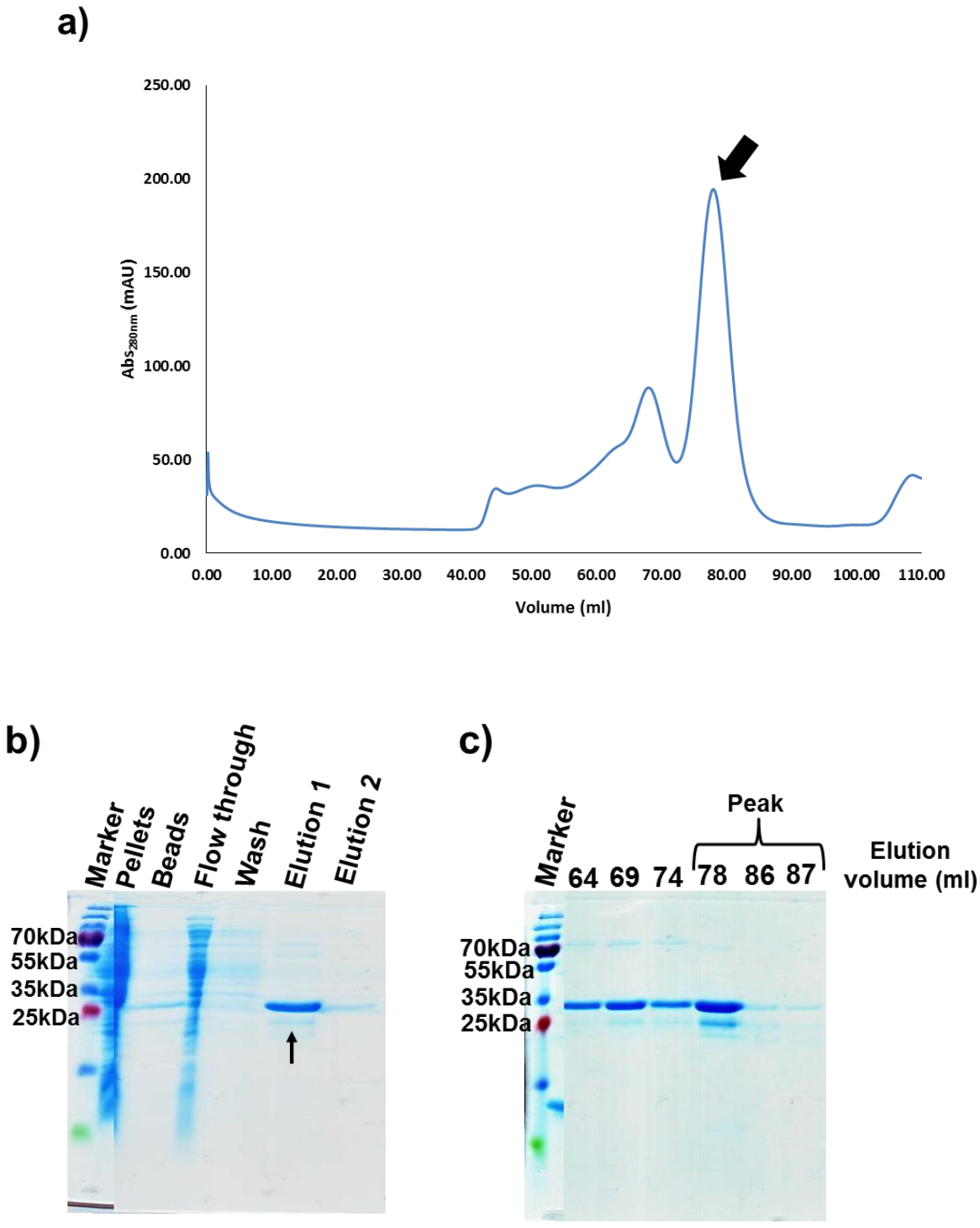


Figure 4.25: Purification of GST PHD2. a) Size- exclusion chromatography (SEC) of the GST PHD2 protein with a molecular weight of 36kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of GST PHD2. **c)** SDS-PAGE analysis of peak fractions from SEC of GST PHD2.

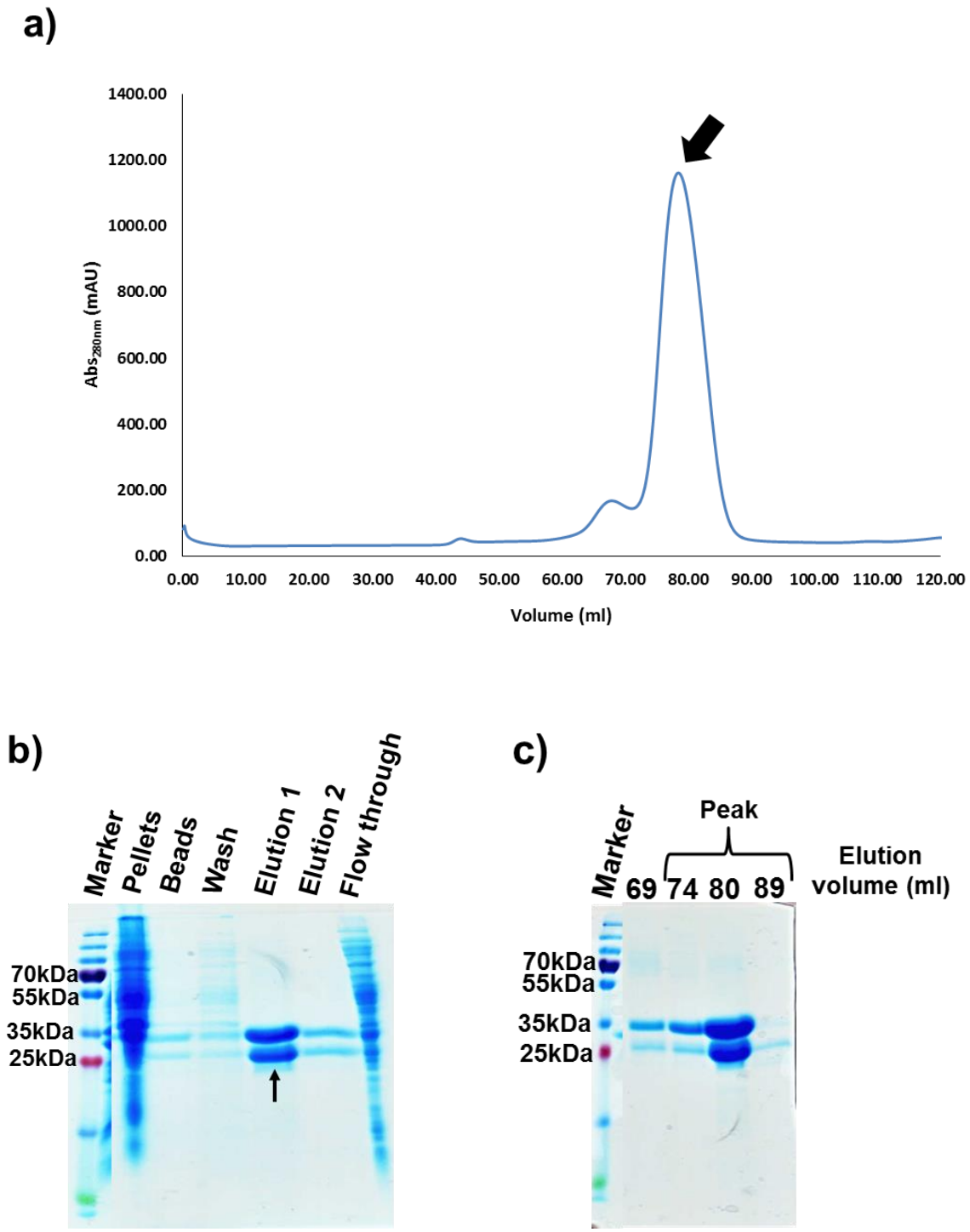


Figure 4.26: Purification of GST PHD3. a) Size- exclusion chromatography (SEC) of the GST PHD3 protein with a molecular weight of 36kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of GST PHD3. **c)** SDS-PAGE analysis of peak fractions from SEC of GST PHD3.

Once the pure proteins were obtained, AUC was performed using GST PHDs with FLINC2, GATA1 NFCF and E47/SCL bHLH (**Figure 4.27**). The apparent sedimentation coefficient distribution (c(s)) profiles of the GST PHDs (red) were superimposed over E47/SCL bHLH, FLINC2, GATA1 NFCF (green) and their mixtures (blue). None of the combinations showed any interaction except GST PHD2 and GATA1 NFCF (shown in **Figure 4.28**) as there is no shift in the sedimentation coefficient peaks of the individual proteins upon mixing them with their binding partners.

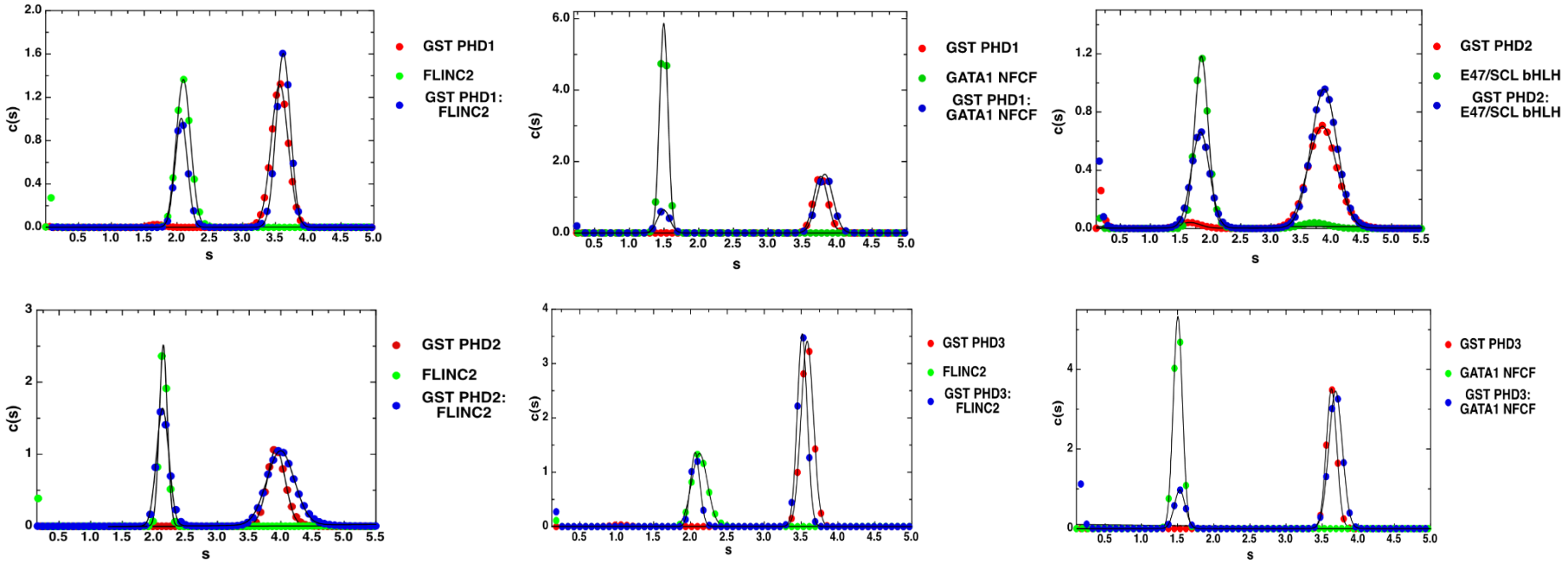


Figure 4.27: AUC sedimentation velocity experiment of GST PHD1/PHD2/PHD3 and E47/SCL bHLH, FLINC2 and GATA1 NFCF. Apparent sedimentation coefficient distribution (c(s)) profiles of GST PHDs (red), their interacting partner in question (green) and their mixture (blue) superimposed. All the graphs were generated using the pro Fit software.

The interaction between GST PHD2 and GATA1 NFCF (**Figure 4.28**) validated the interaction observed previously between His-tagged PHD2 and GATA1 NFCF as shown in **Figure 4.22**. The model independent c(s,ff$_0$) analysis of GST PHD2 absorbance data displayed two peaks (**Figure 4.28a**). The first peak represents GST PHD2 dimers and exhibits a molecular weight of ~75kDa which is in line with the expected theoretical weight of GST PHD2 dimers (70kDa). The second peak displays species with molecular weight of 40kDa and hence likely represents GST PHD2 monomers.

The model independent c(s,ff$_0$) analysis of GATA1 NFCF absorbance data displayed a single peak. The species has a sedimentation coefficient of 1.5S and predicted molecular weight 15kDa, which is the expected theoretical molecular weight of GATA1 NFCF (**Figure 4.28b**).

The superimposition of the apparent sedimentation coefficient profiles of GST PHD2 (red), GATA1 NFCF (green) and GST PHD2+ GATA1 NFCF (blue) showed that GST PHD2 and GATA1 NFCF on their own formed sharp, well defined peaks at 3.8S and 1.5S respectively. The equimolar mixture of both proteins displayed a wide peak centred at 4S, partially overlapping with the GST PHD2 peak but extending further, confirming the interaction between GST PHD2 and GATA1 NFCF. However, it also showed a peak overlapping with GATA1 NFCF alone at 1.5S suggesting that not all the GATA1 NFCF was in complex with GST PHD2 (**Figure 4.28c**).

The model independent c(s,ff$_0$) analysis of the absorbance data obtained from the equimolar mixture of GST PHD2 and GATA1 resulted in three peaks. The first peak has a molecular weight of 15kDa indicating unbound GATA1 and the second peak has a molecular weight of

20kDa which could be GATA1 NFCF bound to DNA or a degradation product of GST PHD2. However, it is difficult to predict the true identity of this peak as it is not visible in the $c(s,ff_0)$ analysis of GST PHD2 or GATA1 NFCF on their own. The third peak represents the complex of GST PHD2+GATA1 NFCF with a molecular weight of 50kDa showing that GST PHD2 is present in its monomeric state when in complex with GATA1 NFCF. So, it is possible that the complex formation between GST PHD2 and GATA1 NFCF leads to the dissociation of the GST PHD2 dimers (**Figure 4.28d**).

In summary, it is fair to say that the above described GST pull-down and AUC experiments indicate that the PHD2 domain of JARID1A interacts with GATA1 NFCF.

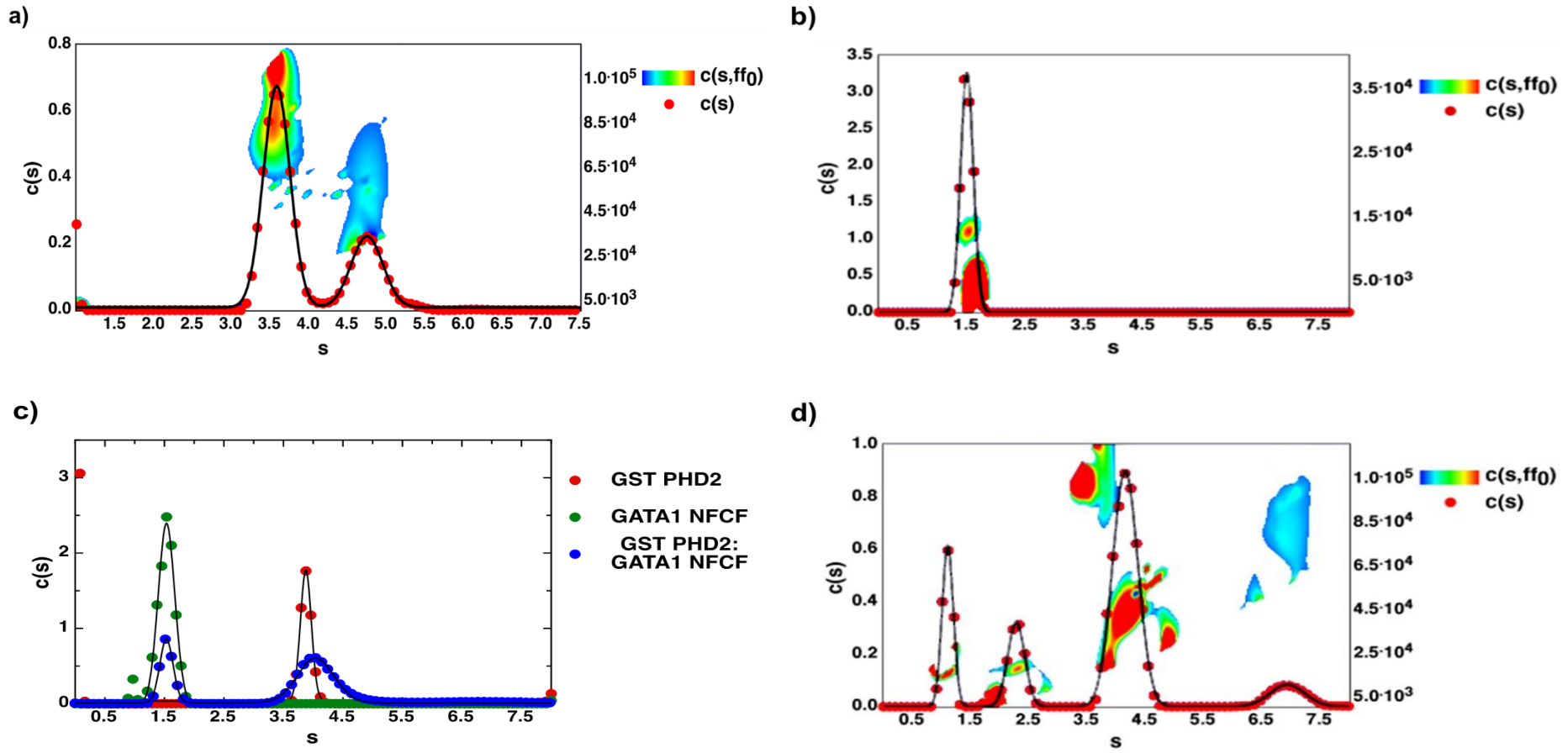


Figure 4.28: AUC sedimentation velocity experiment of GST PHD2 and GATA1 NFCF. a), b) and d) represent c(s,ff$_0$) analysis for GST PHD2, GATA1 NFCF and GST PHD2+GATA1 NFCF absorbance data respectively. The c(s) value has predominant species fitted with a Gaussian function, s=3.8±0.09S for GST PHD2, s= 1.5±0.1S for GATA1 NFCF and s=4±0.3S for GST PHD2+GATA1 NFCF. The Y2 axis shows the molecular weight and the temperature plot represents the distribution of the apparent frictional coefficients in the sample which is plotted over a third dimension (see key for contour plots). **c)** apparent sedimentation coefficient distribution (c(s)) profiles of GST PHD2 (red), GATA1 NFCF (green) and GST PHD2+GATA1 NFCF (blue) mixture superimposed. All the graphs were generated using the pro Fit software.

A summary of interactions obtained between the PHD domains of JARID1A and the SCL complex components in the AUC is presented in **Table 4.2**.

Table 4.2: Summary of interaction studies by AUC.

Protein combination	Inference
PHD2 (N-His): FLINC2	No interaction
PHD2 (N-His): GATA1 NFCF	Interaction
PHD2 (N-His): E47/SCL bHLH	No interaction
GST PHD1: FLINC2	No interaction
GST PHD1: GATA1 NFCF	No interaction
GST PHD1: E47/SCL bHLH	No interaction
GST PHD2: FLINC2	No interaction
GST PHD2: GATA1 NFCF	Interaction
GST PHD2: E47/SCL bHLH	No interaction
GST PHD3: FLINC2	No interaction
GST PHD3: GATA1 NFCF	No interaction
GST PHD3: E47/SCL bHLH	No interaction

To proceed forward, I attempted to quantify the binding affinity of PHD2 for GATA1 NFCF by surface plasmon resonance (SPR) methods using the Biacore system. The assay was performed exploiting the streptavidin-biotin system. However, due to the highly basic nature of these proteins, they were sticking to the negatively charged Biacore chip surface. Several attempts were carried out to resolve this issue including alternating biotin tags on both these

proteins, saturating the chip surface with the biotin-tagged protein and using high and low molecular weight PEI to reverse the charge on the chip surface. But none of these resulted in interpretable results.

Another attempt at measuring affinities was performed by using Nuclear magnetic resonance (NMR) chemical shift titrations experiments. NMR titrations are used to study interactions between two proteins by studying changes in the chemical shifts caused by gradual addition of an unlabelled 'ligand' protein to a fixed concentration of isotopically labelled protein of interest. The PHD2 protein was isotopically labelled (due to its low expression levels) and titrated against increasing concentration of GATA1 NFCF. There were shifts observed in three peaks (Appendix-V, Figure A.7.8) but the shifts were too small to determine the binding affinities. This could be due to the low expression levels of labelled PHD2, the dynamic nature of the PHD2-GATA1 NFCF interaction or the low binding affinities between the two proteins.

Chapter 5: Protein expression, purification and crystallization of JARID1A domains

5 Cloning, protein expression, purification and crystallization of JARID1A domains

5.1 Introduction

In order to gain structural insights into the individual domains of JARID1A, which in turn would give some useful information about their function, we designed a battery of constructs with N- and C- terminal His-tag for crystallization based on the GST tagged constructs described in Chapter 4, section 4.3.1 (**Figure 4.10** and **Figure 4.11**).

5.2 Protein expression and purification in *E. coli*

All the N- and C- terminal His-tagged JARID1A constructs were tested for expression through the high-throughput screening pipeline at the OPPF (Berrow et al. 2007) using two different *E. coli* strains: *BL21 Lemo* and *Rosetta*. Protein expression was triggered using IPTG induction in the Power Prime broth™ (Molecular Dimensions) and the auto-induction in Overnight Express™ Instant TB Medium (EMD Millipore). Protein expression and purification was performed as described in Chapter 2, section 2.16.3. The eluted proteins were analysed by SDS-PAGE on 10% gel. The expression test for the JARID1A constructs in *E. coli BL21 Lemo* strain using both the auto-induction and the classical IPTG induction systems is shown in **Figure 5.1**. Arid, PHD1 and PHD2 domains showed expression in *E. coli BL21 Lemo*, in which Arid domain showed expression in both the induction systems, however, PHD2 domain displayed preference for the auto-induction system and PHD1 showed preference for the classical IPTG induction system (**Figure 5.1a** and **Figure 5.1b**). The expression test for the JARID1A constructs in *E. coli BL21 Rosetta* strain using both the auto-induction and the classical IPTG induction systems is shown in **Figure 5.2**. JmjN+Arid, PHD1 and PHD2 domains showed expression in *E. coli Rosetta*, where JmjN+Arid and

PHD2 constructs expressed only in the auto-induction system. The two PHD1 domain constructs showed differences in their expression preferences, with PHD1 N-His expressing in the auto-induction system and PHD1 C-His expressing in the IPTG induction system (**Figure 5.2a** and **Figure 5.2b**).

