

**Characterisation of the interaction interface of the
MOZ/BRPF1 acetyltransferase complex and its role in
gene regulation**

Bismoy Mazumder

**Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy**

September 2018

Abstract

Monocytic leukaemia zinc-finger protein (MOZ) is a histone lysine acetyltransferase, which plays important roles in haematopoiesis and tissue development. Alterations of the Lysine (K) acetyl-transferase 6A (KAT6A) gene encoding (MOZ) are associated with cancers and in particular, the expression of MOZ-fusion proteins in haematopoietic progenitors has been shown to be causative of acute myeloid leukaemia (AML). More recently, mutations in KAT6A have been associated with developmental syndromes, the clinical features of which are global developmental delay and intellectual disability. MOZ has been shown to associate with a protein complex that includes BRPF1 (bromodomain and PHD finger containing protein 1) and direct association of MOZ and BRPF1 stabilises this complex. Mutations in the BRPF1 gene are also associated with intellectual disability syndromes. Importantly, association with BRPF1 modulates the substrate specificity of MOZ and enhances its acetyltransferase activity for Histone H3 in nucleosomes. However, both MOZ and BRPF1 contain chromatin reader domains that allow them to associate with chromatin by their ability to recognise specific histone modifications. Thus normal MOZ function appears to be regulated by BRPF1, although they may also function independently in different tissues. In animal models, the induction of AML by the MOZ-Transcriptional Intermediary Factors 2 (TIF2) was shown to be dependent on its association with BRPF1. Therefore, in this study, I set out to investigate the molecular interface between human MOZ and BRPF1 proteins, to gain a better insight into how the complex is formed and to understand the functional cooperativity of these proteins.

Firstly, to map the minimal sequences required for the interaction of the MOZ/BRPF1 complex, I performed yeast-two-hybrid mapping of protein-protein interactions. This was achieved by generating a series of deletion constructs of both proteins in yeast two hybrid expression vectors. This identified that the sequence BRPF1 59-190 is sufficient to bind the MOZ sequence 676-763, which encodes helix-turn-helix motif located C-terminus of the MOZ MYST acetyltransferase domain. Site-directed mutagenesis identified key residues in the MOZ 676-763 sequence that when altered/disrupted interaction with BRPF1. These results strongly suggest that MOZ 676-763 is a BRPF1 recruitment domain, consistent with a recent structure of the related HBO1/BRPF2 complex. A dual expression vector has been constructed in which the minimal interaction domains were successfully co-expressed in *E.coli*, which will be taken forward for large scale purification and crystallisation trials.

To extend our study to full-length proteins, we performed co-localisation and co-immunoprecipitation studies with epitope-tagged proteins transiently expressed in mammalian HEK293 cells. Interestingly, while FLAG-MOZ localised to the nucleus, expression of HA-BRPF1 alone showed it to be located to the cytoplasmic compartment. In contrast co-expression of both wild type FLAG-MOZ and HA-BRPF1 resulted in their strong co-localisation to the nucleus, and co-immunoprecipitation as a complex. In contrast, a FLAG-MOZ mutant lacking a key region of the BRPF1 binding site, was unable to induce nuclear localisation of HA-BRPF1, suggesting that the helix turn helix region (676-767) is the major interface between these proteins. Surprisingly, both the wild type and mutant FLAG-MOZ proteins continued to co-precipitate together

from cell extracts, but we show using a BRPF1 bromodomain-selective inhibitor that this may be in part explained by indirect mechanism involving co-localisation to chromatin sites involving the histone reader domains.

To investigate the functional consequences of disrupting the MOZ/BRPF1 complex, rescue experiments were performed using CRISPR-CAS9 edited HEK293 cells developed in our laboratory in which MOZ expression was inactivated. Due to the repressive function of BRPF1, we focussed on genes that we upregulated after MOZ deletion. The RTqPCR data showed YFP-MOZ successfully repressed several of these upregulated genes, whereas a YFP-MOZ mutant deficient in BRPF1 binding was slightly less efficient. These results are consistent with a role for MOZ/BRPF1 complexes in suppression of gene expression, although further studies will be required to investigate this fully.

In summary, the minimal sequences involved in the molecular interface between MOZ and BRPF1 have been defined and shown to be contained within a helix-turn helix domain conserved in MYST acetyltransferase domains. This interaction is essential for co-localisation of MOZ and BRPF1 to the nucleus, and for MOZ-mediated repression of several of its target genes, although histone modification reader domains also play a role in assembling these complexes on chromatin. Future work will focus on how specific MYST/BRPF1 complexes are assembled.

Acknowledgements

I would first and foremost like to thank my supervisor, Professor David M Heery, for the opportunity to work on such an exciting project and for his continuous guidance and support over the past four years. I would also like to thank Hilary Collins for investing copious amounts of time in my training, but also for her advice and guidance throughout my PhD. I am also most grateful to a number of members of the Gene Regulation Group (past and present) for all their help, friendship and for creating an enjoyable work environment. In particular, I would like to thank Jonathan Whitchurch, Cintia Monteiro, Lorenzo Pavanello, Chris Roberts, Mitchel Masterson, Joel Fulton, Ameerul Aalam, Barbara Rampersad and Jane Doughty. I must also thank the Vice Chancellor of University of Nottingham for funding my PhD project.

A special thank you to my mum, dad, my in-laws and last but not the least, my lovely wife Rai (without you it was not possible) for their love and support and their encouragement through everything I have accomplished. To Snehungshu Sekhar Dey whose love and 'jokes' saw me through the good and bad times.

Abbreviations

Acetyl-CoA	Acetyl-CoenzymeA
ACF	ATP-dependent Chromatin assembly and remodelling Factor
AAD	Acidic Activation domain
ADP	Adenosine Di-Phosphate
AIRE	Autoimmune Regulator
ALL	Acute Lymphoblastic Leukaemia
AML	Acute Myeloid Leukaemia
AML1	Acute Myeloid Leukaemia Protein 1
ATP	Adenosine Tri-Phosphate
BD	Bromodomain
β-GAL	β-Galactosidase
bp	Base pairs
BPTF	Bromodomain and PHD domain Transcription Factor
BRPF1	Bromodomain- and PHD Finger-containing Protein 1
BSA	Bovine Serum Albumin
CARM1	Coactivator-associated arginine methyltransferase 1
CBP	CREB Binding Protein
CD	Chromodomain
CGI	CpG Islands
CHD	Chromodomain, Helicase, DNA binding
ChIP	Chromatin Immunoprecipitation
CID	CBP Interacting Domain
CREB	cAMP Response Element Binding Protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas9	CRISPR-associated (Cas) enzyme 9
CSD	ChromoShadow Domain
CSF1R	Macrophage Colony-Stimulating Factor Receptor
CTIP2	COUP-TF-Interacting Protein 2
DDR	DNA Damage Response
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribo Nucleic Acid
DPF	Double PHD Finger

DTT	Dithiothreitol
EB	Elution Buffer
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene Glycol Tetraacetic Acid
EMBL	European Molecular Biology Laboratory
ENL	Eleven Nineteen Leukaemia
E(Z)	Enhancer of Zeste
FACT	Facilitates Chromatin Transcription
FCS	Fetal Calf Serum
GAG	Glycosaminoglycan
GFP	Green Fluorescent Protein
GNAT	Gcn5-related N-acetyltransferase
GTF	General Transcription Factor
H1,2A,2B,3,4	Histone 1,2A,2B,3,4
HAT	Histone Acetyltransferase
HBO1	Histone Acetyltransferase Binding to ORC-1
HBS	HEPES Buffered Saline
HCS	HoloCarboxylase Synthetase
HDAC	Histone Deacetyltransferase
HEPES	4-(2-HydroxyEthyl)-1-PiperazineEthaneSulfonic acid
HF	High-Fidelity
HMT	Histone MethylTransferase
Hox	Homeobox
HP1	Heterochromatin Protein 1
HRP	Horseradish Peroxidase
HSC	Hematopoietic Stem Cells
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ING	Inhibitor of Growth
IP	Immunoprecipitation
ISWI	Imitation SWI
KAT	Lysine Acetyltransferase
KDAC	Lysine Deacetyltransferase
KIX	Kinase-Inducible Domain
KMT	Lysine Methyltransferase
LB	Luria-Bertani

mRNA	Messenger RNA
MART	Mono-ADP-Ribosyltransferase
MBT	Malignant Brain Tumour
MLL	Mixed Lineage Leukaemia
MOF	Males Absent On the First
MORF	Monocytic Leukaemia Zinc-finger Related Factor
MOZ	Monocytic Leukaemia Zinc-finger protein
MYST	MOZ, Ybf2/Sas3, Sas2, TIP60
NAD	ADP-Ribosyl Nicotinamide
NAP1	Nucleosome Assembly Protein 1
NCP	Nucleosome Core Particle
NEB	New England Biolabs
NELF	Negative Elongation Factor
NLS	Nuclear Localisation Signal
NMR	Nuclear Magnetic Resonance
NURF	Nucleosome Remodelling Factor
OD	Optical Density
PBS	Phosphate Buffer Saline
PCAF	p300/CBP Associated Factor
PcG	Polycomb Group
PCL	Polycomblike
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PHD	Plant HomeoDomain
PIC	Pre-Initiation Complex
PML	Promyelocytic leukaemia nuclear bodies
PMSF	PhenylMethaneSulfonylFluoride
PTM	Post-Translational Modification
RIPA	Radiolmmunoprecipitation Assay buffer
RNA	Ribo Nucleic Acid
rRNA	Ribosomal RNA
RNAP	RNA Polymerase
RT	Room Temperature
S phase	Synthesis phase
SAGA	Spt-Ada-Gcn5 Acetyltransferase
SDS	Sodium Dodecyl Sulphate

SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SIM	SUMO Interacting Motif
siRNA	Short interfering RNA
sgRNA	Single guide RNA
SUMO	Small Ubiquitin-related Modifier
SWI/SNF	Switching defective/Sucrose Nonfermenting
tRNA	Transfer RNA
TAQ	Thermus Aquaticus
TBE	Tris-borate-EDTA
TBP	TATA-Binding Protein
TE	Tris- EDTA
TSAP	Thermosensitive Alkaline Phosphatase
TIF2	Transcriptional Intermediary Factor 2
TIP60	Tat Interacting Protein, 60kDa
TSS	Transcription Start Site
UAS	Upstream Activating Sequence
UV	Ultraviolet
WT	Wild-type

List of Publications

Published

Fulton, J., Mazumder, B., Whitchurch, J.B., Monteiro, C. J., Collins, H. M., Chan, C. M., Heery, D. M. (2017). *Heterodimers of photoreceptor-specific nuclear receptor (PNR/NR2E3) and peroxisome proliferator-activated receptor- γ are disrupted by retinal disease-associated mutations*. **Cell Death & Disease**, 8(e2677), DOI:10.1038/cddis.2017.98.

Lek, N., Tadokoro-Cuccaro, R., Whitchurch, J., Mazumder, B., Miles, H., Prentice, P., Bunch, T., Zielinska, K., Metzler, V., Mongan, N.P., Heery, D.M., Hughes, I.A., (2018). *Predicting puberty in partial androgen insensitivity syndrome: use of clinical and functional androgen receptor indices*. **Lancet: EBioMedicine**. (Accepted).

In Preparation

Bismoy Mazumder, Jonathan B Whitchurch, Cintia J Monteiro, Hilary Collins, Joel Fulton, David M Heery. *Characterisation of the interaction interface of the MOZ /BRPF1 acetyltransferase complex and its role in gene regulation*.

Aksoy, O., Pencik, J., Hartenbach, M., Moazzami, A., Mazumder, B., Whitchurch, J.B., Schleder, M., Balber, T., Susani, M., Redl, E., Haitel, A., Hassler, M., Mitterhauser, M., Esterbauer, H., Marculescu, R., Greiner, G., Kramer, G., Baltzer, P.A., Schlangen, K., Schreiber, M., Quehenberger, F., Javaheri, T., Stoiber, D., Horvath, J., Hartenbach, S., Roos, S., Malcolm, T.I., Turner, S.D., Popper, H.H., Egger, G., Moriggl, R., Culig, Z., Hoermann, G.,

Hacker, M., Heery, D.M., Merkel, O. & Kenner, L. *The thyroid hormone/ μ -crystalline axis is a key player in prostate cancer growth and a novel marker for prostate cancer progression in men. Nature (Submitted).*

Sarwar, M., Whitchurch, J.B., Masterson, M., Hernandez-Quiles, M., Collins, H.M., Fulton, J., Robertson, C., Monteiro, C.J., Mazumder, B., Mongan, N.P., Heery, D.M. and Persson, J. *CRISPR-Cas9 targeting of PIP5K1A hampers metastasis in prostate cancer.*

Monteiro, C.J., Collins, H.M., Mazumder, B., Whitchurch, J.B., Fulton, J., Mongan, J. and Heery, D.M. *MOZ/KAT6A regulates genes associated with leukaemia and senescence in K562 cells.*

Whitchurch, J.B., Mazumder, B., Monteiro, C.J., Collins, H.M., Mongan, N.P., Heery, D.M. *Premature Termination Codons in Exon 17 of MOZ Result in Nonsense-Mediated Decay: Implications for Mental Retardation Autosomal Dominant 32 (MRD32) Syndrome.*

Table of Contents

Chapter 1: Introduction	1
1.1 Gene Transcription	2
1.2. Chromatin structure	6
1.3. Histone Code and Post Translation Modifications	10
1.3.1. Histone Acetylation	13
1.3.2. Phosphorylation	14
1.3.3 Methylation	15
1.3.4. Ubiquitination	16
1.3.5. SUMOylation	17
1.3.6. ADP Ribosylation	17
1.3.7. Biotinylation	18
1.3.8. Proline Isomerisation	19
1.3.9. Histone Tail Clipping	19
1.3.10. Crotonylation	20
1.3.11. Chromatin Remodelling	20
1.4. Lysine Acetyltransferase Enzymes	21
1.4.1. CBP/p300 family	23
1.4.2. GNAT family	23
1.4.3. MYST Family	25
1.5. MOZ/ MYST3	28
1.5.1. MOZ resides in multi-subunit complexes	29
1.5.2. Chromatin binding and Modification by MOZ/MYST3	31
1.5.3. MOZ Functions in Complex with BRPF1	32
1.6. Bromodomain and PHD finger containing protein 1 (BRPF1)	34
AIMS and OBJECTIVES	41
Chapter 2: Materials and Methods	42
2.1. Sources of Materials	43
2.1.1. General Suppliers	43
2.1.2. Bacterial reagents	43
2.1.3. Molecular biology reagents	43
2.1.4. Tissue culture reagents	44
2.1.5. Biochemical reagents	44
2.1.6. List of antibodies used	45
2.2. Bacteriological preparation and culture	46
2.2.1. Composition of solutions and media used in bacterial procedures	46
2.2.2. Culture of <i>Escherichia coli</i> strains	46
2.2.3. Preparation of <i>Escherichia coli</i> competent cells	47

2.2.4. Transformation into <i>Escherichia coli</i> cells	47
2.2.5. Long-term storage of bacterial cultures	48
2.3. Molecular biology techniques	48
2.3.1. Composition of solutions used in molecular biology methods.....	48
2.3.2. Small-scale plasmid purification from bacterial cells	48
2.3.3. Large-scale plasmid purification from bacterial cells	49
2.3.4. Phenol/chloroform extraction and ethanol precipitation of DNA.....	49
2.3.5. Ethanol precipitation	50
2.3.6. Spectrophotometric quantification	50
2.3.7. Sequencing of plasmid DNA	50
2.3.8. Agarose gel electrophoresis	50
2.3.9. Isolation and purification of DNA from agarose gels	51
2.3.10. Oligonucleotides	51
2.3.11. Polymerase chain reaction (PCR)	51
2.3.12. Restriction digests	52
2.3.13. Phosphatase treatment	53
2.3.14. Ligation reactions	53
2.3.15. Generation of mutations/deletions by recombinant PCR.....	53
2.3.16. Colony PCR screening.....	54
2.4. Biochemical techniques.....	55
2.4.2. Protein concentration determination	56
2.4.3. Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS- PAGE)	56
2.4.4. Staining of SDS-polyacrylamide gels.....	57
2.4.5. Western blotting and immunodetection	57
2.4.6. Small-scale bacterial expression of HIS-tagged proteins.....	58
2.4.7. Large-scale bacterial expression, purification of HIS-tagged proteins	59
2.4.8. Peptide binding and inhibitor assay	60
2.4.9. Yeast-2-Hybrid (Yeast transformation and β -Galactosidase Assay)....	60
2.4.10. Preparation of cytoplasmic and nuclear extracts	62
2.4.11. Co-immunoprecipitation	63
2.5 Cell culture.....	64
2.5.1. Maintenance of adherent cell lines	64
2.5.2. Cryopreservation of cell lines	64
2.5.3. Calcium phosphate-mediated transfection of adherent cells	65
2.5.4. Immunofluorescence for co-localisation.....	65

Chapter 3: Mapping the Molecular Interface of the MOZ/BRPF1 Complex

3.1. Introduction	68
-------------------------	----

3.2 Detection of MOZ/BRPF1 complexes in HEK293 cells.....	72
3.3. Nuclear co-localisation of HA-BRPF1 and YFP-FLAG-MOZ	74
3.4. BRPF family proteins and sequence alignment of their N-terminal regions	77
3.5. Sequence and structural similarities of MYST domains	81
3.6. Mapping N-terminus BRPF1 for MOZ interacting site	85
3.6.1 Determination of the boundaries	85
3.6.2 Sub-cloning strategy to generate VP16-AAD-BRPF1 constructs for yeast two hybrid assay	88
3.6.3 The BRPF1 BN domains are sufficient to bind MOZ MYST domain in yeast two hybrid assays	94
3.7. Mapping MOZ and other MYST protein (MORF) for BRPF1 interacting site	98
3.7.1 Sub-cloning of MOZ deletion constructs in pBTM116mod vector	98
3.7.2 MOZ MYST domain contains the BRPF1 interaction site.....	105
3.7.3 BRPF1 BN1 domain interacts with MYST HAT domain of MORF.....	108
3.8 <i>In vivo</i> confirmation of the interaction of BRPF1 with the MYST domain of MOZ	110
3.9 Summary	113
Chapter 4: Effect of Mutations on MOZ/BRPF1 Complex.....	118
4.1. Introduction	119
4.2. Site-Directed Mutagenesis of the BRPF1 interaction region of MOZ	121
4.3. Incorporation of mutations using site-directed mutagenesis	125
4.4. Disruption of MOZ/BRPF1 complex in yeast-two-hybrid assays	130
4.5. Attempts to disrupt MOZ/BRPF1 complexes in mammalian cells	133
4.6. Targeting the loop 'finger' region (735-750) in full length MOZ	136
4.7. Co-localisation of BRPF1 and MOZ proteins in mammalian cell line	139
4.8. BRPF1 bromodomain inhibitors partially disrupt MOZ/BRPF1 association	142
4.8.1. Bacterial expression of BRPF1 bromodomain and large scale purification	143
4.8.2. Detection of 6xHis-BRPF1 bromodomain and evaluating the specificity of bromodomain to histone peptide by western blotting	146
4.8.3. Effects of GSK6835A and GSK9311A on BRPF1 bromodomain binding to H4K12ac.....	149
4.9. Effects of GSK6835A on MOZ/BRPF1 complex formation in mammalian cells	153
4.10. Summary	156
Chapter 5: Functional Investigation of MOZ/BRPF1 Complexes.....	161
5.1. Introduction	162
5.2. CRISPR edited MOZ-deficient HEK293 cell line	165
5.3. MOZ Regulation of Gene Expression: Rescue Experiments in HEK293 G10 Cells	169

5.4. Co-expression of MOZ and BRPF1 minimal interacting sites in <i>E.coli</i>	173
5.5.1. Gibson assembly cloning	173
5.4.2. Cloning BRPF1 into first multiple cloning site of pACYC Duet-1	176
5.4.3. Cloning of MOZ 676-763 into MCS-2 of pACYC Duet-1 containing 6xHis-BRPF1 90-139	180
5.5. Purification of His-BRPF1 90-139 complexes	182
5.6. Summary	185
Chapter 6: Concluding Remarks	189
Appendix	194
References	199

List of Figures

Figure 1. 1: Cartoon for transcriptional initiation by targeting to the Upstream Activation Sequence (UAS).	3
Figure 1. 2: X-Ray structure of the nucleosome core molecule..	7
Figure 1. 3: Schematic model for the histone tails and their respective modifications.....	12
Figure 1. 4: Showing human MYST family KATs domain structures.	27
Figure 1. 5: Schematic representation of variety of multi-subunit complexes, where MOZ functions as the catalytic subunit.	30
Figure 1. 6: Schematic illustration of MOZ/MORF complex formation with BRPF1, EAF6 and ING5.....	33
Figure 1. 7: Schematic showing the multi-domain structure of BRPF1.....	36
Figure 1. 8: Showing the subcellular localisation of BRPF1.....	38
Figure 3. 1. Co-immunoprecipitation of MOZ and BRPF1 complex <i>in vivo</i>	73
Figure 3. 2. Co-localisation of MOZ with BRPF1.	76
Figure 3. 3. Domain organisation and sequence comparison of different BRPF family proteins.	79
Figure 3. 4. Sequence and structural comparison of different MYST family proteins.....	83
Figure 3. 5. Secondary structure prediction of N-terminal BRPF1.	87
Figure 3. 6. Diagram for sub-cloning strategy.....	90
Figure 3. 7. Cloning of several N-terminal constructs into pASV3mod vector.	91
Figure 3. 8. β -galactosidase reporter assay from yeast two hybrid.....	96
Figure 3. 9. Secondary structure prediction of MOZ 450-900.....	100
Figure 3. 10. Cloning of several MOZ deletion constructs into pBTM116mod vector.....	103

Figure 3. 11. β -galactosidase reporter assay from yeast two hybrid for MOZ deletion constructs and BRPF1 1-222 interaction.....	107
Figure 3. 12. β -galactosidase reporter assay from yeast two hybrid for MORF 541-626 and BRPF1 1-222 interaction.	109
Figure 3. 13. Co-immunoprecipitation of shorter MOZ and BRPF1 complex <i>in vivo</i>	112
Figure 3. 14. Schematic for MOZ/BRPF1 complex and the tertiary structure of MOZ MYST domain and location of BRPF1 interacting site.	116
Figure 4. 1. Positions of the point mutations on MOZ 676-763.....	124
Figure 4. 2. Sequencing data to confirm the mutations.	127
Figure 4. 3. Confirmation of expression of both BRPF1 1-222 and MOZ mutations.	129
Figure 4. 4. β -galactosidase reporter assay from yeast two hybrid for all the LexA-DBD MOZ mutations vs VP16-AAD BRPF1 1-222.....	132
Figure 4. 5. Co-immunoprecipitation result for Q743P and V745P point mutation in FLAG-MOZ 1-779 and FLAG-MOZ 1-810.....	135
Figure 4. 6. Creating MOZ deleted double β -sheet ($\Delta\beta$) and co-immunoprecipitation result for deleted double β -sheet ($\Delta\beta$) on YFP-FLAG-MOZ full length and FLAG-MOZ 1-810.....	138
Figure 4. 7. Co-localisation of YFP-FLAG-MOZ FL $\Delta\beta$ and HA-BRPF1 in HEK293 cells.....	141
Figure 4. 8. Expression and purification of 6xHis-bromodomain in BL21 <i>E.coli</i>	145
Figure 4. 9. Western blot result for detection and specificity assay of His-bromodomain.....	148
Figure 4. 10. Flow diagram for incubation stages for peptide binding and inhibition assay.	151
Figure 4. 11. Western blot result for BRPF1 bromodomain inhibition assay.....	152
Figure 4. 12. Co-immunoprecipitation result for deleted double β -sheet ($\Delta\beta$) on YFP-FLAG-MOZ full length in presence of bromodomain inhibitor.	155
Figure 4. 13. Crystal structure of HBO1/BRPF2 complex.	158

Figure 5. 1. Schematic for CRISPR/Cas9 activity on MOZ gene (KAT6A). .	167
Figure 5. 2. Rescuing the gene expression by exogenous MOZ expression in MOZ knockout HEK293 cells.....	172
Figure 5. 3. Gibson assembly cloning steps in flow diagram.	175
Figure 5. 4. Cloning of BRPF1 59-222 into pACYC Duet-1 vector at MCS-1 and mini scale induction.	178
Figure 5. 5. Cloning of BRPF1 90-139 into pACYC Duet-1 vector at MCS-1 and mini scale induction.	179
Figure 5. 6. Cloning of MOZ 676-763 into pACYC Duet-1 vector containing BRPF1 90-139 and mini scale induction of both the protein.....	181
Figure 5. 7. Purification of MOZ 676-763 and BRPF1 90-139 by His-tag purification and western blot confirmation of His-BRPF1 90-139.....	184

Chapter 1: Introduction

1.1 Gene Transcription

Mammalian genes consist of intronic and exonic sequences and vary greatly in size from a few thousand to millions of base pairs. Transcription factors and their coregulators control the rate of gene transcription by binding to specific sites within promoter or enhancer regions. RNA Pol II transcribes protein encoding genes, and the primary transcript is processed to produce a 5' capped polyA tailed spliced mRNA. mRNAs are then transported to the cytoplasm for translation by ribosomes, and eventual degradation or storage.