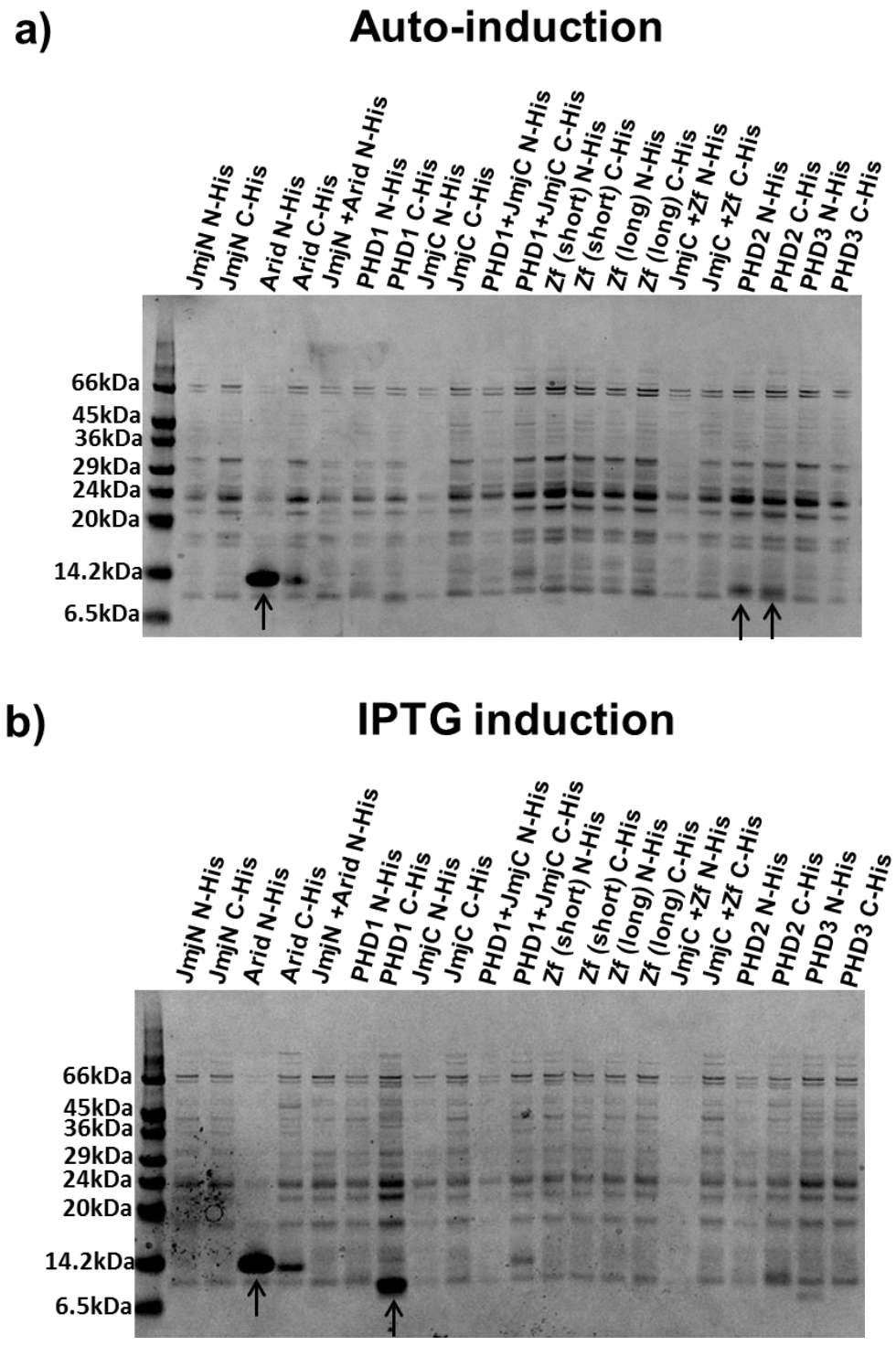


Figure 5.1: Small scale expression test of N- and C-terminal His-tagged JARID1A clones in *E.coli BL21 Lemo*. a) Expression test for the N- and C-terminal His-tagged JARID1A clones in *E. coli BL21 Lemo* strain using the auto-induction system. **b)** Expression test for the N- and C-terminal His-tagged JARID1A clones in *E. coli BL21* strain using the IPTG induction system. The black arrows indicates the proteins which showed expression.

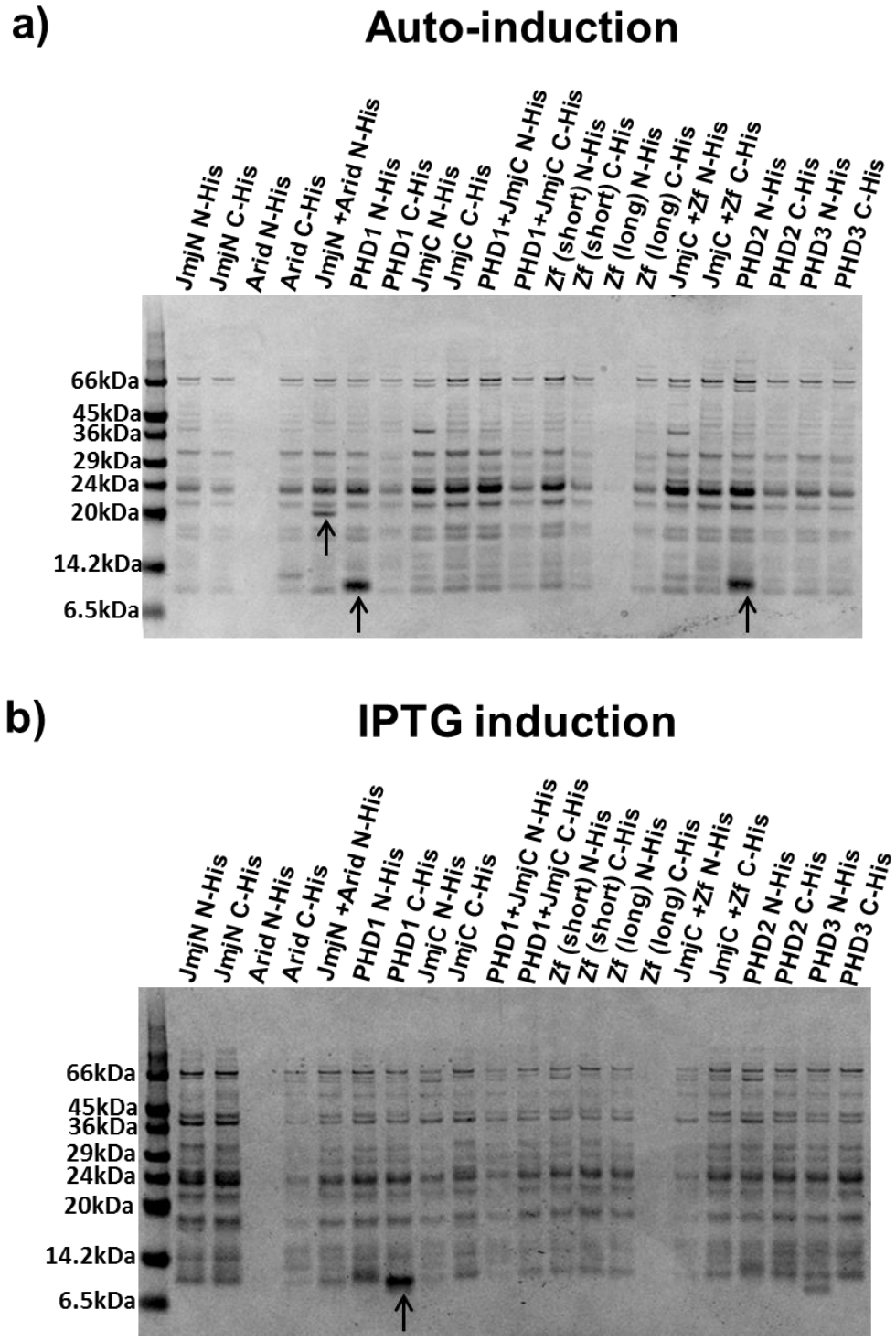


Figure 5.2: Small scale expression test of N- and C-terminal His-tagged JARID1A clones in *E.coli Rosetta*. a) Expression test for the N- and C-terminal His-tagged JARID1A clones in *E. coli BL21 Rosetta* strain using the auto-induction system. **b)** Expression test for the N- and C-terminal His-tagged JARID1A clones in *E. coli Rosetta* strain using the IPTG induction system. The black arrows indicates the proteins which showed expression.

A list of JARID1A constructs showing expression in the small-scale expression tests is displayed in **Table 5.1**.

Table 5.1: List of JARID1A clones showing expression in the small scale expression test.

	Auto-induction	**IPTG induction**
E. coli BL21 Lemo	Arid N-His PHD2 N-his PHD2 C-His	Arid N-His PHD1 C-His
E. coli Rosetta	JmjN+Arid N-His PHD1 N-His PHD2 N-His	PHD1 C-His

PHD1, PHD2 and JmjN+Arid expression and purification were then optimized for scale-up. The Arid domain construct was designed to be used only for co-crystallization with the SCL complex components if it showed any interaction with the latter, because its three-dimensional structure is already published. However, since it did not show any interaction with the SCL complex components in GST pull down assays (Refer Chapter 4, section 4.3, Figure 4.12), it was not scaled-up.

The PHD1 and PHD2 (both N- and C-terminal clones) protein expression was further optimized using LB (Luria Bertani), LB auto-induction, Power broth and TB (Terrific Broth) auto-induction media. Of all media, the PHD1 N- His and PHD2 N- His constructs showed optimal expression in TB auto-induction media and hence these were scaled up. The PHD1 and PHD2 proteins were purified as described in Chapter 2, section 2.13.2. The eluted protein fractions were concentrated using a 3kDa MWCO concentrator and further purified by Size exclusion chromatography (SEC) using a Superdex 75 16/60 column (GE Healthcare).

Protein peak fractions were pooled together and concentrated using a 3kDa MWCO concentrator. The purification of PHD1 N-His is shown in **Figure 5.3**. The SEC of PHD1 N-His showed two peaks (**Figure 5.3a**). The first peak suggests protein aggregation as it elutes close to the void volume whilst the second peak, indicated by a black arrow, represents mono-dispersed PHD1 N-His. The SDS-PAGE analysis of samples from affinity chromatography was carried out, where pellets, beads post elution, flow-through post binding, wash, elution1 (first concentrated fraction) and elution2 (rest of the eluted fractions) fractions were analysed (**Figure 5.3b**). The elution1 fraction showed good expression of the PHD1 N-His construct and is highlighted by a black arrow. The SDS-PAGE analysis of peak fractions from SEC showed that the second small peak fractions contain a protein band of the correct size (**Figure 5.3c**). The protein fractions which were pooled together for concentration are highlighted.

The PHD2 N-His construct was purified similarly (**Figure 5.4**). The SEC profile showed several protein peaks (**Figure 5.4a**). The first peak just after the void volume probably corresponds to protein aggregation, whilst the last peak is likely mono-dispersed PHD2 N-His, which is highlighted by a black arrow. The protein eluting before the void volume of the column is probably because the column was not washed well by previous user. The SDS-PAGE analysis of samples from affinity chromatography was performed, where pellets, beads post elution, flow-through post binding, wash, elution1 (first concentrated fraction) and elution2 (rest of the eluted fractions) fractions were analysed (**Figure 5.4b**). The elution1 fraction showed good expression of the PHD2 N-His construct and is highlighted by a black arrow. The SDS-PAGE analysis of peak fractions from SEC showed the presence of protein aggregation, contaminants and of protein bands of the right size, which are highlighted (**Figure 5.4c**). The peak fractions which were pooled together are indicated.

The purified PHD2 N-His protein was used for AUC experiments. The PHD1 N-His expression levels were very low (270µg from 6L culture) and hence there was not enough to set up crystallization screens or AUC experiments. PHD2 N-His was also used for co-crystallization with GATA1 NFCF. The latter was purified as described in chapter 4, section 4.4.1.

The JmjN+Arid N-His construct also showed promising soluble protein expression in the small-scale expression test and therefore was chosen for large-scale purification. The purification protocol was similar to the one used for the PHD1 and PHD2 N-His constructs. (**Figure 5.5**).

The major peak in the SEC chromatogram profile indicated by the black arrow represents JmjN+Arid (**Figure 5.5a**). The SDS-PAGE analysis of samples from affinity chromatography was performed, where pellets, beads post elution, flow-through post binding, wash, elution1 (first concentrated fraction) and elution2 (rest of the eluted fractions) fractions were examined (**Figure 5.5b**). Although JmjN+Arid N-His construct showed expression (indicated by the black arrow), a stronger contaminating protein band was observed just above. The identity of the JmjN+Arid protein band was confirmed by western blotting using anti-His antibodies (data not shown). The SDS-PAGE analysis of peak fractions from the SEC showed that the contaminating protein co-eluted in gel filtration chromatography, as shown in **Figure 5.5c** where the contaminant band is seen running very closely to the protein band (indicated by a black arrow). Elimination of the contaminating protein was attempted by

using more stringent washes and ion exchange chromatography with no success. Therefore, crystallization of JmjN+Arid construct could not be performed.

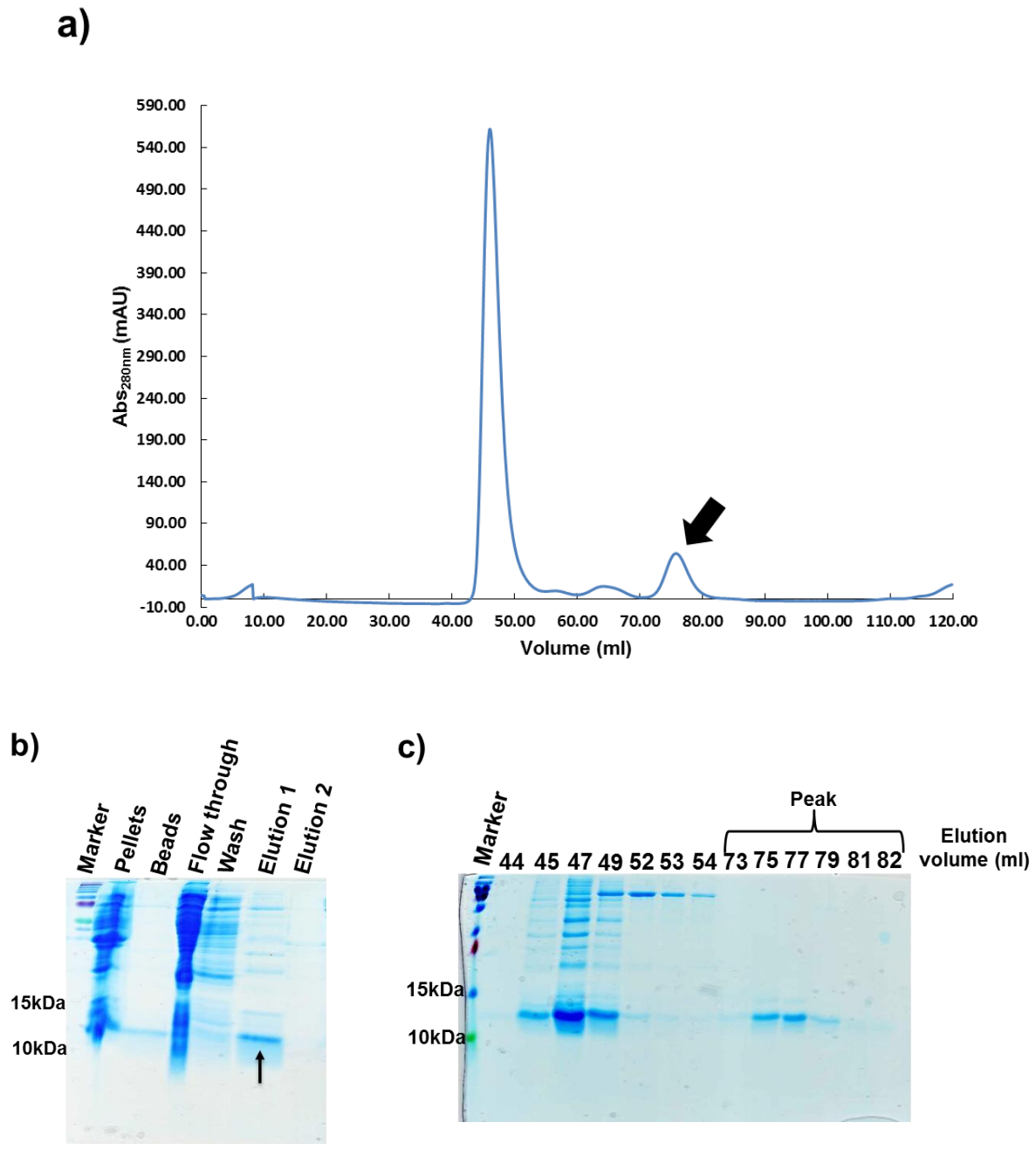


Figure 5.3: Purification of PHD1 N-His. a) Size- exclusion chromatography (SEC) of the PHD1 N-His protein with a molecular weight of 10kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of PHD1 N-His (the protein band is indicated by a black arrow). **c)** SDS-PAGE analysis of peak fractions from SEC of PHD1 N-His.

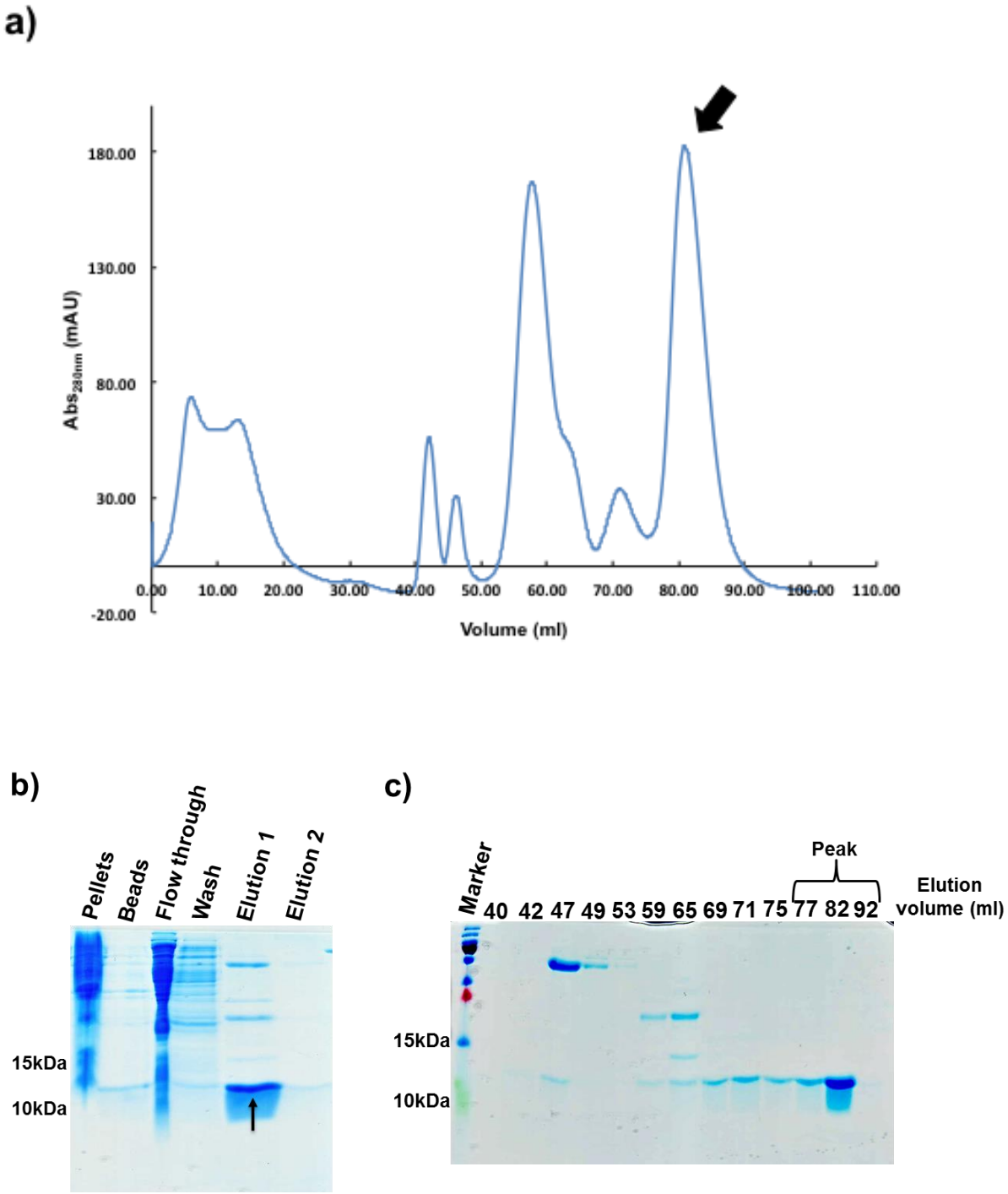


Figure 5.4: Purification of PHD2 N-His. a) Size- exclusion chromatography (SEC) of the PHD2 N-His protein with a molecular weight of 10kDa (a black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of PHD2 N-His (protein band indicated by black arrow). **c)** SDS-PAGE analysis of peak fractions from SEC of PHD2 N-His.

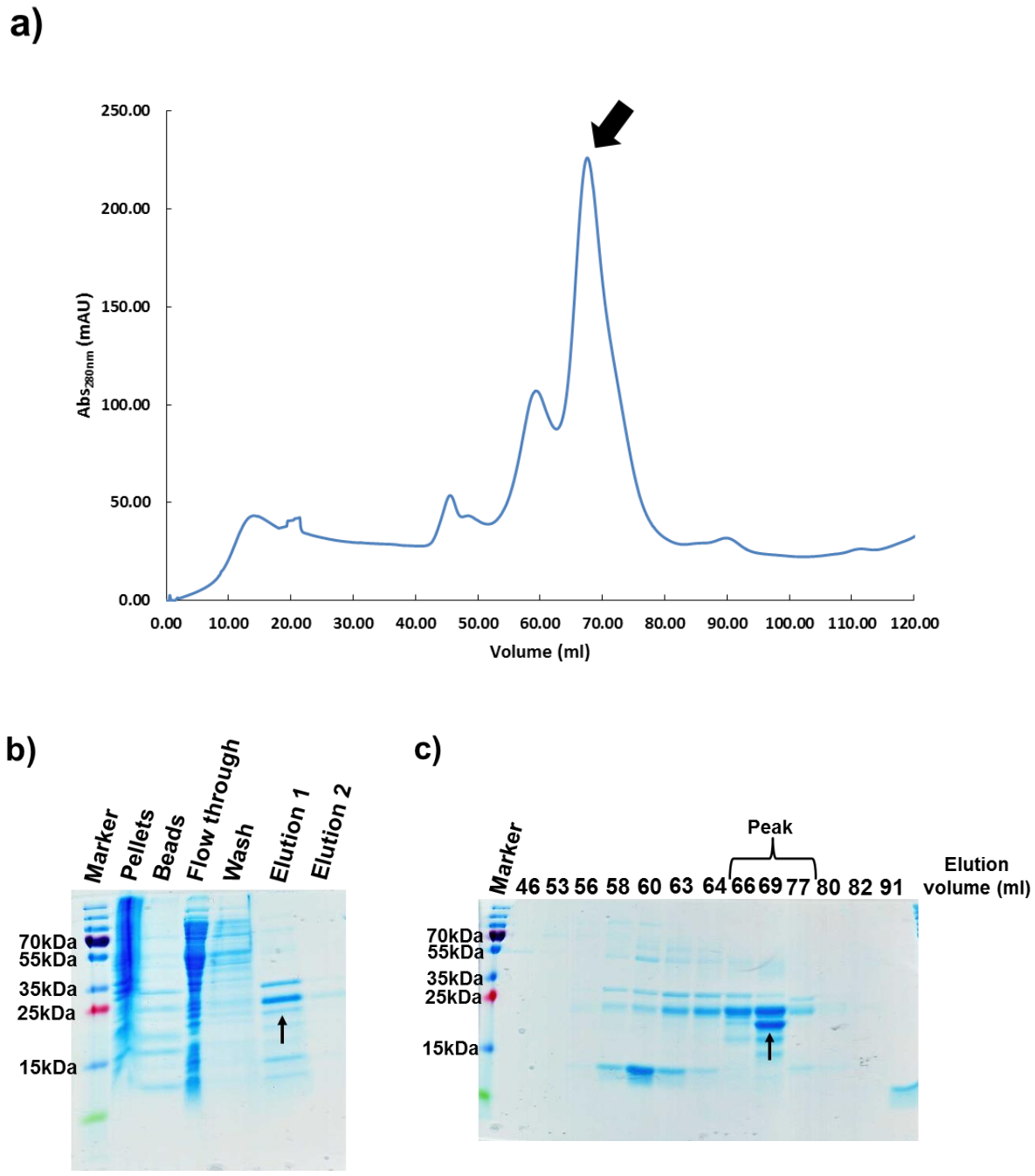


Figure 5.5: Purification of JmjN+Arid N-His. a) Size- exclusion chromatography (SEC) of the JmjN+Arid N-His protein with a molecular weight of 20kDa (a black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of JmjN+Arid N-His (protein band indicated by a black arrow). **c)** SDS-PAGE analysis of peak fractions from SEC of JmjN+Arid N-His (protein band indicated by a black arrow).

5.3 Protein expression and purification in Sf9 cells

As most JARID1A constructs did not show signs of soluble protein expression in *E.coli* expression systems, expression was attempted in Sf9 cells. Sf9 cells were cultured to a density of 1×10^6 cells were seeded in 35mm well of a 6-well plate and co-transfected with bacmid DNA and plasmid DNA coding for the protein of interest using the Fugene transfection kit for generation of recombinant baculovirus. After 5 days of incubation at 27.5°C, recombinant baculovirus (P1 stock) was harvested from the culture media by centrifugation. The remaining cells were resuspended in 1×PBS buffer and mixed with SDS-loading buffer followed by SDS-PAGE analysis and western blotting to check for the expression of proteins. Western blotting was performed using anti-His antibodies. Although it is not always necessary to see protein expression at this stage, 13 out of 17 clones showed expression as listed in **Figure 5.6a** and shown in the western blotting analysis of cells in **Figure 5.6b** and **Figure 5.6c.**

a)

Sr No.	Clones	Expression
1	JmjN N-His	Negative
2	JmjN C-His	Negative
3	JmjN+Arid N-His	Positive
4	PHD1 N-His	Positive
5	PHD1 C-His	Positive
6	JmjC N-His	Positive
7	JmjC C-His	Positive
8	PHD1+JmjC N-His	Positive
9	PHD1+JmjC C-His	Positive
10	Zf (short) N-His	Negative
11	Zf (short) C-His	Positive
12	Zf (long) N-His	Positive
13	Zf (long) C-His	Positive
14	JmjC+Zf N-His	Positive
15	JmjC+Zf C-His	Positive
16	PHD2 N-His	Negative
17	PHD2 C-His	Positive

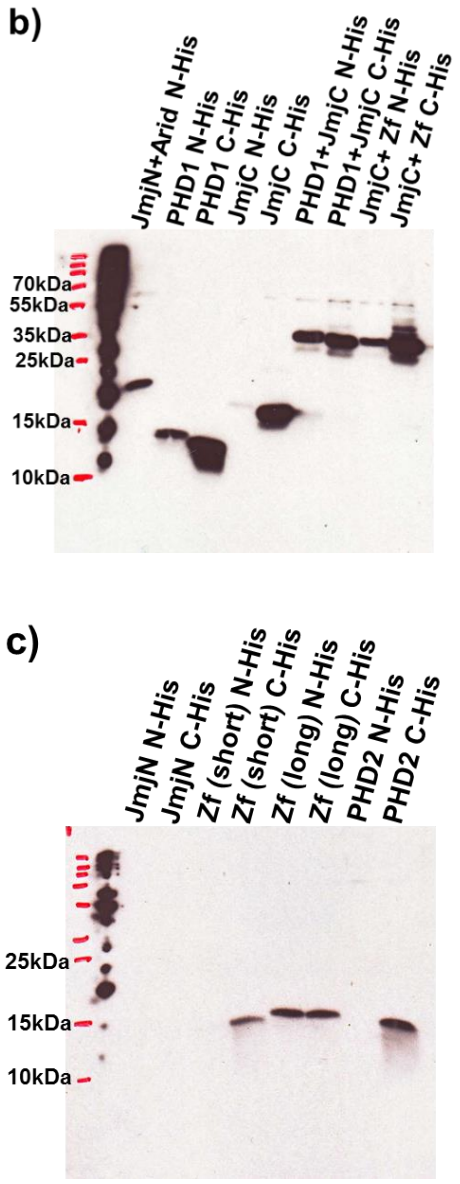
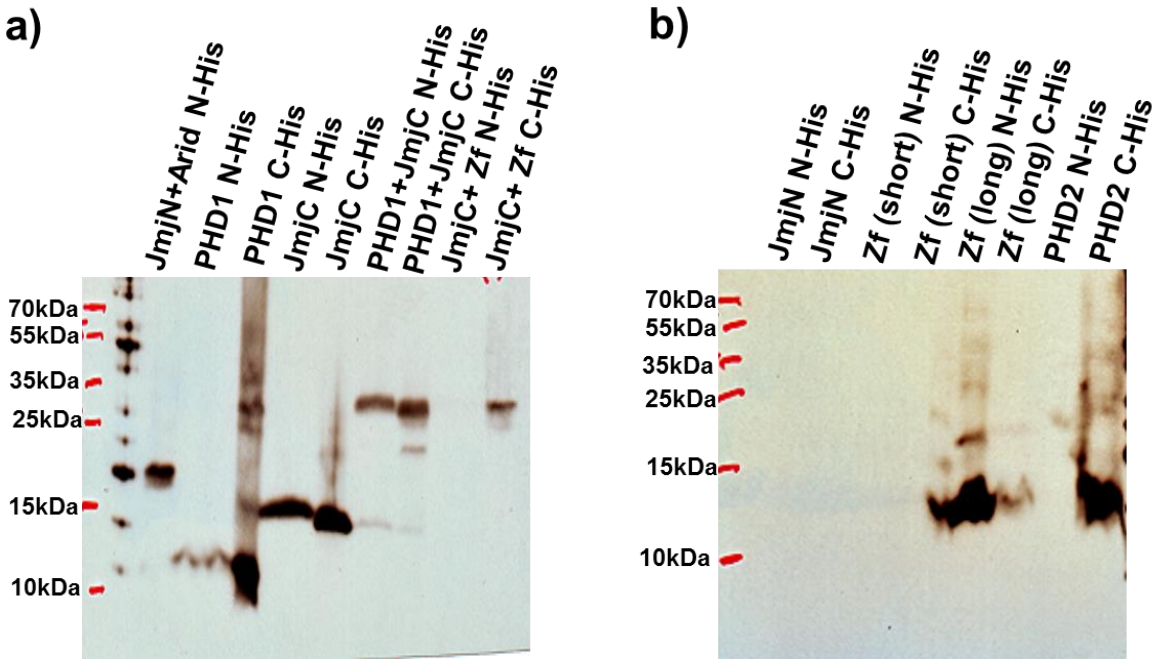


Figure 5.6: Expression test for N- and C-terminal His-tagged clones of JARID1A in SF9 cells. a) List of JARID1A clones with their expression status after first infection in SF9 cells. **b)** and **c)** Western blotting analysis on the cells at the P1 stage of infection to check expression of the JARID1A constructs.

For amplification of the recombinant baculovirus, 50ml of Sf9 suspension culture with a cell density of 2×10^6 cells/ ml was infected with 1ml of recombinant baculovirus from the first infection (P1) and incubated for three days at 27.5°C in a shaking incubator at 100 RPM. Post infection, the amplified recombinant virus from second infection (P2) was harvested by centrifugation. The remaining cells were resuspended in 1×PBS buffer and mixed with SDS-

loading buffer followed by SDS-PAGE analysis and western blotting to check for the expression of proteins. Western blotting was performed using anti-His antibodies. All clones which showed expression after the first infection also showed expression after the second infection by recombinant baculovirus as shown in **Figure 5.7a** and **Figure 5.7b**. Of these, expression of all clones except the PHD2 construct was scaled-up for protein production as listed in **Figure 5.7c**. Expression of the PHD2 construct was not scaled up as it was already optimized in the bacterial system. The protein expression and purification was carried out as described in Chapter 2, section 2.13.2.



c)

Sr No.	Clones
1	JmjN+Arid N-His
2	PHD1 N-His
3	PHD1 C-His
4	JmjC N-His
5	JmjC C-His
6	PHD1+JmjC N-His
7	PHD1+JmjC C-His
8	Zf (short) C-His
9	Zf (long) N-His
10	Zf (long) C-His
11	JmjC+Zf N-His
12	JmjC+Zf C-His

Figure 5.7: Expression test for N- and C-terminal His-tagged clones of JARID1A in SF9 cells. a) and **b)** Western blotting analysis on cells at the P2 stage of infection to check expression of JARID1A clones. **c)** List of clones whose expression was scaled-up for protein production.

Of all constructs, only PHD1 C-His expressed at appropriate levels (1.1mg per litre of culture). The SDS-PAGE analysis of fractions from affinity purification of PHD1 C-His showed that the protein was being eluted with 50mM Imidazole (**Figure 5.8b**) and hence for further purifications beads were only washed with the wash buffer containing 20mM Imidazole followed by elution with the elution buffer containing 500mM Imidazole. The eluted protein was concentrated using a 3kDa MWCO concentrator and further purified by Size exclusion chromatography using a Superdex 75 16/60 column. Protein peak fractions were pooled together and concentrated using a 3kDa MWCO concentrator. The purification of PHD1 C-His is shown in **Figure 5.8**. The expected size of PHD1 C-His is 10kDa. The SEC chromatogram profiles showed two protein peaks (**Figure 5.8a**). The first peak corresponds to aggregated proteins as it elutes close to the void volume and the second peak, indicated by a black arrow, represents mono-dispersed PHD1. SDS-PAGE analysis of samples from affinity chromatography was performed, where pellets, beads post elution, flow-through post binding, wash1 (20mM Imidazole), wash2 (50mM Imidazole), wash3 (250mM Imidazole) and elution fractions were analysed. The elution fractions contained bands of molecular weight compatible with that of the PHD1 C-His protein, as indicated by a black arrow (**Figure 5.8b**). The SDS-PAGE analysis of peak fractions obtained from SEC showed that PHD1 was eluted in the second peak. The peak fractions that were pooled together are highlighted (**Figure 5.8c**). The purified protein was used for crystallization.

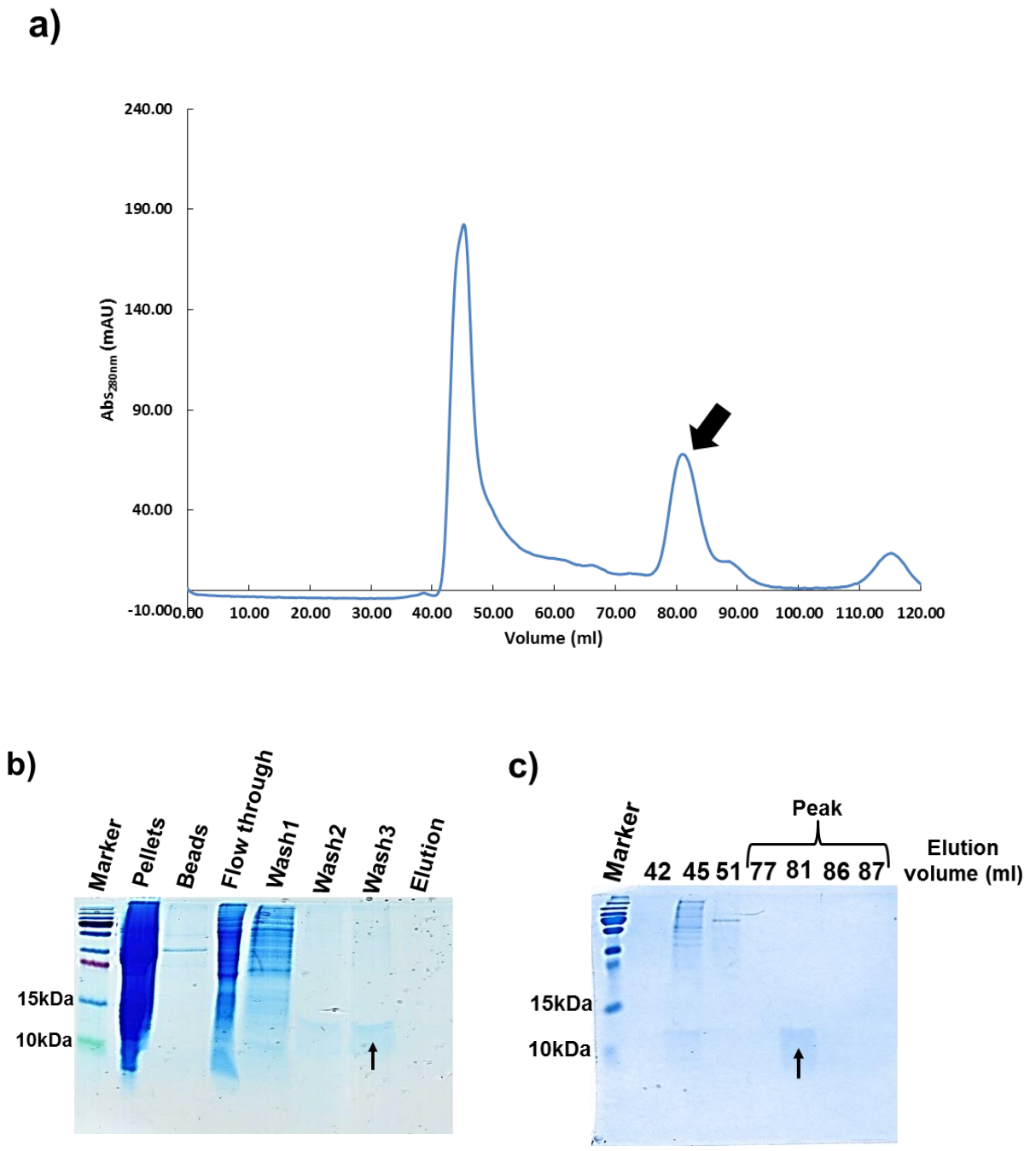


Figure 5.8: Purification of PHD1 C-His. a) Size-exclusion chromatography (SEC) of the PHD1 C-His protein with a molecular weight of 10kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of PHD1 C-His (the protein band is indicated by a black arrow). **c)** SDS-PAGE analysis of peak fractions from SEC of PHD1 C-His (protein band indicated by a black arrow).

5.4 Purification of JmjN-JmjC

GST JmjN-JmjC (N-terminal half of full length JARID1A consisting of the domains JmjN, Arid, PHD1 and JmjC) was initially cloned for GST pull-down assays. As the small-scale expression tests showed good levels of expression, we attempted to purify this construct for the purpose of crystallization. GST jmjN-jmjC was expressed in *E. coli BL21* cells by IPTG induction. The protein was purified and the GST tag was cleaved as described in Chapter 2, section 2.13.2. After tag cleavage, pellets, flow-through (post binding), wash, beads (after cleavage) and elution (after cleavage) fractions were analysed by SDS-PAGE in which the JmjN-JmjC protein is indicated by the star and the GST JmjN-JmjC protein is indicated by the horizontal black arrow (**Figure 5.9b**). However, after cleaving off the GST tag, the protein precipitated. GST tag cleavage was then attempted in presence of three different components in the cleavage buffer: high salt (700mM NaCl), 2% glycerol and 1M NDSB-201 (Non-Detergents Sulfobetaines), on their own and in all possible combinations (**Figure 5.9a**). In all the conditions, the protein precipitated without the GST tag which is highlighted by stars in **Figure 5.9c** which displays beads and elution fractions post cleavage. The precipitate was eliminated by centrifugation and was analysed by SDS-PAGE along with the supernatant. All the protein was present in the precipitated fraction in each condition as indicated by black stars in **Figure 5.9d**. Hence, crystallization could not be performed for the JmjN-JmjC construct.

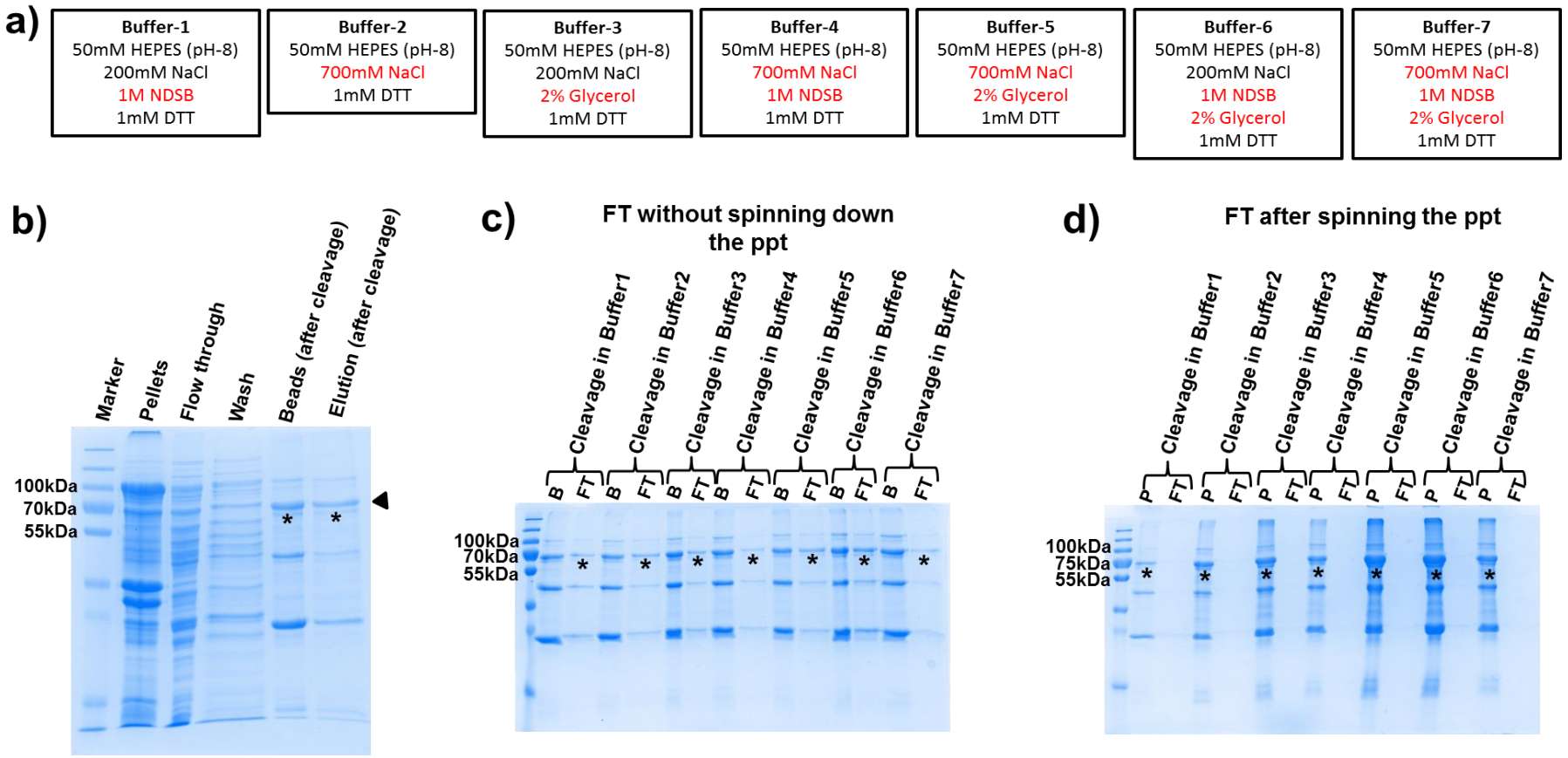


Figure 5.9: Purification of GST JmjN-JmjC. a) List of buffers used for optimization of GST tag cleavage. **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of GST JmjN-JmjC (100kDa). **c)** SDS-PAGE analysis of fractions after the GST tag cleavage in different buffers prior to the separation of precipitates (size of JmjN-JmjC is 70kDa). **d)** SDS-PAGE analysis of fractions after the GST tag cleavage in different buffers after removal of precipitates by centrifugation.

A summary of all proteins successfully purified along with their host systems is presented in **Table 5.2**.

Table 5.2: Summary of all the proteins purified.

Construct	Host system
PHD1 N-His	*E.coli*
PHD2 N-His	*E.coli*
PHD2 C-His	Sf9 insect cells

5.5 Crystallization of JARID1A domains

5.5.1 Crystallization of PHD2 domain of JARID1A

The PHD2 N-His protein was expressed and purified as described earlier in section 5.2 from *E.coli.* The purified protein was concentrated using a 3kDa MWCO concentrator to a concentration of 6.67mg/ml in buffer (50mM HEPES, pH 8, 200mM NaCl, 1mM DTT). Pre-crystallization tests were performed using Pre-crystallization test (PCT™) kit (Hampton Research) to determine an appropriate starting protein concentration. The PCT kit contains four reagents, which consists of salts and polymers. PCT gives an idea if the protein concentration is appropriate for crystallization. For example, if the protein forms heavy amorphous precipitates in all solutions, it needs to be diluted whereas if it forms light granular precipitates in any of the four solutions, it can be used for performing screens. For PCT, 100µl of each of the four reagents (A1, A2, B1, B2) were pipetted into the reservoir well. 1µl of protein was added in the crystallization well of the plate, mixed with 1µl of the reservoir and the plate was sealed. After 15 minutes, drops were observed under the microscope and the protein concentration was decided based on the 'PCT results and recommended action' table provided in the kit. For PHD2, 6.67mg/ml concentration was found to be appropriate. Six crystallization screens were performed including Block1, Block3, PACT premier, Index, SaltRx and PEGRx. These are all commercially available

sparse matrix screens from the Hampton Research and comprise of a large combination of solutions with varying conditions such as precipitant concentration, pH etc. Crystallization trays were set up using the Robbins Hydra 96 micro-dispenser (Matrix Technologies Ltd, UK) which dispenses 100µl of the precipitant solutions into the Greiner plate reservoir chambers and a Cartesian robot (Cartesian Technologies Microsys MIC4000), which dispenses 100nl of protein as a sitting drop in the central well of the raised platform followed by addition of 100nl of reservoir to each protein drop from the corresponding reservoir well. Plates were then sealed, stored at 20°C in a Homebase automated storage vault and imaged automatically on a regular schedule. Crystals were obtained after about 3 months in two conditions, however, when tested on IO2 beamline at the Diamond Light Source, they were shown to be salt crystals.

5.5.2 Crystallization of PHD1 domain of JARID1A

The PHD1 C-His protein was expressed and purified as described earlier in section 5.3 from Sf9 cells. The purified protein was concentrated using a 3kDa MWCO concentrator to a concentration of 5.26 mg/ml in buffer (50mM HEPES, pH 8, 200mM NaCl, 1mM DTT). The pre-crystallization test was performed using the PCT™ kit (Hampton Research) in the same way as that of the PHD2 domain to determine the appropriate protein concentration for crystallization. For crystallization, 5.26 mg/ml concentration was used and crystal screens Block2, Block4, NucPro and KeraFast were set up. Crystallization was carried out as described in previous section (5.6.1). However, no crystals were obtained in any of the conditions.

5.6 Crystallization of PHD2 in complex with GATA1 NFCF

Purification of the PHD2-GATA1 NFCF complex was attempted by Size-exclusion chromatography but both proteins eluted as separate peaks, thus no complex was formed. The absence of a complex under SEC conditions suggests that this interaction might be dynamic or concentration dependant. Nevertheless, crystallization was attempted as the process of crystallization itself leads to a gradual increase in the protein concentration in the drop and hence could aid in complex formation. Crystallization of the PHD2-GATA1 NFCF complex was carried out by mixing PHD2 (3.3mg/ml in 50mM HEPES, pH 8, 200mM NaCl, 1mM DTT) and GATA1 NFCF (7.7mg/ml in 50mM Tris, pH 8, 200mM NaCl, 1mM DTT) in equimolar ratio (1:1). Crystallization screens which were set up included Block1, Block2, Block3, Block4, Index, PACT Premier, $PEGR_X$, KeraFast, NucPro. Crystals were obtained after about 3 months in two conditions, however, when tested on IO2 beamline at the Diamond Light Source they were shown to be salt crystals.