Regulation of gene expression is highly complex and underlies the development, differentiation and growth of multi-cellular organisms. It is regulated at many different levels, and can involve activators and repressors, epigenetic modifications, microRNAs and other small regulatory RNAs, non-coding and antisense transcripts, deadenylation, regulated transport storage, translational control and turnover. These processes work in unison to achieve controlled expression in different tissues, in response to cellular signals or during development. The transcriptional process is dependent upon formation of an initiation complex of evolutionary conserved transcription factors and polymerases. There are three functionally distinct RNA polymerases (RNAP) found in mammalian cell nucleus, they are RNA polymerase I, II and III. RNAP II synthesises all the eukaryotic messenger (m)RNA and most small nuclear (sn)RNA and microRNA. Synthesis of RNA is the primary step to determine cellular gene expression level and RNA polymerase II is the end target for various signal transduction pathways (Hahn, 2004). To start the transcription RNAP II needs general transcription factors (GTFs), because it is unable to activate transcription on its own (Kornberg, 2007).

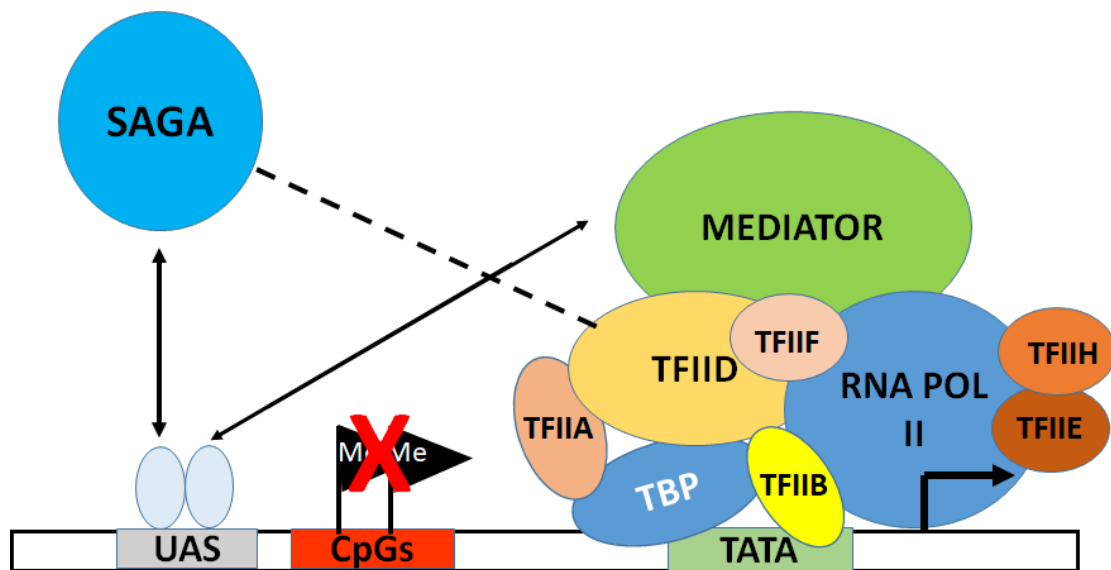


Figure 1. 1: Cartoon for transcriptional initiation by targeting to the Upstream Activation Sequence (UAS), where activators recruit coactivators (SAGA, Mediator, Swi/Snf). This recruitment of activators triggers hyperacetylation of histones and the subsequent remodelling of nucleosomes and unmethylated CpG Islands (CGIs) and also provides a permissive chromatin state to expose the entire core promoter and the SAGA/Mediator complexes facilitate the assembly of GTFs and RNAP II initiate transcription. (Figure adapted and recreated from Hahn, 2004).

Transcription requires a very high level of control, which is guided by regions of eukaryotic protein coding genes. The activation or repression of transcription initiation depends on the recognition of the core promoter region and a unique array of proximal and distal regulatory elements by sequence specific DNA binding factors. Activators bind the Upstream Activation Sequence (UAS), situated near or far from gene promoters, and initiate a cascade recruitment of coactivator complexes, which includes histone modification enzymes or chromatin remodelers and mediators (Figures 1.1). The nucleosomal positioning at the promoter for many genes, which is transcribed by RNAP II, is well defined and because of that SAGA histone acetyltransferase activity,

histone acetylation and the Swi/Snf remodeler act prior to Pre-Initiation Complex (PIC) formation, which leads to overcome the nucleosomal barrier and make DNA more accessible (Reinke et al., 2001). A large multi-subunit complex called mediator is highly conserved within eukaryotes, which interacts with enhancer bound activator proteins directly and contacts RNAP II and GTFs at the promoter to begin the initiation of transcription (Kornberg, 2005). Mediators is not only a coactivator but it also functions as a corepressor and GTF and it is also a signal processor, relaying both positive and negative regulatory information from enhancer to promoters (Takagi and Kornberg, 2006).

The identification of CpG islands (CGIs) made our understanding about transcriptional control more complex. Mammalian genomes consist with high G/C rich sequences and high density of CpG dinucleotides, which are found in mammalian genome very frequently. The transcription start site (TSS) and most of the promoter regions (60-70%) of human genes are enclosed by the CGIs (Illingworth and Bird, 2009) but they often lack DNA methylation. These elements sometimes indicate active gene promoters, however they are refractory to epigenetic silencing by DNA methylation. The connection between unmethylated CGIs and transcriptional activation was established before but it was found recently that the underlying sequence of CGIs influences the chromatin modification by recruiting CpG binding proteins. Therefore, the DNA sequence at CGIs controls the genome function directly by chromatin structure manipulation (Thomson *et al.*, 2010).

As soon as the specific positive regulatory events of activator recruitment and nucleosomal rearrangement occurs, the promoter DNA becomes available for a short period of time to interact with the transcription machinery. Generally the

RNAP II transcription cycle starts with gene specific regulatory factor association at the sequence element within the core promoter and these elements include the TATA box. Assembly of RNAP II and GTFs to form the PIC at the correct position requires recognition from the core promoter (Figures 1.1).

Transcription Factor II D (TFIID) containing the TATA box binding protein (TBP) allows approximately 14 TBP associated factors (TAFs) to bind with promoter sequence for promoter recognition. The TFIID-promoter attachment is stabilised by sequential binding of the TFIIA and TFIIB transcription factors. The pre-formed RNAP II-TFIIF complex binds with TFIIIE and TFIIH. TFIIH works as a helicase to expose the template strand at the transcription start points and it also plays an important role in nucleotide base excision repair, which leads to clearance of the promoter at the start of the elongation phase (Zhovmer *et al.*, 2010). The transcriptional initiation starts to move with the formation of the first phosphodiester bond of RNA.

There are several steps involved in regulation of transcription, which regulate the elongation and termination as well as initiation. RNAP II release contacts with core promoter and the rest of the transcriptional machinery to clear the promoter and enter early elongation phase. The distribution of transcriptionally engaged RNAP II across the genome were analysed by using ChIP-chip and permanganate foot printing (Core and Lis, 2008) and it was found that RNAP II positioned within the 5' body of the gene in high density. This concept defines how transcriptionally engaged RNAP II transcribes TSS 20-50bp downstream to produce RNA and pause. This process is called promoter-proximal pausing (Core and Lis, 2008). To release RNAP II from its paused state, activator

proteins recruit all the factor and promote RNA chain synthesis, processing, export and chromatin modification. Hyperphosphorylated RNAP II and its associated pausing factors allow the enzyme to escape the pause to either enter productive elongation or terminate RNA synthesis. Nucleosomes are displaced and redeposited during elongation to allow the RNAP II to transcribe the entire gene body and after completion of transcribing the entire gene, the transcription terminates and the RNAP II is removed from the DNA and the RNA released. The freed RNAP II can be reused in future transcription initiation (Fuda *et al.*, 2009).

We know more about the transcriptional regulation and maintenance of genome integrity. Transcription also correlates with mutagenesis and breakage within chromosomal region at the coding region of an individual gene, which leads to disrupt the genome stability (Aguilera, 2002). More research is needed to enable us to understand the role of transcription in other cellular process such as DNA damage and repair pathways and chromatin remodelling (Svejstrup, 2010). Understanding how gene transcription is disrupted in human diseases like cancers will provide new insights into the molecular basis and treatment of disease.

1.2. Chromatin structure

The chromosomal DNA of a single cell of all the eukaryotic organisms if unravelled would be approximately two meters long and it is packaged non-randomly into only within few cubic microns in volume, i.e. the cell nucleus. This is achieved by the association of DNA with histones and other nuclear proteins.

This complex of nuclear histones and DNA is called chromatin (Peterson and Laniel, 2004), which is necessary for cell cycle, gene transcription, DNA replication and DNA repair along with DNA packaging in the nucleus.

In mammalian cells transcription does not occur on naked DNA, it occurs in the context of chromatin. The basic repeating structure of this nucleoprotein complex is the evolutionary conserved nucleosome. This nucleosome includes two copies of each

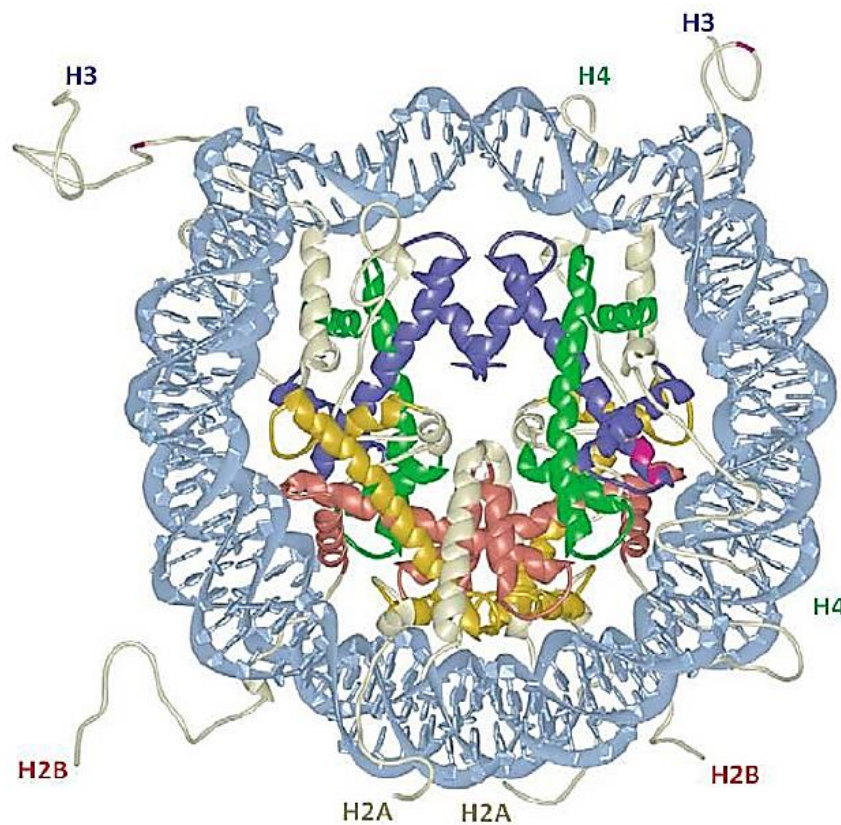


Figure 1. 2: X-Ray structure of the nucleosome core. The core histone octamer consists of a central heterotetramer of histones H3 (blue) and H4 (green), twisted with two heterodimers of histones H2A (yellow) and H2B (red). The DNA helix (light blue) is shown and protruding histone N-terminal tails (white). (Luger *et al*, 1997).

of the four core histones H2A, H2B, H3 and H4 wrapped by 147 bp of DNA, as shown by the crystal structure (Luger *et al.*, nature 1997) (Figures 1.2).

Globular core histones are synthesised at S-phase and N-terminal 20-35 residue segment extended to the surface of the nucleosome (Peterson and Laniel, 2004). Both the inter-nucleosomal histone-histone interactions and the direct contact with the DNA backbone is provided by the central globular domain, which allows consequent organisation of the two wraps of nucleosomal DNA. These protein-DNA complexes are stable under physiological conditions because of these multiple interactions (Li *et al.*, 2007). Histone proteins have crucial structural and functional roles in regulation of many nuclear processes regulation such as transcriptional control. The reversible modifications of histones on their N-terminal tails provide an epigenetic code, signalling for changes in chromatin architecture.

In addition there is a fifth histone H1 (linker histone) present in nuclei, other than the four core histones from the nucleosome core particle (NCP). Histone H1 within the nucleosome array associates with non-constrained linker DNA for partial nuclease protection (Woodcock and Ghosh, 2010). The linker histones are also known as chromatin architectural proteins, which convert nucleosomal arrays into higher and stable 30nm structures with enhanced compaction and regularity (Routh *et al.*, 2008). Linker histones also can play role as a recruitment hub for many non-histone nuclear and cytosolic proteins.

The nucleosomal array is further stabilised by forming a condensed helical coil (30nm fibre) by core associated linker histone H1 (Routh *et al.*, 2008). The stability of the nucleosome array is also governed by the introduction of

disulphide crosslink, that's how the chromatin fibre comprises two stacks of nucleosomes with zig zag back and forth in a two-start helix organisation (Dorigo *et al.*, 2004). The modulation of chromatin structure consists with an array of histone variants and accessory proteins with chromatin fibre. The stability of nucleosomes and nucleosomal arrays could be adjusted by histone variants, which are non-allelic isoforms of the canonical core histones (Park and Luger, 2008). Further, structural studies revealed that these variants contribute to the alteration of higher-order chromatin structure by internucleosomal interaction (Luger 2003). The main difference between histone variants and core histones is that, the variants are expressed outside of S-phase and incorporated into chromatin (Li *et al.*, 2007). The distribution of the variants in chromatin is not random and it is expressed in developmentally constrained or cell type specific patterns (Verdone *et al.*, 2006). H2A.Z is a variant of histone H2A and the deposition of H2A.Z into nucleosome can be possible by either ATP-dependent histone exchange reactions (Mizuguchi *et al.*, 2004) or by the help of the histone chaperone NAP1 (Park *et al.*, 2005).

The higher chromatin structure and nucleosome prevent the access of enzymes and other factors, which assist transcription, DNA damage and repair. Therefore it is highly important to maintain the chromatin structure for eukaryotic gene regulation. The genome integrity is maintained by another group of proteins called histone chaperones by retaining a balance between the assembly and the partial or complete disassembly of nucleosomes (Park and Luger, 2008). Higher order chromatin structures play a great role for packaging all genetic information within cell nuclei. They also provide a ground by assisting

many fundamental nuclear processes and controlling the expression of an individual eukaryotic gene.

1.3. Histone Code and Post Translation Modifications

In multicellular organisms, all cell types have identical genomes but still they can differentiate to have specific functions and morphology. These differences come from the gene expression patterns during development and subsequently are retained through mitosis. This is achieved in part through epigenetic modifications which are heritable but have potential to influence gene expression without altering the DNA sequence. In mammalian cells DNA methylation and histone modifications are the major mechanisms known to establish epigenetic regulation.

Post-translational modifications, i.e. various chemical alterations in polypeptides after translation, occur in the cells to regulate the stability and function of both histone and nonhistone proteins. In histones numerous types of modifications have been identified, the best studied of which are lysine acetylation and methylation, or arginine methylation. Two major impacts of histone modification mechanisms have been characterised, which are a) the disruption of DNA-histone or histone-histone contacts to decondense chromatin and b) recruitment of a series of nonhistone cofactors or chromatin regulators. Many of these chromatin associated proteins carry out enzymatic activities, which are capable of further modifying chromatin and its structure (Kouzarides, 2007). These alterations in the chromatin structure lead to functional alteration in transcriptional regulation, DNA damage repair, and replication. In addition to

acetylation and methylation, lysine side chains are known to be subject to SUMOylation, Ubiquitination, biotinylation and butylation (Kouzarides, 2007). In addition, other amino acid side chain modifications like serine, threonine or tyrosine phosphorylation, proline isomerisation and arginine deamination to citrulline also play a part in histone tail signalling.

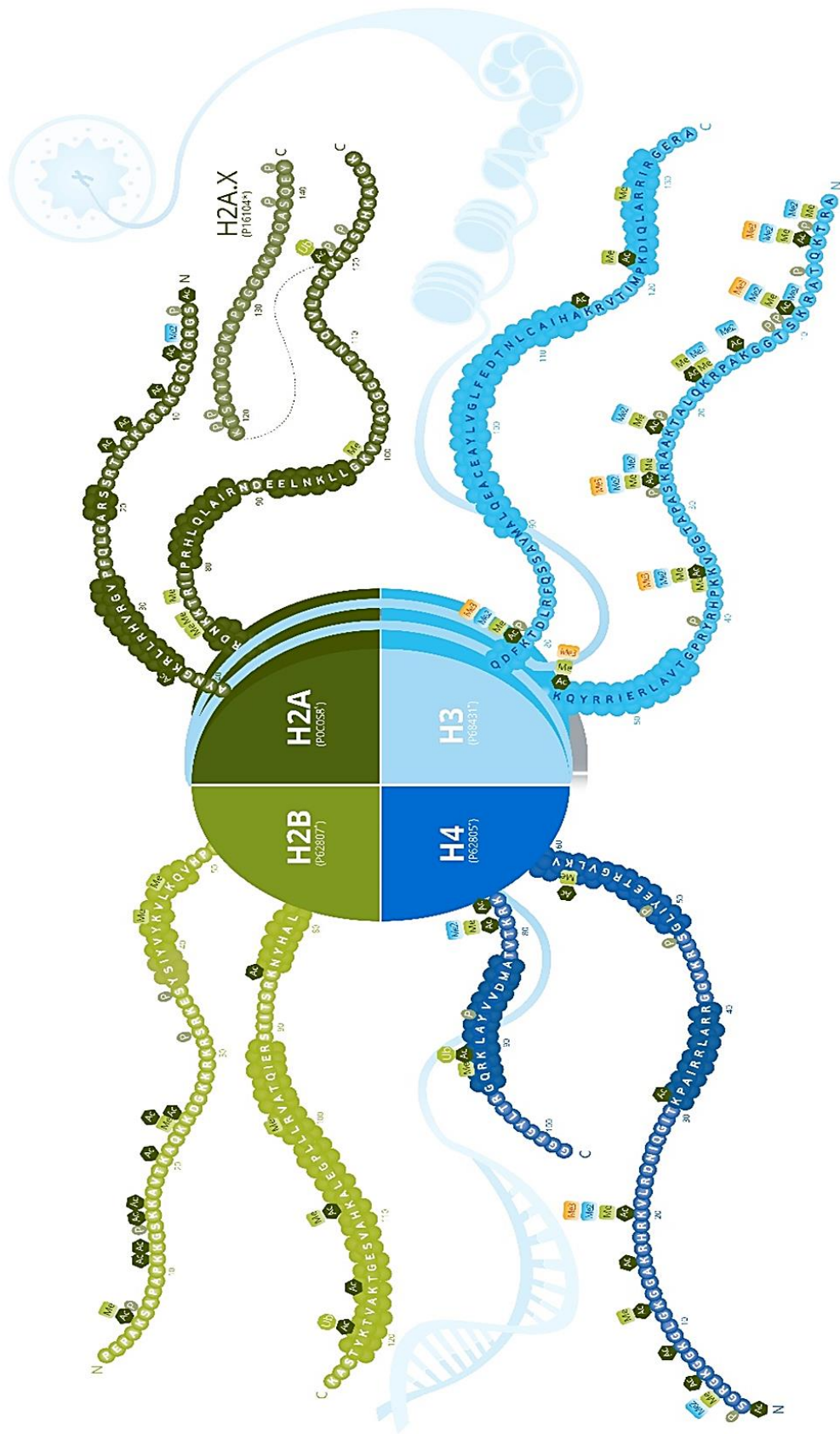


Figure 1. 3: Schematic model for the histone tails and their respective modifications (reproduced from Merck Millipore Histone Modification Poster)

1.3.1. Histone Acetylation

Acetylation of histone at lysine residues is one of the most thoroughly studied covalent modifications, which has a major influence on gene expression in eukaryotic cells. Reversible acetylation of histones at specific lysine residues occurs over the whole eukaryotic genome, and is generally associated with the activation of gene transcription. Histone acetylation is dynamically regulated by an enzyme called, histone acetyl transferases (HAT) and histone deacetylases (HDAC). HATs, which are also known Lysine (K) acetyltransferases (KAT), transfer the acetyl-group from acetyl-CoA to specific lysine residues the ϵ -amino group of specific lysine residues within the N-terminal tails of histones (Berndsen *et al.*, 2007). Removal of the acetyl group is achieved by histone deacetylases (HDAC or KDACs) (Allis *et al.*, 2007).

Histone acetylation is generally associated with the activation of transcription and this acetylation take place throughout the whole genome of eukaryotes. Acetylation has the potential to neutralise the electrostatic interactions between the positively charged histone lysines and negatively charged phosphate backbone of DNA, thus preventing the compaction and folding of higher order chromatin structures that also reduce inter-nucleosomal contacts and promote accessibility for the transcriptional machinery. Furthermore, like other PTMs, lysine acetylation plays major roles as a direct signal and binding platform for trans-acting factors (Verdone *et al.*, 2006) and antagonistically KDAC plays major role in repression at specific genes by deacetylation. The deacetylation promotes the association of repressor proteins, which recognise unacetylated histones. Chromatin assembly is another major function of histone acetylation patterns. Newly synthesised histones are deposited into nucleosomes correctly

by the chaperones, because they remain acetylated, which helps them to be recognised (Shahbazian and Grunstein, 2007).

So far 3 major types of HATs have been characterised based on sequence comparisons (Heery & Fischer, 2007). Those are Gcn5 N-acetyltransferases (GNAT), MYST, and p300/CBP. Our group is investigating the structure and function of MOZ (MYST3) protein and its role in cancer, more specifically its role as an oncogenic fusion protein in acute myeloid leukaemia (AML).

1.3.2. Phosphorylation

Phosphorylation is a reversible process of adding a phosphate group to a serine or threonine or tyrosine residue in a protein. The addition of a phosphate group is catalysed by protein kinases whereas protein phosphatases remove those groups to maintain the balance between substrate phosphorylation and dephosphorylation within the cell. A bulky and negatively charged phosphate group is added to the hydroxyl side chain of the serine, threonine and tyrosine residue, which leads to electrostatic ion-pairing increase and hydrogen bonding capability (Taverna *et al.*, 2007). Alternatively, protein interaction can be greatly induced by phosphorylation, which is read by specialised domains (Komander, 2009). Histone phosphorylation can control the cell-cycle, DNA damage repair, PTM-regulated signal transduction and transcriptional regulation depending on the site of phosphorylation. Along with other histone marks, phosphorylation of chromatin was recently been found to have a direct effect on gene expression. As an example, phosphorylation of H3T6 by Protein Kinase C β 1 prevents LSD1

and JARID1B from demethylating H3K4 during androgen receptor gene activation (Metzger *et al.*, 2010).

1.3.3 Methylation

Methylation is the process of transferring a methyl group to the ϵ -amino group of lysine residue or the guanidine nitrogen of an arginine residue from S-adenosylmethionine (AdoMet) (Bannister and Kouzarides, 2005). Methylation is generally a much more complex process than acetylation due to the presence of different types of methylation such as mono-, di- or tri-methylation. This process is catalysed by three types of lysine methyltransferase (KMT). The arginine methylation is governed by the Class I family of protein arginine N-methyltransferases (PRMT), whereas methylation of lysine has found to be catalysed by the conserved SET domain family of proteins.

A series of nuclear processes, such as heterochromatin formation, X-chromosome inactivation and transcriptional regulation, are controlled by histone methylation (Martin and Zhang, 2005). There are many studies that have taken place to understand methylation code as a transcriptional code and it was found that methylation can result in activation or repression whereas acetylation is only associated with transcriptional activation. The effect of methylation depends on the sequence specific methylation as well as the methylation state of the arginine or lysine residue (Qian and Zhou, 2006). Twenty four methylation sites in histones have been reported so far (Bannister and Kouzarides, 2005) but understandings of the functional consequences of

these modification is still very limited. It is still remain unclear, how these signals work in combination to impact on gene regulation.

The arginine demethylases are yet to be identified, but the presence of lysine demethylases indicate that there might be arginine demethylase as well. The lysine demethylase or LSD1 works through an amine oxidase reaction and it is predicted that this reaction might be compatible with arginine demethylation as well. It has been found that adjacent pre-existing PTMs near to the methylation site can mask the recognition motif of protein methyltransferases (Bedford and Richard, 2005).

1.3.4. Ubiquitination

Protein ubiquitination is the process of incorporation of a small protein ubiquitin, a polypeptide chain of 76 amino acids, to the lysine residue for its modification. Ubiquitin and lysine residue are covalently attached together by three-step enzymatic reaction. In first step the ubiquitin activating enzyme attaches ubiquitin to the E2 enzyme through an ATP-dependent reaction then E2 transfers the ubiquitin and finally E3 ligase ubiquitin to the substrate through formation of an isopeptide bond (Komander, 2009). Ubiquitin itself contains seven lysines in its primary sequence, which is also a substrate for ubiquitination. There are various combination of ubiquitination available such as mono- and poly- ubiquitination, which can occur on a single substrate. Like other histone modifications, ubiquitination is also a reversible modification and deubiquitinating enzymes remove or edit of the ubiquitin chain (Sowa *et al.*, 2009). Ubiquitination of a protein signals for protein turnover and then the

ubiquitinated protein would undergo degradation. Ubiquitination has also other functions such as DNA repair (Kouzarides, 2007), DNA-damage response, endocytosis and the mediation of protein-protein interaction (Komander, 2009).