Chapter 6: Expression, Purification and crystallization of SCL-LMO1 complex

6 Expression, purification and crystallization of SCL-LMO1 complex

6.1 Introduction

LMO1 is an oncogene that, like LMO2, is involved in chromosomal translocations in T-cells leading to T-ALL (Rabbitts 1998). It also forms a complex with SCL and LDB1 and these interactions have been shown to play a role in T-ALL (Valge-Archer et al. 1994; Valge-Archer et al. 1998). Previously, we have successfully assembled *in vitro* and solved the crystal structure of a complex consisting of (SCL: E47) bHLH: LMO2:$LDB1_{LID}$. The crystal structure gave useful insights into how LMO2 plays a crucial role in sequestration of E47 by SCL thereby blocking the normal progression of T cell differentiation and causing T-cell leukaemia (El Omari et al. 2013). Since LMO1-SCL interaction has also been implicated in T-cell malignancies (Valge-Archer et al. 1994; Rabbitts 1998), and LMO1/LMO2, share more than 50% sequence identity, it would be interesting to determine if there are any differences in the way in which LMO1 interacts with SCL when compared to LMO2. To investigate this, crystallization of the LMO1-SCL complex was carried out. In parallel, crystallization of LMO1 on its own was also attempted with the aim of comparing its structure with that of LMO2 (El Omari et al. 2011). Comparison of the two structures could give us insights into their functional differences.

6.2 Expression and purification of LMO1

LMO1 was cloned as FLINC1 as described in Chapter 4, section 4.4.1. and expressed in *E. coli Rosetta* cells, which were cultured in LB media using IPTG induction. The protein expression and purification was carried out as described in Chapter 2, section 2.13.2 except

that different lysis buffer (50mM HEPES, pH 8, 500mM NaCl, 20mM Imidazole, 34µl of sigma protease inhibitor, 12.5 units/L of DNaseI and a pinch of lysozyme), wash buffer (50mM HEPES, pH 8, 500mM NaCl, 30/50mM Imidazole) and elution buffer (50mM HEPES,pH 8, 200mM NaCl, 500mM Imidazole) were used. The FLINC1 protein was purified from 6L cell pellets which were disrupted in a cell disruptor at 30 kpsi at 4°C. Elution fractions from affinity chromatography were concentrated using a 10kDa MWCO concentrator and further purified by Size exclusion chromatography (SEC) using a Superdex-75 16/60 column (GE Healthcare). The protein peak fractions were pooled together and concentrated using a 10kDa MWCO concentrator.

The purification of FLINC1 is shown in **Figure 6.1**. SEC of FLINC1 displayed two protein peaks (**Figure 6.1a**). The first protein peak contains aggregated FLINC1 as it is eluting just after the void volume of the column (40ml) and the second peak contains mono-dispersed FLINC1, which is indicated by a black arrow. The SDS-PAGE analysis of samples from affinity chromatography is represented by **Figure 6.1b,** where pellets (after disruption), beads post elution, flow-through post binding, wash1 (30mM Imidazole), wash2 (50mM Imidazole), elution1 (first concentrated fraction) and elution2 (rest of the eluted fractions) fractions were examined. The SDS-PAGE analysis of peak fractions from SEC showed that the second peak represents the mono-dispersed FLINC1 protein (**Figure 6.1c**). The FLINC1 peak fractions that were pooled together are highlighted.

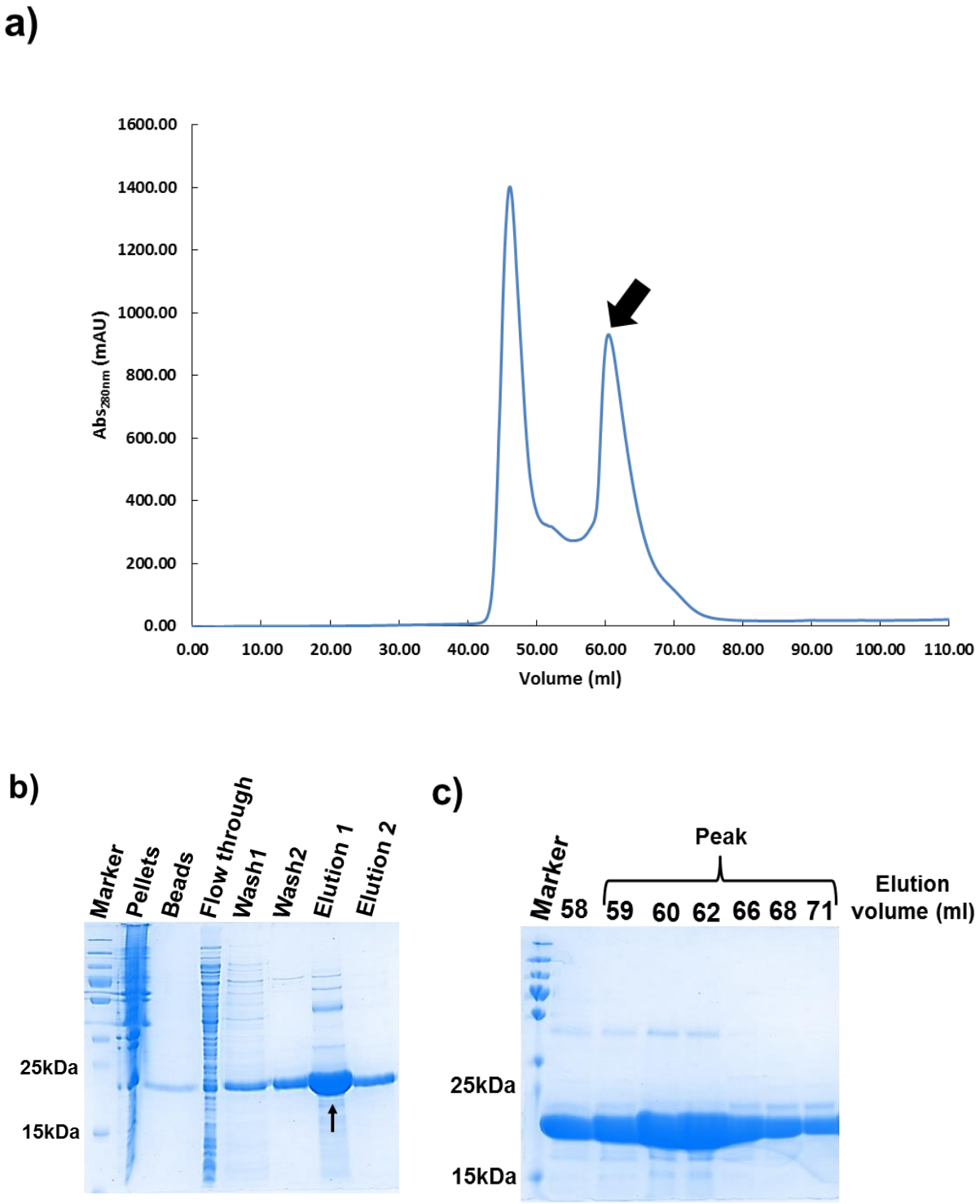


Figure 6.1: Purification of FLINC1. a) Size- exclusion chromatography (SEC) of the FLINC1 protein with a molecular weight of 20kDa (the black arrow represents the protein peak). **b)** SDS-PAGE analysis of samples from different stages of affinity chromatography purification of FLINC1 (protein band indicated by a black arrow). **c)** SDS-PAGE analysis of peak fractions from SEC of FLINC1.

6.3 Crystallization of FLINC1

The FLINC1 protein was expressed and purified as described earlier in section 6.2 from *E.coli.* The purified protein was concentrated using a 10kDa MWCO concentrator to a concentration of 8.1mg/ml in buffer (50mM HEPES, pH 8, 200mM NaCl, 1mM DTT). A pre-crystallization test was performed using the PCT™ kit (Hampton Research) in order to determine an appropriate protein concentration, as described in Chapter 5, section 5.5.1. Following this, nine crystallization screens were performed including Block1, Block2, Block3, Block4, Index, PEGRx, NucPro, KeraFast and Morpheus. Crystallization was set up using a Cartesian robot (Cartesian Technologies Microsys MIC4000) and stored at 20°C in a Homebase automated storage vault and imaged as described in Chapter 5, section 5.5.1. Crystals were obtained in PEGRx, Morpheus, Block3 and KeraFast conditions. Only the PEGRx and KeraFast conditions however gave crystals suitable for diffraction (**Figure 6.2**).

These crystals were tested for diffraction on IO4 beamline at the Diamond Light Source. Unfortunately, only diffraction of 12Å could be obtained from crystals obtained from the PEGRx condition.

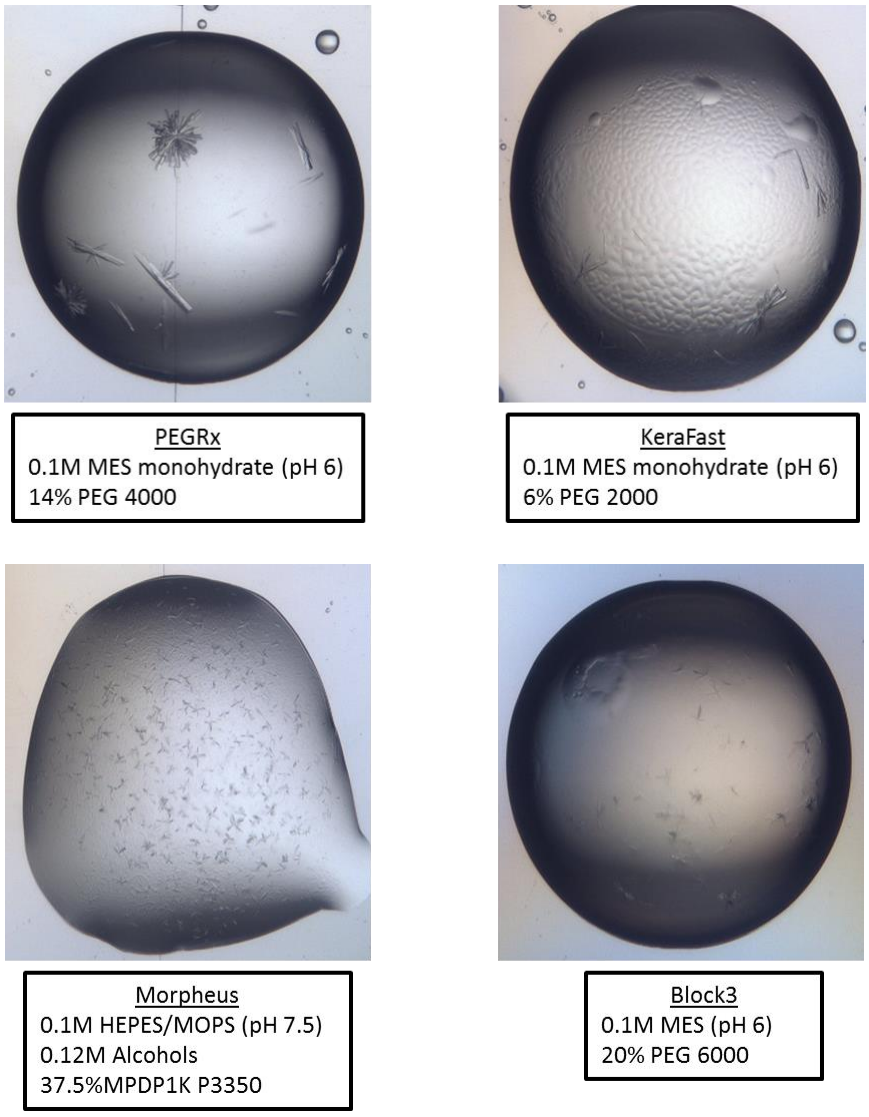


Figure 6.2: Crystallization of FLINC1. Crystallization of FLINC1 was carried out in various conditions of which crystals were obtained in PEGRx, KeraFast, Morpheus and Block3 screens

All these conditions were subjected to further optimization to obtain bigger crystals. Optimization included additive screen, different protein: reservoir ratios (1:1, 2:1 and 1:2) and different dilutions of the reservoir. However, none of the conditions produced any crystals except the PEGRx condition with 8% Pentaerythyritol ethoxylate (3/4 EO/OH) (**Figure 6.3**). The PEGRx condition was further optimized by varying the screen across various PEGs, NaCl concentrations and pHs. The following screens were performed:

1) PEG4k (8-20%) Vs pH (4-9)
2) Seeding
3) PEG4k (0-20%) Vs NaCl (0-0.5M)
4) PEG4k (0-20%) Vs NaCl (0-0.5M) with additive
5) PEG400 (0-20%) Vs NaCl (0-0.5M)
6) PEG1.5k (0-20%) Vs NaCl (0-0.5M)
7) PEG6k (0-20%) Vs NaCl (0-0.5M)
8) PEG10k (0-20%) Vs NaCl (0-0.5M)

In screens 3-8, 0.1M MES (pH 6), which was present in the original PEGRx condition, was added to all wells. Interestingly, crystals were obtained in screens 5, 6, 7 and 8 in the form of needles, which grew only in 0.1M MES (pH 6) without any PEG or NaCl (**Figure 6.3**). These crystals were very unstable and dissolved back in the solution possibly due to the lack of any precipitant. Hence, they were frozen immediately within two days of their growth. Crystals were also obtained in screens 5 and 8 in 0.1M MES (pH 6) +42-84mM NaCl. All crystals were tested for diffraction on beamline IO4 at the Diamond Light Source and they diffracted to 9.1Å as shown in **Figure 6.4**.

Each diffraction spot in the diffraction pattern obtained from a crystal corresponds to a wave consisting of an amplitude and a phase. The X-ray detector can only measure intensities of the diffracted X-rays but not phases. The latter contain important information required for the computation of electron densities, which in turn is essential for the determination of the protein structure to atomic resolution. There are a few different methods available for phase determination one of which is by utilizing anomalous scattering. Most of the elements absorb as well as emit X-rays. The absorption of X-rays by an element decreases just below its

characteristic emission wavelength. This is called absorption edge of the element. When the wavelength of X-rays is close to the absorption edge of an element, it exhibits anomalous scattering. However, only heavier atoms contribute to anomalous scattering as the absorption edge of the lighter elements, such as carbon, oxygen, or nitrogen, does not fall in the range of wavelength of X-rays. When the wavelength of X-ray is close to the absorption edge of the heavy atom, a fraction of the radiation is absorbed by the heavy atom and reemitted with altered phase.

As LMO1 is a zinc finger protein, we wanted to take advantage of the presence of the endogenous Zinc (Zn) and solve the structure by single anomalous wavelength (SAD) methods. In order to determine the peak wavelength, an X-ray absorption-edge scan was recorded close to the absorption edge of Zn (**Figure 6.4a**). Once the peak absorption wavelength was determined to be 1.28Å, diffraction data was collected at that wavelength. The diffraction pattern obtained from LMO1 crystal is shown in **Figure 6.4b** where the diffraction spot of 9.1Å resolution is indicated by a black arrow. To improve on the diffraction limit, we attempted to grow bigger crystals by increasing the size of the crystal drop, however this was not successful.

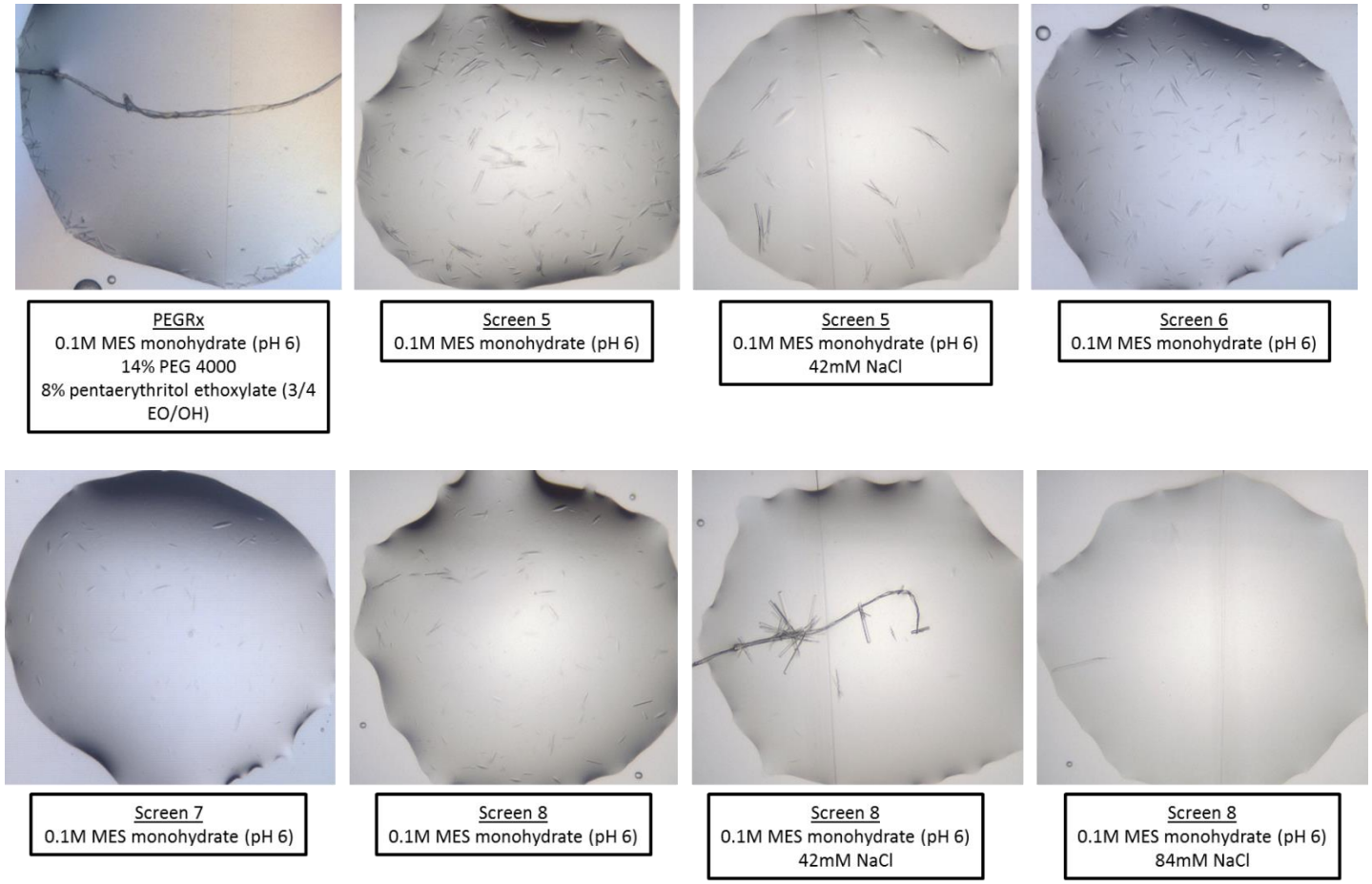
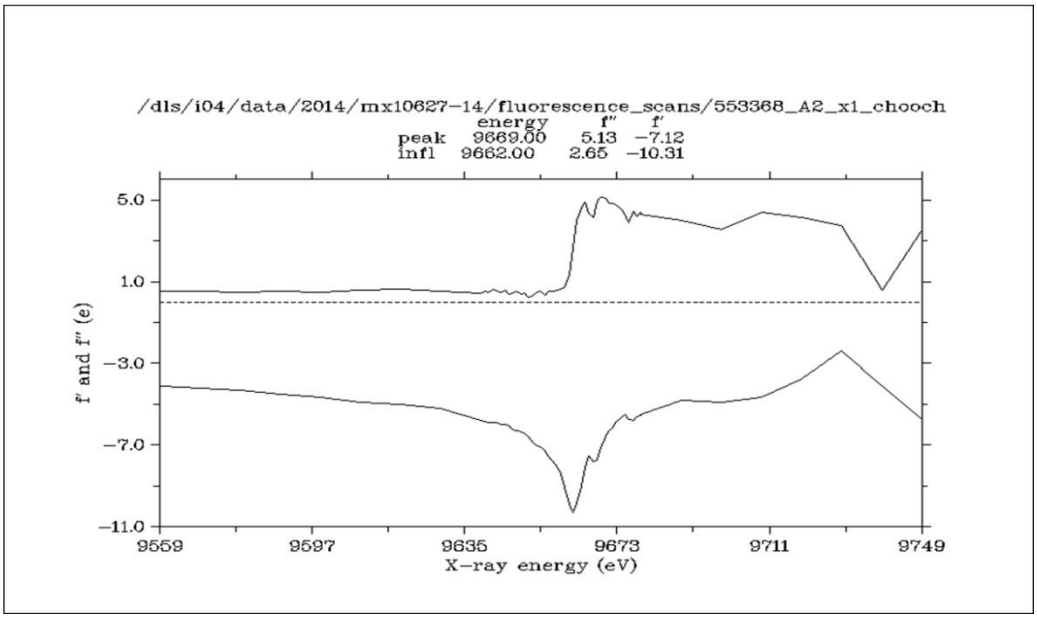


Figure 6.3: Optimization of FLINC1 crystallization. PEGRx condition for FLINC1 crystals was optimized by additive screen and varying PEGs (0-20%), NaCl (0-0.5M) and pHs, of which crystals in the form of needles were obtained with additive and only 0.1M MES (pH 6).

a)



b)

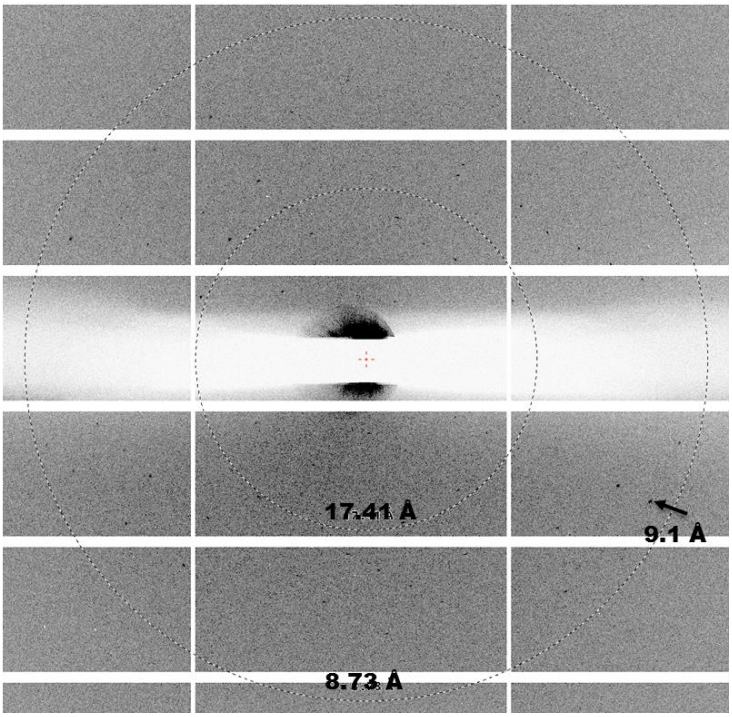


Figure 6.4: Diffraction of FLINC1: a) An anomalous scan of FLINC1 crystals was performed to determine the peak absorption wavelength of Zn. Moreover, the absorption peak of Zn obtained in the zinc scan confirmed that the crystal was likely to be containing FLINC1, as it is a zinc finger protein with four Zn atoms. b) Diffraction image of the FLINC1 crystal which diffracted to 9.1Å. The dotted rings represent the resolution limits. The diffraction spot is indicated by a black arrow.

6.4 Crystallization of SCL-FLINC1 complex

To study the complex of FLINC1 and SCL, constructs E47/SCL bHLH and FLINC1 were purified as described in Chapter 4, section 4.4.1 and section 6.2 of this chapter respectively. For complex formation, proteins were mixed in 1:1 ratio and incubated overnight at 4°C. Next day, the mixture was incubated for 45 minutes at room temperature. Incubation was followed by SEC of the mixture on a Superdex-75 16/60 column. Superimposition of chromatograms from SEC of FLINC1, E47/SCL bHLH and FLINC1+E47/SCL bHLH (hereafter referred to as FLINC1+bHLH) showed that there was a shift in the peak towards lower elution volumes with FLINC1 eluting at 67.29ml, E47/SCL bHLH at 72.7ml and FLINC1+bHLH at 63.48ml thereby indicating that a complex was formed between E47/SCL bHLH and FLINC1 proteins (**Figure 6.5**, **top**). SDS-PAGE analysis of peak fractions from SEC of FLINC1+bHLH confirms the presence of both proteins: FLINC1 (20kDa) and E47/SCL bHLH (19kDa) (**Figure 6.5**, **bottom**).