1.3.5. SUMOylation

SUMO, small ubiquitin-related modifier, is a small protein with 11kDa in size, which associates with other proteins to control their function. All four histones undergo SUMOylation, however the specific sites for this modification have been identified only on H2A, H2B and H4 (Nathan *et al.*, 2006). In spite of structural similarities between SUMO and ubiquitin their function is totally different such as ubiquitin tags proteins for degradation whereas SUMO affects the intracellular localisation, interaction, stability of the target proteins (Meulmeester and Melchior, 2008). SUMOylation regulates the fundamental cellular processes such as cell-cycle regulation, cellular differentiation and gene transcription for well-being of most of the organisms. In mammals there are four types of SUMO proteins SUMO 1, 2, 3 and 4, distinguished by their choice of target proteins and differential functions (Meulmeester and Melchior, 2008). Like ubiquitination, SUMOylation is also starts with E1 activating enzyme Aos1-Uba2 then SUMO is transferred to E2 enzyme Ubc9 and finally E3 ligase transfer the SUMO to a lysine residue of its target protein via isopeptide bond.

1.3.6. ADP Ribosylation

ADP ribosylation is the process of transferring the ADP-ribose molecule from ADP ribosyl nicotinamide (NAD⁺) to an arginine or glutamic acid and as a bi-

product it releases nicotinamide. This is a reversible process with either mono- or poly- form of modification. The enzymes involved in this process are Mono-ADP-ribosyltransferases (MARTs) for mono-form and Poly-ADP-ribose polymerases (PARPs) for poly-form. These two enzymes are possibly represent a novel intracellular MARTs family (Hassa *et al.*, 2006). The function of this modification still remain unclear, although there are several reports of ADP-ribosylation of histones. There are few links established between ADP-ribosylation and some fundamental cellular processes such as DNA repair, apoptosis, transcription and signalling. In chromatin structure, poly ADP ribosylation relax the site of modification, which leads to an opening for factors involved in DNA mediated processes (Hassa *et al.*, 2006).

1.3.7. Biotinylation

It is the process of attaching the vitamin biotin to the lysine residue of histone. This process is catalysed by biotinidase and holocarboxylase synthetase (HCS). There are eleven biotinylation sites has been identified within histone H2A, H3 and H4. These biotinylation has several biological functions such as cell proliferation, gene silencing and cellular response to DNA damage. Biotinylation of H4K12 is specifically enriched in heterochromatin region and it plays a major role in gene repression (Hassan and Zemleni, 2008). A link has been observed between histone biotinylation and other modification such as methylation or phosphorylation. Biotinylation is also a reversible process, however the enzyme or mechanism involved in debiotinylation still remain unclear. The absence of biotinylated histones depends on the exogenous biotin

supply and availability at the specific genomic loci (Hassan and Zemleni, 2008).

1.3.8. Proline Isomerisation

In a polypeptide chain, proline exists in either *cis* or *trans* forms and this conformation has crucial role in protein folding. The changes between these two conformations can change the function of the protein completely. Proline isomerisation is a non-covalent modification of histones as well as non-histone proteins. This modification acts as a regulatory switch in signalling pathways. The enzyme that accelerate the interconversion of the proline is named FPR4, which is a member of the FK506 binding protein. This enzyme binds the amino terminal proline residue P30/38 of the tails of Histone H3 and H4. P38 isomerisation antagonises the H3K36 methylation by SET2. This is a link between methylation and proline isomerisation in the transcriptional regulation (Nelson *et al.*, 2006). The isomerisation of peptidyl-prolyl bonds are catalysed by the parvulin (Pin1) family enzymes in various transcription factors in a phosphorylation-dependent manner and the isomerisation occurs to control the conformation, the stability and the activity of the various proteins, which take part in transcription and cell-cycle regulation (Shaw, 2007).

1.3.9. Histone Tail Clipping

A novel endopeptidase activity has been discovered on histone H3 in *S. cerevisiae*. This identification gave us a new insight into histone modification, which is called as 'clipping'. This modification plays role in gene expression.

The enzyme involved in this clipping of the histone tails, cleaves the N-terminal tails of histone H3 soon after alanine residue 21 showing a bias to those tails carrying repressive signals (Santos-Rosa *et al.*, 2009). The function of clipping found at the promoter bound nucleosomes, therefore it is suggesting a localised expulsion of repressive marks and protein promoter at the time of expression. The nucleosomes are displaced during transcriptional event through eviction of histone by histone tail clipping (Santos-Rosa *et al.*, 2009).

1.3.10. Crotonylation

Histone lysine crotonylation is a similar modification like acetylation. This modification is associated with active transcription. This modification enhance the active genes by enriching at the promoters in human somatic cells. The crotonylation takes place on the ϵ -amino group of lysine. The crotonyl group (C_4H_5O) has two extra carbon and hydrogen atoms that acetyl group (C_2H_3O) (Tan *et al.*, 2011). The coactivator p300 plays a role in histone modifications as it is a crotonyltransferase as well as acetyltransferase, however it has more activity as crotonyltransferase than acetyltransferase for histone modification on a potent transcriptional activator. The availability of crotonyl-CoA and acetyl-CoA determines the histone modification will be either crotonylated or acetylated (Sabari *et al.*, 2015).

1.3.11. Chromatin Remodelling

All the modifications on histones provide a code of signalling to change the chromatin architecture for allowing the progression of fundamental nuclear

processes. The structural alteration of nucleosomes is directed by the histone tails themselves by ATP-dependent chromatin-remodelling complexes. Alteration at histone tails of the histone proteins directly affects several processes such as histone dimer exchange, nucleosome sliding and DNA wrapping (Ferreira *et al.*, 2007). As an example, truncation of the different histone tails affects the nucleosome mobility. The deletion of the H4 and H2B tails reduces the nucleosome mobility but, truncation of the H2A tail promotes nucleosome mobility (Ferreira *et al.*, 2007). On the other hand, deletion of tails of H3 has great influence on nucleosome structure. Truncation on H3 tails leads to the reduction of the stability of histone-DNA contacts as a result the wrapping of nucleosomal DNA is altered within the nucleosome core. Due to this altered stability, the DNA from the octamer structure has been removed and destabilise the H2A/H2B histone dimer to allow the access the underlying genetic information (Ferreira *et al.*, 2007). There are four types of chromatin remodelers classified by their domain distribution. They are the Switching defective/Sucrose Nonfermenting (SWI/SNF), Imitation SWI (ISWI), Chromodomain, Helicase, DNA binding (CHD) and Inositol requiring 80 (INO80) remodelers. All four families share a common catalytic subunit, ATPase domain, which breaks the histone-DNA contacts by using ATP dependent DNA translocation to redefine promoter architecture (Clapier and Cairns, 2009).

1.4. Lysine Acetyltransferase Enzymes

The importance of histone acetylation by HATs is well studied for chromatin biology as it is important for transcriptional regulation. Several proteins have

been identified to possess the intrinsic acetyltransferase activity and some of those are directed towards non-histone lysine residues as well. The ChIP-Sequencing mapping and high-resolution mass spectrometry data have revealed how protein acetylation has a regulatory effect on cellular processes (Choudhary *et al.*, 2009). Exploring the entire set of acetylated proteins is called 'acetylome' (Smith and Workman, 2009). Lysine acetyltransferases (KATs) are highly conserved from yeast to humans and characterised into different groups depending on the sequence comparisons. Three main groups have been classified such as Gcn5-related N-acetyltransferase (GNAT), p300/CBP and MYST families. These groups of KATs are intensely studied due to their high conservation and functional links with disease. These proteins also play a role as transcriptional coactivators in various nuclear and cytoplasmic processes.

KATs catalyse the transfer of an acetyl moiety from an acetyl-CoA molecule to the ϵ -amino group of a lysine residue of a protein. This acetyltransferase enzymatic activity of KATs is regulated in various ways such as post-translation modification, degradation or altering the subcellular localisation (Yang, 2004). Also most of the KATs found as a part of a multi-subunit complex with different combinations which leads to the distinctive characteristics of each KAT complex (Lee and Workman, 2007). KAT complexes show higher activity than their counterpart enzymes, which demonstrate their particular substrate targeting and implicating the non-catalytic proteins in the regulation of activity and substrate specificity (Yang, 2004), however the specificity of KAT still remains elusive, even though intensive characterisation of different complexes already explored.

1.4.1. CBP/p300 family

The CREB-Binding Protein (CBP) is a nuclear protein, which was first discovered as an interacting partner of the activated phosphorylated form of CREB, cAMP response element binding protein, transcription factor (Chrivia *et al.*, 1993), however its paralogue p300 was identified as a cellular factor bound with the adenovirus E1A protein (Eckner *et al.*, 1994). Both of these proteins are found in all the higher eukaryotes, even including worms, flies or plants, however they are not expressed in lower eukaryotes like yeast (Yuan and Giordano, 2002). CBP and p300 proteins consist of sequence identity and close homology region between themselves such as cysteine-histidine rich regions (CH1, CH2 and CH3), bromodomain, and the acetyltransferase domain.

The CBP and p300 act on chromatin related process such as DNA replication, repair and transcription as well as on non-chromatin linked process such as regulation of p53 turnover (Grossman *et al.*, 1998) and nuclear import (Ryan *et al.*, 2006). The HAT domain of p300/CBP protein consist two functional motifs, the coenzyme A (CoA) binding and catalytic motif and zinc finger. The crystal structure of p300 resolved the ambiguity of the catalytic mechanism of p300/CBP mediated acetylation.

1.4.2. GNAT family

Homologues of GCN5 have been found in most organisms from higher eukaryotes to lower eukaryotes. There are two related GCN5 acetyltransferases that have been classified in mammal such as GCN5 and PCAF (Verdone *et al.*, 2006). Both of these KATs are absolute necessary for

normal cellular function and they are considered global coactivators. These two proteins in vertebrates possess highly similarity in their sequences which is 73% of sequence identity and it's mainly focused on their highly homologous domain structure. The N-terminal half consist of PCAF homology domain and the C-terminal half contains the bromodomains and acetyltransferase domains.

GCN5 and PCAF plays role in cellular processes such as transcriptional activation, cell-cycle, differentiation and apoptosis. PCAF also works as a coactivator in myogenesis (Puri *et al.*, 1997). Various substrates have been discovered for GCN5 and PCAF and their major targets are mainly histones and nucleosomes. These enzymes also exist like MYST as part of multi-subunit complexes, such as the GCN5 complex containing ySAGA, hTFTC, hSTAGA and the PCAF complex containing hPCAF. Presence within the protein complexes has been found to modify their nucleosomal substrate specificity; for instance when incorporated into a complex, the acetylation target of GCN5 changes from H3K14 to H3K9, K18, K23, and to a lesser extent H2B and H4 (Guelman *et al.*, 2006). As well as locus-specific coactivator functions GCN5 and PCAF are able to acetylate non-histone substrates including p53 (Poux and Marmorstein, 2003), TFIIIE/F (Imhof *et al.*, 1997) and MyoD (Dilworth *et al.*, 2004). The acetyltransferase domain has been very well studied for various GNAT family members. This domain can be classified into two distinct regions; the structurally conserved central protein core associated with Acetyl-CoA binding, and a pronounced cleft flanked by N-terminal and C-terminal regions. These N-terminal and C-terminal segments have a structural divergence from other family members, and are implicated for histone substrate binding and catalysis (Trievel *et al.*, 1999).

1.4.3. MYST Family

The name MYST came from its four founding members in yeast and mammals, MOZ, Ybf2, Sas2, and Tip60. So far five human KATs come under this family, which are TIP60 (KAT5), MOZ (KAT6A), MORF (KAT6B), HBO1 (KAT7) and MOF (KAT8). This MYST family KATs are highly conserved in higher eukaryotes and orthologues have been identified in many lower eukaryotes. Sequence and structural comparisons show their catalytic domain consists of an acetyl-CoA binding motif and a zinc finger motif and is usually accompanied by a chromatin binding domain such as plant homeodomain-linked (PHD) zinc fingers, or chromodomains (Yang, 2004).

The consistency of the catalytic activity of MYST family proteins is very poorly defined so far. The Essential Sas2-related acetyltransferase 1 (Esa1) is a yeast homologue of TIP60, which is used as a prototypical enzyme to understand the structure and mechanism of MYST family (Berndsen *et al.*, 2007). After additional structural and functional investigations Yan *et al.*, (2002) re-assessed the mechanism and proposed the new double displacement model (ping-pong catalysis). This moves forward through transfer of the acetyl group to a cysteine from the acetyl-CoA cofactor, which is strictly conserved among the MYST family members. Then the acetyl group is transferred to histone lysine substrate from this acetyl-cysteine enzyme (Yan *et al.*, 2002). The ϵ -amino group of the target lysine among all the KATs is highly conserved despite their differing catalytic mechanism.

MYST family enzymes play a well-established role in histone acetylation by conducting a considerable proportion of all nuclear acetylation. The MYST KATs are known to be important for physiological functions such as

haematopoiesis (MYST3), neural development (MYST4), X chromosome dosage compensation (MOF) DNA replication (HBO1) and DNA damage response (TIP60), although they are likely to play a role in other functions as well. The understanding for the mode of action of KAT enzymes might play crucial role to design of inhibitors. Besides the MYST domain some of the MYST family proteins also consist with chromodomains, Plant homeodomain-linked PHD zinc finger domain and a region with high quantity of a particular amino acid residue (Figures 1.4)

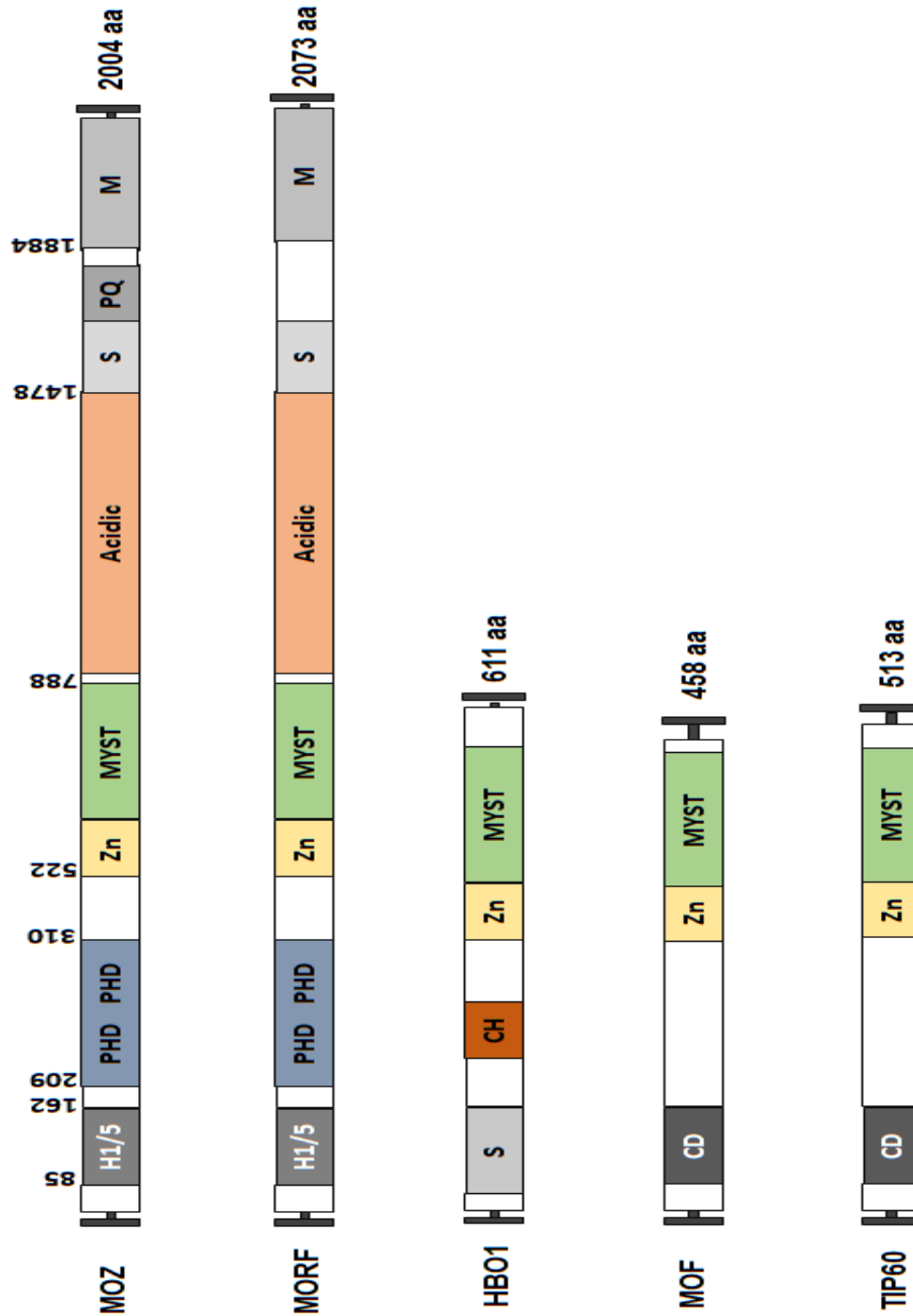


Figure 1. 4: Human MYST family KATs domain structures. The MYST domain of all the proteins of this family is highly conserved. Other than MYST domain there are chromodomain, double PHD-zinc finger (DPF) domain and acidic region is present.

1.5. MOZ/ MYST3

Monocytic leukaemia zinc finger protein aka MOZ, MYST3 is encoded by the KAT6A gene in human. This gene was first discovered at the breakpoint of a reciprocal chromosomal translocation t(8;16)(p11;p13), generating a fusion between MOZ and the gene encoding CREB Binding protein CBP or KAT8, in an aggressive form of acute myeloid leukaemia (Borrow *et al.*, 1996). Consistent with these knockout studies MOZ/MYST3 is essential for haematopoiesis, as ablation of the gene leads to loss of hematopoietic stem cells self-renewal and defective B cell function.

The modular structure of MOZ contains several discrete domain such as central MYST histone acetyltransferase domain, the amino-terminal domain, double PHD zinc finger domain and c-terminal acidic and other specific amino acid rich domain. The central MYST domain is highly conserved with the MYST family proteins, which plays the catalytic function. The crystal structure of MOZ MYST domain revealed that the conserved central core region is associated with the binding of acetyl CoA (Holbert *et al.*, 2007). The divergent N-terminal and C-terminal region contains TFIIIA type zinc finger and helix-turn-helix motifs respectively and these motifs helps MOZ to bind with DNA (Holbert *et al.*, 2007).

Subsequent studies identified MOZ in AML associated translocations with other genes including p300, TIF2, NCOA3, ASXL2 and LEUTX. In all cases the N-terminal region of MOZ including the double PHD fingers and MYST domain are expressed as a fusion with the C-terminus of the partner protein. Our laboratory has shown that MOZ-TIF2 function is dependent on interaction of the TIF2 portion with CBP and p300 (Kindle *et al.*, 2005) via a conserved helical motif termed AD1 (Sheppard *et al.*, 2001). This leads to changes in the

expression of genes that regulated by retinoic acid receptors and p53 (Kindle *et al.*, 2005; Collins *et al.*, 2006). Interaction of MOZ-TIF2 with CBP/p300 leads to their destruction by a calpain-dependent pathway and blocking this event can rescue the phenotype (Kindle *et al.*, 2010). Interestingly, the HAT activity of the MYST domain is not essential for the leukaemogenicity of MOZ-TIF2 (Deguchi *et al.*, 2003; Kindle *et al.*, 2005)

1.5.1. MOZ resides in multi-subunit complexes

MOZ is found as a catalytic subunit in several stable, nuclear multi-subunit complexes (Figure 1.5). Kitabayashi *et al.*, first identified MOZ as part of the AML1-CBF β transcription factor complex, whereas MOZ plays a role to stimulate AML1-mediated transcription (Kitabayashi *et al.*, 2001). After that, a number of interacting partners have been discovered and demonstrated the function in complex with MOZ. The most thoroughly studied and comparatively newly discovered multi-subunit complex containing MOZ along with the homolog of Esa1-associated factor 6 (EAF6) and Bromodomain- and PHD finger-containing (BRPF) proteins 1/2/3; BRPF family proteins are rich in modules

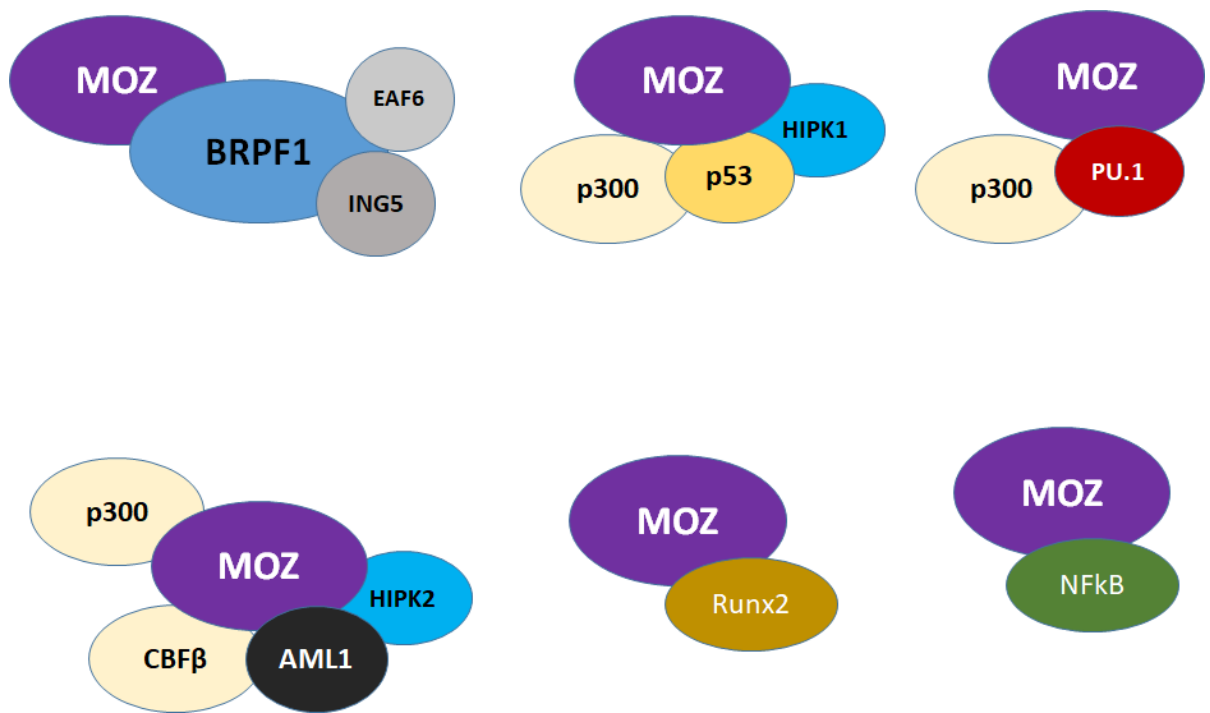


Figure 1. 5: Schematic representation of variety of multi-subunit complexes, where MOZ functions as the catalytic or repressive subunit.

commonly found in chromatin-associated factors such as PHD fingers, bromodomain and the chromo/Tudor-related PWWP domain. The presence of these domains suggests that the BRPF family protein is responsible for recruiting the MOZ complex to its site of action and consequently in chromatin regulation (Yang and Ullah, 2007). The inhibitor of growth 5 (ING5) member from the ING family of putative tumour suppressors is also found as part of the MOZ complex *in vivo*. The ING family of tumour suppressors are known for their involvement in many critical processes, but they do not possess intrinsic enzymatic activity. They may be functioned as and adaptors, providing a powerful link between chromatin-modifying activities and their genomic sites. More specifically, ING subunits have been found to be crucial for the KAT

acetylation of chromatin substrates, where they are might be responsible for the initial binding to the chromatin through their PHD fingers (Doyon *et al.*, 2006).

Further detailed studies to the molecular architecture of the quartet MOZ KAT complex has been provided, which is revealing the associated subunits regulate both the acetyltransferase and other functions of MOZ. BRPF1 binds to the MOZ MYST domain, which plays role as a scaffold for the association of ING5 and EAF6 subunits and the overall assembly of the complex (Ullah *et al.*, 2008). At the functional level, complex formation significantly increase the acetyltransferase activity of the MOZ MYST domain and further influenced its preferred target. The native complexes have different substrate specificity than to their enzyme counterpart, primarily acetylating H3 and H4 free histones and nucleosomal histone H3 (Ullah *et al.*, 2008). Additionally, interaction with BRPF1 was also found to stimulate the transcriptional activation potential of MOZ towards Runx2. Complex reconstitution studies are providing vital importance to our understanding of the molecular mechanisms underlying MOZ function; where it is proven that in physiological conditions non-catalytic subunits plays crucial role to regulate acetyltransferase and coactivator activities and refine substrate specificity.