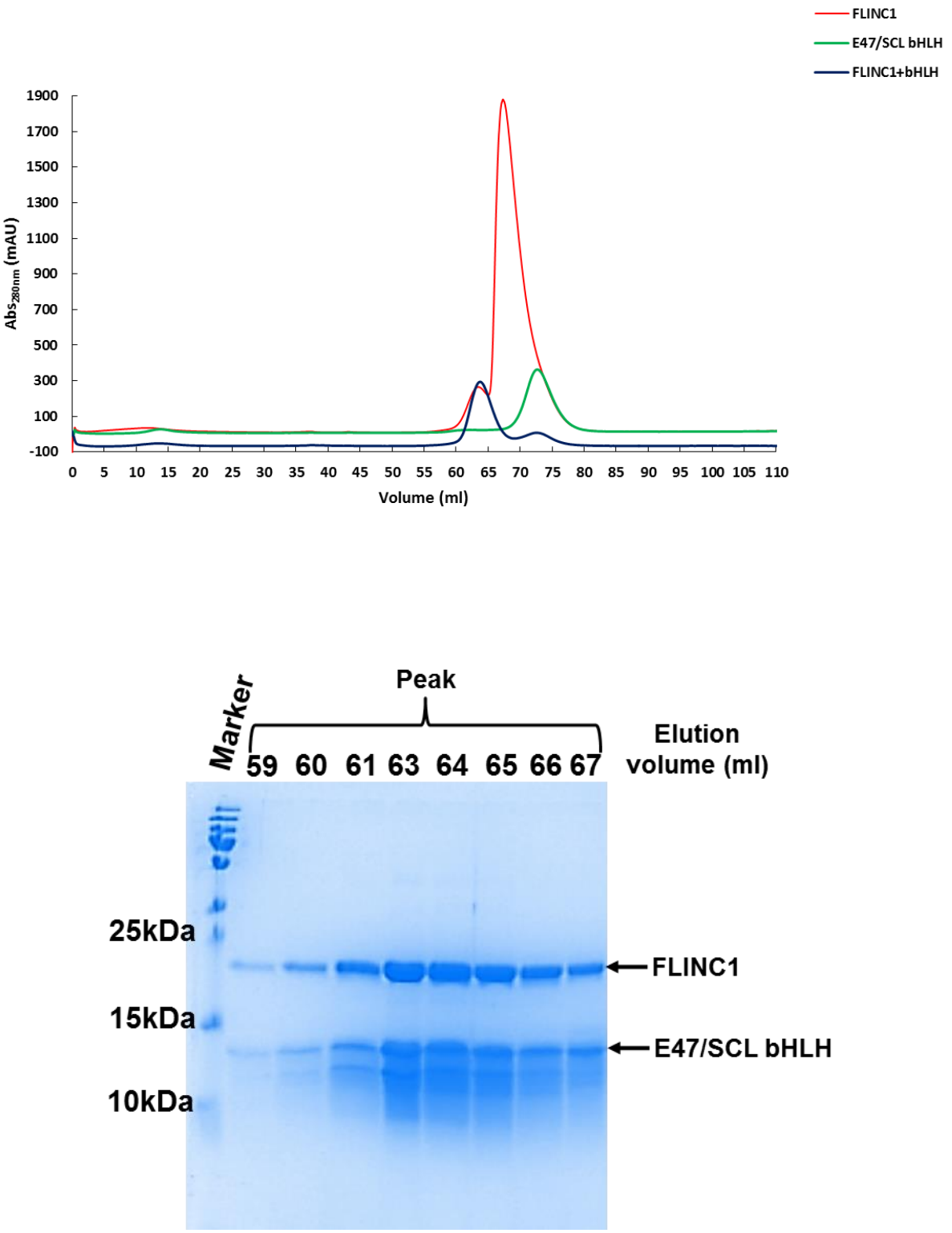


Figure 6.5: Size exclusion chromatography of FLINC1, E47/SCL bHLH and FLINC1+E47/SCL bHLH. SEC chromatograms of FLINC1 (red), E47/SCL bHLH (green) and FLINC1+bHLH (blue) proteins were superimposed. The shift in the elution volume of the complex towards lower edge confirmed complex formation between FLINC1 and E47/SCL bHLH (top). SDS-PAGE analysis of peak fractions from SEC of FLINC1+bHLH is shown in the gel at bottom where both proteins are indicated.

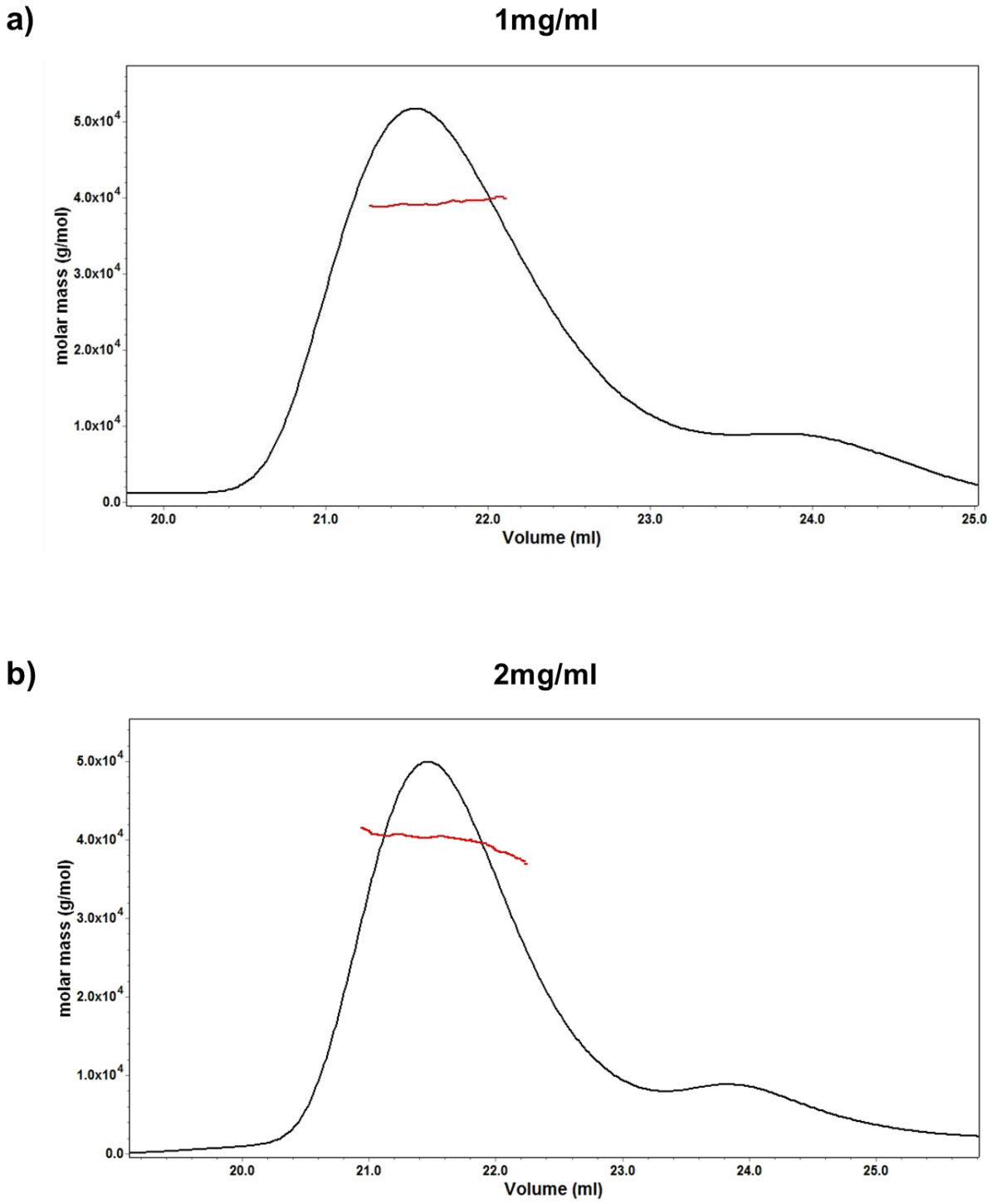


Figure 6.6: Multi-angle light scattering (MALS) of FLINC1-E47/SCL bHLH complex. **a)** MALS of the FLINC1+bHLH complex at 1mg/ml concentration. **b)** MALS of the FLINC1+bHLH complex at 2mg/ml concentration. Samples were run through a Superdex-75 10/300 size exclusion column. The UV_{280} absorbance identifies protein elution (black) and points under the peak (red) correspond to the molar mass identified by MALS.

Complex formation was further confirmed by Multi-angle light scattering (MALS). MALS was carried out in flow mode meaning that the MALS detector was coupled downstream to SEC. MALS is a technique commonly used for the determination of the absolute molar mass of particles in solution. In MALS, a protein sample is loaded onto SEC and the polarized laser light is incident on the protein eluting from the SEC column. The light scattered by particles in the protein solution is simultaneously detected by several detectors fixed at different angles. The intensity of the scattered light at each angle is proportional to the molar mass and concentration of the protein in solution.

MALS was performed using a Superdex-75 10/300 column with an injection volume of 90µl. First the column was equilibrated in buffer [50mM HEPES (pH 8), 200mM NaCl, 1mM DTT]. Ovalbumin with a molecular weight of 44kDa was run on the column as a standard to determine the elution volume at which the complex could be expected to elute. FLINC1+bHLH was run at two different concentrations, 1mg/ml (**Figure 6.6a**) and 2mg/ml (**Figure 6.6b**) to check that the complex formation is not concentration dependant. At both the concentrations, a species with a molecular weight of 40kDa was detected which is close to the expected molecular weight of the FLINC1+bHLH complex (39kDa). In both figures, X-axis represents the elution volume in ml and Y-axis represents the molar mass in daltons. The black curve shows the UV_{280} indicating protein elution and the red line corresponds to the molar mass identified by MALS (**Figure 6.6**).

For crystallization, 8.2mg/ml of FLINC1+bHLH complex was mixed with DNA (E-box) in 1:1.2 protein: DNA ratio and incubated at RT for 1 hour prior to crystallization. Previously, lab colleagues Olga Platonova and Sam Mallinson had managed to crystallize

FLINC1+bHLH+E-box complex in one of the conditions of Morpheus crystallization screen (**Figure 6.7**). The crystal diffracted to 3.9Å but poor completeness of the data prevented determination of the structure to atomic resolution. I carried out optimization of this condition using additive screen, different protein: reservoir ratios (1:1, 1:2, 2:1 and 3:1) and different dilutions of the reservoir. However, no protein crystals could be obtained.

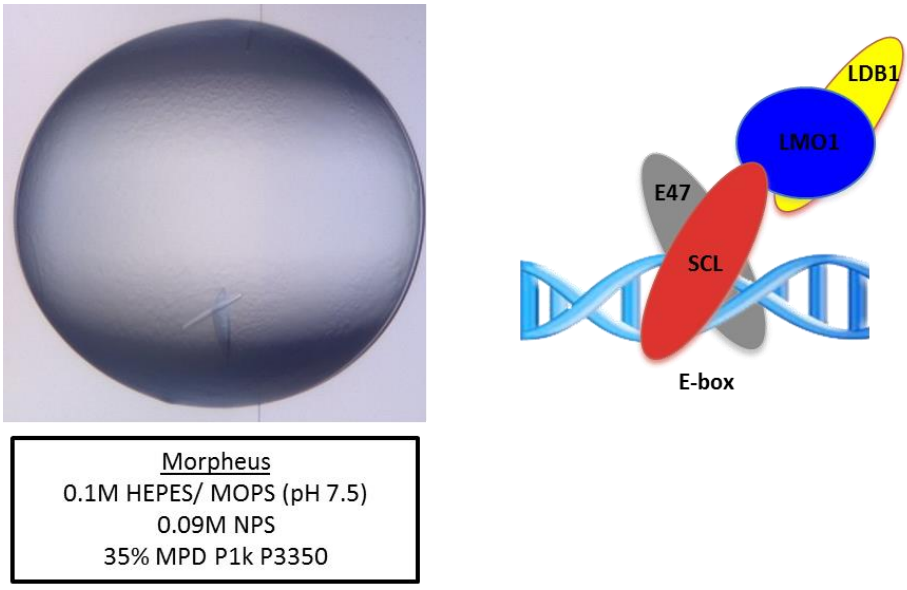


Figure 6.7: Optimization of FLINC1+bHLH+E-box crystallization. The Morpheus condition in which the first form of FLINC1+bHLH+E-box crystals were obtained is shown on the left and the schematic diagram representing FLINC1+bHLH is shown on the right.

Chapter 7: Conclusion and Future work

7 Conclusion and Future work

Several cellular processes are governed by the action of multiprotein complexes. Regulation of gene expression is one such process which is controlled by multiprotein complexes involving transcription factors, cofactors and chromatin remodeling proteins. In erythroid cells, the SCL multiprotein complex comprised of SCL-E47-LMO2-LDB1-GATA1 plays a key role in the regulation of expression of haematopoietic-specific genes. It does so by recruiting various cofactors and chromatin remodelers such as ETO2, mSin3A, P300, PCAF and LSD1. Earlier studies suggested that a histone demethylase, JARID1A, could have a function in normal haematopoiesis as well as at the onset of leukaemia. Moreover, an interaction between JARID1A and one of the members of the SCL complex, LMO2, was described in erythroid cells. The work presented in this thesis explores the possible interactions between the chromatin remodeling protein JARID1A and the oncogenic DNA-binding complex nucleated by the transcription factor SCL, using a combination of biochemical, functional, biophysical and crystallographic methods. In this concluding chapter, I will summarise and discuss the work carried out in this thesis and also the future work.

7.1 Interaction of endogenous JARID1A with the SCL multiprotein complex

To determine if JARID1A binds to the SCL multiprotein complex, various techniques were employed such as Size-exclusion chromatography (SEC) and Co-immunoprecipitation (Co-IP). First, protein fractionation experiments were carried out using SEC in which nuclear extracts obtained from Mouse ErythoLeukaemia (MEL) cells were used. MEL cells are undifferentiated erythroid precursors which serve as a valuable model for studying the erythropoiesis process *in vitro* as upon induction of differentiation by DMSO, these cells

undergo biochemical and morphological changes that closely resemble normal erythroid differentiation (Friend et al. 1966).

SEC of MEL cell nuclear extracts showed JARID1A co-eluting in high molecular weight protein fractions together with SCL and components of the SCL complex suggesting that it could be a part of this complex (See Chapter 3, Figure 3.1).

The potential JARID1A and the SCL complex interactions were further examined by Co-IP assays. JARID1A IP protocol was optimized using three different commercial antibodies (See Chapter 3, Figure 3.2a) and different ratios of Protein G beads to nuclear extracts (See Chapter 3, Figure 3.2b). Once an optimal ratio was obtained, JARID1A Co-IP was performed and the co-immunoprecipitated proteins were subjected to western blotting using antibodies against SCL, LMO2, LDB1 and GATA1. Since the molecular weight of heavy chain of antibodies is similar to that of SCL, LDB1 and GATA1, JARID1A antibodies were cross-linked to Protein G beads in order to reduce the background signal. Cross-linking did help in the reduction of the background but at the same time it also reduced the efficiency of JARID1A IP. Moreover, it also resulted in the appearance of a non-specific band in the negative control (See Chapter 3, Figure 3.4). To maintain optimal IP efficiency, antibody cross-linking was eliminated and Protein G beads were blocked using salmon sperm DNA and BSA, and nuclear extracts pre-cleared prior to IP. This helped with the reduction of the background but despite working efficiently in immunoprecipitating JARID1A, the monoclonal antibodies failed to co-immunoprecipitate SCL, GATA1, LDB1 or LMO2. This could have been because of the monoclonal antibodies blocking the binding site on JARID1A. So, the purified polyclonal antibodies, which were initially rejected, were checked

for their ability to co-immunoprecipitate the SCL complex components. These successfully and reproducibly pulled down JARID1A from nuclear extracts along with SCL and GATA1 but not LDB1 or LMO2 (See Chapter 3, Figure 3.5).

To further detail interactions between JARID1A and the SCL complex, reverse Co-IPs were performed using antibodies against SCL, LDB1, GATA1 and LMO2. SCL antibodies reproducibly co-immunoprecipitated JARID1A but LDB1 antibodies did not, as expected. In case of LMO2, a faint band was detected in the JARID1A blot in the IP lane, however, LMO2 was not detected in the JARID1A Co-IP (See Chapter 3, Figure 3.6). In the past JARID1A has been shown to interact with LMO2 by Co-IP using radiolabelled proteins but bands observed were very weak (Mao et al. 1997). The weak band of JARID1A observed in our LMO2 IP could be because antibodies were used for detection of proteins, which are less sensitive than radioactivity. It could also mean that not all the JARID1A protein present in MEL cells could bind to LMO2 or probably the expression levels of JARID1A is much lower than that of LMO2 and so there is not enough JARID1A to bind all the LMO2. The fact that no LMO2 could be pulled down by JARID1A antibodies could be because of technical issues or the interaction is very weak and hence could not be detected by LMO2 antibodies. This could be overcome by labelling LMO2 with radioactive isotope and repeating the assay as radiolabelling also solves the issue of non-specific binding of antibodies.

The GATA1 IP required optimization, hence, three different GATA1 antibodies and two Protein G beads to nuclear extracts ratios were examined (See Chapter 3, Figure 3.8). Once the optimal conditions were obtained, GATA1 Co-IP was performed which effectively and reproducibly pulled down JARID1A. Thus, it can be concluded that JARID1A interacts with

SCL and GATA1 but not LDB1. Although JARID1A may interact with LMO2, this requires further investigation. Therefore, there is a possibility that these regulatory/ adaptor proteins (LMO2 and LDB1) might be dispensable for the role of JARID1A in modulating functions of transcription factors SCL and GATA1 in erythropoiesis. Furthermore, *Jarid1a, Scl* and *Gata1* also showed similar expression pattern during different stages of haematopoiesis in mouse bone marrow and fetal liver in line with the possibility of these proteins interacting (See Chapter 3, Figures 3.13 and 3.14).

GATA1 is a transcription factor which plays a critical role in erythroid differentiation and GATA1 null mice display a block in erythropoiesis and mouse embryos die at E10.5 (Pevny et al. 1995; Fujiwara et al. 1996). It has been shown that the interaction between GATA1 and the retinoblastoma protein (pRB) is essential for normal erythroid differentiation. However, GATA1 interacts with pRB only in the presence of E2F-2, a transcription factor that plays a crucial role in the control of cell cycle. The GATA1/pRb/E2F-2 trimeric complex induces cell cycle arrest during differentiation thereby inhibiting proliferation of erythroid precursors by altering G1 to S phase progression (DeGregori 2002; Kadri et al. 2009). JARID1A has also been shown to interact with pRB (Benevolenskaya et al. 2005), so it is possible that JARID1A interacts with the GATA1/pRb/ E2F-2 complex and establishes a repressive state on the cell cycle genes promoter through its demethylase activity thereby inducing cell cycle arrest and promoting erythroid differentiation. Moreover, GATA1 associates with a transcriptional coactivator CBP which is a HAT and inhibition of this interaction leads to abrogation of expression of GATA1 target genes like globin (Blobel et al. 1998). So, perhaps during the early stages of haematopoiesis, GATA1 recruits JARID1A to silence globin genes which is then replaced by CBP following differentiation resulting in activation of globin

genes in differentiated erythroid cells. In this way, JARID1A could play a role in erythroid differentiation via its interaction with GATA1.

SCL also plays a critical role in erythropoiesis and SCL null mice die during embryogenesis due to lack to red blood cells (Porcher et al. 1996). It has been reported that SCL represses transcription of the target genes by altering the H3K4 methylation pattern on the promoter. It does so by recruiting LSD1, which is an H3K4 demethylase, like JARID1A. Specifically, in undifferentiated MEL cells which are erythroid precursors, SCL interacts with LSD1 possibly to repress SCL target genes that promote cell proliferation. The interaction occurs via serine 172 in SCL, which in the sequence of the protein just precedes the bHLH domain region. Hence, the demethylase activity of LSD1 negatively regulates SCL mediated transcription in erythroid cells (Hu et al. 2009a). This interaction is dynamic in nature, which might serve a regulatory function during haematopoiesis (Li et al. 2012). JARID1A is also a H3K4 demethylase and could be interacting with SCL in a similar way but with different interacting residues and affinities. This could give rise to further complexity in the negative regulation of SCL during haematopoiesis, thereby adding a spatial or temporal dimension to the process. It remains to be shown whether LSD1 and JARID1A are part of the same SCL complex at the same stage of haematopoiesis.

In summary, SCL and GATA1 may be recruiting JARID1A to carry out repression of cell cycle genes by modulating H3K4 methylation levels in erythroid precursors, thus inducing erythroid differentiation for the production of RBCs. These interactions could be regulating the expression of globin genes as both SCL and GATA1 play crucial roles in terminal erythroid differentiation.

7.2 Determination of direct interactions between JARID1A and the SCL complex

After establishing the existence of interactions between endogenous JARID1A and SCL/ GATA1, it was important to verify if these interactions were direct as this could shed light on how the complex is formed and how it is regulated during haematopoiesis. Initially, full-length JARID1A protein was divided into two parts: JmjN-JmjC (N-terminal) and Zf-PHD3 (C-terminal) based on disorder prediction by RONN software (Yang et al. 2005). Another construct consisting of the third PHD domain (PHD3) of JARID1A was made as it had been reported to interact with LMO2 (Mao et al. 1997). Mammalian two-hybrid assays were performed using all three JARID1A's constructs with SCL, LMO2, E47 and GATA1. However, none of the constructs showed interaction with the SCL complex components. This could be due to the requirement of other proteins which could be mediating these interactions.

Despite the negative initial results, we decided to further our analysis with a different technique, GST pull-down assays. For GST pull-down assay, the N- and C- terminal constructs of JARID1A were tagged with a GST tag and GST pull-down assays were performed with *in vitro* transcribed and translated (IVTT) SCL, E47, LMO2, LDB1 and ETO2. Although ETO2 is not a part of the SCL core complex, had previously been shown to interact with SCL in erythroid cells (Schuh et al. 2005), and so was also screened for its interaction with JARID1A. However, IVTT of GATA1 did not work and hence it could not be tested. Of all proteins tested, SCL and LDB1 showed interaction with JARID1A and with both N- and C- terminal constructs (See Chapter 4, Figures 4.7 and 4.8). Therefore, a battery of constructs containing individual domains of JARID1A fused to GST tag was designed. Three double domain constructs were also designed for domains that did not have any

unstructured regions between them (See Chapter 4, Figure 4.11). All constructs were expressed in *E. coli Rossetta* cells and purified by affinity chromatography. Of all the constructs tested, Arid, JmjN+Arid, Zf and all three PHD domains expressed very well (See Chapter 4, Figure 4.11). Of these, the JmjN+Arid construct showed interaction with LDB1 and all three PHD domains showed interaction with both SCL and LDB1 (See Chapter 4, Figures 4.12, 4.13, 4.14, 4.15). Overall the PHD domains, which are in general believed to be responsible for mediating protein-protein interactions, were shown to be involved in the interaction of JARID1A with the SCL complex.

To further confirm these interactions, Analytical Ultracentrifugation (AUC) was performed with all three PHD domains of JARID1A and the SCL complex components. The E47/ SCL bHLH, FLINC2 and GATA1 NFCF constructs (See Chapter 4, Figure 4.19, 4.20 and 4.21) were used instead of full-length SCL, LMO2 and GATA1 as the latter are very unstable and hence cannot be purified. All proteins were purified by two step purification including affinity and gel filtration chromatography. Initially, His-tagged proteins were used in which case the PHD2 domain of JARID1A showed interaction with GATA1 NFCF (See Chapter 4, Figure 4.22). A combination of E47/SCL bHLH and FLINC2 was used as positive control (See Chapter 4, Figure 4.23). Since the expression levels of His-tagged PHD1 and PHD3 proteins were very low and not enough for AUC, these were expressed as GST-tagged proteins followed by tag cleavage. However, both proteins were unstable without the tag. So, we decided to use the GST-tagged PHD constructs were used for further AUC experiments. Again, GST PHD2 interacted with GATA1 NFCF, further confirming the interaction observed previously by co-IP (See Chapter 3, Figure 3.8 and Chapter 4, Figure 4.28). Hence, it was confirmed that PHD2 domain of JARID1A interacts with GATA1.