1.5.2. Chromatin binding and Modification by MOZ/MYST3

MOZ possesses a double plant homeodomain (PHD) finger domain, which has strong binding affinity for the unmodified H3 tail. The interaction of the MOZ DPF with H3 tail peptides give rise to a conformational change in the sequence

of H3K4-T11 to form an α -helix in order to facilitate acetylation and/or methylation of histone H3 (Dreveny *et al.*, 2014). This enables the MOZ protein to sense for the presence of modifications to specific residues, including H3R2, H3K4, H3R8 and H3S10 and H3K14. For example, tri-methylation of H3K4 blocks the binding of H3 to the DPF, whereas H3K9me3 has no effect as it is facing away from the DPF. We have proposed that the helical tail structure positions H3K9 and H3K14 so that they can be more easily available for acetylation by the MYST domain. Indeed, comparison in kinetic assays of H3 acetylation showed that the purified DPF-MYST domain is 10-fold more efficient at acetylation of H3 than the MYST domain alone. Moreover, structures show that H3K14 acetylation stabilises the interaction by forming another contact with the DPF. (Dreveny *et al.*, 2014) Thus, the DPF has a very important role to determine whether H3 will be subjected to HAT activity.

1.5.3. MOZ Functions in Complex with BRPF1

Ullah *et al.*, (2008) have reported that MOZ associates with BRPF1, EAF6 and ING5 and that this complex stimulates the acetyl transferase domain activity in acetylation of nucleosomal histones. In fact, it has been suggested that the complex has a different substrate specificity than the isolated MOZ, targeting histone H3>H4. However, our recent studies showed that while the isolated MYST domain targets H4>H3, the function of the DPF domain is in substrate selection, as it binds H3 with higher affinity (Dreveny *et al.*, 2014). In fact the MOZ DPF induces an alpha helical fold of the H3 tail from residues 1-11, which allows it to probe for specific modifications. We have shown that H3K4 tri-methylation prevents the binding of MOZ to the H3 tail. Subsequently

MOZ can acetylate H3K9 and H3K14, and the acetylated form of H3K14 promotes enhanced interaction with MOZ by docking to binding pocket on the DPF surface (Dreveny *et al.*, 2014)

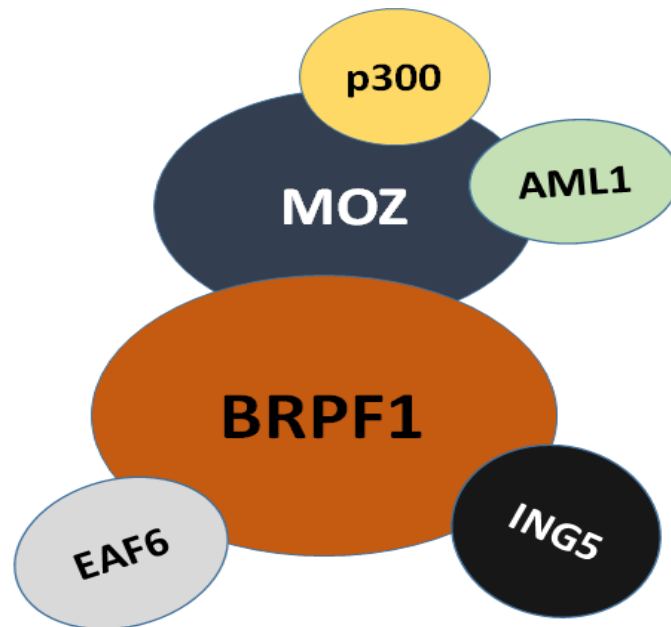


Figure 1. 6: Schematic illustration of MOZ/MORF complex formation with BRPF1, EAF6 and ING5 to enhance the acetylation on histone H3.

Although MOZ-TIF2 fusion protein retains the chromatin binding and modification functions, our group showed previously that MOZ-TIF2 expression in different cell lines causes major alterations in histone modifications and deregulate transcription by CBP-dependent transcription factors like p53 and retinoic acid receptors (Kindle *et al.*, 2005; Collins *et al.*, 2006; Kindle *et al.*, 2010). Interestingly, it was also shown that the transcriptional activity of Runx transcription factors, which are involved in regulation of blood and bone differentiation, was enhanced by MOZ-TIF2 fusion protein (Collins *et al.*, 2006). The alteration of MOZ activity in oncogenic fusion proteins leads to aberrant

gene expression that blocks the differentiation process of myeloid cell and leads to AML.

1.6. Bromodomain and PHD finger containing protein 1 (BRPF1)

As mentioned above, MOZ appears to exist in multi-protein complexes that influence its histone acetylation substrate specificity, in particular due to a bridging protein called BRPF1, Bromodomain and PHD finger containing protein 1 (You *et al.*, 2015). The BRPF1 gene is expressed ubiquitously and at the highest level in testes and spermatogonia. It is located at the chromosome 3p25.3, and was previously called BR140 (bromodomain 140kDa). The BRPF1 protein is localized within nuclei in presence of MOZ, and it is very similar in structure to two zinc finger proteins, AF10 and AF17, which are often involved in chromosomal translocations in AML.

BRPF1 contains a PZP domain domain, a bromodomain and proline-tryptophan-tryptophan-proline (PWWP) domain with the molecular weight of 140 kDa. The PZP is a PHD-Zinc Knuckle PHD motif, as found in NSD1, GASC1 and JADE proteins. Interestingly, the PZP (PHD fingers, zinc knuckle) domain of BRPF1 has similar H3 preference to MOZ double PHD finger domain in that it recognises unmodified H3K4 and H3K14ac. Humans contain three paralogous genes encoding closely related proteins BRPF2/BRD1 and BRPF3. BRPF1 has an additional domain in its N-terminus, which shows weak (<25%) identity to the Cactin C domain and also TAF1. However, the unique N-terminus of BRPF1 is conserved across species, even *Drosophila*.

BRPF1 plays role in the epigenetic regulation of gene expression and have been involved in human cancer (Yang *et al.*, 2015). Recently Shima *et al.* (2014) has shown that BRPF1 is required for leukemogenesis in association with MOZ-TIF2 fusion protein. BRPF1 and MOZ-TIF2 enhance the promotion of the Histone H3 acetylation in HOXA9 and HOXA10 gene that leads to the leukaemia.



Figure 1. 7: Schematic showing the multi-domain structure of BRPF1 with denoted different interaction and binding sites on the BRPF1. Multi-domain structure of other two BRPF family protein, BRPF2 and BRPF3, where brown CH domain stands for Cysteine-Histidine rich domain, grey BN1/2 domains are known as BRPF1 N-terminal domain 1/2, light orange EPC-I/II domains are known as Enhance polycomb-I/II, the blue PHD1/2 finger domains are known as Plant homeodomain 1/2, bright orange one is Bromodomain and the yellow PWWP is Proline-Tryptophan-Tryptophan-Proline rich domain.

The bromodomain from BRPF1 is required along with PWWP domain for chromatin targeting in fusion with MOZ (Laue *et al.*, 2008). Bromodomains are very specific to read acetylated lysine (KAc) residues on the histone tails of chromatin as well as KAc residues from other proteins (Palmer *et al.*, 2015). The bromodomains are known for its conserved structural motifs which recognise acetylated Histones. Recently X-ray crystal structure for BRPF1 bromodomain was revealed in complex with H2AK5ac histone peptides (Lubula *et al.*, 2014). Two alternatively spliced transcript variants of BRPF1 encoding different isoforms have been found in humans, one of which lacks a short sequence that is key to forming the acetyl-lysine binding pocket of the bromodomain. The function of this isoform *in vivo* is unknown.

Various subcellular localisation studies for BRPF1 have shown that BRPF1 remain in the cytoplasm after expression (Figures 1.8). On the other hand MOZ is a nuclear protein. In presence of MOZ cytoplasmic BRPF1 moves to the nucleus as well as a form of complex (Ullah *et al.*, 2008; Laue *et al.*, 2008). From this finding it is evident now that, to undergo any chromatin regulation by BRPF1, it needs to form a complex with MOZ to enter into the nucleus.

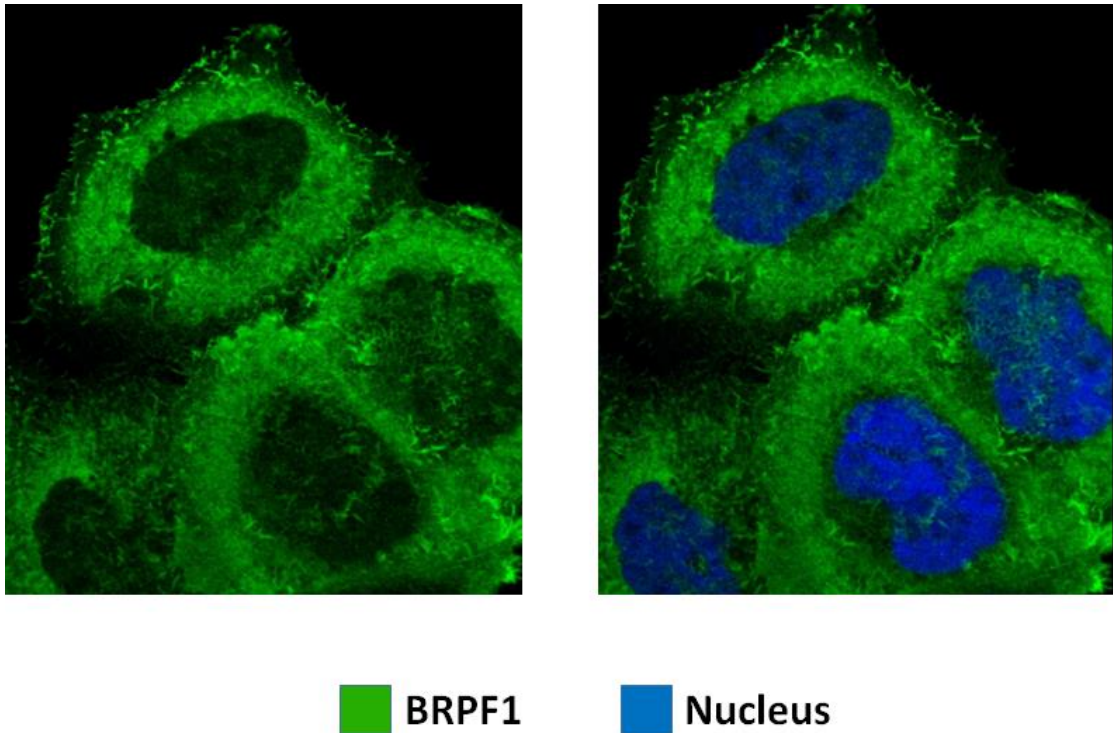


Figure 1. 8: Showing the subcellular localisation of BRPF1 in A451 cell lines. This image is evident that BRPF1 remain in the cytoplasm after expression. Green colour is indication presence of BRPF1 and blue colour is indicating the nucleus. The images obtained from: <https://www.proteinatlas.org/ENSG00000156983-BRPF1/cell#img>.

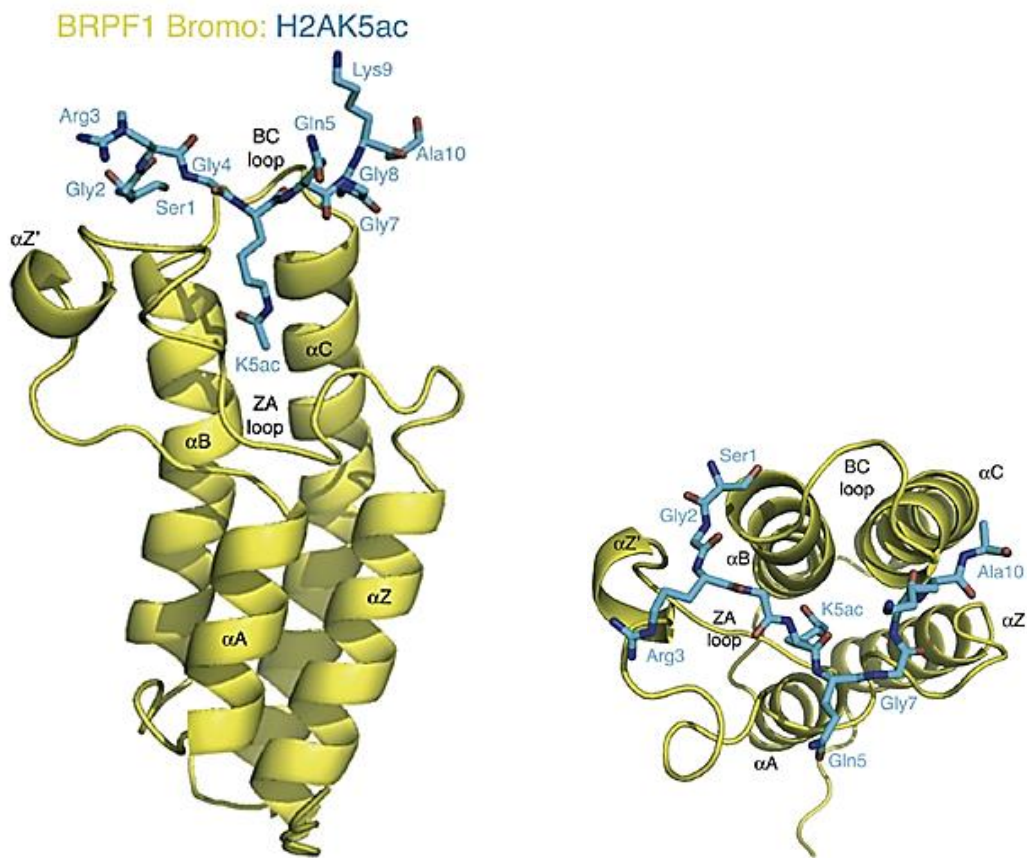


Figure 1.9: Structure of BRPF1 bromodomain in complex with H4K5ac. A) side view of the bromodomain and B) top view of bromodomain (adopted from Lubula *et al.*, 2014).

A systematic study of histone PTM binding preference by Poplawski *et al.* (2013) has shown that BRPF1 bromodomain has very high binding affinity to H4K12ac and H2AK5ac.

As bromodomains form very specific complexes with acetylated histones, several studies have developed small molecule inhibitors to target these complexes. One well-studied inhibitor is i-BET151, a so-called BET-bromodomain inhibitors block the BET bromodomain interactions acetylated

histones. It has been shown that i-BET151 blocks recruitment of BRD2 protein to H3K14ac and can successfully prevent MLL-fusion induced leukaemia in a mouse model (Dawson *et al.*, 2009). Recently Demont *et al.*, (2014) described another small molecule, 1,3-Dimethyl Benzimidazolones, which compete with acetylated lysine chain and inhibits binding of lysine chain by setting itself into the bromodomain pocket.

Though it has been demonstrated that BRPF1 bromodomain binds to several histone peptides such as H2AK5ac, H4K12ac, H4K8ac, H4K5ac, and H3K14ac; but the biological function of the BRPF1 bromodomain and the importance of these binding is not very clear (Palmer *et al.*, 2015). So there is a need to understand role of BRPF1/MOZ complexes in histone tail acetylation and using novel BRPF1 bromodomain inhibitors as experimental tools can give a new dimension to this

AIMS and OBJECTIVES

In this project we are going to investigate the structure and functions of the MOZ/BRPF1 complex. A thorough study on the MOZ/BRPF1 molecular interface will be performed using deletion mapping and yeast-two-hybrid interactions. Once minimal sequences required for interaction of MOZ and BRPF1 are identified, the effect of mutagenesis of specific residues or regions will be explored. It may then be possible to assess the effect of mutations in MOZ that disable BRPF1 requirement on its leukaemogenic potential, as described in previous studies from our group (Monteiro *et al.*, unpublished). We can also investigate whether BRPF1 interactions occur with other MYST family proteins. If a co-complex can be expressed, we may be able to initiate crystallisation trials for structure determination. We will also explore the function of MOZ on gene expression. We will try to create MOZ knocked out cell line and will validate how MOZ controls the expression of various genes.

Chapter 2: Materials and Methods

2.1. Sources of Materials

2.1.1. General Suppliers

All general laboratory chemicals were purchased from Fisher Chemicals or Sigma Aldrich unless otherwise stated and all are analytical grade. PBS (Phosphate Buffer Saline) tablets, purchased from OXOID Ltd, were used to prepare PBS solution. All solutions were made in double deionised water and if RNase free water was required and then deionised water was purified by passage through a Pur1te Neptune water purification system. The pH of solutions was maintained by checking the pH using a pH meter (Jenway 3510) against known pH standards. Where sterilisation required, the materials were sterilised by filtration through 0.2 micron filter or by autoclaving for 20 minutes at 120 °C.

2.1.2. Bacterial reagents

The bacteriological growth medium ingredients, tryptone and yeast extract were purchased from OXOID Ltd. *Escherichia. coli* strain DH5 α (Hanahan, 1983) was purchased from Stratagene.

2.1.3. Molecular biology reagents

Enzymes for restriction digestion and their compatible buffers were purchased from Roche Applied Science or New England Biolabs (NEB). DNA polymerase with their suitable buffers and supplements, molecular size markers and dNTPs were purchased from New England Biolabs (NEB). Gel purification kits were

purchased from York Biosciences, Nucleospin® plasmid kits from Macherey-Nagel and maxi-DNA kits from QIAGEN Ltd.

2.1.4. Tissue culture reagents

All tissue culture supplements were purchased from Lonza Biowhittaker® and plastic-ware was obtained from Helena Biosciences.

2.1.5. Biochemical reagents

SDS-PAGE apparatus, Precision Protein Standards, western blot transfer apparatus and Nitrocellulose transfer membrane were purchased from Bio-rad. 30 % (w/v) acrylamide solution was obtained from National Diagnostics. Complete Protease Inhibitor cocktail tablets were purchased from Roche. All primary antibodies, secondary antibodies conjugated with Horseradish peroxidase (HRP) were purchased from Abcam, Sigma-Aldrich, Roche and Millipore. Biotin conjugated histone peptides were purchased from Millipore and Anaspec Ltd. Protein A or G sepharose and magnetic beads were purchased from Abcam and Thermo Fisher respectively, Sigma-Aldrich, Roche and Millipore. Biotinylated histone peptides were purchased from Millipore or custom made by Peptide Protein Research and fluorescein-labelled histone peptides were also custom produced by Peptide Protein Research.

2.1.6. List of antibodies used

Name of the antibodies	Purpose	Dilution	Manufacturer	Product number
Rabbit anti-LexA primary antibody	Western blot from yeast	1:500	Merck Millipore	Cat# 06-719
Mouse anti-VP16 primary antibody	Western blot from yeast	1:500	Santa cruz Biotechnology	Cat# sc-7576
Mouse anti-FLAG-M2 antibody	Co-IP and western blot	Co-IP: 20µg/sample WB: 1:500	Sigma-Aldrich	Cat# F1804
Rat anti-HA antibody	Co-IP / western blot / immune-fluorescence	Co-IP: 20µg/sample WB: 1:500 Immuno-fluorescence: 1:300	Roche	Cat# 11867923001
Mouse anti-β actin antibody	Western blot	1:500	Sigma-Aldrich	Cat# A5441
Mouse penta-His antibody	Western blot	1:500	Qiagen	Cat# 34660
Chicken anti-mouse IgG-HRP secondary antibody	Western blot	1:5000	Santa Cruz Biotechnology	Cat# sc-2954
Chicken anti-Rat IgG-HRP secondary antibody	Western blot	1:5000	Santa Cruz Biotechnology	Cat# sc-2006
Goat anti-Rabbit IgG-HRP secondary antibody	Western blot	1:5000	Santa Cruz Biotechnology	Cat# sc-2004
Donkey anti-Rat IgG (H+L), Alexa Fluor 594	Immuno-fluorescence	1:500	Thermo Fisher Scientific	Cat# A-21209

Table 2.1. List of antibodies used in this whole project along with purpose, dilution, manufacturer's name and respective catalogue number.

2.2. Bacteriological preparation and culture

2.2.1. Composition of solutions and media used in bacterial procedures

Luria-Bertani (LB) medium: 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl at pH 7.0. To solidify the medium 2 % (w/v) bacteriological agar was added.

Solid Growth Medium: For LB agar plates 1 L of LB agar medium was sterilised. The appropriate antibiotic was added to the melted agar medium and 20-25 ml was poured into each petri-dish to cover up the surface and allowed to solidify at room temperature. Plates were stored at 4 °C for up to 1 month.

Ampicillin: 1000x stock solution: 100 mg/ml in 50 % ethanol.

Chloramphenicol: 1000x stock solution: 34 mg in 100 % ethanol.

Kanamycin: 1000x stock solution: 50 mg/ml in H₂O.

2.2.2. Culture of *Escherichia coli* strains

The *E. coli* strain DH5 α was used for all DNA cloning work. BL21 *E. coli* strains were used in protein expression experiments. Liquid bacterial culture was streaked onto LB agar plates containing the appropriate antibiotic and grown at 37 °C to obtain single colonies of bacteria. Liquid cultures were grown by inoculating a single bacterial colony into LB medium containing appropriate antibiotic. Cultures were then allowed to grow in universal tubes (Bibby Sterilin, Staffordshire, UK) or conical flasks at 37 °C with shaking at 140 rpm.

2.2.3. Preparation of *Escherichia coli* competent cells

Calcium chloride-magnesium chloride method was used to make *E. coli* DH5 α cells competent for DNA transformation. A single colony was inoculated in to 2.5 ml LB medium and grown overnight at 37 °C. The overnight grown culture was diluted 1:100 in fresh LB medium and allowed to grow at 37 °C until the Optical Density (OD_{600 nm}) reached 0.3. The cells were cooled down on ice for 20 minutes and followed by centrifugation at 4 °C, 3500 rpm for 5 minutes to harvest the cells. Cell pellets were resuspended in 20 ml of chilled 100 mM MgCl₂ and incubated on ice for 5 minutes. After another centrifugation, 10 ml chilled 100 mM CaCl₂ cells were added to resuspended the cells and incubated on ice for 30 minutes. Then cells were spun down and resuspended in 5 ml 70 mM CaCl₂ and 15 % glycerol solution mixture. 200 μ l cell suspensions were aliquoted into microfuge tubes were placed in liquid nitrogen for snap freezing. Competent cells were stored at -80 °C. The transformation efficiency of the cells was checked using control plasmid and the transformation serially diluted to allow the determination of the number of transformants: 1 μ g DNA should give $\sim 1 \times 10^6$ transformants.

2.2.4. Transformation into *Escherichia coli* cells

Plasmid DNA was transformed into competent DH5 α or BL21 *E. coli* strains. The competent cells were thawed on ice. 0.5 μ g of plasmid DNA was mixed gently with a 50 μ l cell suspension volume for each of the transformations. The cells were given heat shock at 42°C for 2 minutes followed by incubation for 2 minutes on ice. After that 500 μ l of LB medium was added and the reactions were incubated at 37°C for 1 hour. The culture was spun down at 7,000 rpm for

1 minute to pellet the cells and 450 µl of supernatant was discarded. Cells were resuspended in the remaining medium and spread onto LB agar plates containing the specific antibiotic. The plates were allowed to grow overnight at 37°C.

2.2.5. Long-term storage of bacterial cultures

To preserve transformed bacterial strains for long periods of time, glycerol stocks were prepared. Stocks consisted of a mixture of the bacteria in broth with sterile glycerol at an 80 % to 20 % ratio respectively. The contents were vortexed and stored at -80 °C.

2.3. Molecular biology techniques

2.3.1. Composition of solutions used in molecular biology methods

Gel loading buffer (5x): 15 g Ficoll400 dissolved in 10 ml 10 mM Tris-HCl-pH7.5 and 50 mM EDTA-pH8.0, containing 0.4 % orange G, 0.03 % bromophenol blue and 0.03 % xylene cyanol FF.

Tris-borate-EDTA (TBE): 40 mM Tris-base, 40 mM Boric Acid, 0.5 mM EDTA (pH8.0).

2.3.2. Small-scale plasmid purification from bacterial cells

5 ml of *E. coli* culture was grown overnight in at 37 °C with 140 rpm shaking. Cells were harvested by centrifugation and the plasmid DNA was extracted

from the cells using the NucleoSpin® Plasmid kit (Macherey-Nagel) using manufacturer's protocol.

2.3.3. Large-scale plasmid purification from bacterial cells

To extract plasmid DNA in large-scale, 1/100 fold dilution of suspended cells were inoculated into 300 ml LB media with specific antibiotic and allowed to grow overnight at 37 °C, 140 rpm. Plasmid DNA was extracted from these cells using the QIAGEN Plasmid Maxi Kit and its protocol, involving alkaline lysis of the cells, followed by immobilisation of the plasmid DNA onto a QIAGEN anion-exchange resin. The plasmid DNA was eluted and precipitated using isopropanol, followed by in 75 % ethanol and dried under vacuum. Finally the DNA pellet was resuspended in sterile water.

2.3.4. Phenol/chloroform extraction and ethanol precipitation of DNA

To purify nucleic acids from contaminating proteins, a phenol/chloroform extraction was performed. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 saturated with 10 mM Tris pH8.0, 1 mM EDTA) solution was added to the nucleic acid solution to eliminate protein contaminants. The sample was vigorously mixed until it emulsified and spun down at 13,000 rpm for 2 minutes in a desk-top centrifuge (Heraeus Biofuge pico). The aqueous layer was taken out in to a new microfuge tube and a further treatment of phenol:chloroform extraction performed to remove all the protein contaminants. Followed by an ethanol precipitation was carried out to collect the DNA.

2.3.5. Ethanol precipitation

DNA was precipitated by adding 2 volumes of 100 % ethanol and 0.1 volumes of 3 M NaOAc (pH5.2). The preparation was then centrifuged for 30 minutes at 13,000 rpm, 4 °C (ALC PK131R, rotor T528). The pellet was washed with 75 % ethanol and the centrifugation repeated to remove excess salt ions. Followed by drying the pellet under vacuum in a heated speed-vac (Savant DNA120 speed-vac) for 5 minutes and then was resuspended in sterile water.

2.3.6. Spectrophotometric quantification

The concentration and quality of DNA was measured by optical density measurement at 260 nm using the NanoDrop® ND-1000 UV-Vis Spectrophotometer.

2.3.7. Sequencing of plasmid DNA

Plasmid DNA was sent for sequencing to Source Bioscience Ltd Nottingham, UK and sequenced using Big Dye V3.1 chemistry and an AB13730xl Automated Sequencer.

2.3.8. Agarose gel electrophoresis

DNA fragments were separated according to their size on neutral agarose gels ranging between 0.7 % and 2 % (w/v) agarose. 0.5 µg/ml ethidium bromide was added to the gel and run in 0.5x TBE buffer according to the standard protocols (Sambrook J, 1989). DNA samples were mixed with 5x loading dye and loaded, along with NEB standard 1kb DNA ladder (0.2 µg/µl). Separated DNA bands

were then visualised by using an ultraviolet transilluminator (BioRad Gel Doc 2004) and BioRad Quantity One computer software. The size (kb) of DNA fragments were determined with comparison of standard molecular weight markers.

2.3.9. Isolation and purification of DNA from agarose gels

DNA bands were visualized using an ultraviolet light box and cut out using a sharp scalpel from the gel after separating by electrophoresis. DNA fragments were purified using the York Biosciences Gel/PCR purification kit according to manufacturer's protocol.

2.3.10. Oligonucleotides

Oligonucleotides were designed using ApE computer software and purchased from Sigma Genosys as lyophilised and desalted pellets. They were resuspended in sterile water according to the manufacturer's described volume to obtain 100 μ M concentration and stored at -20 °C.

2.2.11. Polymerase chain reaction (PCR)

DNA fragments were generated using PCR which is needed for the construction of recombinant plasmids. The DNA which was used for cloning purposes, were synthesized by DNA polymerase enzyme with proofreading 3' to 5' exonuclease activity (Phusion® High-Fidelity Polymerase). A typical reaction mixture contained 150 ng of DNA template and final 1x Phusion HF reaction buffer from 5x stock, 0.5 μ M of both the forward and reverse PCR primers, 0.2 mM of each

dNTP, 2 % DMSO, 1.25 units of Phusion® High-Fidelity DNA polymerase and made up to a final volume of 50µl with nuclease-free sterile water. Reactions were then set for 'hot start' PCR method in the Applied Biosystems 2720 thermal cycler. The DNA was initially denatured at 98°C for 30 seconds and then typical thermal cycling parameters were as follows: denaturation, 90°C for 10 seconds, annealing, 55°C for 30s and extension, 72°C for 30 seconds per kb of target DNA, for a total of 25-35 cycles. The temperature was maintained finally at 72°C for 10 minutes. After completion of the PCR cycle, 10 % of the PCR reaction was subjected into agarose gel electrophoresis for separation to verify that correctly sized DNA fragment had been generated.

2.3.12. Restriction digests

Digestion of plasmid DNA with restriction endonucleases was used to generate particular gene of interest with compatible ends for cloning. It was also used to verify the insertion of the gene of interest into a newly created plasmid. The selection of the enzymes was directed by the primer set used in the cloning. The buffers and required temperature was used according to the manufacturer (Roche) protocol.

In a typical digestion reaction the working volume was used 50 µl containing 1x reaction buffer, 2 units restriction endonuclease, template DNA (concentration depending on purpose) and sterile water to make up final volume to 50 µl. Digestion reactions were generally performed at 37 °C for 2 hours.

2.3.13. Phosphatase treatment

To prevent reannealing and self-ligation of linearized cloning vector DNA before preparing for ligation, it was treated by 1.0 μ l of Thermosensitive Alkaline Phosphatase (Promega, 1 U/ μ l - TSAP) in 1x Promega Multicore™ buffer for 30 minutes at 37 °C. TSAP catalyses the removal of the terminal phosphate groups of DNA. Following the reactions were incubated at 74 °C for 20 minutes to irreversibly inactivate TSAP.

2.3.14. Ligation reactions

Cut fragments were annealed in empty vector to generate recombinant plasmids using T4 DNA ligase (Promega). Ligation reactions were performed incorporating a vector: insert ratio of 5:1. Typically, 100 ng of vector DNA and 20 ng of insert DNA was added along with T4 DNA ligase (1 μ l) and 1x T4 DNA ligase reaction buffer and sterile water to make up the volume to 40 μ l and incubated for ~12 hours at 4°C. Vector only (without inserts) was used for control with same set up like above. The ligated DNA was transformed in DH5 α *E. coli* (section 2.2.4) and confirmed ligation by PCR.

2.3.15. Generation of mutations/deletions by recombinant PCR

To mutate/delete specific nucleotides within a DNA sequence two oligonucleotides were designed (Appendix table A.2) and used as primers for PCR. The primers were complementary, except for the nucleotide mutations and spanned the region to be mutated. PCR reactions were carried out to extend and incorporate the mutagenic primers, resulting in nicked circular strands of mutated plasmid

DNA. PCR reactions were carried out with the following amendments; 200 ng of DNA template was used and the thermal cycling extension time was increased to 20 minutes at 72 °C. PCR products were then treated with 0.5 µl Dpn1 restriction enzyme to digest the methylated, nonmutated parental DNA template for 2 hours at 37 °C. Control (no oligonucleotides used) and sample reactions were separated by agarose gel electrophoresis to check for complete digestion of the parental template. The nicked, circular dsDNA was transformed into DH5K and the resulting single colonies used to prepare small-scale plasmid DNA preparations, which were then sent for sequencing to check for the incorporation of the mutated nucleotide sequence.

2.3.16. Colony PCR screening

Ligation colonies were picked from LB agar plates and used to inoculate 200 µl of LB (+antibiotic) cultures. These cultures were grown at 37 °C for 2 hours or until the culture was turbid. The turbid culture was used as template, in place of DNA in proceeding PCR reactions. TaqTM (*Thermus aquaticus*) DNA polymerase

(Invitrogen) was used to screen for bacterial cultures containing the appropriate DNA insert. Reactions consisted of the same reagents as previously stated with the following amendments: 1.25 units of Taq DNA polymerase and its corresponding buffer (1x), 1.5 mM MgCl₂ and 2.5 µl of bacterial culture instead of DNA template. Positive clones were identified by running the entire PCR reaction on an agarose gel.

2.4. Biochemical techniques

2.4.1. Composition of solutions used in biochemical methods

Lysis buffer: 20mM Tris-HCl pH 8, 200mM NaCl and 10mM imidazole.

Wash buffer (for His-tag protein purification): 20mM Tris-HCl pH 8, 200mM NaCl and 30mM imidazole

Elution buffer (for His-tag protein purification): 20mM Tris-HCl pH 8, 200mM NaCl and 250mM imidazole

SDS-PAGE loading buffer (4x): 62.5 mM Tris-HCl pH6.8, 40 % glycerol, 2 % SDS, 14.5 mM 2-β-Mercaptoethanol and a trace of bromophenol blue.

Tris-glycine-SDS transfer buffer (for high molecular weight protein transfer): 48 mM Tris base, 380 mM glycine, 0.037 % (w/v) SDS, 20 % (v/v) methanol.

Brilliant Blue® staining solution: 0.25% brilliant blue® (w/v) in 45% methanol, 10% acetic acid.

Destaining solution: 45 % methanol, 10 % acetic acid.

ECL developing solution: 1 ml Tris-HCl pH8.5, 6 µl H₂O₂, 22 µl solution A (90 mM p-Coumari Acid in DMSO), 50 µl solution B (250 mM Luminol in DMSO) to a final volume of 10 ml with ddH₂O.

Cytosol/Nuclear (C/N) Fractionation Buffer A: 10 mM HEPES pH7.9, 10 mM KCl, 0.1 mM EGTA, 1 mM PMSF, 1 mM DTT, 1x Complete protease inhibitors.

C/N Fractionation Buffer B: 20 mM HEPES pH7.9, 400 mM NaCl, 1 mM EGTA, 1 mM PMSF, 1 mM DTT, 1x Complete protease inhibitors, 10% glycerol.

Binding buffer: 50 mM Tris-HCl pH7.5, 250 mM NaCl, 0.1% NP-40, 1 mM PMSF, 1x Complete Protease Inhibitors.

Blocking solution (for western blotting): 3 % (w/v) BSA in 1x PBS.

Z buffer: 16.1 g Na₂HPO₄·7H₂O (60 mM final), 5.5 g NaH₂PO₄·H₂O (40 mM final), 0.75 g KCl (10 mM final), 0.246 g MgSO₄·7H₂O (1 mM final), 3.5 ml β-mercaptoethanol (50 mM final) in 1L water, adjusted to a final pH of 7.0. Store at 4°C.

Phosphate Lysis buffer for MOZ: 20 mM sodium phosphate (pH7), 250 mM NaCl, 30 mM sodium pyrophosphate, 0.1 % NP-40, 10 mM NaF, 0.1 mM Na₃VO₄, 1 mM PMSF, 1x Complete protease inhibitors.

2.4.2. Protein concentration determination

Concentrations of the extracted protein were determined using the Bradford assay method. For a typical protein concentration reactions 2 µl of sample was used along with 800 µl of water and 200 µl of Bradford reagent. The samples were thoroughly mixed and incubated at RT for 20 minutes. Absorbance of the sample was read at 595 nm and the concentration calculated relative to values of diluted Bovine Serum Albumin (BSA) standards of known concentrations.

2.4.3. Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

One-dimensional polyacrylamide gel electrophoresis is generally used for protein separation according to their molecular weight. SDS-PAGE gels consist of the appropriate percentage resolving and stacking gel, according to the

molecular weight of the protein being analysed. Protein extracts were boiled to denature for 5 mins at 95°C in 4x SDS-PAGE loading buffer before to resolve on an 8-15 % SDS-polyacrylamide gel. Gels electrophoresis was performed using a mini-gel system (Protean II, Bio-Rad) in 1x running buffer at a constant voltage.

2.4.4. Staining of SDS-polyacrylamide gels

After completion of the electrophoresis the gels were submerged in 0.25 % Brilliant blue solution for 20 minutes and then two rounds of 20 minute treatment in destain solution. To remove the excess stain completely from the gel, it was submerged in distilled water for overnight. For preserving the stained gels, they were placed on filter paper, covered with Saran wrap and dried (80 °C, 1 hr) under vacuum (Biorad Hydrotech Vacuum Pump and Gel Dryer).

2.4.5. Western blotting and immunodetection

For western blotting proteins were separated according to their molecular weight on SDS-PAGE gels. The separated proteins were then transferred to nitrocellulose membranes by electrophoresis in 1x transfer buffer using wet transfer apparatus. Membranes were stained in 1 % Ponceau S solution and then washed with sterile water to confirm the transfer.

After transfer the proteins on the nitrocellulose membranes were submerged in blocking solution for 1 hour at room temperature to block non-specific antibody binding sites. Membranes were then incubated with primary antibody at 4°C overnight. Primary antibody concentration was 1:500 in blocking solution. Next

day the nitrocellulose membranes were washed three times (3 x 5 min) in 1x PBS and 0.1 % Tween 20 solution, followed by second incubation in secondary antibody for 60 minutes at room temperature. HRP-conjugated secondary antibodies concentration was 1:5000 in 5% milk in PBS solution. Non-specific secondary antibody was washed (3 x 5 min) in 1x PBS containing 0.1 % Tween 20. Membranes were covered for 1 minute with 5ml of ECL developing solution, after 1 minute the membrane was taken out from the solution and wrapped in saran wrap and then protein:antibody complexes were detected by ECL chemiluminescence. The chemiluminescent signal was captured by the Luminescent image analyser (Fujifilm LAS-4000).

2.4.6. Small-scale bacterial expression of HIS-tagged proteins

Single colonies after transformation were inoculated into 2.5 ml of LB-Kanamycin media and allowed to grow for 2 hours at 37 °C. After that 1 ml was placed into 1.5 ml microfuge tubes. As a control 25 µl of water was added to one and 25 µl of 20 mM IPTG inducer (final concentration 0.5 mM) to the other. The rest 0.5 ml was stored at 4 °C for glycerol stock and future large-scale cultures. After another 2 hours growth at 37 °C, cells were spun down to harvest at 13,000 rpm for 5 minutes and the pellets was treated with 15 µl 4x SDS loading buffer. Samples were boiled for 5 minutes at 95°C and performed electrophoresis on SDS-PAGE gel. Then the gel was stained with brilliant blue solution to detect the presence of an induced band.

2.4.7. Large-scale bacterial expression, purification of HIS-tagged proteins

20mL of start-up culture was grown overnight at 37°C from the glycerol stock of the transformed BL21 cells. Next day the culture was expanded into 1 litre culture at 37°C containing kanamycin as selecting agent and after obtaining the OD (at OD₆₀₀) 0.8, 500 µl of 1M IPTG was added to obtain a final concentration of 0.5mM and the culture was then grown overnight at 18°C. After overnight incubation the cells were centrifuged. 10 ml Lysis buffer was added to the cell pellet and incubated on ice for 30 minutes followed by another incubation at -80°C for 1 hour. After thawing the cells on ice the cells were sonicated for 9 minutes on ice with 50% power (30sec on 30 sec off). The lysed cells were then centrifuged for 50 minutes at 14000 rpm at 4°C. The supernatant was collected into a new tube and proceeded for His-tagged protein purification step by adding 2 ml of pre-washed Nickle beads into the supernatant followed by an incubation for 1 hour at 4°C with continuous circular inversion of the tube. After that the beads were centrifuged at 2000 rpm for 5 minutes at 4°C and washed 3 times (3x 10 minutes) with 5 ml wash buffer to remove other unwanted proteins. The His-tagged protein was then eluted by incubating the beads in the 2 ml elution buffer for 1 hour at 4°C with continuous circular inversion of the tube. The tube was then centrifuged at 2000 rpm at 4°C for 5 minutes and the purified His-tagged protein was collected as the form of supernatant and quantified the concentration by Bradford assay.

2.4.8. Peptide binding and inhibitor assay

15µg of purified His-tagged protein was added to 375µl of binding buffer and incubated 1 hour at 4°C with rotation. In other hand 1.5µg of biotin conjugated histone peptide (Millipore/Anaspec) was incubated with 15 µl slurry of pre-washed Dynabeads® (M-280 Streptavidin-Invitrogen) in 375 µl of binding buffer at 4°C for 1 hour with rotation along with a negative control which containing no peptides. After 1 hour the protein solution and the peptide-beads solution mixed together and incubated overnight. Next day the beads were harvested using a magnet and washed 3 times at 4°C for 10 minutes each time. Then the beads were subjected to SDS-PAGE and proceeded to western blot to detect the peptide binding with the His-tagged protein.

For inhibitor assay, the inhibitors with required concentration were added in the binding buffer along with the protein during first incubation and then same method was followed as in the binding assay.

2.4.9. Yeast-2-Hybrid (Yeast transformation and β-Galactosidase Assay)

L40 strain of yeast was streaked and a single colony from a fresh plate with an inoculation loop was inoculated into ~10ml YPD in a universal tube. The culture was allowed to grow overnight at 30°C with shaking. The following morning, the optical density at OD600 was checked and diluted down to ~0.2-0.3 the allowed to grow at 30°C until it reached OD 0.8. After that the cells were spun down at 4000rpm for 5min and resuspended in 10ml sterile water followed by another centrifugation at 4000rpm for 5 minutes. Again the pellet was resuspended in 1.0 ml sterile water and transferred into a 1.5ml microfuge tube and spun down

at maximum speed for 15 sec and removed supernatant. 250µl 100mM LiAc (diluted from 1.0M stock) was added to the pellet to resuspend and incubated at 30°C for 15 minutes. For each transformation, 173µl of transformation mix was taken out (freshly prepared by mixing 1.2ml 50% PEG, 180µl 1.0M LiAc, 250µl carrier DNA [2mg/ml], 100µl water) into labelled 1.5ml microfuge tubes. The relevant volume of both transforming DNA was distributed into each tube (100 ng) and mixed vigorously. A control was prepared in a tube without plasmid DNA. The suspension of yeast cells was mixed properly and transferred 25µl to each transformation tube and mixed well by vortex. Then the transformation tubes were incubated at 30°C for 30min with shaking. After completion of incubation the tubes were subjected for heat-shock at 42°C in a water bath for 20 minutes. After that the whole transformation mix was spread on YSP-LW (Yeast selection plate containing all the amino acids except leucine (L) and tryptophan (W) for selection of the positive transformation) and allowed to grow 2-4 days at 30°C and re-streaked on YSP-LW plates for the β -Galactosidase assay.

For β -Galactosidase assay single colony was inoculate in 15ml YSM-LW (Yeast selection media containing all the amino acids except leucine and tryptophan) medium and allowed to grow overnight. Next day the samples were spun down at 4000 rpm for 5min at 4°C. The cell pellet was resuspended in 1.0ml sterile water and transferred the cell suspension to a 2ml micro-centrifuge tube on ice. Followed by a brief centrifugation for 15 seconds was done and discarded supernatant. 150µl Z buffer supplemented with β - mercaptoethanol was added to the pellet keeping the tubes on ice. One scoop of glass beads was added to each sample. Cells were then disrupted using disruptor to shake each sample

twice for 1 minute (1minute shake followed by 1minute incubation on ice). Then the tubes were spun down at 14000 rpm at 4°C for 30 minutes. The protein extracts were removed as supernatant into a new tube and the protein concentration was determined by Bradford assay. Next 20µl protein extract (or Z buffer for blank) was transferred to a cuvette and adjusted to 500µl with Z buffer. The preparation was mixed properly and allowed to equilibrate for 10 minutes at 30°C. The reaction was initiated by adding 100µl of 4mg/ml ONPG then mixed well and time counting was started. When the pale yellow colour appeared in the preparation, 250µl 1M Na₂CO₃ was added to stop the reaction the total time was noted. Then the absorbance was measured at OD 420nm.

The Specific activity was calculated from the β-Galactosidase assay for the protein-protein interaction using following formula.

$$\text{Specific Activity (nmol/mg/min)} = \frac{\text{OD}_{420} \times 0.85}{0.0045 \times \text{Protein} \times \text{Volume} \times \text{Time}}$$

2.4.10. Preparation of cytoplasmic and nuclear extracts

Cells were harvested by aspirating the DMEM media and scraping the cell layer off in 1 ml sterile 1x PBS. Cells were collected by centrifugation at 13,000 rpm, the supernatant removed and washed twice in PBS. The cell pellet was resuspended in ice-cold C/N buffer A (approximately 3x cell compact volume) and incubated on ice for 10 minutes. A 0.5% final concentration of NP-40 was added to each sample, vortexed for 20 s on ice for 5 minutes and finally

vortexed for a further 20 s. Samples were then spun at 13,000 rpm for 30 s and the supernatant (cytosolic fraction) was transferred to a fresh eppendorf tube. The resulting pellet was resuspended in C/N Buffer B and incubated on ice for 40 minutes. During this time period samples were vortexed for 15 s every 5 minutes. Again, samples were spun at 13,000 rpm for 20 minutes at 4 °C and the supernatant (nuclear fraction, nucleic acid-binding fraction) transferred into an ice-cold eppendorf tube. In order to extract any remaining insoluble proteins from the cellular debris, the pellet was finally resuspended in phosphate buffer. Samples were sonicated for 8 minutes 15 s on/ 15 s off and the cellular debris pelleted by centrifugation at 13,000 rpm, 20 minutes, 4°C. The supernatant (nuclear fraction, insoluble nuclear protein fraction) was transferred to a fresh tube and all fractions stored at -80 °C.

2.4.11. Co-immunoprecipitation

Protein extracts (750 µg) obtained from mock and co-transfected cells were diluted to 500 µl with lysis buffer. In order to pre-clear and remove unspecific extract binding proteins, 10 µl of Protein G Agarose (Roche) was added to each sample and incubated for 1 hour 4 °C with rotation. Samples were spun at 2,000 rpm for 2 minutes to pellet Protein-G beads and the extracts transferred to a fresh tube. Next, 2 µg of antibody was added and samples were incubated overnight at 4 °C with rotation (a no antibody control was also carried out with the co-transfected sample). Protein-G beads were pre-washed in lysis buffer, 25 µl added to each IP and samples incubated for a further 14 hours or overnight at 4 °C with rotation. Immunoprecipitations were washed thrice in lysis

buffer and subsequently subjected to SDS-PAGE and immunodetection alongside a 10% input as a control.

2.5 Cell culture

2.5.1. Maintenance of adherent cell lines

The human embryonic kidney, HEK293 cell lines were routinely maintained in Dulbecco's Modified Eagle Medium (DMEM) containing phenol red and supplemented with 1% (v/v) 0.2 M L-glutamine, 1% Streptomycin antibiotics and 10% (v/v) heat inactivated fetal calf serum (FCS). Cultures were grown in sterile plasticware at 37 °C in a humidified atmosphere containing 5 % CO₂. Adherent cells were passaged at 90% confluency for a maximum of 30 passages, at which point the cells were discarded. To split cells, DMEM was aspirated and the cells washed twice with sterile 1x PBS. Cells were resuspended by pipetting up and down and subsequently diluted in fresh DMEM in 10 cm² plates (for maintenance) or 6/12-well plates (for transient transfections).

2.5.2. Cryopreservation of cell lines

Cells were resuspended into fresh medium and followed by centrifugation and wash twice in PBS. Cells were then resuspended in 2 ml of appropriate media (for a 10cm² dish), counted (section 2.4.2) and diluted to the appropriate concentration. 400 µl (40%) of cell suspension in complete media was transferred to labelled cryovials containing 100 µl (10%) DMSO and a supplement of 500 µl (50%) FCS. The cell suspension was then gently pipetted

to ensure even mixing of DMSO, FCS and media. The vials were then stored at -20 °C until suspensions were frozen solid. The cryovials were then placed at -80 °C overnight, after which the vials were stored in liquid nitrogen.

2.5.3. Calcium phosphate-mediated transfection of adherent cells

Adherent cells were seeded at 20% confluency in 10 cm plates 24 hours prior to transfection. The following day, DMEM was replaced with fresh media 3-4 hours prior to transfection. For transient transfections, 50-80% subconfluent cells were transfected by using the calcium phosphate co-precipitation method. Cells were transfected with 10µg of the appropriate plasmid DNA (quantity requires optimisation for each plasmid) and empty vector to standardise the quantity of DNA in each sample. To each sample 12.5% (v/v) 2M calcium chloride solution was added and made up to a final volume of 800 µl with sterile nuclease-free water pH7.0. This mixture was added dropwise to an equivalent volume of 2x Hepes Buffered Saline (HBS) solution whilst vortexing gently. After a 20 minute incubation at room temperature, 1.5 ml was added dropwise to the appropriate cell plate. Following a 16 hour incubation at 37°C, 5% CO₂, cells were washed thrice with 1x PBS and maintained in DMEM for a further 24 hours. Cells were lysed at the appropriate time post transfection (24-48 hours).

2.5.4. Immunofluorescence for co-localisation

3 x 10⁵ cells were seeded onto 8 well µ-slide and when fully adherent, were transfected with the appropriate plasmid DNA via the calcium-phosphate mediated method (section 2.5.3). 48 hours post-transfection the following steps

were performed at room temperature in the well. Cells were gently washed twice with PBS and fixed by incubation for 10 minutes with 4 % (w/v) paraformaldehyde. Cells were washed four times in PBS to remove excess fixative and permeabilised by incubation for 2 minutes with 0.2 % (v/v) Triton-x-100 (in PBS). Cells were washed as above, and non-specific binding sites blocked by incubation in 3 % (w/v) BSA (in PBS) for 30 minutes. Primary antibodies were prepared in 3 % (w/v) BSA (in PBS) blocking reagent at 1:300 dilutions to a final volume of 250 µl per well. To prevent the µ-slide drying out, a large sheet of blue towel was soaked in water and placed in a sealed container and incubated for 90 minutes in the sealed container creating a semi-humidified environment at 4°C overnight. The wells were washed as above and Alexa-594-conjugated secondary antibodies were prepared in blocking reagent at a 1:500 dilution and 300 µl used per well. Cells were incubated for 30 minutes and covered with an aluminium foil box. Control cells were incubated in solely secondary antibody to determine the level of background fluorescence. The cells were washed again in PBS and finally incubated with 1:10,000 diluted Hoechst stain for 5 minutes. After a final set of washes in PBS from the wells, drained of excess solution and placed 250 µl of PBS into the well. Slides were stored at 4 °C, protected from light. Immunofluorescence images were captured using a Zeiss LSM510 Meta confocal microscope and images were processed in the LSM Zeiss Image Browser

Chapter 3: Mapping the Molecular Interface of the MOZ/BRPF1 Complex

3.1. Introduction

3.1.1. Background

The BioGRID database currently lists 44 unique proteins that are potentially capable of direct or indirect interaction with the MOZ/MYST3 protein. These are based on global interaction studies in mammalian two-hybrid screens, yeast two-hybrid screens or co-purification of protein complexes. At the outset of this study, several reports in the literature (particularly from the Cote and Hammerschmidt groups) provided robust evidence that MOZ associates tightly with BRPF1, Eaf6 and ING5, and that together these proteins act as co-regulators for haematopoietic transcription factors such as Runx1 and Pu.1 (Doyon *et al.*, 2006; Laue *et al.*, 2008; Ullah *et al.*, 2008; Lalonde *et al.*, 2013). Later studies also found roles for BRPF1/MOZ complexes in embryogenesis and development (You *et al.*, 2015). BRPF1 is essential for embryo survival and conditional knockout studies in mice showed it is essential for brain development (You *et al.*, 2014; You *et al.*, 2015) presumably as a consequence of its roles in activation of acetyltransferase activity, histone substrate choice and transcriptional regulation by MYST proteins MOZ, MORF and HBO1. Mutations or any abnormalities in BRPF1 in the MOZ complex along with ING5 and MEAF6 have been reported to lead to loss of acetylation at histone H3K23ac, and that is associated to the developmental anomalies such as smaller brain size, facial dysmorphisms and language delay (Yan *et al.*, 2016).