JARID1A is part of the JARID1 family of demethylases, sharing 52% sequence identity to JARID1B as well as a similar domain organization. JARID1B is known to bind its substrate through its PHD fingers. More specifically, its first and third PHD fingers (not the second) bind to H3K4me0 and H3K4me3 marks respectively (Klein et al. 2014). Similarly, the first and third fingers of JARID1A also bind to H3K4me0 and H3K4me3 marks respectively (Wang et al. 2009; Upadhyay et al. 2011). It is therefore possible that the second PHD fingers of these proteins do not possess histone-binding capability and instead are involved in protein-protein interaction. This would be in line with our data, which suggests an interaction between PHD2 of JARID1A and GATA1.

It remains to be determined how JARID1A interacts with the SCL complex. Considering the direct interactions observed between JARID1A and GATA1 NFCF, it is likely that JARID1A is recruited to the SCL complex via GATA1 NFCF. The possibility persists however that a second direct interaction occurs between JARID1A and SCL. We have previously generated SCL mutants, which are disrupted in their ability to bind LMO2 or E47 (El Omari et al. 2013). These mutants could be used to confirm if JARID1A is recruited to the SCL complex via SCL or GATA1.

A summary of the interactions obtained between JARID1A and the SCL complex components using Co-immunoprecipitation, GST pull-down assay and AUC is presented in **Table 7.1**.

Table 7.1: Summary of interactions obtained between JARID1A and the SCL complex components.		
Co-immunoprecipitation		
JARID1A	SCL	Positive
JARID1A	GATA1	Positive
JARID1A	LMO2	Negative
JARID1A	LDB1	Negative
GST Pull-down assay		
JmjN	SCL	Negative
Arid	SCL	Negative
JmjN+Arid	SCL	Negative
PHD1	SCL	Positive
JmjC	SCL	Positive
PHD1+JmjC	SCL	Positive
Zf (short)	SCL	Negative
Zf (long)	SCL	Negative
JmjC+Zf	SCL	Positive
PHD2	SCL	Positive
PHD3	SCL	Positive
JmjN	LDB1	Negative
Arid	LDB1	Negative
JmjN+Arid	LDB1	Positive
PHD1	LDB1	Positive
JmjC	LDB1	Positive
PHD1+JmjC	LDB1	Positive
Zf (short)	LDB1	Negative
Zf (long)	LDB1	Negative
JmjC+Zf	LDB1	Positive
PHD2	LDB1	Positive
PHD3	LDB1	Positive
AUC		
PHD2	FLINC2	Negative
PHD2	GATA1	Positive
PHD2	E47/SCL bHLH	Negative
GST PHD1	FLINC2	Negative
GST PHD1	GATA1	Negative
GST PHD1	E47/SCL bHLH	Negative
GST PHD2	FLINC2	Negative
GST PHD2	GATA1	Positive
GST PHD2	E47/SCL bHLH	Negative
GST PHD3	FLINC2	Negative
GST PHD3	GATA1	Negative
GST PHD3	E47/SCL bHLH	Negative

7.3 Expression, Purification and crystallization of JARID1A

Apart from studying interactions between JARID1A and the SCL complex, we also investigated the JARID1A protein on its own. For this, individual domains of JARID1A and three double domain constructs with 6x His tag at N- and C- terminal were designed using the high through-put In-Fusion cloning (See Chapter 4, Figure 4.11). Following successful cloning, these constructs were screened for expression using high-throughput screening pipeline at OPPF. Two different *E. coli* strains, *BL21 Lemo and Rosetta*, and two induction systems, IPTG induction and auto-induction, were tested (See Chapter 5, Figures 5.1 and 5.2). Arid, JmjN+Arid, PHD1 and PHD2 constructs showed expression which was further optimized for scale-up using different media. Following optimization, the PHD2 protein was purified (See Chapter 5, Figure 5.4) and used for AUC and co-crystallization with GATA1 NFCF. PHD1 protein expression levels were however very low (270µg protein from pellet of 6L culture) and hence it could not be used for crystallization. The expression of JmjN+Arid construct also resulted in expression of a contaminant protein co-eluting with the protein of interest in gel filtration (See Chapter 5, Figure 5.5). Elimination of the contaminant protein was attempted by using more stringent washes and ion-exchange chromatography with no success. Hence, the crystallization of JmjN+Arid could not be performed.

Since expression of most of the JARID1A constructs failed in the bacterial system, expression tests were carried out in SF9 cells. 13 out of 17 constructs showed expression (See Chapter 5, Figure 5.7) and 12 constructs' expression was scaled-up. However, only the PHD1 construct yielded enough soluble protein for crystallization purpose (1.1mg per litre of culture) (See Chapter 5, Figure 5.8).

Furthermore, JmjN-JmjC (N terminal half of JARID1A full length protein comprising of JmjN, Arid, PHD1 and JmjC domains) which was initially cloned as GST fused protein for GST pull-down assay showed promising levels of expression. Hence, we attempted to purify this protein and to cleave the GST tag with the aim of using it for crystallization. However, the protein became unstable and precipitated upon cleavage of the GST tag. Several attempts were made to stabilize the protein by performing tag cleavage in the buffer containing stabilizing agents such as glycerol, NDSB and high NaCl concentration. But none of these conditions helped to stabilize the protein and as a result crystallization of JmjN-jmjC construct could not be performed.

Crystallization of PHD1, PHD2 and PHD2+GATA1 NFCF was carried out using the commercially available screens. PHD2-GATA1 NFCF complex could not be purified by size exclusion chromatography which could possibly mean that the interaction is dynamic in nature or concentration dependant. As crystallization is a process leading to gradual increase in the protein concentration, it could help in complex formation. So, crystallization was carried out by mixing PHD2 and GATA1 NFCF proteins in equimolar ratio. PHD2 and PHD2+GATA1 NFCF yielded some crystals but when tested on IO2 beamline at Diamond Light Source, they turned out to be salt crystals. PHD1 construct did not result in the formation of any crystals.

New constructs for the PHD domains could be generated which could give better expression and may produce protein crystals. Alternatively, NMR could be used to determine the structure of PHD1 and PHD2 proteins. The structure of ARID and PHD3 domains of JARID1A have been determined using NMR (Tu et al. 2008; Wang et al. 2009).

7.4 Expression, Purification and Crystallization of SCL-LMO1 complex

While studying the connection between JARID1A and the SCL complex, we also initiated studies of a different variation of the SCL complex itself containing LMO1 instead of LMO2 protein. In the past we have successfully solved crystal structure of the SCL-LMO2 complex which gave us an understanding of the mechanism by which this complex could be functioning in T-ALL (El Omari et al. 2013). Like LMO2, LMO1 is also involved in T-ALL and also interacts with SCL and LDB1 (Valge-Archer et al. 1994; Rabbitts 1998; Valge-Archer et al. 1998). This led us to investigate the difference of SCL binding to LMO1, when compared to LMO2, by solving the structure of the SCL-LMO1 complex. In parallel, crystallization of the LMO1 protein on its own was also attempted in order to compare its structure with that of LMO2 (El Omari et al. 2011)**.** Expression and purification of FLINC1 (both LIM domains of LMO2 fused to LID domain of LDB1) from *E. coli* yielded milligram quantities of protein (4mg from 6L culture) (See Chapter 6, Figure 6.1). Crystallization was then carried out first by performing a coarse screen using commercially available screens which resulted in small needle shaped crystals in four screens: PEGRx, Morpheus, Block3 and KeraFast (See Chapter 6, Figure 6.2). Crystals from PEGRx condition diffracted the best of all to 12Å. These crystallization conditions were further subjected to optimization. In the fine screening, only PEGRx condition yielded crystals but again in the form of thin needles. Hence, further optimization was attempted by screening different PEGs and NaCl concentrations. This resulted in formation of slightly bigger crystals (See Chapter 6, Figure 6.3) which displayed improved diffraction when tested on IO4 beamline at Diamond Light Source with highest resolution being 9.1Å (See Chapter 6, Figure 6.4). More optimization is required to improve the resolution of the diffraction.

To obtain the SCL-LMO1 complex, the individual proteins were purified as E47/SCL bHLH and FLINC1 constructs and complex formation was carried out. The complex was successfully purified using size exclusion chromatography (See Chapter 6, Figure 6.5). Complex formation was further confirmed by performing Multi-angle light scattering that determined the molecular weight of the complex to be 40kDa, in line with the theoretical molecular weight of the complex. In the past, our group members had managed to obtain crystals of FLINC1- E47/SCL bHLH complex in one of the conditions in the Morpheus screen (See Chapter 6, Figure 6.7). Those crystals diffracted to 3.9Å, however, poor completeness of the diffraction data was an obstacle in obtaining the structure of the complex. In line with this, I carried out optimization of the potential condition that yielded crystals at first. Unfortunately, none of the optimization screens produced any crystals.

7.5 Final Remarks and Future Work

It is well known that the chromatin remodeling proteins coordinate with transcription factors to accomplish regulation of target gene expression. The work presented in this thesis focuses on the investigation of interactions between a chromatin remodeling protein, JARID1A and the oncogenic DNA binding multiprotein complex nucleated by a transcription factor, SCL. In erythroid cells, JARID1A interacts with SCL and GATA1. And more precisely, the second PHD domain of JARID1A interacts directly with GATA1. Since both SCL and GATA1 play an indispensable role in the process of erythropoiesis, this interaction could play an important part in regulation of the process. However, to ascertain this, it is essential to determine if knock-down of JARID1A has any effects on erythropoiesis. This could be achieved by knock-down of JARID1A using shRNA in MEL cells followed by induction of differentiation. Since none of our efforts were fruitful in knock-down of JARID1A, more

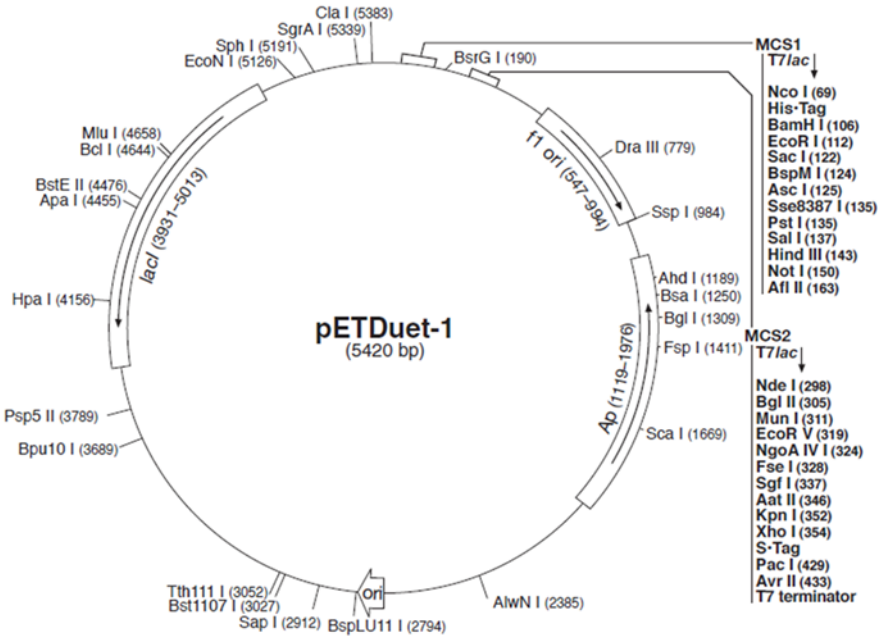
siRNAs need to be screened. Alternatively, CRISPR/Cas9 technology can be used to achieve the same.

Furthermore, ChIP-sequencing using antibodies against JARID1A from erythroid cells would reveal the target genes which are controlled by JARID1A. And comparison of this set with the data obtained from ChIP-sequencing of SCL (Kassouf et al. 2010) would also disclose if there are any common target genes, thereby further aiding us in understanding how these proteins work together.

Appendices

Appendix-I: Vector Maps

pETDuet-1



pET-15b

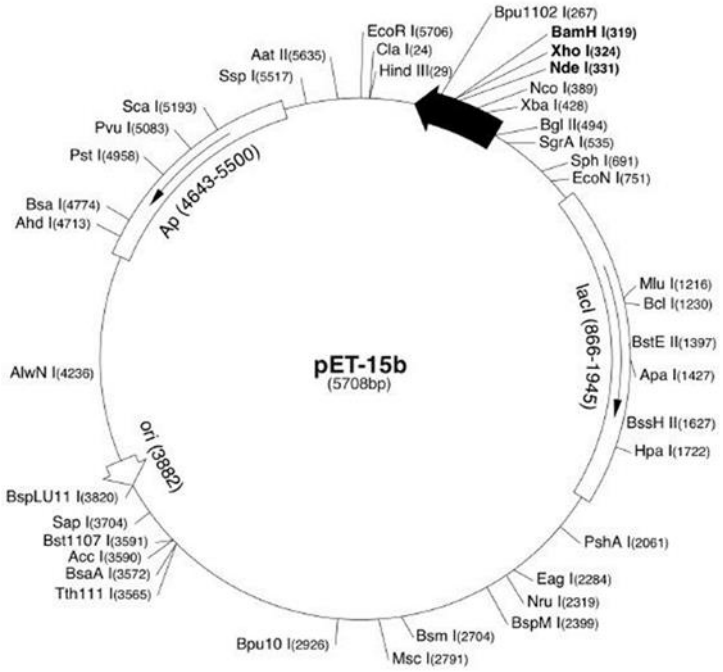
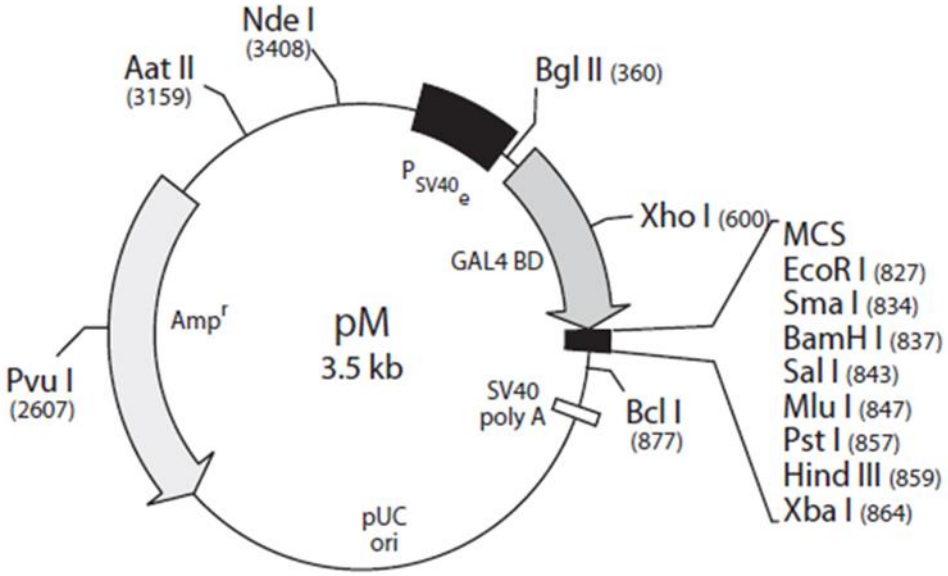


Figure A.7.1: Vector maps of pETDuet-1 and pET-15b. E47-SCL bHLH, LMO1 and LMO2 constructs were cloned into pETDuet-1 vector (top) using Nco-1 and EcoRI restriction sites. GATA1 NFCF construct was cloned into pET-15b (bottom) utilizing NdeI and EcoRI restriction sites.

pM (Gal4 vector)



pVP16

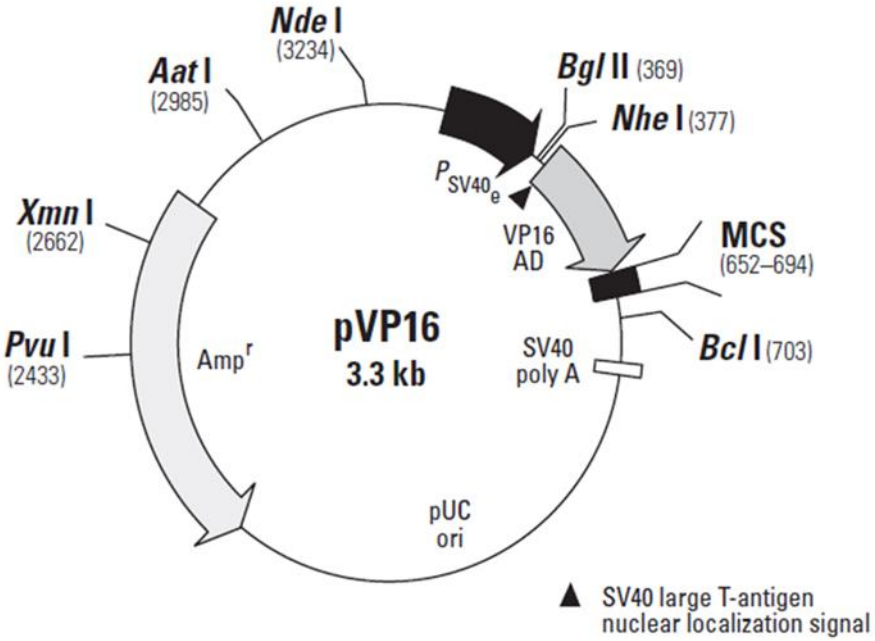
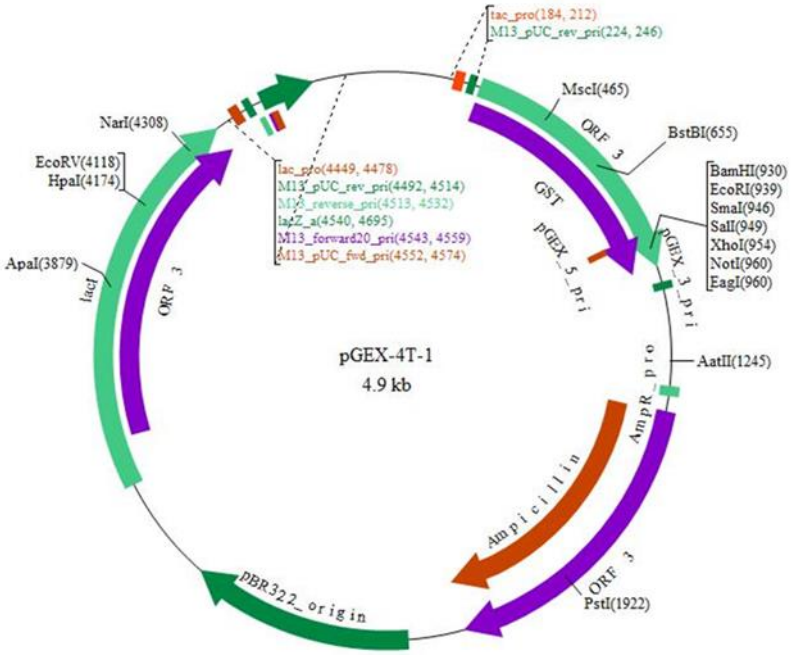


Figure A.7.2: Vector maps of pM and pVP16. N- and C- terminal parts of JARID1A full length protein along with SCL, E47, LMO2 and GATA1 were cloned into pM (top) and pVP16 (bottom) vectors utilizing EcoRI and XbaI restriction sites for mammalian two-hybrid assay. Multiple cloning site (MCS) in pVP16 vector is same as that of pM vector.

pGEX-4T-1



pGEX-6P-1

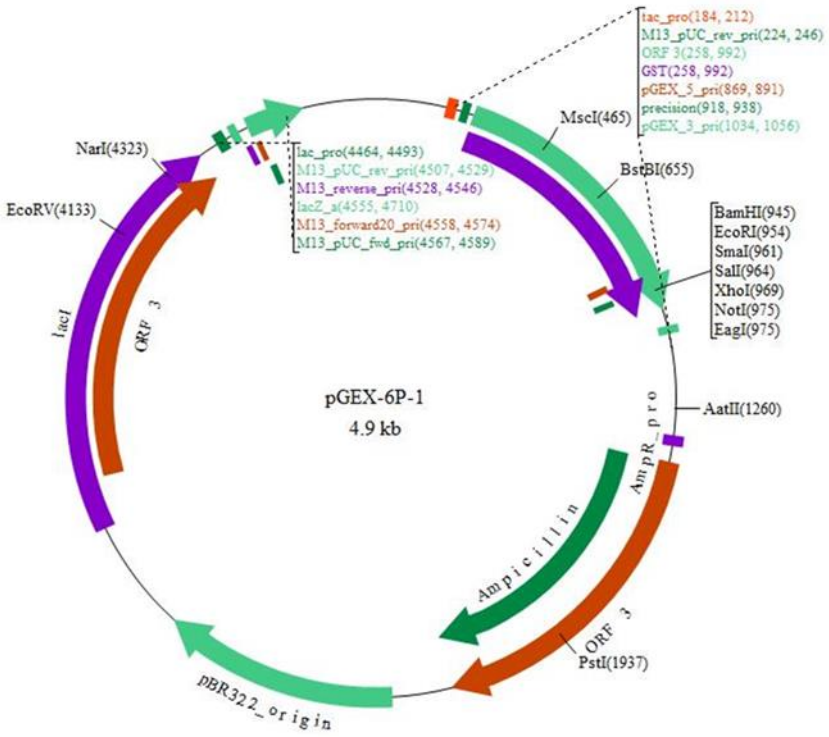
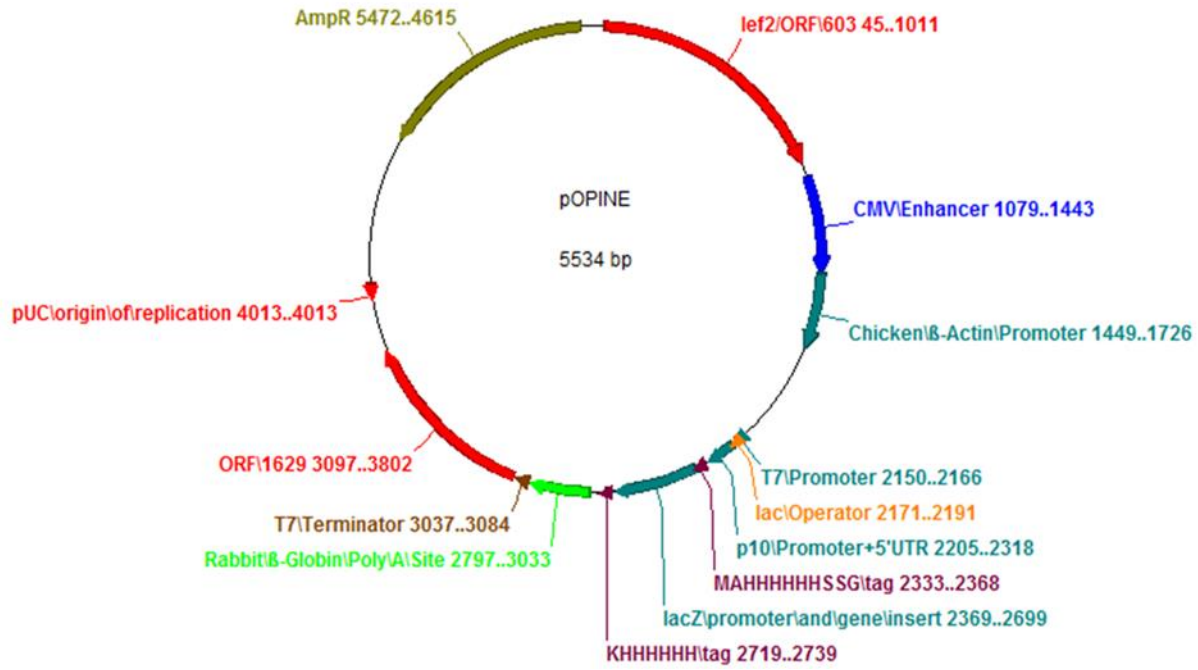