Mechanistically, it has been proposed that BRPF1, an enhancer of polycomb (EPC)-related protein, is essential to recruit MOZ and MORF HAT proteins to chromatin thus functioning as an essential bridging protein (Ullah *et al.*, 2008; Yan *et al.*, 2017). However, evidence to support this as a critical requirement is rather minimal, as our

group and others have shown that both BRPF1 and MOZ contain domains that enable them to bind avidly to histone tail modifications, suggesting they may have independent abilities to bind chromatin and thus can have independent functions (Dreveny *et al.*, 2014; Poplawski *et al.*, 2013; Lubula *et al.*, 2014). Gene expression databases indicate that not all cell types co-express these proteins, again suggesting that co-dependence is not a prerequisite to their function. Further, it has been proposed that BRPF1 modulates the substrate specificity of MOZ by making it more selective for H3 tails in chromatin rather than as free histones. However, we have shown that the MOZ Double PHD Finger (DPF) domain can bind the H3 tail, impose an alpha helical structure on the tail and both sample for existing modifications (H3R2me2, H3K4me3) and facilitate acetylation of H3K9 and H3K14 (Dreveny *et al.*, 2014). We showed that while the isolated MYST (HAT) domain acetylates H4>H3, a DPF domain facilitates substrate selection, as it binds H3 with much higher affinity. This induces an alpha helical fold of the H3 tail from residues 1-11, which is incompatible if specific modifications such as H3K4me3 are already present, as this modification prevents the binding of MOZ to the H3 tail. Subsequently MOZ can acetylate H3K9 and H3K14, and the acetylated form of H3K14 promotes enhanced interaction with MOZ by docking to binding pocket on the DPF surface (Dreveny *et al.*, 2014).

The precise roles of MOZ and BRPF1 proteins in regulating gene expression were unknown at the outset of this study. Subsequent reports found that knockout of BRPF1 resulted in differentially expressed genes, the majority of which were upregulated suggesting a repressor function of BRPF1 (Yan *et al.*, 2015). In contrast the MOZ acetyltransferase is thought of as a coactivator of gene expression.

CRISPR-CAS9 genome editing has been used in our group to target KAT6A and generate cell lines lacking MOZ expression (Monteiro *et al.* in preparation; Whitchurch *et al.* in preparation). RNA sequencing analysis performed on these cell lines indicate both upregulated and downregulated genes, indicating that the functions of MOZ in gene regulation are complex. Thus to try to elucidate the functions of MOZ/BRPF1, the major aim of this study was to define the molecular interaction interface between MOZ and BRPF1 to try to better understand the functional roles of these proteins by introducing mutations that disable the complex without impacting on other functions of MOZ. To achieve this we decided to use yeast two hybrid mapping, a system of which the laboratory has extensive experience (Heery *et al.*, 1997; Heery *et al.*, 2001; Coulthard *et al.*, 2004; Waters *et al.*, 2006; Chan *et al.*, 2013; Dreveny *et al.*, 2014; Fulton *et al.*, 2017).

A number of previous studies have attempted to define the molecular interface between MOZ and BRPF1 (Laue *et al.*, 2008, Ullah *et al.*, 2008, Shima *et al.*, 2014) with partially consistent conclusions. Laue *et al.*, (2008) used transfected tagged deletion constructs and co-immunoprecipitation studies in HEK293 cells to define amino acids 1-245 of BRPF1 as necessary and sufficient for binding full-length MOZ, whereas 1-149 was not. Similarly Ullah *et al.*, (2008) reported that the region 59-222 is capable of binding to full length MOZ/MORF, although the data was not quantitative and an intermediate score (+/-) was assigned to the region 192-543. Regarding the region of MOZ or MORF required to bind BRPF1, Ullah *et al.*, (2008) reported MOZ 507-810 and MOZ 507-783 were capable of interaction with BRPF1 full length but MOZ 507-760 was not. Therefore, they concluded that MOZ 760-810 is likely to contain the BRPF1 interaction site, although they did not demonstrate this region was

sufficient for BRPF1 binding, or required in the context of full-length MOZ. Moreover, Shima *et al.*, (2014) provided evidence that MOZ deleted for the sequence 760-783 retains binding to BRPFs; the same study identified 488-664 as the BRPF1 binding region. Both these studies have contradictions with each other and moreover, the experimental approaches used will detect both direct and indirect interactions, as other proteins present in human cell lines co-immunoprecipitate and potentially influence the results.

3.1.2. Objective of the chapter

To define in detail the MOZ/BRPF1 interface, yeast two hybrid mapping was undertaken, supported by other protein-protein interaction assays. The advantages of the yeast system are several fold; firstly there is no obvious homolog of BRPF1 in yeast cells; secondly while yeast do contain the MYST proteins ESA1P, SAS2 and SAS3, apart from the MYST domain none of these homologs shares a similar domain structure as MOZ and MORF; thirdly, the absence of other mammalian proteins that may contribute to complex formation allows analysis of direct interactions; finally the studies reported using mammalian cells based on immunoprecipitation of truncated proteins in cell extracts, have produced inconsistent data, possibly due to interference by other proteins that may reside in these complexes or the use of different cell lines. Moreover, we have previously used LexA-MOZ fusion proteins in yeast two hybrid assays and they are well expressed. Therefore we embarked on attempting to map interactions of MOZ and BRPF1 in yeast cells, but before moving to more direct interaction studies in yeast two hybrid, we attempted to validate the

detection of the MOZ/BRPF1 complex in mammalian cell lines using co-immunoprecipitation as well as co-localisation experiments.

3.2 Detection of MOZ/BRPF1 complexes in HEK293 cells

To validate that we can detect MOZ/BRPF1 complexes in human cells, we performed co-immunoprecipitation experiments using YFP-FLAG or HA-tagged recombinant proteins in HEK293 cells. Expression vectors for YFP-FLAG MOZ full-length was co-transfected with vector for HA-BRPF1 full-length protein. At 48 hours post-transfection, cells were harvested and lysed in phosphate lysis buffer as described in Materials and Methods to release soluble proteins. This MOZ specific phosphate lysis buffer was used for extraction of MOZ from cells because this buffer contains various protease inhibitor which keeps very delicate protein like MOZ stable. YFP-FLAG-MOZ constructs was efficiently precipitated with anti-FLAG.

As shown in Figure 3.1A, western blots confirmed the expression and successful immunoprecipitation of the YFP-FLAG MOZ full-length protein in the lysates which is detected above the 245kDa marker (1st lane is for input and 2nd lane is for Co-IP). HA-BRPF1, which is detected as a band of approximately 140kDa (Figure 3.1B). This band for BRPF1 at ~140kDa indicated that MOZ is interacting with BRPF1 and when MOZ was immunoprecipitated using anti-FLAG antibody, BRPF1 was also co-immunoprecipitated as a form of complex. Thus, these results confirmed the complex formation of MOZ and BRPF1 proteins in mammalian cells.

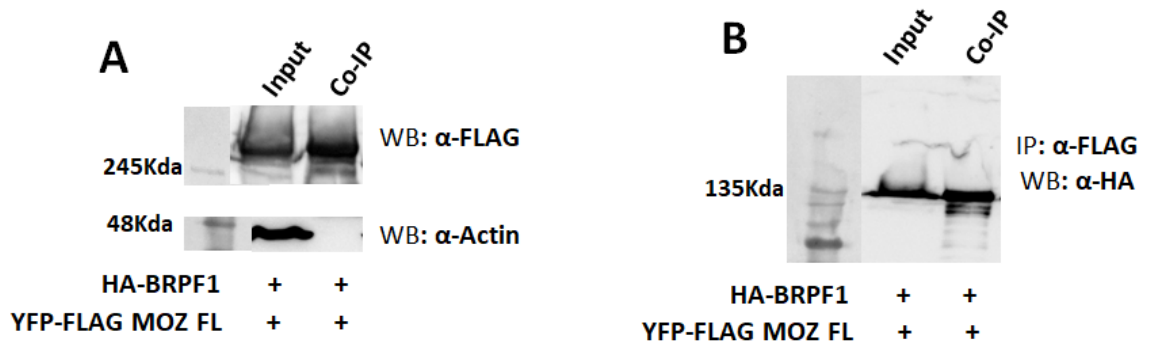


Figure 3. 1. Co-immunoprecipitation of MOZ and BRPF1 complex *in vivo*.

A & B) Western blot result is showing, YFP-FLAG MOZ full length was immunoprecipitated from transfected HEK293 cells by anti-FLAG antibody and HA-BRPF1 was detected by using anti-HA antibody. YFP-FLAG-MOZ full length was detected at ~245kDa and HA-BRPF1 was detected at ~140kDa. This result had shown that when MOZ is present in cell lysate and immunoprecipitated by using anti-FLAG antibody, BRPF1 was also co-immunoprecipitated because of their interaction, which supports the complex formation of MOZ and BRPF1. Expression of transfected proteins were confirmed in the input lanes.

3.3. Nuclear co-localisation of HA-BRPF1 and YFP-FLAG-MOZ

Having confirmed the co-immunoprecipitation of recombinant MOZ and BRPF1 full length proteins, we next examined the subcellular localisation of exogenous YFP FLAG MOZ and HA-BRPF1 proteins using confocal microscopy. MOZ was detected directly using its YFP tag and BRPF1 was detected using an anti-HA antibody. HEK293 cells were seeded into eight well glass bottom μ -slides and transfected with YFP FLAG MOZ full-length and/or HA-BRPF1 expression vectors (individually or together using empty vectors as controls). At 48 hours post-transfection, cells were fixed, permeabilised and incubated with anti-HA antibody followed by an anti-rabbit Alexa594 conjugated secondary antibody. As YFP-FLAG MOZ fluoresces directly, antibody detection was not necessary to visualise this protein. The nucleus was visualised with DRAQ-5 which stains the DNA (purple). YFP MOZ (green) and HA-BRPF1 (red) were detected, and images were merged (yellow) to check for co-localisation.

As shown in Figure 3.2 A, YFP FLAG MOZ was present in the nucleus of transfected HEK293 cells. A non-uniform distribution pattern in the nucleus was observed consistent with our previous studies (Kindle *et al.*, 2005; Collins *et al.*, 2006). Merging with nuclear stain confirmed that YFP-FLAG MOZ was entirely nuclear. As shown in the zoomed images the staining pattern suggested the accumulation of YFP-FLAG-MOZ in nuclear speckles and perinucleolar regions, which we have previously shown are distinct from PML bodies.

In cells transfected with HA-BRPF1 only, HA-BRPF1 was localised to the cytoplasm and appeared excluded from the nucleus (Figure 3.2 B), as observed by merging the image with the nuclear stain. This finding also suggested that BRPF1 may lack its

own nuclear localisation sequence, may be subject shuttling, or may be dependent on MOZ or other proteins for translocation to or retention in the nucleus. The distribution of HA-BRPF1 was not uniform and appeared to be localised to perinuclear regions or cytoplasmic regions close to the nucleus.

As shown in Figure 3.2C, HEK293 cells co-transfected with YFP FLAG MOZ and HA-BRPF1 showed a strong co-localisation of both proteins. Interestingly, in this case, both MOZ and BRPF1 proteins were both found to be exclusively nuclear, and the YFP-FLAG-MOZ distribution was similar to before (Figure 3.2 C). Merging the images confirmed the strong co-localisation with HA-BRPF1, consistent with reports from other groups (Laue *et al.*, 2008; Ullah *et al.*, 2008), and suggesting that interaction with MOZ or other MYST proteins may be essential for nuclear translocation of BRPF1. Interestingly several commercial antibodies raised against BRPF1 report cytoplasmic staining of the protein in cell lines, and similar findings are recorded on the Human protein atlas, with the majority of stain being cytoplasmic (Figure 1.8 in introduction). Similar results were reported by Ullah *et al.* (2008) using NIH3T3 cells, in which HA-BRPF1 was detected in cytoplasmic foci, unless co-expressed with MOZ, EAF6 or ING5.

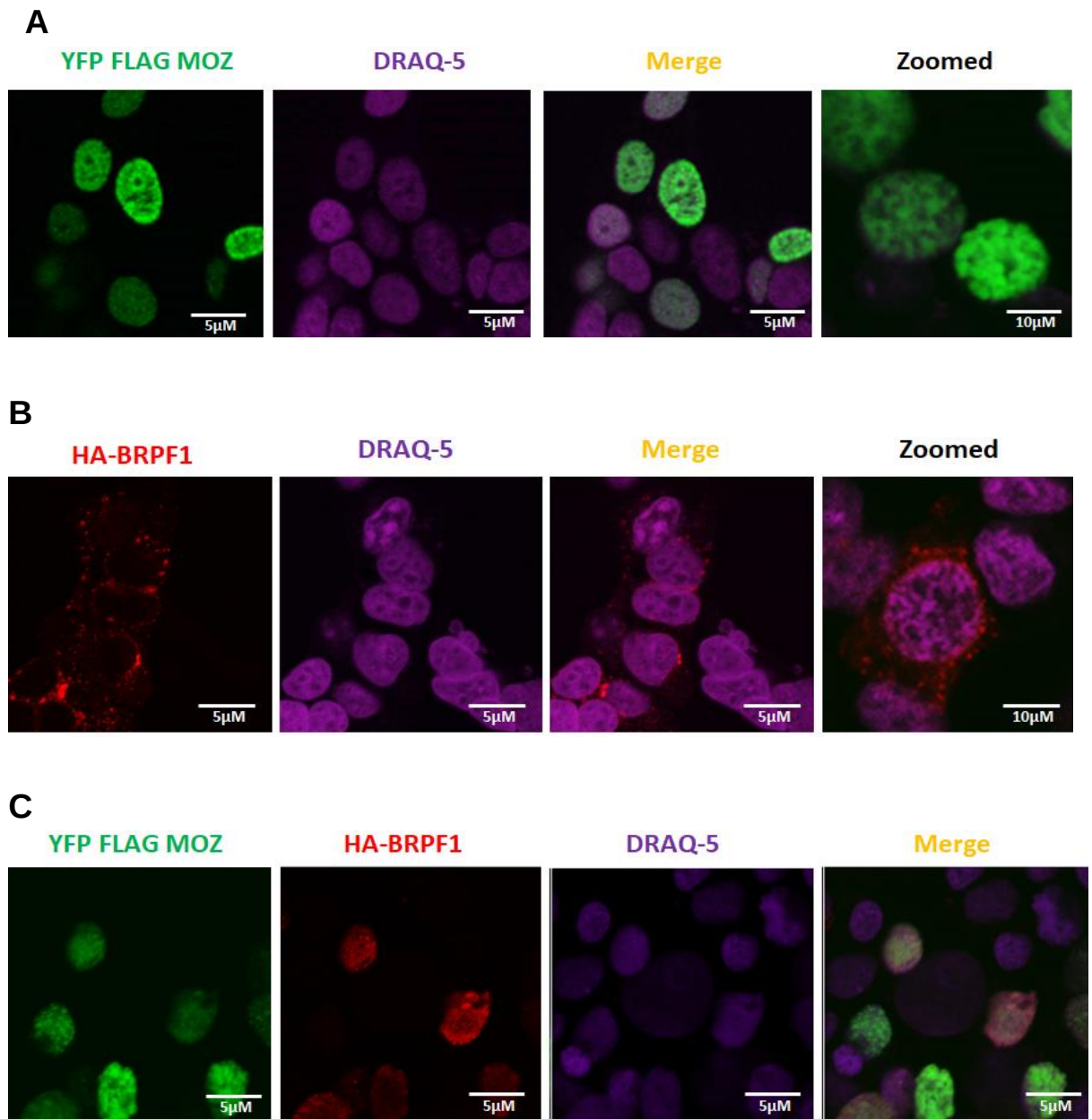


Figure 3. 2. Co-localisation of MOZ with BRPF1.

A) Immunofluorescence imaging showing HEK293 cells transfected with YFP-FLAG-MOZ full length and HA-BRPF1 together or alone and when YFP FLAG MOZ full length (green) transfected into HEK293, it was found in the nucleus of the cell due to nuclear localisation sequence of MOZ. Location of nucleus (purple) was also confirmed. **B)** When HA-BRPF1 (red) transfected individually, it expressed and remained in the cytoplasm only. It did not move into nucleus at all. **C)** Interestingly, when YFP FLAG MOZ and HA-BRPF1 co-transfected together into HEK293 cells, both the proteins found in the nucleus, which confirms the fact that, MOZ forms complex with BRPF1 and moved into nucleus together.

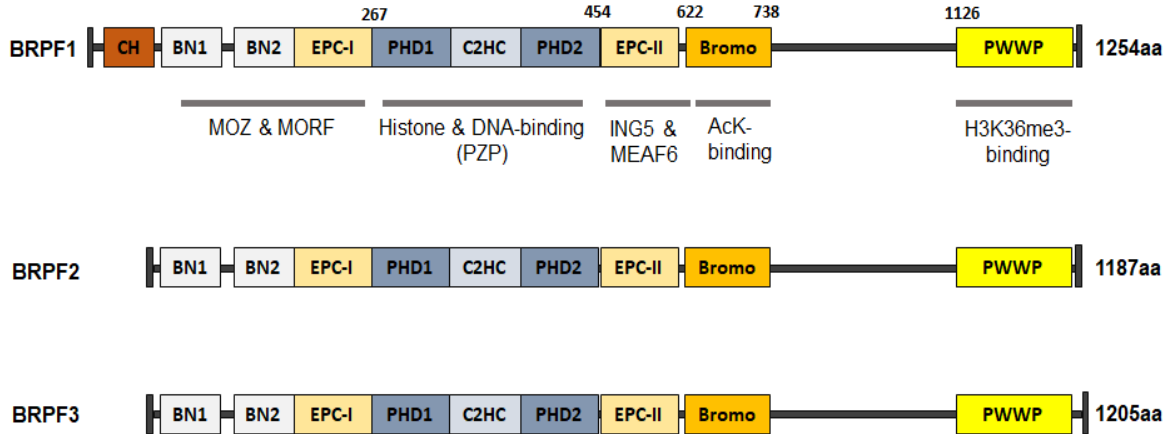
3.4. BRPF family proteins and sequence alignment of their N-terminal regions

The BRPF family of proteins share conserved domains that function as chromatin recognition/ binding domains, as shown in Figure 3.3 A. The PZP domain comprises two PHD finger domains linked with a C2HC zinc finger domain. The first PHD finger domain recognises histone H3 whereas the second PHD finger along with the zinc knuckle finger domain forms a non-specific DNA binding domain (Doyon *et al.*, 2006). The bromodomain recognises acetylated lysine sidechains from histone tails and the PWWP domain targets histone H3 especially when it is trimethylated (Lalonde *et al.*, 2012). The EPC-II, enhancer of polycomb-like motif (C-terminal to the PZP domain) contains the ING5 and EAF6 interacting site. The N-terminal half of BRPF1, prior to the first PHD finger contains four small motifs such as C2H2 zinc finger (CH), BRPF N-terminal 1 (BN1), BN2 and EPC-I (Liu *et al.*, 2012; Klein *et al.*, 2016 and Yan *et al.*, 2017).

The N-terminal BRPF1 sequence was aligned with other two BRPF family proteins, BRPF2 and BRPF3, to determine if any structural features would give a clue to BRPF and MYST proteins interaction (Figure 3.3 B). The multiple alignment was conducted using the *MultAlin*, Expasy based program (Multiple sequence alignment with hierarchical clustering. F. CORPET, 1988, Nucl. Acids Res., 16 (22), 10881-10890). The N-terminal regions of the BRPF family were compared, and the sequence alignments showed that residues 1-60 (approximately) are unique to BRPF1 (Figure 3.3 B), whereas the BN1, BN2 and EPC-1 domains are common to all three members of the family, consistent with the observation that they can each bind to MYST proteins (Mishima *et al.*, 2011; Yan *et al.*, 2015). The molecular mechanisms underlying apparent specificity in these interactions is unknown, and we hoped to

address this by defining the interaction interface in better detail. All three BRPF family proteins have very similar domain distribution, only BRPF1 has an extra N-terminal (CH) region (Figure 3.3 A).

A



B

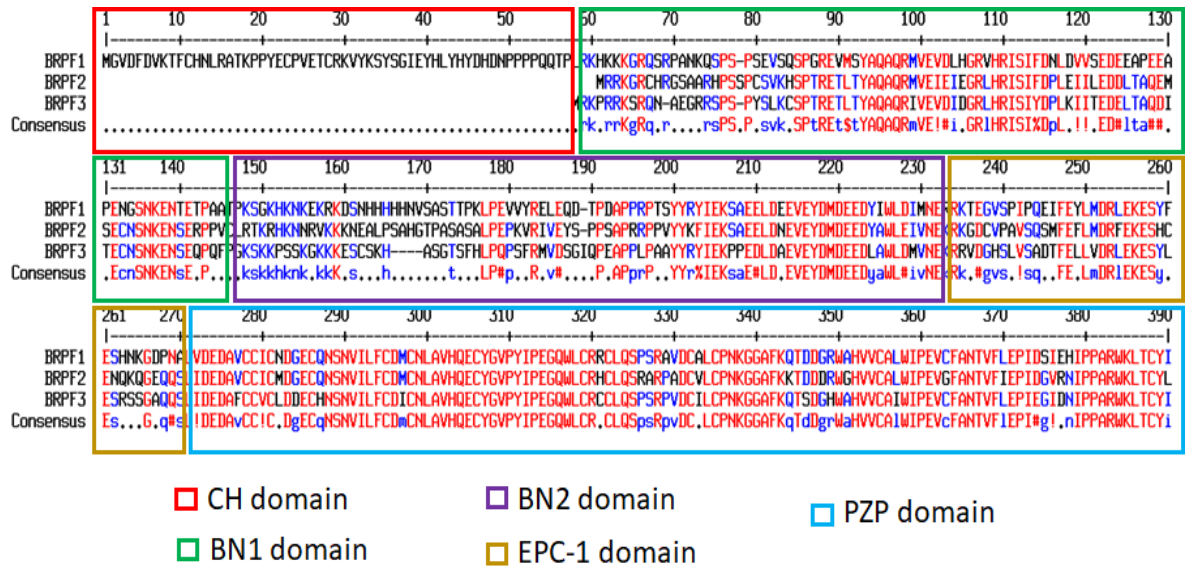


Figure 3. 3. Domain organisation and sequence comparison of different human BRPF family proteins.

A) Schematic representation of BRPF family proteins, showing multiple domains including two PHD finger domains linked with C2HC zinc finger (PZP domain), bromodomain and PWWP domain. The first PHD finger domain recognises histone H3 and the second PHD finger along with the zinc knuckle finger domain forms a non-specific DNA binding domain.

The bromodomain recognises acetylated lysine from histone tails and the PWWP domain binds H3K36Me3. Apart from the unique N-terminal region (CH) of BRPF1, domain structure is very similar to the other family members, BRPF2 and BRPF3. **B)** Sequence Alignments of Human BRPF1, BRPF2 and BRPF3 protein N-terminal regions where, red indicates identical amino acids, blue indicates similar side chains and the black is for non-conserved residues.

3.5. Sequence and structural similarities of MYST domains

MYST family proteins have been extensively studied both biochemically as well as structurally. This MYST family KATs are highly conserved in higher eukaryotes and orthologues have been identified in many lower eukaryotes. Sequence and structural comparisons show their catalytic MYST domain consists of an acetyl-CoA binding motif and a zinc finger motif and is usually accompanied by a chromatin binding domain such as plant homeodomain-linked (PHD) zinc fingers, or chromodomains (Yang, 2004). The MYST domain is also conserved even in yeast which contains the orthologues Sas3, Sas2 and Esa1 though Esa1 lacks three out of four amino acids for zinc binding (Figures 3.4 A).

The function of the MYST region comprises not only the acetyltransferase activity, which is responsible for the lysine acetylation, but it has also been implicated in interactions with BRPF1 proteins (Mishima *et al.*, 2011, Ullah *et al.*, 2008, Tao *et al.*, 2017 and Shima *et al.*, 2014). For example, crystal structures have been reported which show that the N-terminal side of the MOF MYST domain is required for interaction with MSL1 in flies and human cells to regulate dosage compensation. (Huang *et al.*, 2012, Kadlec *et al.*, 2011).