Figure A.7.3: Vector maps of pGEX-4T-1 and pGEX-6P-1. JmjN-JmjC (N-terminal half of JARID1A) construct was cloned into pGEX-4T-1 (top) using EcoRI and XhoI restriction sites. The construct was then cloned into pGEX-6P-1 (bottom) using same restriction sites in order to utilize 3C protease for GST tag cleavage post protein purification as opposed to thrombin cleavage site in pGEX-4T-1.

pOPINE



pOPINF

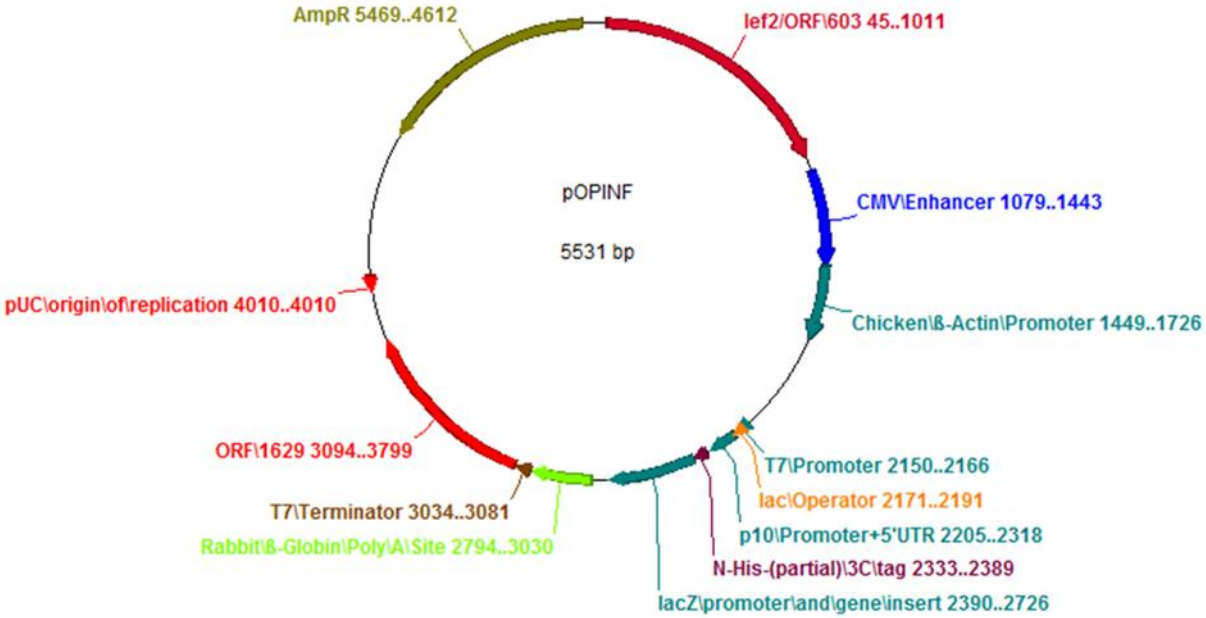


Figure A.7.4: Vector maps of pOPINE and pOPINF. JARID1A constructs were cloned into pOPINE (top) and pOPINF (bottom) vectors using In-fusion cloning for C- and N-terminal His tags respectively.

pOPINJ

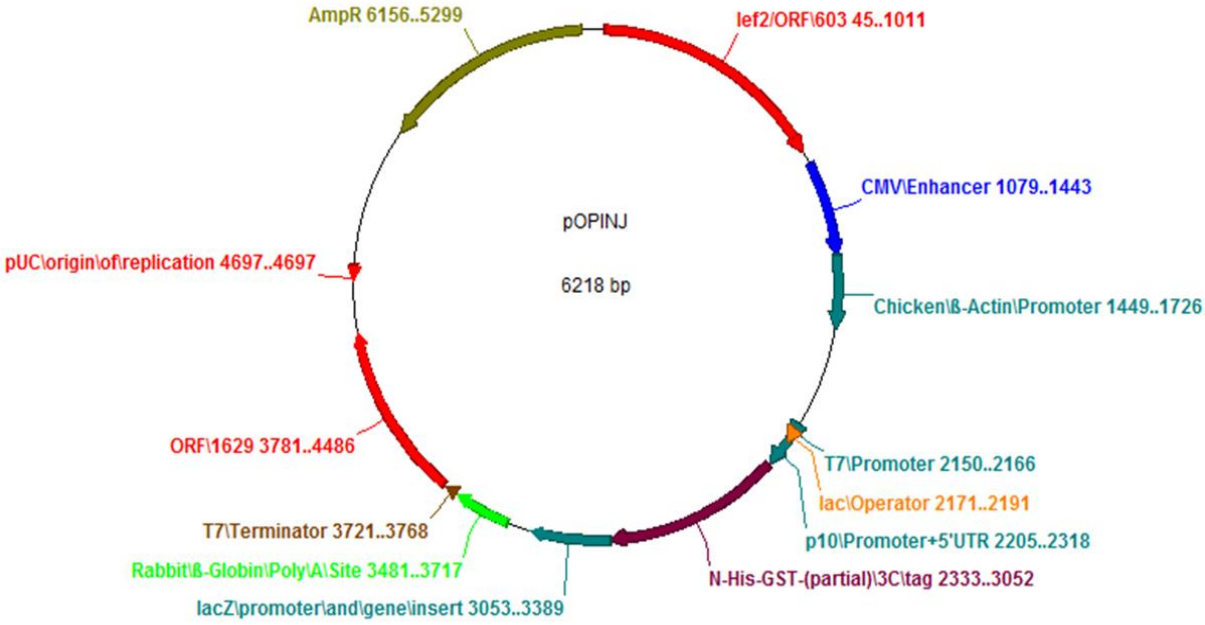


Figure A.7.5: Vector map of pOPINJ. JARID1A constructs were cloned into pOPINJ vector using In-fusion cloning for N- terminal GST tag.

Appendix-II: List of JARID1A constructs

Table A.7.2: List of JARID1A constructs. All the constructs were cloned into pOPINF, pOPINE and pOPINJ vectors respectively for N-, C- terminal His tags and N- terminal GST tag respectively.

Sr No.	Construct	Protein Sequence
1	JmjN	PECPVFEPSWEEFTDPLSFIGRIRPFAEKTGICKIRPPKDWQPPFACEVKTFRFTPRVQRLN E
2	Arid	EAMTRVRLDFLDQLAKFWELQGSTLKIPVVERKILDLYALSKIVASKGGFEIVTKEKKWSK VGSRLGYLPGKGTGSLLKSHYERILYPYELFQSG
3	JmjN+Arid	PECPVFEPSWEEFTDPLSFIGRIRPFAEKTGICKIRPPKDWQPPFACEVKTFRFTPRVQRLN ELEAMTRVRLDFLDQLAKFWELQGSTLKIPVVERKILDLYALSKIVASKGGFEIVTKEKKWS KVGSRLGYLPGKGTGSLLKSHYERILYPYELFQSG
4	PHD1	RKGTLSVNFVDLYVCMFCGRGNNEDKLLLCDGCDDSYHTFCLLPPLPDVPKGDWRCPKC VAEECNKPREAFGFEQAVRE
5	JmjC	EQSVLAHINVDISGMKVPWLYVGMCFSSFCWHIEDHWSYSINYLHWGEPKTWYGVPS HAAEQLEEVMRELAPELFESQPDLLHQLVTIMNPNVLMEHGVPVYRTNQCAGEFVVTF PRAYHSGFNQGYNFAEAVNFCTADWLPIGRQCVNHYRRLRRHCVFS
6	PHD1+JmjC	QRKGTLSVNFVDLYVCMFCGRGNNEDKLLLCDGCDDSYHTFCLLPPLPDVPKGDWRCP KCVAEECNKPREAFGFEQAVREYTLQSFGEMADNFKSDYFNMPVHMVPTELVEKEFW RLVSSIEEDVIVEYGADISSKDFGSGFPKKDGQRKMLPEEEEYALSGWNLNNMPVLEQSV LAHINVDISGMKVPWLYVGMCFSSFCWHIEDHWSYSINYLHWGEPKTWYGVPSHAAE QLEEVMRELAPELFESQPDLLHQLVTIMNPNVLMEHGVPVYRTNQCAGEFVVTFPRAY HSGFNQGYNFAEAVNFCTADWLPIGRQCVNHYRRLRRHCVFS
7	Zf (short)	DDERQCSACRTTCFLSALTCSCNPERLVCLYHPTDLCSCPMQNKCLRYRYPLEDLPSLLYG VKVR
8	Zf (long)	DDERQCSACRTTCFLSALTCSCNPERLVCLYHPTDLCSCPMQNKCLRYRYPLEDLPSLLYG VKVRAQSYDTWVNRVTEALSASFNHKK
9	JmjC+Zf	EQSVLAHINVDISGMKVPWLYVGMCFSSFCWHIEDHWSYSINYLHWGEPKTWYGVPS HAAEQLEEVMRELAPELFESQPDLLHQLVTIMNPNVLMEHGVPVYRTNQCAGEFVVTF PRAYHSGFNQGYNFAEAVNFCTADWLPIGRQCVNHYRRLRRHCVFSHEELIFKMAADP ECLDVGLAAMVCKELTLMTEEETRLRESVVQMGVVMSEEEVFELVPDDERQCSACRTT CFLSALTCSCNPERLVCLYHPTDLCSCPMQNKCLRYRYPLEDLPSLLYGVKVRAQSYDTW VNRVTEALSASFNHKK
10	PHD2	ERIEEVKFCICRKTASGFMLQCELCKDWFHNSCVPLPKSSSQKKGSSWQAKDVKFLCPLC MRSRRPR
11	PHD3	AVCAAQNCQRPCKDKVDWVQCDGGCDEWFHQVCVGVSAEMAENEDYICINCA

Appendix-III: List of SCL complex components constructs

Table A.7.3: List of SCL complex components constructs. First three constructs were cloned in pETDUet-1 vector and fourth construct was cloned in pET15b vector. Both vectors possess His tag at the N-terminal.

Sr No.	Construct	Protein Sequence
1	E47 bHLH SCL bHLH	MGSSHHHHHHSQDPSLEEKDLRDRERRMANNARERVRVRDINEAFRELGRMCQLHLKS DKAQTKLLILQQAVQVILGLEQQVRERNLNPKAA MADLQISDGPHTKVVRRIFTNSRERWRQQNVNGAFAELRKLIPTHPPDKKLSKNEILRLAM KYINFLAKLLNDQEEEGTQR
2	LMO1	MGSSHHHHHHSQDGKQKGCAGCNRKIKDRYLLKALDKYWHEDCLKCACCDCRLGEVGS TLYTKANLILCRRDYLRLFGTTGNCAACSKLIPAFEMVMRARDNVYHLDCFACQLCNQRFC VGDKFFLKNNMILCQMDYEEGQLNGTGGSGGHMGSGGDVMVVGEPTLMGGEFGDED ERLITRLENTQFDAANGIDDE
3	LMO2	MGSSHHHHHHSQDPSLLTCGGCQQNIGDRYFLKAIDQYWHEDCLSCDLCGCRLGEVGRR LYYKLGRKLCRRDYLRLFGQDGLCASCDKRIRAYEMTMRVKDKVYHLECFKCAACQKHFCV GDRYLLINSDIVCEQDIYEWTKINGGGSGGHMGSGGDVMVVGEPTLMGGEFGDEDERLI TRLENTQFDAANGIDDE
4	GATA1 NFCF	MGSSHHHHHHSSGLVPRGSHMARECVNCGATATPLWRRDRTGHYLCNACGLYHKMNG QNRPLIRPKKRMIVSKRAGTQCTNCQTTTTTLWRRNASGDPVCNACGLYFKLHQVNRPLT MRKDGIQTRNRKASGKG

Appendix-IV: Analysis of multi-protein complexes by size exclusion chromatography

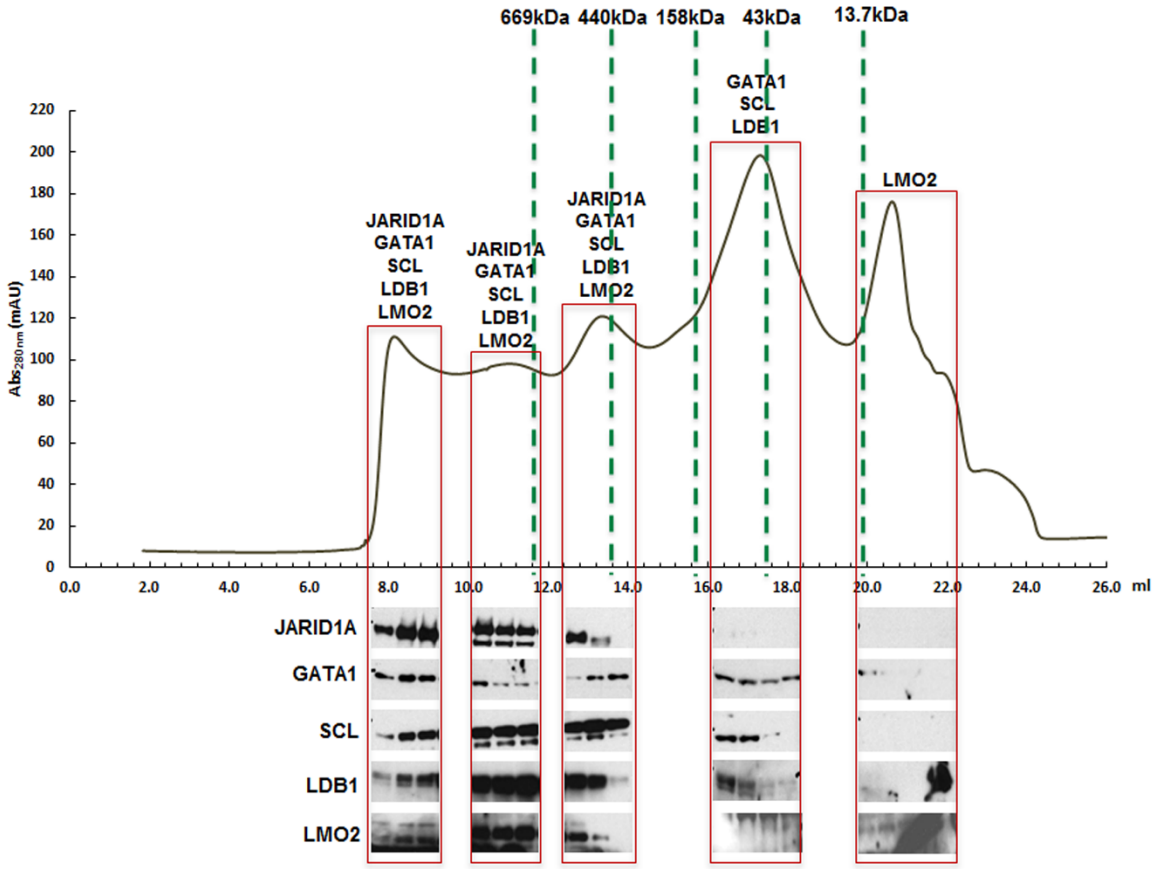


Figure A.7.6: Size exclusion chromatography. MEL cell nuclear extracts were fractionated by size exclusion chromatography and the eluted fractions were Tri-chloro acetic acid precipitated and analysed by western blotting. The X-axis represents UV absorption of proteins at 280nm wavelength and the Y-axis represents the elution volume in ml. Proteins detected at different elution volumes are indicated at the top of the blue boxes. The green vertical dotted lines highlight the elution volumes at which the standard molecular weight markers corresponding to molecular weights indicated at the top eluted under same experimental conditions.

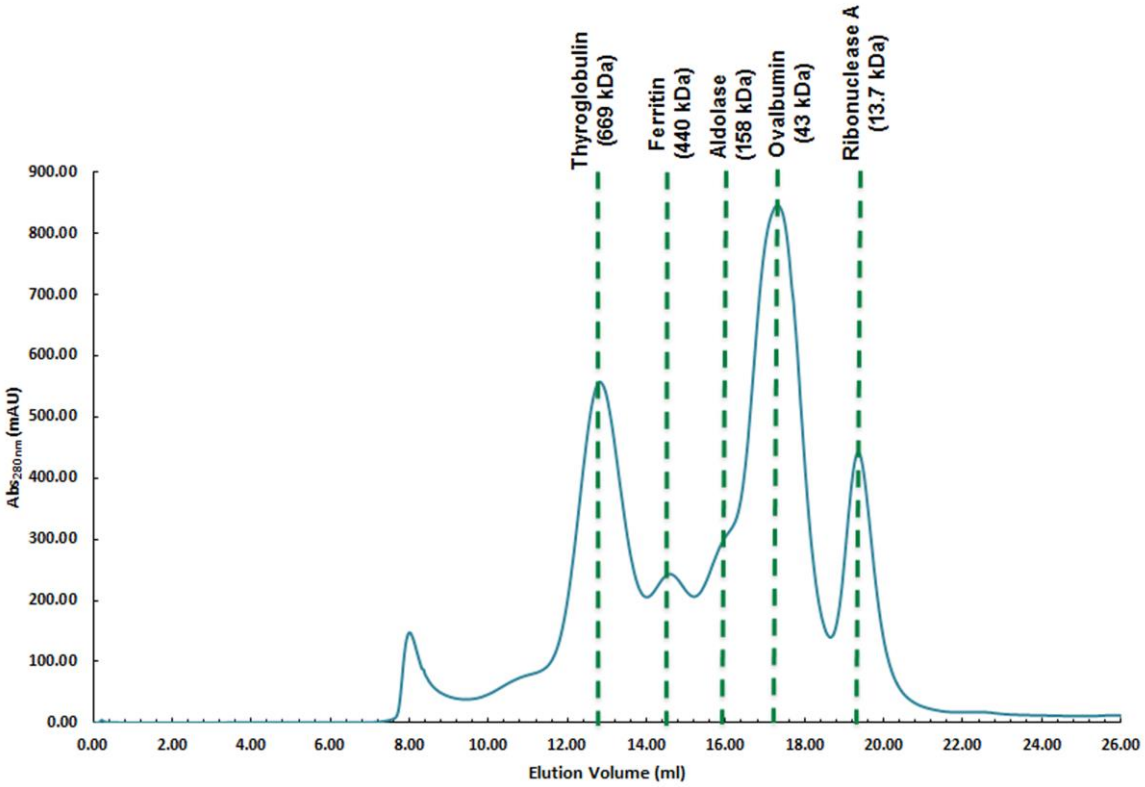


Figure A.7.7: Calibration curve of Superose6 10/300 column. Five different high and low molecular weight standards- Thyroglobin (669 kDa), Ferritin (440 kDa), Aldolase (158 kDa), Ovalbumin (43 kDa) and Ribonuclease A (13.7 kDa) were run on Superose6 10/300 column for determination of the elution volume at which they elute. The X-axis represents UV absorption of proteins at 280nm wavelength and the Y-axis represents the elution volume in ml. The green vertical dotted lines highlight the elution volumes at which the standard molecular weight markers eluted under same experimental conditions in which MEL nuclear extracts were fractionated.

Appendix-V: NMR titration of JARID1A + GATA1

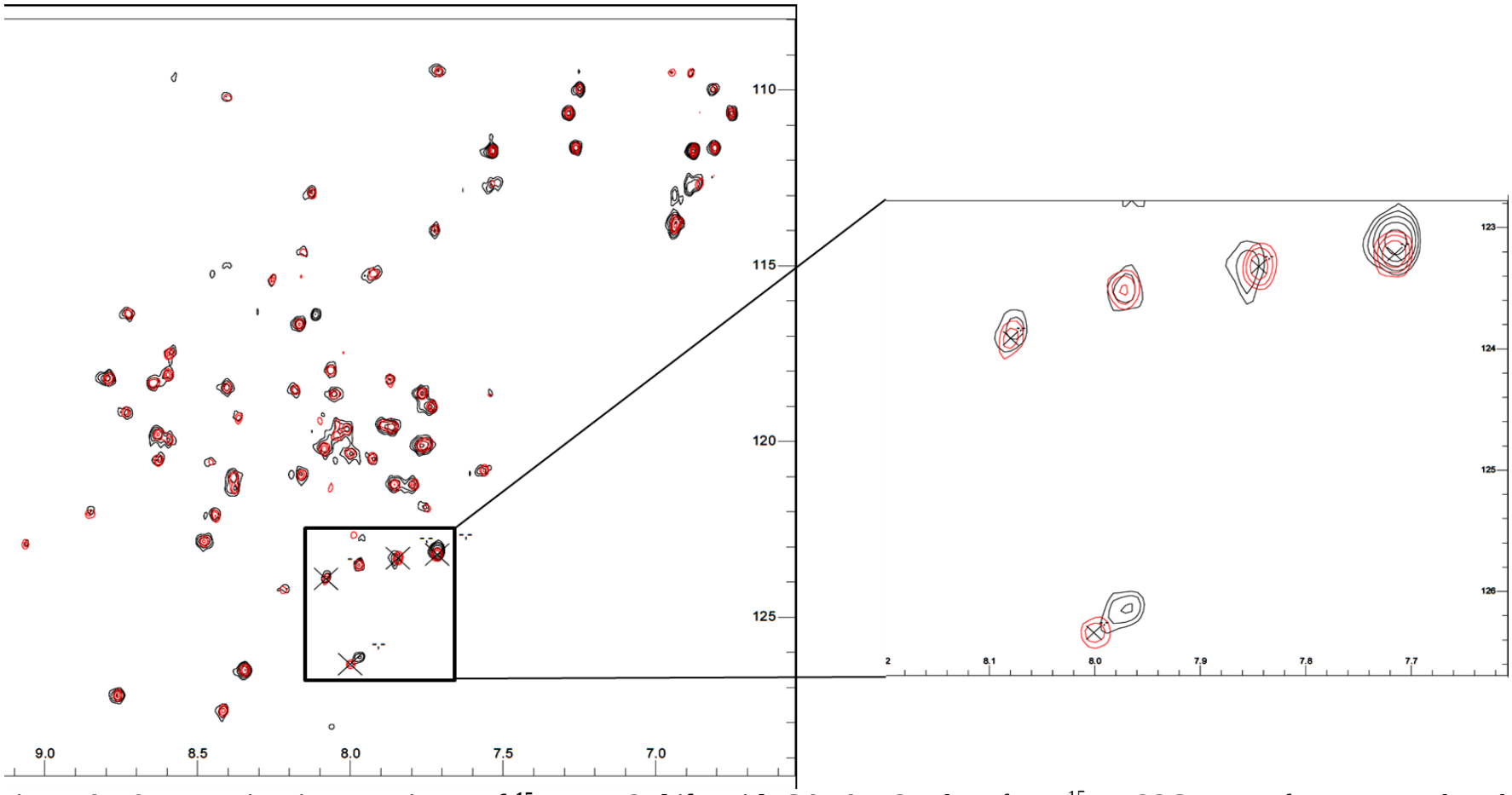


Figure A.7.8: NMR titration experiment of ^{15}N- PHD2 shifts with GATA1. Overlay of two ^{15}N-HSQC spectra for PHD2, each with different GATA1 concentrations. Red peaks represent PHD2 without any GATA1 and black peaks represent PHD2 with 2.6mM GATA1. Shifts in the peak confirms the interaction between the two proteins.

References

Alter BP. 1994. Biology of erythropoiesis. *Annals of the New York Academy of Sciences* **731**: 36-47.

Aplan PD, Jones CA, Chervinsky DS, Zhao X, Ellsworth M, Wu C, McGuire EA, Gross KW. 1997. An scl gene product lacking the transactivation domain induces bony abnormalities and cooperates with LMO1 to generate T-cell malignancies in transgenic mice. *The EMBO journal* **16**: 2408-2419.

Ball P. 2003. Portrait of a molecule. *Nature* **421**: 421-422.

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res* **21**: 381-395.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**: 120-124.

Barber BA, Rastegar M. 2010. Epigenetic control of Hox genes during neurogenesis, development, and disease. *Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft* **192**: 261-274.

Benevolenskaya EV. 2007. Histone H3K4 demethylases are essential in development and differentiation. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **85**: 435-443.

Benevolenskaya EV, Murray HL, Branton P, Young RA, Kaelin WG, Jr. 2005. Binding of pRB to the PHD protein RBP2 promotes cellular differentiation. *Molecular cell* **18**: 623-635.

Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, Owens RJ. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic acids research* **35**: e45.

Beshiri ML, Holmes KB, Richter WF, Hess S, Islam AB, Yan Q, Plante L, Litovchick L, Gevry N, Lopez-Bigas N et al. 2012. Coordinated repression of cell cycle genes by KDM5A and E2F4 during differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 18499-18504.

Blackledge NP, Klose RJ. 2010. Histone lysine methylation: an epigenetic modification? *Epigenomics* **2**: 151-161.

Blobel GA, Nakajima T, Eckner R, Montminy M, Orkin SH. 1998. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 2061-2066.

Cao J, Liu Z, Cheung WK, Zhao M, Chen SY, Chan SW, Booth CJ, Nguyen DX, Yan Q. 2014. Histone demethylase RBP2 is critical for breast cancer progression and metastasis. *Cell reports* **6**: 868-877.

Cedar H, Bergman Y. 2011. Epigenetics of haematopoietic cell development. *Nature reviews Immunology* **11**: 478-488.

Chan SW, Hong W. 2001. Retinoblastoma-binding protein 2 (Rbp2) potentiates nuclear hormone receptor-mediated transcription. *The Journal of biological chemistry* **276**: 28402-28412.

Chayen NE. 2004. Turning protein crystallisation from an art into a science. *Current opinion in structural biology* **14**: 577-583.

Chayen NE, Saridakis E. 2008. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature methods* **5**: 147-153.

Chi P, Allis CD, Wang GG. 2010. Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nature reviews Cancer* **10**: 457-469.

Chicas A, Kapoor A, Wang X, Aksoy O, Evertts AG, Zhang MQ, Garcia BA, Bernstein E, Lowe SW. 2012. H3K4 demethylation by Jarid1a and Jarid1b contributes to

retinoblastoma-mediated gene silencing during cellular senescence. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 8971-8976.

Christensen J, Agger K, Cloos PA, Pasini D, Rose S, Sennels L, Rappsilber J, Hansen KH, Salcini AE, Helin K. 2007. RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* **128**: 1063-1076.

Clapier CR, Cairns BR. 2009. The biology of chromatin remodeling complexes. *Annual review of biochemistry* **78**: 273-304.

Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C et al. 2010. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**: 360-363.

Dawid IB, Breen JJ, Toyama R. 1998. LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends in genetics : TIG* **14**: 156-162.

de Rooij JD, Hollink IH, Arentsen-Peters ST, van Galen JF, Berna Beverloo H, Baruchel A, Trka J, Reinhardt D, Sonneveld E, Zimmermann M et al. 2013. NUP98/JARID1A is a novel recurrent abnormality in pediatric acute megakaryoblastic leukemia with a distinct HOX gene expression pattern. *Leukemia* **27**: 2280-2288.

Deane JE, Ryan DP, Sunde M, Maher MJ, Guss JM, Visvader JE, Matthews JM. 2004. Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex. *The EMBO journal* **23**: 3589-3598.

Deane JE, Sum E, Mackay JP, Lindeman GJ, Visvader JE, Matthews JM. 2001. Design, production and characterization of FLIN2 and FLIN4: the engineering of intramolecular ldb1:LMO complexes. *Protein engineering* **14**: 493-499.

Defeo-Jones D, Huang PS, Jones RE, Haskell KM, Vuocolo GA, Hanobik MG, Huber HE, Oliff A. 1991. Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene product. *Nature* **352**: 251-254.

DeGregori J. 2002. The genetics of the E2F family of transcription factors: shared functions and unique roles. *Biochimica et biophysica acta* **1602**: 131-150.

DiTacchio L, Le HD, Vollmers C, Hatori M, Witcher M, Secombe J, Panda S. 2011. Histone lysine demethylase JARID1a activates CLOCK-BMAL1 and influences the circadian clock. *Science (New York, NY)* **333**: 1881-1885.

El Omari K, Hoosdally SJ, Tuladhar K, Karia D, Hall-Ponsele E, Platonova O, Vyas P, Patient R, Porcher C, Mancini EJ. 2013. Structural basis for LMO2-driven recruitment of the SCL:E47bHLH heterodimer to hematopoietic-specific transcriptional targets. *Cell reports* **4**: 135-147.

El Omari K, Hoosdally SJ, Tuladhar K, Karia D, Vyas P, Patient R, Porcher C, Mancini EJ. 2011. Structure of the leukemia oncogene LMO2: implications for the assembly of a hematopoietic transcription factor complex. *Blood* **117**: 2146-2156.

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

Ferreira R, Ohneda K, Yamamoto M, Philipsen S. 2005. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Molecular and cellular biology* **25**: 1215-1227.

Fraser PJ, Curtis PJ. 1987. Specific pattern of gene expression during induction of mouse erythroleukemia cells. *Genes & development* **1**: 855-861.

Friend C, Patuleia MC, De Harven E. 1966. Erythrocytic maturation in vitro of murine (Friend) virus-induced leukemic cells. *National Cancer Institute monograph* **22**: 505-522.

Friend C, Scher W, Holland JG, Sato T. 1971. Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide. *Proceedings of the National Academy of Sciences of the United States of America* **68**: 378-382.

Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 12355-12358.

Ge Z, Li W, Wang N, Liu C, Zhu Q, Bjorkholm M, Gruber A, Xu D. 2010. Chromatin remodeling: recruitment of histone demethylase RBP2 by Mad1 for transcriptional repression of a Myc target gene, telomerase reverse transcriptase. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **24**: 579-586.

Goardon N, Lambert JA, Rodriguez P, Nissaire P, Herblot S, Thibault P, Dumenil D, Strouboulis J, Romeo PH, Hoang T. 2006. ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *The EMBO journal* **25**: 357-366.

Greenbaum S, Zhuang Y. 2002. Regulation of early lymphocyte development by E2A family proteins. *Seminars in immunology* **14**: 405-414.

Harrison PR. 1976. Analysis of erythropoeisis at the molecular level. *Nature* **262**: 353-356.

Hong W, Nakazawa M, Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA. 2005. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *The EMBO journal* **24**: 2367-2378.

Hou J, Wu J, Dombkowski A, Zhang K, Holowatyj A, Boerner JL, Yang ZQ. 2012. Genomic amplification and a role in drug-resistance for the KDM5A histone demethylase in breast cancer. *American journal of translational research* **4**: 247-256.

Howlett GJ, Minton AP, Rivas G. 2006. Analytical ultracentrifugation for the study of protein association and assembly. *Current opinion in chemical biology* **10**: 430-436.

Hsu HL, Cheng JT, Chen Q, Baer R. 1991. Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. *Molecular and cellular biology* **11**: 3037-3042.

Hu X, Li X, Valverde K, Fu X, Noguchi C, Qiu Y, Huang S. 2009a. LSD1-mediated epigenetic modification is required for TAL1 function and hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 10141-10146.

Hu X, Ybarra R, Qiu Y, Bungert J, Huang S. 2009b. Transcriptional regulation by TAL1: a link between epigenetic modifications and erythropoiesis. *Epigenetics : official journal of the DNA Methylation Society* **4**: 357-361.

Huang S, Brandt SJ. 2000. mSin3A regulates murine erythroleukemia cell differentiation through association with the TAL1 (or SCL) transcription factor. *Molecular and cellular biology* **20**: 2248-2259.

Huang S, Qiu Y, Shi Y, Xu Z, Brandt SJ. 2000. P/CAF-mediated acetylation regulates the function of the basic helix-loop-helix transcription factor TAL1/SCL. *The EMBO journal* **19**: 6792-6803.

Huang S, Qiu Y, Stein RW, Brandt SJ. 1999. p300 functions as a transcriptional coactivator for the TAL1/SCL oncoprotein. *Oncogene* **18**: 4958-4967.

Hung HL, Lau J, Kim AY, Weiss MJ, Blobel GA. 1999. CREB-Binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. *Molecular and cellular biology* **19**: 3496-3505.

Ingley E, Tilbrook PA, Klinken SP. 2004. New insights into the regulation of erythroid cells. *IUBMB life* **56**: 177-184.

Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, Whetstine JR, Bonni A, Roberts TM, Shi Y. 2007. The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* **128**: 1077-1088.

Kadri Z, Shimizu R, Ohneda O, Maouche-Chretien L, Gisselbrecht S, Yamamoto M, Romeo PH, Leboulch P, Chretien S. 2009. Direct binding of pRb/E2F-2 to GATA-1 regulates maturation and terminal cell division during erythropoiesis. *PLoS biology* **7**: e1000123.

Kamakaka RT, Biggins S. 2005. Histone variants: deviants? *Genes & development* **19**: 295-310.

Kassouf MT, Chagraoui H, Vyas P, Porcher C. 2008. Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis. *Blood* **112**: 1056-1067.

Kassouf MT, Hughes JR, Taylor S, McGowan SJ, Soneji S, Green AL, Vyas P, Porcher C. 2010. Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. *Genome research* **20**: 1064-1083.

Kim YW, Otterson GA, Kratzke RA, Coxon AB, Kaye FJ. 1994. Differential specificity for binding of retinoblastoma binding protein 2 to RB, p107, and TATA-binding protein. *Molecular and cellular biology* **14**: 7256-7264.

Klein BJ, Piao L, Xi Y, Rincon-Arano H, Rothbart SB, Peng D, Wen H, Larson C, Zhang X, Zheng X et al. 2014. The histone-H3K4-specific demethylase KDM5B binds to its substrate and product through distinct PHD fingers. *Cell reports* **6**: 325-335.

Klose RJ, Yan Q, Tothova Z, Yamane K, Erdjument-Bromage H, Tempst P, Gilliland DG, Zhang Y, Kaelin WG, Jr. 2007. The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128**: 889-900.

Kooistra SM, Helin K. 2012. Molecular mechanisms and potential functions of histone demethylases. *Nature reviews Molecular cell biology* **13**: 297-311.

Krantz SB. 1991. Erythropoietin. *Blood* **77**: 419-434.

Lacombe J, Herblot S, Rojas-Sutterlin S, Haman A, Barakat S, Iscove NN, Sauvageau G, Hoang T. 2010. Scl regulates the quiescence and the long-term competence of hematopoietic stem cells. *Blood* **115**: 792-803.

Lankhorst CE, Wish JB. 2010. Anemia in renal disease: diagnosis and management. *Blood reviews* **24**: 39-47.

Lazorchak A, Jones ME, Zhuang Y. 2005. New insights into E-protein function in lymphocyte development. *Trends in immunology* **26**: 334-338.

Lebowitz J, Lewis MS, Schuck P. 2002. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein science : a publication of the Protein Society* **11**: 2067-2079.

Lecuyer E, Hoang T. 2004. SCL: from the origin of hematopoiesis to stem cells and leukemia. *Experimental hematology* **32**: 11-24.

Lecuyer E, Lariviere S, Sincennes MC, Haman A, Lahlil R, Todorova M, Tremblay M, Wilkes BC, Hoang T. 2007. Protein stability and transcription factor complex assembly determined by the SCL-LMO2 interaction. *The Journal of biological chemistry* **282**: 33649-33658.

Leder A, Leder P. 1975. Butyric acid, a potent inducer of erythroid differentiation in cultured erythroleukemic cells. *Cell* **5**: 319-322.

Lee N, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. 2009. The H3K4 demethylase lid associates with and inhibits histone deacetylase Rpd3. *Molecular and cellular biology* **29**: 1401-1410.

Li G, Reinberg D. 2011. Chromatin higher-order structures and gene regulation. *Current opinion in genetics & development* **21**: 175-186.

Li Y, Deng C, Hu X, Patel B, Fu X, Qiu Y, Brand M, Zhao K, Huang S. 2012. Dynamic interaction between TAL1 oncoprotein and LSD1 regulates TAL1 function in hematopoiesis and leukemogenesis. *Oncogene* **31**: 5007-5018.

Liefke R, Oswald F, Alvarado C, Ferres-Marco D, Mittler G, Rodriguez P, Dominguez M, Borggrefe T. 2010. Histone demethylase KDM5A is an integral part of the core Notch-RBP-J repressor complex. *Genes & development* **24**: 590-601.

Lin W, Cao J, Liu J, Beshiri ML, Fujiwara Y, Francis J, Cherniack AD, Geisen C, Blair LP, Zou MR et al. 2011. Loss of the retinoblastoma binding protein 2 (RBP2) histone demethylase suppresses tumorigenesis in mice lacking Rb1 or Men1. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 13379-13386.

Lopez-Bigas N, Kisiel TA, Dewaal DC, Holmes KB, Volkert TL, Gupta S, Love J, Murray HL, Young RA, Benevolenskaya EV. 2008. Genome-wide analysis of the H3K4 histone demethylase RBP2 reveals a transcriptional program controlling differentiation. *Molecular cell* **31**: 520-530.

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.

Mao S, Neale GA, Goorha RM. 1997. T-cell oncogene rhombotin-2 interacts with retinoblastoma-binding protein 2. *Oncogene* **14**: 1531-1539.

Martin C, Zhang Y. 2005. The diverse functions of histone lysine methylation. *Nature reviews Molecular cell biology* **6**: 838-849.

Matthews JM, Bhati M, Craig VJ, Deane JE, Jeffries C, Lee C, Nancarrow AL, Ryan DP, Sunde M. 2008. Competition between LIM-binding domains. *Biochemical Society transactions* **36**: 1393-1397.

Matthews JM, Visvader JE. 2003. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO reports* **4**: 1132-1137.

Nishibuchi G, Shibata Y, Hayakawa T, Hayakawa N, Ohtani Y, Sinmyozu K, Tagami H, Nakayama J. 2014. Physical and Functional Interactions between the Histone H3K4 Demethylase KDM5A and the Nucleosome Remodeling and Deacetylase (NuRD) Complex. *The Journal of biological chemistry* **289**: 28956-28970.

O'Neil J, Billa M, Oikemus S, Kelliher M. 2001. The DNA binding activity of TAL-1 is not required to induce leukemia/lymphoma in mice. *Oncogene* **20**: 3897-3905.

Ono Y, Fukuhara N, Yoshie O. 1998. TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3. *Molecular and cellular biology* **18**: 6939-6950.

Oram SH, Thoms J, Sive JI, Calero-Nieto FJ, Kinston SJ, Schutte J, Knezevic K, Lock RB, Pimanda JE, Gottgens B. 2013. Bivalent promoter marks and a latent enhancer may prime the leukaemia oncogene LMO1 for ectopic expression in T-cell leukaemia. *Leukemia* **27**: 1348-1357.

Orkin SH. 1995. Hematopoiesis: how does it happen? *Current opinion in cell biology* **7**: 870-877.

Orkin SH, Zon LI. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**: 631-644.

Osada H, Grutz G, Axelson H, Forster A, Rabbitts TH. 1995. Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 9585-9589.

Pabo CO, Sauer RT. 1992. Transcription factors: structural families and principles of DNA recognition. *Annual review of biochemistry* **61**: 1053-1095.

Pasini D, Hansen KH, Christensen J, Agger K, Cloos PA, Helin K. 2008. Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes & development* **22**: 1345-1355.

Perinchery G, Sasaki M, Angan A, Kumar V, Carroll P, Dahiya R. 2000. Deletion of Y-chromosome specific genes in human prostate cancer. *The Journal of urology* **163**: 1339-1342.

Pevny L, Lin CS, D'Agati V, Simon MC, Orkin SH, Costantini F. 1995. Development of hematopoietic cells lacking transcription factor GATA-1. *Development (Cambridge, England)* **121**: 163-172.

Pierce BA. 2010. *Genetics: A Conceptual Approach*. W.H. Freeman & Company.

Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin SH. 1996. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86**: 47-57.

Qi L, Zhu F, Li SH, Si LB, Hu LK, Tian H. 2014. Retinoblastoma binding protein 2 (RBP2) promotes HIF-1alpha-VEGF-induced angiogenesis of non-small cell lung cancer via the Akt pathway. *PloS one* **9**: e106032.

Quong MW, Romanow WJ, Murre C. 2002. E protein function in lymphocyte development. *Annual review of immunology* **20**: 301-322.

Rabbitts TH. 1998. LMO T-cell translocation oncogenes typify genes activated by chromosomal translocations that alter transcription and developmental processes. *Genes & development* **12**: 2651-2657.

Ralston G, ed. *Introduction to Analytical Ultracentrifugation*. Beckman.

Rice KL, Hormaeche I, Licht JD. 2007. Epigenetic regulation of normal and malignant hematopoiesis. *Oncogene* **26**: 6697-6714.

Robb L, Begley CG. 1996. The helix-loop-helix gene SCL: Implicated in T-cell acute lymphoblastic leukaemia and in normal haematopoietic development. *The International Journal of Biochemistry & Cell Biology* **28**: 609-618.

Ryan DP, Duncan JL, Lee C, Kuchel PW, Matthews JM. 2008. Assembly of the oncogenic DNA-binding complex LMO2-Ldb1-TAL1-E12. *Proteins* **70**: 1461-1474.

Sang M, Ma L, Sang M, Zhou X, Gao W, Geng C. 2014. LIM-domain-only proteins: multifunctional nuclear transcription coregulators that interacts with diverse proteins. *Molecular biology reports* **41**: 1067-1073.

Schuh AH, Tipping AJ, Clark AJ, Hamlett I, Guyot B, Iborra FJ, Rodriguez P, Strouboulis J, Enver T, Vyas P et al. 2005. ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Molecular and cellular biology* **25**: 10235-10250.

Secombe J, Eisenman RN. 2007. The function and regulation of the JARID1 family of histone H3 lysine 4 demethylases: the Myc connection. *Cell cycle (Georgetown, Tex)* **6**: 1324-1328.

Secombe J, Li L, Carlos L, Eisenman RN. 2007. The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. *Genes & development* **21**: 537-551.

Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA et al. 2010. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**: 69-80.

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**: 941-953.

Shivdasani RA, Fujiwara Y, McDevitt MA, Orkin SH. 1997. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *The EMBO journal* **16**: 3965-3973.

Shivdasani RA, Orkin SH. 1996. The transcriptional control of hematopoiesis. *Blood* **87**: 4025-4039.

Stavropoulos P, Blobel G, Hoelz A. 2006. Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat Struct Mol Biol* **13**: 626-632.

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**: 207-234.

Sun XH, Baltimore D. 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell* **64**: 459-470.

Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F, Li J, Rao A, Shi Y. 2007. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* **447**: 601-605.

Tanaka M, Levy J, Terada M, Breslow R, Rifkind RA, Marks PA. 1975. Induction of erythroid differentiation in murine virus infected eythroleukemia cells by highly polar compounds. *Proceedings of the National Academy of Sciences of the United States of America* **72**: 1003-1006.

Teng YC, Lee CF, Li YS, Chen YR, Hsiao PW, Chan MY, Lin FM, Huang HD, Chen YT, Jeng YM et al. 2013. Histone demethylase RBP2 promotes lung tumorigenesis and cancer metastasis. *Cancer research* **73**: 4711-4721.

Terada M, Fujiki H, Marks PA, Sugimura T. 1979. Induction of erythroid differentiation of murine erythroleukemia cells by nicotinamide and related compounds. *Proceedings of the National Academy of Sciences of the United States of America* **76**: 6411-6414.

Tremblay M, Tremblay CS, Herblot S, Aplan PD, Hebert J, Perreault C, Hoang T. 2010. Modeling T-cell acute lymphoblastic leukemia induced by the SCL and LMO1 oncogenes. *Genes & development* **24**: 1093-1105.

Tu S, Teng Y-C, Yuan C, Wu Y-T, Chan M-Y, Cheng A-N, Lin P-H, Juan L-J, Tsai M-D. 2008. The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. *Nat Struct Mol Biol* **15**: 419-421.

Upadhyay AK, Horton JR, Zhang X, Cheng X. 2011. Coordinated methyl-lysine erasure: structural and functional linkage of a Jumonji demethylase domain and a reader domain. *Current opinion in structural biology* **21**: 750-760.

Valge-Archer V, Forster A, Rabbitts TH. 1998. The LMO1 and LDB1 proteins interact in human T cell acute leukaemia with the chromosomal translocation t(11;14)(p15;q11). *Oncogene* **17**: 3199-3202.

Valge-Archer VE, Osada H, Warren AJ, Forster A, Li J, Baer R, Rabbitts TH. 1994. The LIM protein RBTN2 and the basic helix-loop-helix protein TAL1 are present in a complex in erythroid cells. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 8617-8621.

van Oevelen C, Wang J, Asp P, Yan Q, Kaelin WG, Jr., Kluger Y, Dynlacht BD. 2008. A role for mammalian Sin3 in permanent gene silencing. *Molecular cell* **32**: 359-370.

van Zutven LJ, Onen E, Velthuizen SC, van Drunen E, von Bergh AR, van den Heuvel-Eibrink MM, Veronese A, Mecucci C, Negrini M, de Greef GE et al. 2006. Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene. *Genes, chromosomes & cancer* **45**: 437-446.

Verrier L, Vandromme M, Trouche D. 2011. Histone demethylases in chromatin cross-talks. *Biology of the cell / under the auspices of the European Cell Biology Organization* **103**: 381-401.

Vitelli L, Condorelli G, Lulli V, Hoang T, Luchetti L, Croce CM, Peschle C. 2000. A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. *Molecular and cellular biology* **20**: 5330-5342.

Voronova AF, Lee F. 1994. The E2A and tal-1 helix-loop-helix proteins associate in vivo and are modulated by Id proteins during interleukin 6-induced myeloid differentiation.

Proceedings of the National Academy of Sciences of the United States of America **91**: 5952-5956.

Wadman I, Li J, Bash RO, Forster A, Osada H, Rabbitts TH, Baer R. 1994. Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. *The EMBO journal* **13**: 4831-4839.

Wadman IA, Osada H, Grutz GG, Agulnick AD, Westphal H, Forster A, Rabbitts TH. 1997. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *The EMBO journal* **16**: 3145-3157.

Walter TS, Diprose JM, Mayo CJ, Siebold C, Pickford MG, Carter L, Sutton GC, Berrow NS, Brown J, Berry IM et al. 2005. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta crystallographica Section D, Biological crystallography* **61**: 651-657.

Walter TS, Mancini EJ, Kadlec J, Graham SC, Assenberg R, Ren J, Sainsbury S, Owens RJ, Stuart DI, Grimes JM et al. 2008. Semi-automated microseeding of nanolitre crystallization experiments. *Acta crystallographica Section F, Structural biology and crystallization communications* **64**: 14-18.

Wang GG, Song J, Wang Z, Dormann HL, Casadio F, Li H, Luo JL, Patel DJ, Allis CD. 2009. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**: 847-851.

Warren AJ, Colledge WH, Carlton MBL, Evans MJ, Smith AJH, Rabbitts TH. 1994. The Oncogenic Cysteine-rich LIM domain protein Rbtn2 is essential for erythroid development. *Cell* **78**: 45-57.

Weiss MJ, Orkin SH. 1995. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 9623-9627.

Woodcock CL, Ghosh RP. 2010. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology* **2**: a000596.

Yamane K, Tateishi K, Klose RJ, Fang J, Fabrizio LA, Erdjument-Bromage H, Taylor-Papadimitriou J, Tempst P, Zhang Y. 2007. PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Molecular cell* **25**: 801-812.

Yang ZR, Thomson R, McNeil P, Esnouf RM. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics (Oxford, England)* **21**: 3369-3376.

Zhang Y, Reinberg D. 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & development* **15**: 2343-2360.

Zhuang Y, Jackson A, Pan L, Shen K, Dai M. 2004. Regulation of E2A gene expression in B-lymphocyte development. *Molecular immunology* **40**: 1165-1177.