A sequence alignment of MYST HAT domains of human MOZ, MORF, HBO1, MOF, TIP60 and yeast ESA1P show a high similarity in their sequence, which is an overall 38% identity between them (Figure 3.4 B). This would predict that all these MYST HAT domains are likely to have a similar tertiary structure and available structures for MYST HAT domains were downloaded and visualised in UCSF Chimera, (Figure 3.4 C). This revealed that MYST HAT domains from MOZ (PDB code: 2RC4), ESA1P (PDB code: 3TO9), MOF (PDB code: 5J8C) and TIP60 (PDB code: 2OU2) have very

similar structure conformations (Holbert *et al.*, 2007; Wu *et al.*, 2007; Yuan *et al.*, 2011; McCullough *et al.*, 2016). Most of the MYST HAT domains consist of eight α -helices and eleven β -strands. Also there are four β -strands and an α -helix present in the central region of all the MYST family proteins. The MYST domains superpose very well with each other with R^2 value of 62.85 (Figures 3.4 C). The acetyl-CoA substrate binds within a characteristic motif at the central bottom region (Figure 3.4 C), which consists the Arg/Gln-X-X-Gly-XGly/Ala sequence at $\beta 9$ - $\alpha 4$ loop (Holbert *et al.*, 2007). This was noted that same HAT domain structures revealed a helix-turn-helix conformation, followed by a double β -strands loop on the C-terminal side of the domain (opposite the MSL1 binding site in MOF or ESA1P). In contrast in the MOZ MYST domain no structure was predicted for that region (amino acid residue from 739 to 747) due to low electron density, however the helix-turn-helix region before that unstructured loop area is well conserved structurally within these MYST proteins.

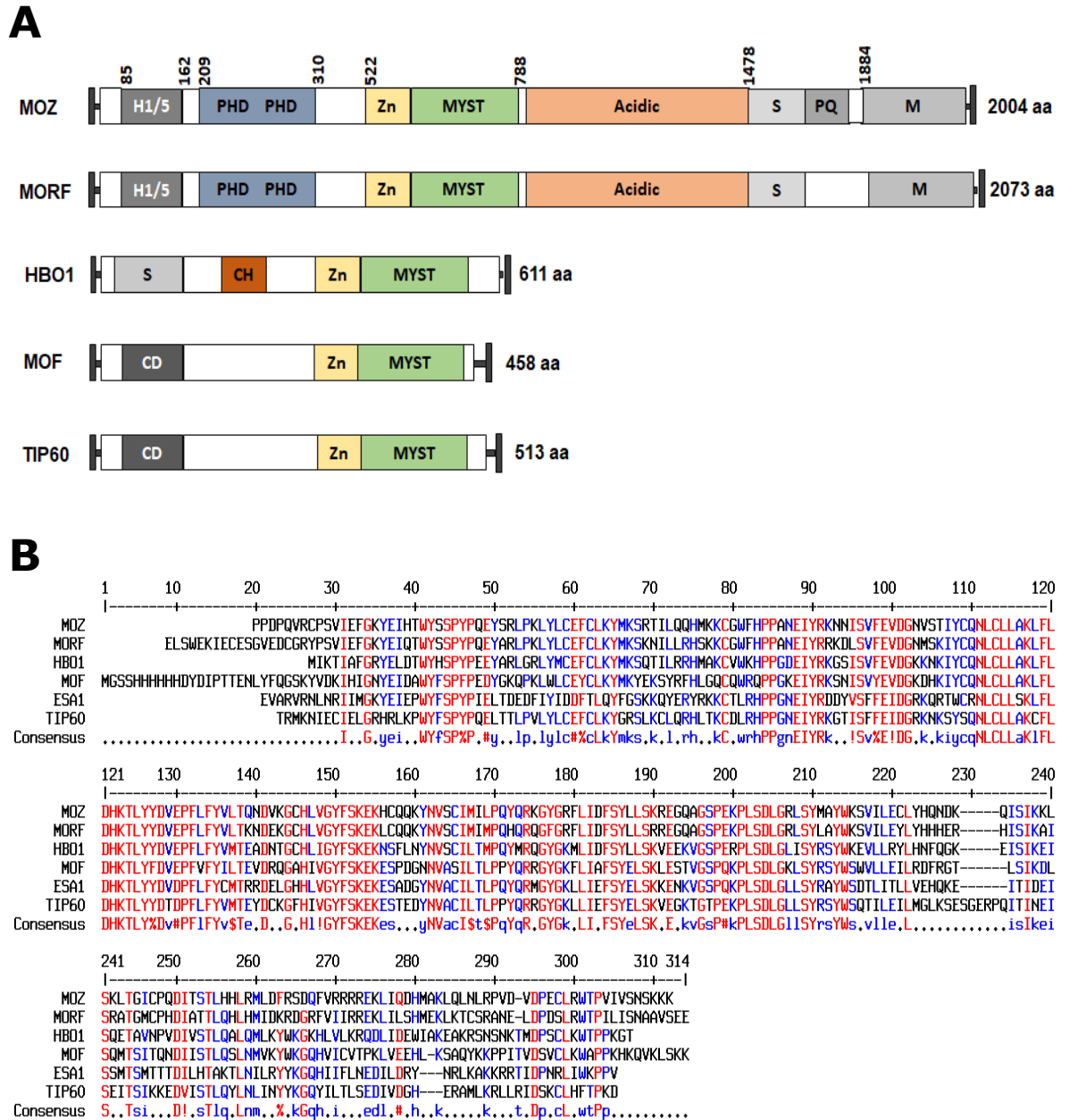


Figure 3. 4 (a). Sequence comparison of different MYST family proteins.

A) Schematic representation of MYST family proteins MOZ, MORF, HBO1, MOF and TIP60 showing multiple domains and all of them have common MYST domain. **B)** This figure is showing *MultAlin* Sequence Alignments of Human MOZ, MORF, HBO1, MOF and TIP60 along with yeast orthologue ESA1P MYST domain where, red colour indicates the complete matching of the amino acid sequences between MYST family members; blue colour indicates partial matching or similar side chain classified amino acid and the black colour is for no match at all. The MYST domains from the above MYST proteins have very high similarities, which is almost 38% sequence identity.

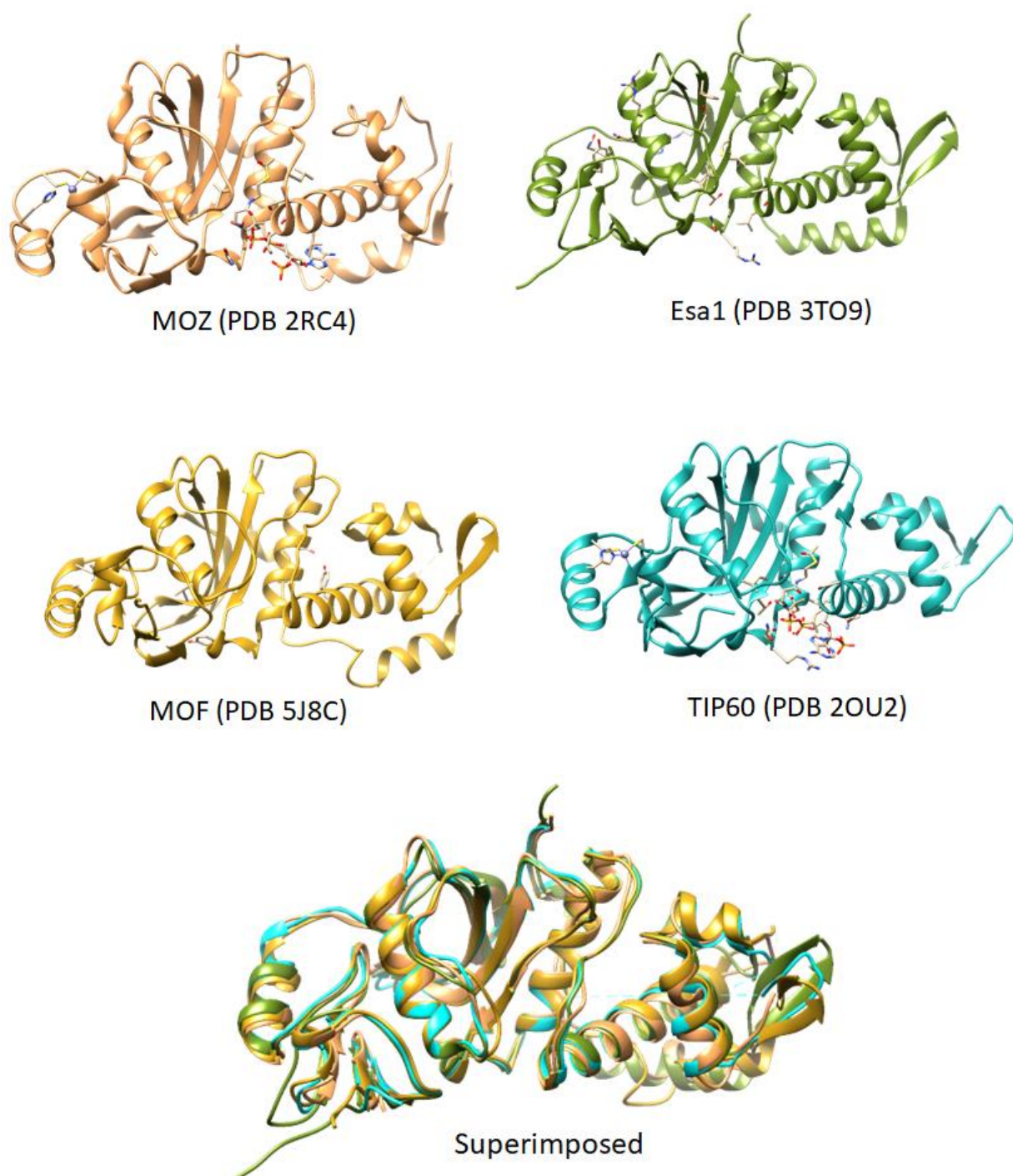


Figure 3. 5 (b). Structural comparison of different MYST family proteins.

This is showing the ribbon structures of the available MYST domains in Protein Data Bank, where the orange coloured structure is for MOZ MYST domain, olive green is for ESA1P, purple for HBO1, golden yellow for MOF and cyan is for TIP60. The last structure at the bottom is showing the alignment of all the MYST domains from MYST family proteins.

3.6. Mapping N-terminus BRPF1 for MOZ interacting site

With regard to the region of MOZ that binds to BRPF1, there is conflicting evidence in the literature such as according to Ullah *et al.*, (2008) the N-terminus of BRPF1 1-222 interacts with MOZ within the 760-810 residues, whereas Shima *et al.*, (2014) reported that the BRPF1 interaction site is located within residue 488 to 664, although neither study showed that these regions are necessary and sufficient. Understanding how complexes are formed between these proteins, and the nature of selective interactions between MYST and BRPF proteins, would enable us to generate mutant versions of BRPF1 or MOZ that are defective for complex formation, enabling us to study their functions. Therefore we undertook mapping the BRPF1 interaction site of MOZ using yeast two hybrid assays.

3.6.1 Determination of the boundaries

To determine whether we could reproduce the published findings and further define the MOZ/BRPF1 interface, we began by generating a series of VP16 AAD-BRPF1 constructs using human BRPF1 cDNA as the template, which was provided by the Yang group, McGill University, Canada. The boundary of those constructs were determined by the secondary structure prediction (Figure 3.5) using Phyre2 online secondary structure prediction software (The Phyre2 web portal for protein modelling, prediction and analysis, Protein Homology/analogy Recognition Engine V 2.0) to avoid the construct boundaries disrupting any alpha helix or beta strand. For cloning into the VP16 vector (pASV3modified, figure 3.7 A) (LeDouarin *et al.*, 2001 Methods in Molecular Biology) we utilised XhoI (forward primer) and SacII/KspI (reverse primer). Sequences were assessed using ApE software

(<http://jorgensen.biology.utah.edu/wayned/ape/>). The ApE program was used to check that the required restriction sites used for sub-cloning are absent within the sequences of interest. BRPF1 primers were designed to amplify fragments comprised within the region BRPF1 1-222.

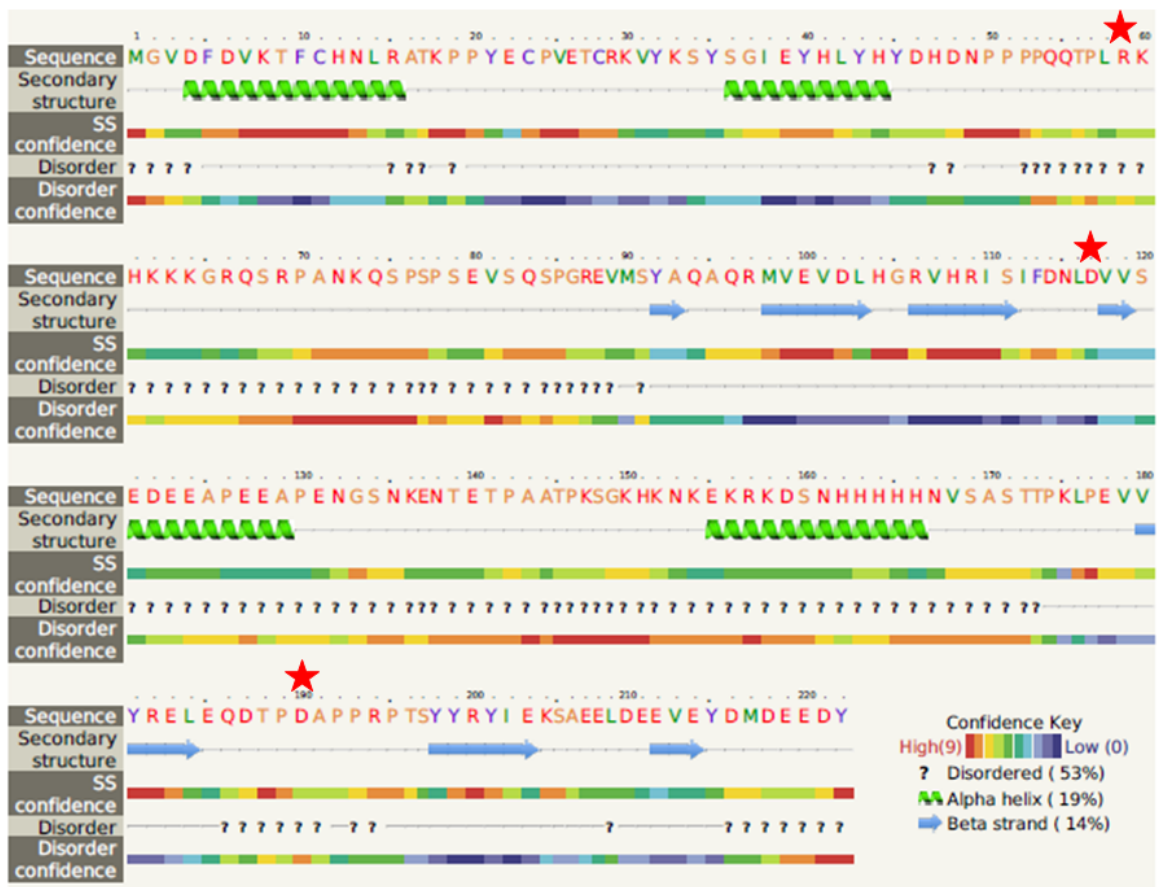


Figure 3. 6. Secondary structure prediction of N-terminal BRPF1.

Phyre2 prediction showing the predicted secondary structure of N-terminal BRPF1 (1-222), where the green coiled structure represents the alpha helices and the blue arrows are for beta strands. The red asterisks show the breakpoints for the BRPF1 constructs to avoid the disruption of an alpha helix or beta strand.

3.6.2 Sub-cloning strategy to generate VP16-AAD-BRPF1 constructs for yeast two hybrid assay

A selection of N-terminal BRPF1 sequences were fused with the yeast transcriptional activation domain of VP16 (AAD). The target vector map, pASV3mod, is shown in Figure 3.7 A. The bait protein sequence, for this project the MOZ constructs are fused with the bacteria repressor domain LexA (DBD). The boundaries with amino acid residue number of the N-terminal BRPF1 constructs, which are B1, B2, B3, B4, B5, B6 and B7, are listed in Table 3.1 and the domain distribution within these constructs are shown in the Figure 3.7 E. All these BRPF1 constructs fall within the CH domain to BN2 domain of BRPF1.

The N-terminal BRPF1 constructs were generated using the sub-cloning process, mentioned in Figure 3.6. The human full length BRPF1 cDNA was used as a template for the PCR reaction with the primers containing unique restriction sites, which were designed according to the multiple cloning site of the pASV3mod vector. The PCR products were run on a 1% agarose gel to confirm the right size of the constructs (Figure 3.7 C). The constructs and the target vector pASV3mod were then subjected to XhoI and SacII restriction enzymes to release the inserts and to cut and open the vector (Figure 3.7 B). After ligation of the digested inserts into the MCS of cut pASV3mod vector, the plasmid ligation mix was transformed into *E. coli* and single colonies for each construct were picked and colony PCR was performed for them. The products from the colony PCR, seen on a 1% agarose gel (Figures 3.7 D), confirmed the plasmids contained their respective N-terminal BRPF1 constructs and subsequently all constructs/inserts were confirmed by sequencing.

This sub-cloning strategy was the general cloning process, which was used to generate all the plasmids over the course of this project. The primer sequences and

sequences for all the constructs are used in this project are documented in Appendix I.

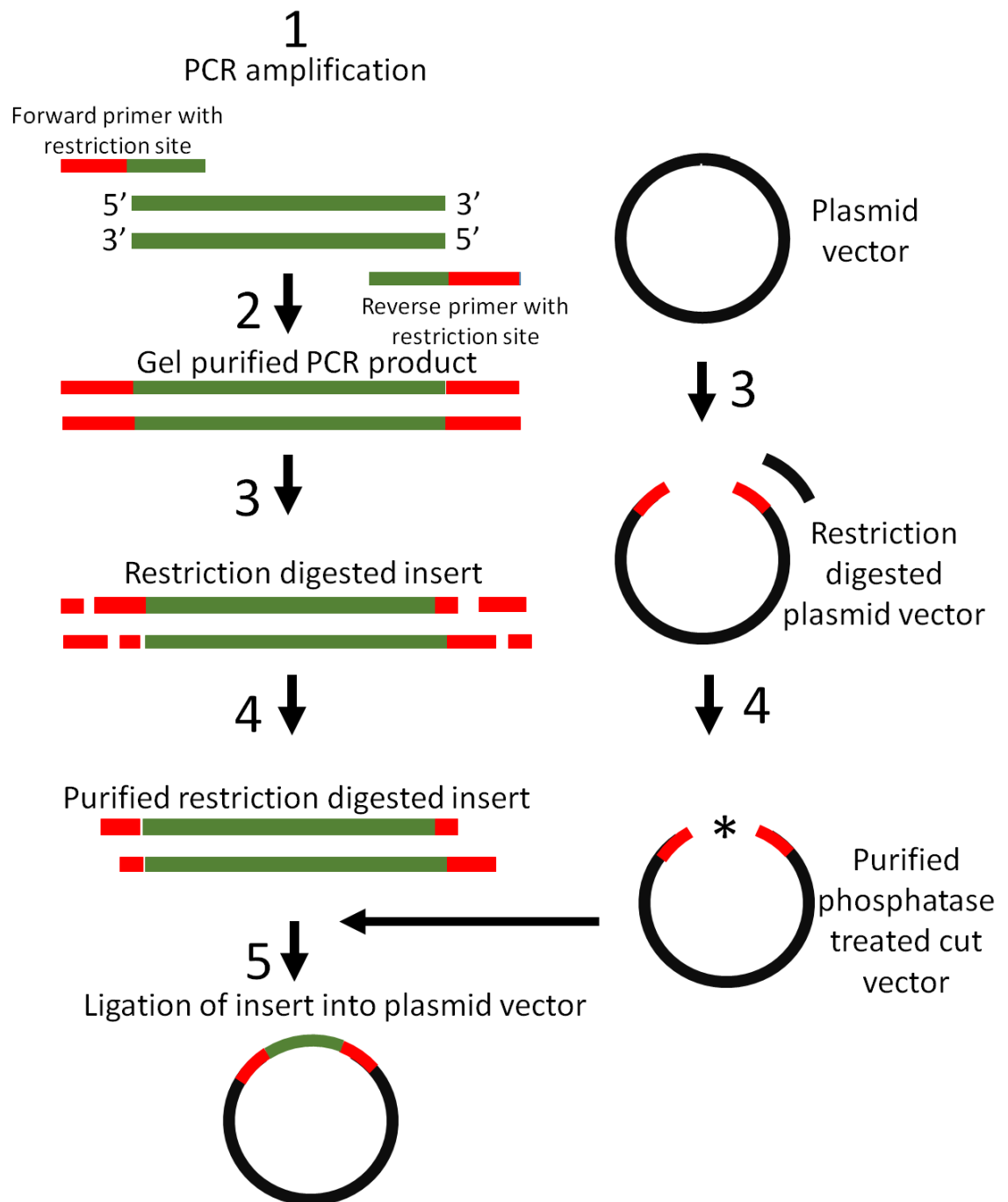


Figure 3. 7. Diagram for sub-cloning strategy.

Step1, PCR amplification of N-terminal BRPF1 cDNA inserts. **Step 2**, purification of PCR products. **Step 3**, Restriction digestion of vector plasmid and purified inserts. **Step 4**, Phosphatase treatment of digested vector plasmid. **Step 5**, Restriction digested vector and inserts were purified. **Step 6**, Ligation of purified and digested inserts into restriction digested and phosphatase treated cut vector.

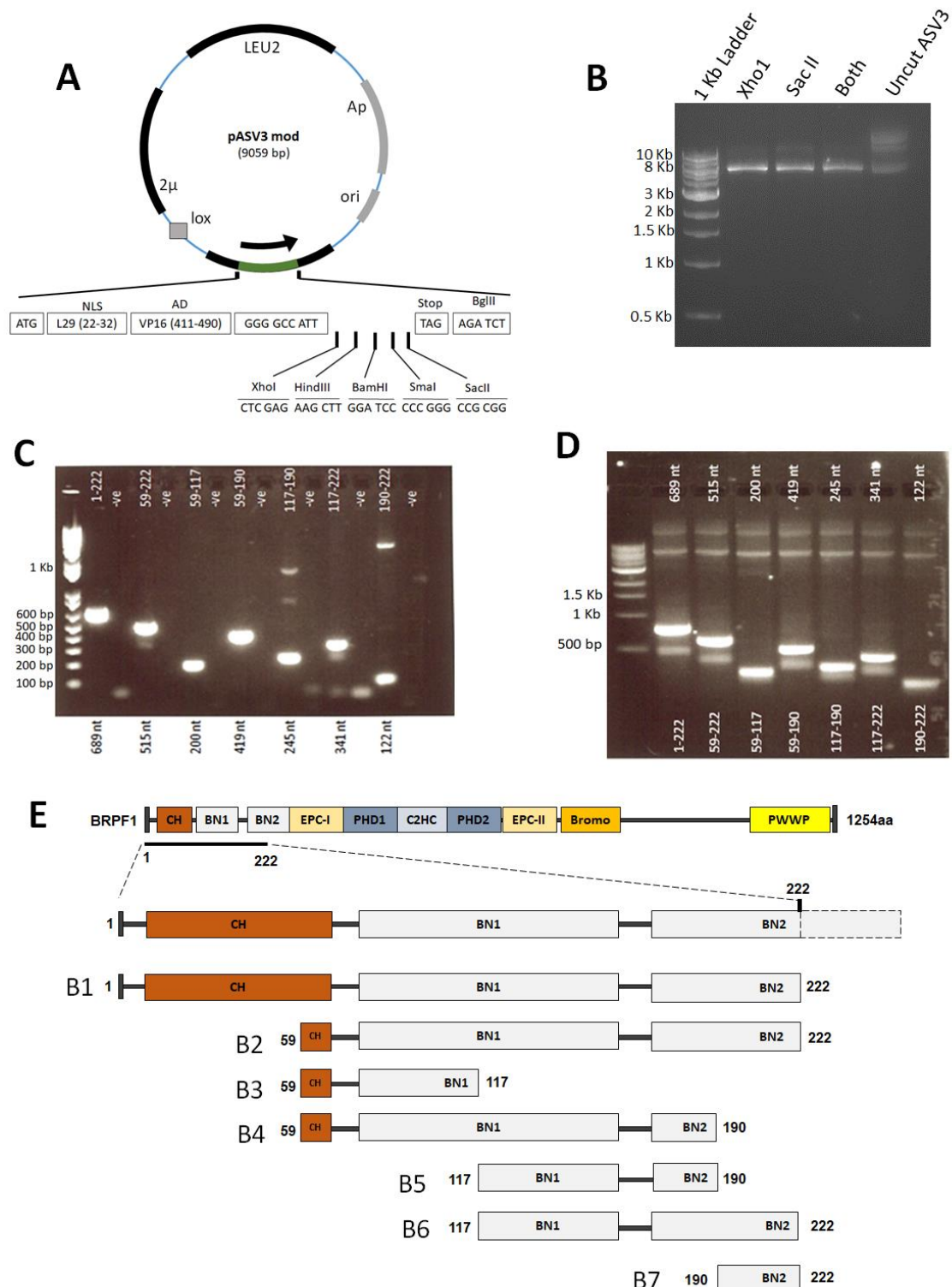


Figure 3. 8. Cloning of several N-terminal constructs into pASV3mod vector.

A) The target vector pASV3mod and its multiple cloning site (MCS), as described by Le Douarin *et al.* This pASV3mod vector is a yeast *E. coli* shuttle vector, which synthesises the AD-fusion proteins under the PGK promoter control. The start codon, nuclear localisation

sequence (NLS) from the ribosomal protein L29 (AA 22-32) of yeast, VP16 AD (AA 411-490) and the polylinker sequence are represented. **B)** Restriction digestion of the vector pASV3mod showing digestion with XhoI and SacII individually as well as both together along with a suitable negative control which is undigested vector. The restriction enzymes linearize the vector completely. **C)** PCR amplification of a series of N-terminal BRPF1 constructs from the template cDNA. The boundaries and the sizes of the constructs are shown. The negative controls for this PCR amplification contained only the primer pair but no template cDNA. **D)** Colony PCR of transformed *E. coli* colonies post-ligation for N-terminal BRPF1 constructs in XhoI and SacII digested pASV3mod. The bands represent the positive colonies. **E)** Showing the schematic representation of the N-terminal BRPF1 constructs and their boundaries (denoted by the start and the end amino acid numbers). The schematic also shows the domain distribution within the BRPF1 constructs.

BRPF1 Constructs	Boundary amino acids number	Nucleotide numbers	Total number of amino acids
B1	1-222	689	222
B2	59-222	515	163
B3	59-117	200	58
B4	59-190	419	131
B5	117-190	245	73
B6	117-222	341	105
B7	190-222	122	32

Table 3.1. List of N-terminal BRPF1 constructs into pASV3mod vector.

N-terminal BRPF1 constructs are listed with their boundary amino acid numbers, their size in nucleotides number and also the number of amino acids contained in those constructs.

3.6.3 The BRPF1 BN domains are sufficient to bind MOZ MYST domain in yeast two hybrid assays

Reporter strain *S.cerevisiae* L40 (Genotype: MATa his3 Δ 200 trp1-901 leu2-3112 ade2 LYS:: (4lexAop-HIS3) URA3:: (8lexAop-LacZ) GAL4) was co-transformed with VP16-AAD-BRPF1 1-222 in combination with LexA-DBD-MOZ 1-321, LexA-DBD MOZ 450-900 or LexA-DBD-MOZ 901-1150 separately after confirming correct sequence of the VP16-AAD-BRPF1 deletion constructs. The transformants were selected for growth on yeast selection media lacking leucine and tryptophan. LexA-DBD-MOZ plasmids are in pBTM116mod primary vector and these three constructs of MOZ were already available in our lab from previous investigation on MOZ interaction (Phil Troke Thesis 2003). After selection of double transformants as described in Materials and Methods section, single colonies were selected and inoculated into 15 ml of selective media YNB glucose lacking tryptophan and leucine. Cell-free extracts were prepared and β -galactosidase specific activities measured as described in methods. The BRPF1 construct BRPF1 1-222 (B1) showed a strong interaction with MOZ 450-900 but MOZ 1-321 and MOZ 901-1150 did not show any interaction with BRPF1 1-222 (Figure 3.8 A). This initial β -galactosidase reporter assay result indicates that the N-terminal of BRPF1 1-222 has the MOZ binding site and that MOZ 450-900 contains the BRPF1 interacting site. The control LexA-MOZ 450-900 co-transformed with empty pASV3 plasmid showed only background reporter activity, confirming that LexA MOZ 450-900 alone was not able to activate the reporter (Figure 3.8 B).

We next performed further mapping to identify the minimal MOZ interacting site of BRPF1. The LexA-MOZ 450-900 construct was co-transformed with the AAD-BRPF1 constructs, B2 to B7, and reporter activation determined. The result showed

that BRPF1 59-222 (B2) and 59-190 (B4) activated the reporter, albeit at a fourfold lower level than that of AAD-BRPF1 1-222 (B1). However, no reporter activation was observed using BRPF1 59-117 (B3), 117-190 (B5), 117-222 (B6) and 190-222 (B7) together with LexA-MOZ 450-900 (Figure 3.8 B). Western blots on yeast cell-free extracts confirmed the expression of the LexA-MOZ constructs in yeast (Figure 3.8 C). In addition, we know from previous studies that LexA MOZ 1-321 is expressed and capable of undergoing functional interactions with HIPK1 (Sirirsriro 2010, PhD thesis), or histones (Dreveny *et al.*, 2014) in Y2H assays.

Using the anti-VP16 AAD antibody to detect the BRPF1 fusion proteins, only BRPF1 1-222 and BRPF1 59-222 expression was confirmed in yeast, as we were unable to detect the other shorter AAD-BRPF1 constructs (Figure 3.8 D). This may indicate these fusion proteins are unstable in yeast, perhaps due to inability to fold into globular domains. It is notable that although a HA-tagged BRPF1 59-191 construct was expressed in mammalian cells, it was expressed at around 5-10 fold lower level than HA-BRPF1 1-222 (Ullah *et al.*, 2008), similar to our results.

Overall these results are consistent with the previous studies in mammalian cells that the N-terminal BRPF1 1-222 can bind directly to a MOZ construct containing the MYST HAT domain and adjacent sequences, but does not bind the MOZ N-terminal regions containing the NEMM and DPF domains. The minimal sequence sufficient to bind MOZ is BRPF1 59-190 comprising the BN1 domain, and indicates that the adjacent conserved region (BN2) is not essential for MOZ recruitment. However, Ullah *et al.*, 2008 also reported a weak interaction of MOZ with the BRPF1 192-543 region potentially indicating a second binding surface involving PHD knuckle Zn finger region.

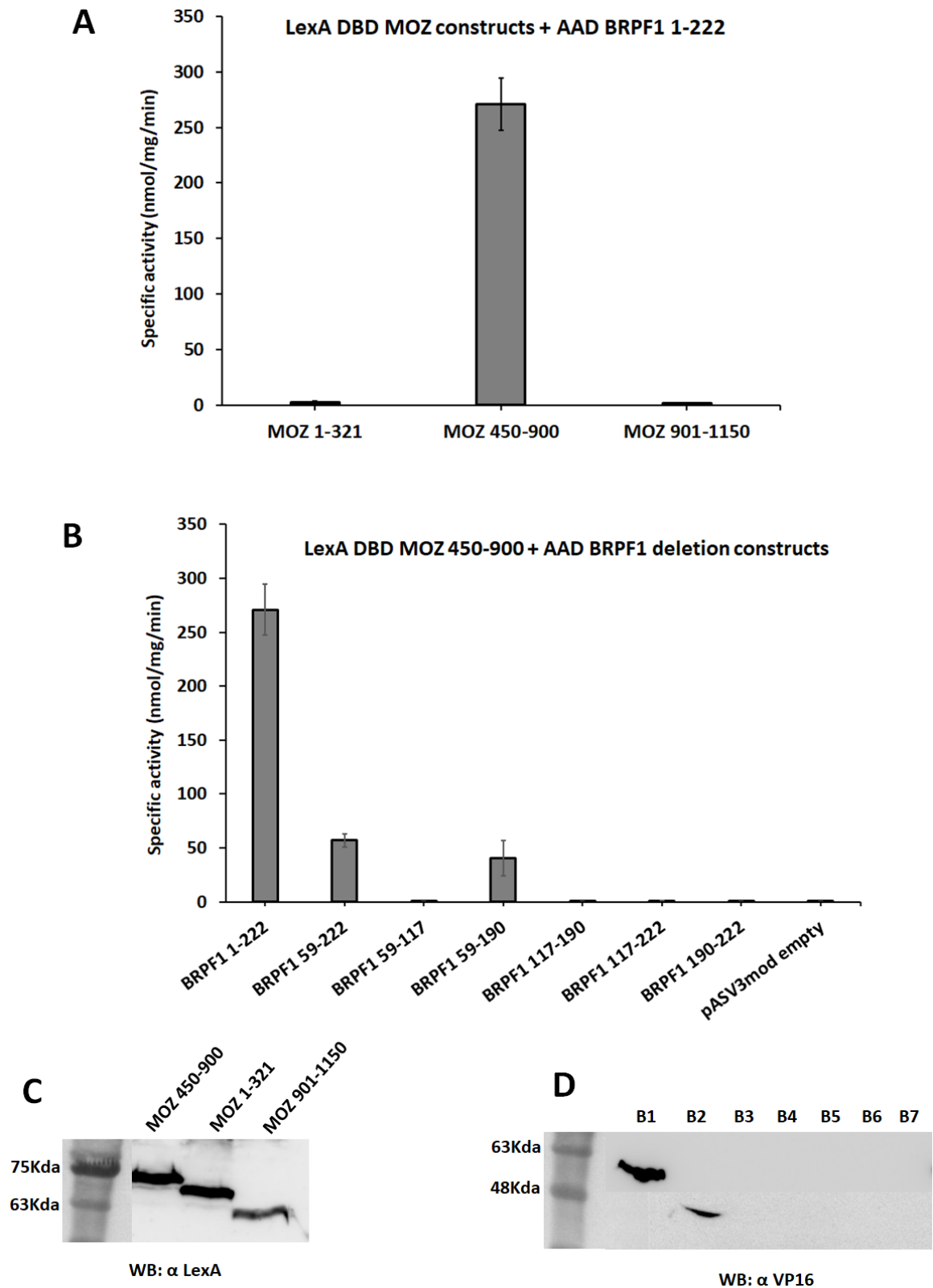


Figure 3. 9. β -galactosidase reporter assay from yeast two hybrid.

A) Showing specific activity for the MOZ and BRPF1 interaction from β -galactosidase reporter assay, where it indicates BRPF1 1-222 interacts with MOZ 450-900 but not with

MOZ 1-321 or MOZ 901-1150. **B)** Showing the specific activity from the β -galactosidase reporter assay, where further deletion constructs of BRPF1 1-222 were tested against MOZ 450-900. Among them B2 and B4 activate the reporter but rest failed. **C & D)** Western blot results, using anti-LexA antibody for MOZ constructs and anti-VP16 antibody for BRPF1 constructs, showing expression of MOZ and BRPF1 constructs in yeast cells. Expression of the BRPF1 constructs B3 to B7 were not detected by western blot, which might be due to stability issue of those constructs.

3.7. Mapping MOZ and other MYST protein (MORF) for BRPF1 interacting site

Having detected an interaction of MOZ 450-900 with the N-terminus of BRPF1 BN1 and the adjacent region, we started looking for the finer mapping of MOZ in the search of the BRPF1 interaction site. The MYST HAT region of another MYST family protein HBO1 was also reported by Mishima *et al.* (2011) to interact with the N-terminal BRPF2. Therefore, further mapping of this region became essential to find the minimal region, which is responsible for BRPF1 interaction. Thus, to achieve the aim of finding the minimal region of MOZ responsible for BRPF1 interaction, we next generated several deletion constructs within MOZ 450-900 to define the minimal sequence in MOZ responsible for BRPF1 interaction. From the PDB structural analysis (Figure 3.4 C) we already knew that this MYST HAT domain of MOZ contains several alpha helices and beta strands, so in defining the boundary of these constructs extra care was taken. The boundaries were defined by the secondary structure prediction of the MOZ 450-900 using Phyre2 online programme (Figure 3.9). The PCR fragments for the finer deletion mapping of MOZ 450-900 were sub-cloned into pBTM116mod (LexA DBD) yeast expression vector as the BRPF1 coding sequence is in the VP16 AAD vector.

3.7.1 Sub-cloning of MOZ deletion constructs in pBTM116mod vector

Various deletions constructs, listed in Table 3.2, were generated by PCR amplification for use in sub-cloning, the MOZ 450-900 construct was used as the template. The sub-cloning method was used as described in Figure 3.6. Following confirmation of size by agarose gel electrophoresis (Figures 3.10 C) the PCR

products (and the pBTM116mod vector) were digested with SacII and BamHI restriction enzymes. After ligating the restriction digested and purified inserts into pre-cut vector they were transformed and single colony was selected for colony PCR to confirm the presence of the respective inserts (Figures 3.10 D). The sequences were confirmed via sequencing.

are showing the breakpoints for the MOZ deletion constructs to avoid the disruption of an alpha helix or beta strand.

MOZ Constructs	Boundary amino acids number	Nucleotide number	Total number of amino acids
M1	478-900	1286	422
M2	478-675	611	197
M3	498-675	551	177
M4	561-675	362	114
M5	586-675	287	89
M6	603-675	236	72
M7	676-900	692	224
M8	676-806	410	130
M9	676-763	281	87
M10	807-900	299	93
M11	764-900	428	136
M12	764-806	146	42
M13	764-779	65	15
M14	676-707	113	31
M15	707-763	188	56

Table 3.2. List of further deletion constructs of MOZ 450-900 into pBTM116mod vector.

MOZ 450-900 deletion constructs are listed with their boundary amino acid numbers, their size in nucleotides number and also number of amino acids contained in those constructs.

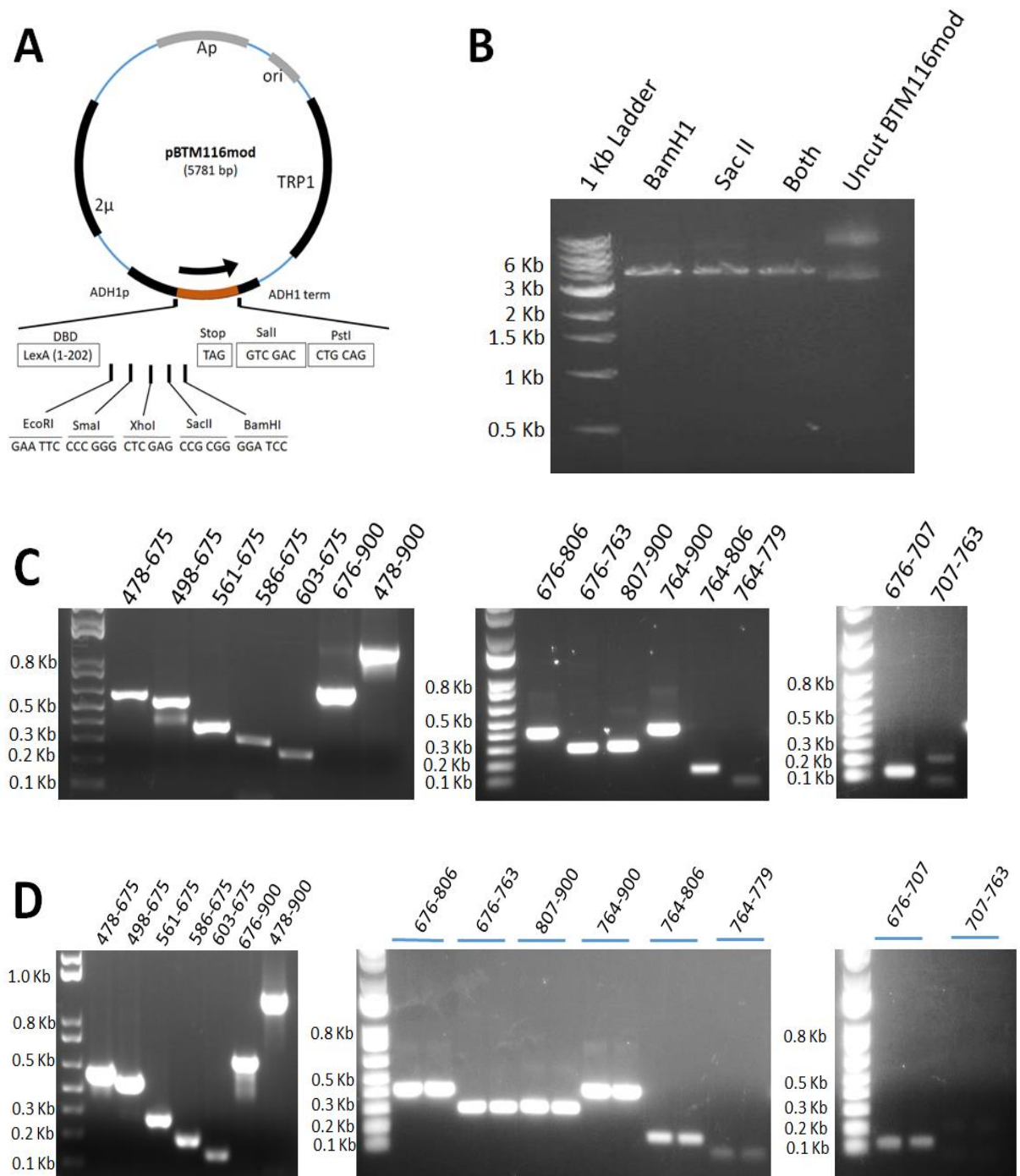


Figure 3. 11. Cloning of several MOZ deletion constructs into pBTM116mod vector.

A) The target vector pBTM116mod and its multiple cloning site (MCS), as described by Le Douarin *et al.* This pBTM116mod vector is a yeast *E. coli* shuttle vector, which synthesises the DBD-fusion proteins. **B)** The restriction digestion of the vector pBTM116mod showing digestion with SacII and BamHI individually as well as both together along with a suitable negative control which is undigested vector. The restriction enzymes linearize the vector completely. **C)** PCR amplification of a series of MOZ deletion constructs from the template

cDNA. The negative controls for this PCR amplification contained only primer pair but no template cDNA. **D)** Colony PCR of transformed *E. coli* colonies post-ligation for MOZ deletion constructs in SacII and BamHI digested pBTM116mod. The bands represent positive colonies.

3.7.2 MOZ MYST domain contains the BRPF1 interaction site

The LexA-DBD-MOZ deletion constructs M1-M15 series were used to transform L40, reporter strain *S.cerevisiae* in combination with VP16-AAD-BRPF1 1-222 (B1) or pASV3mod empty vector. The transformants were selected for growth on yeast selection media lacking leucine and tryptophan. After selection of double transformants as described in Material and Methods section, single colonies were selected and inoculated into 15 ml of selective media YNB glucose lacking tryptophan and leucine. Cell-free extracts were prepared and β -galactosidase specific activities measured as described.

A series of β -galactosidase reporter assays revealed that constructs MOZ 478-900 (M1), MOZ 676-900 (M7), MOZ 676-806 (M8) and MOZ 676-763 (M9) show an interaction with BRPF1 1-222. M9, which is MOZ 676-763, is found to be the minimum sequence for the interaction (Figure 3.11 A). Other N-terminal MYST domain constructs within the boundary of 478-675 amino acid residues (M2 to M6) and constructs after amino acid residue number 764 to 900 (M10 to M13) did not show any interaction. Further smaller deletion constructs MOZ 676-707 and MOZ 707-763 (M14 and M15) of MOZ 676-763 failed to show any interaction. Therefore, these findings suggested that, the C-terminus of the MYST domain is sufficient for BRPF1 binding and the whole MOZ 676-763 is required for interaction but smaller fragments than MOZ 676-763 disturb the folding of the protein and disrupt the binding. The LexA-DBD-MOZ 450-900 and VP16-AAD-BRPF1 1-222 transformation was used as the positive interaction control. Western blots on yeast cell-free extracts confirmed the expression of all the LexA-DBD-MOZ fusion proteins as well as VP16-AAD-BRPF1 1-222 fusion protein (Figures 3.11 B). LexA-DBD-MOZ fusion proteins

from all the constructs were detected by the anti-LexA-DBD antibody and VP16-AAD-BRPF1 1-222 expression was confirmed in yeast by using anti-VP16-AAD antibody.

Overall this result suggested that a MOZ construct containing the C-terminus of the MYST HAT domain, is sufficient to bind BRPF1 1-222 which is quite far away from the region Shima *et al.* (2014) reported (MOZ 488-664). However, the minimal sequence (MOZ 760-810) Ullah *et al.* (2008) concluded for BRPF1 to interaction with MOZ, slightly falls within the region we mapped but the constructs had shown positive interaction in their mapping, all of them contained the whole region we found to have the minimal sequence for BRPF1 interaction. The further deletion of MOZ 676-763 (M9) failed to show any interaction, therefore we can conclude from this yeast-two-hybrid result that, to interact with BRPF1 MOZ requires this whole region from amino acid residue 676 to 763 residue.

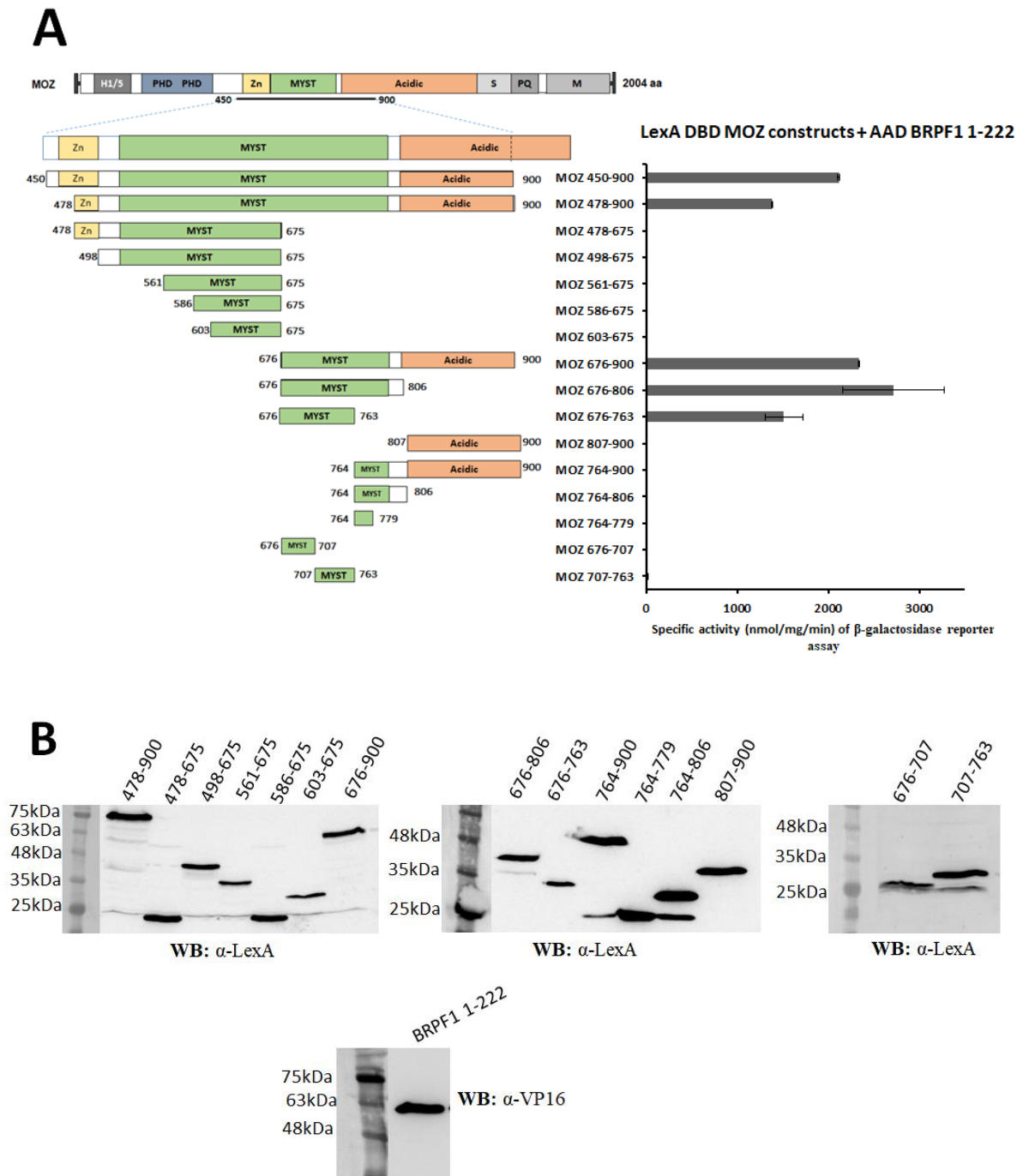


Figure 3. 12. β-galactosidase reporter assay from yeast two hybrid for MOZ deletion constructs and BRPF1 1-222 interaction.

A) Showing specific activity for the MOZ deletion constructs and BRPF1 1-222 interaction from β-galactosidase reporter assay, where it indicates that MOZ 676-763 is the minimum sequence required for BRPF1 recruitment. **B)** Western blot result showing the confirmation of the expression of both VP16 BRPF1 1-222 and all the LexA MOZ fusion proteins.

3.7.3 BRPF1 BN1 domain interacts with MYST HAT domain of MORF

After defining the minimal region of MOZ required for BRPF1 recruitment, we investigated the interaction between the same N-terminal portion of BRPF1 (1-222) with the nearest homolog of MOZ in humans, MORF. Taking the previous result which defined the minimal sequence of MOZ required for BRPF1 interaction into account, this sequence MOZ 676-763 with the corresponding region in MORF (541-626) was aligned. This MORF sequence was then sub-cloned into the pBTM116mod vector. After generation and sequence confirmation of the LexA-DBD-MORF 541-626 construct, it was co-transformed into the yeast reporter strain *S.cerevisiae* L40 with VP16-AAD-BRPF1 1-222. The transformants were selected for growth on yeast selection media lacking leucine and tryptophan. After selection of double transformants, single colonies were selected and inoculated into 15 ml of selective media YNB glucose lacking tryptophan and leucine. Cell-free extracts were prepared and β -galactosidase specific activities measured.

The β -galactosidase reporter assay showed that LexA-DBD-MORF 541-626 interacts with VP16-AAD-BRPF1 1-222 (Figures 3.12 A). This indicates that MORF also interacts with BRPF 1 (1-222) through a portion of its MYST HAT domain similar to the interaction specified for MOZ. The expression of LexA-DBD-MORF 541-626 was confirmed by western blotting using the anti-LexA antibody. Ullah *et al.* (2008) mapped MORF 460-716 as the minimal sequence for BRPF1 binding, however they also reported that even if they deleted residue 553 to 588 of MORF, the positive interaction with BRPF1 was retained.

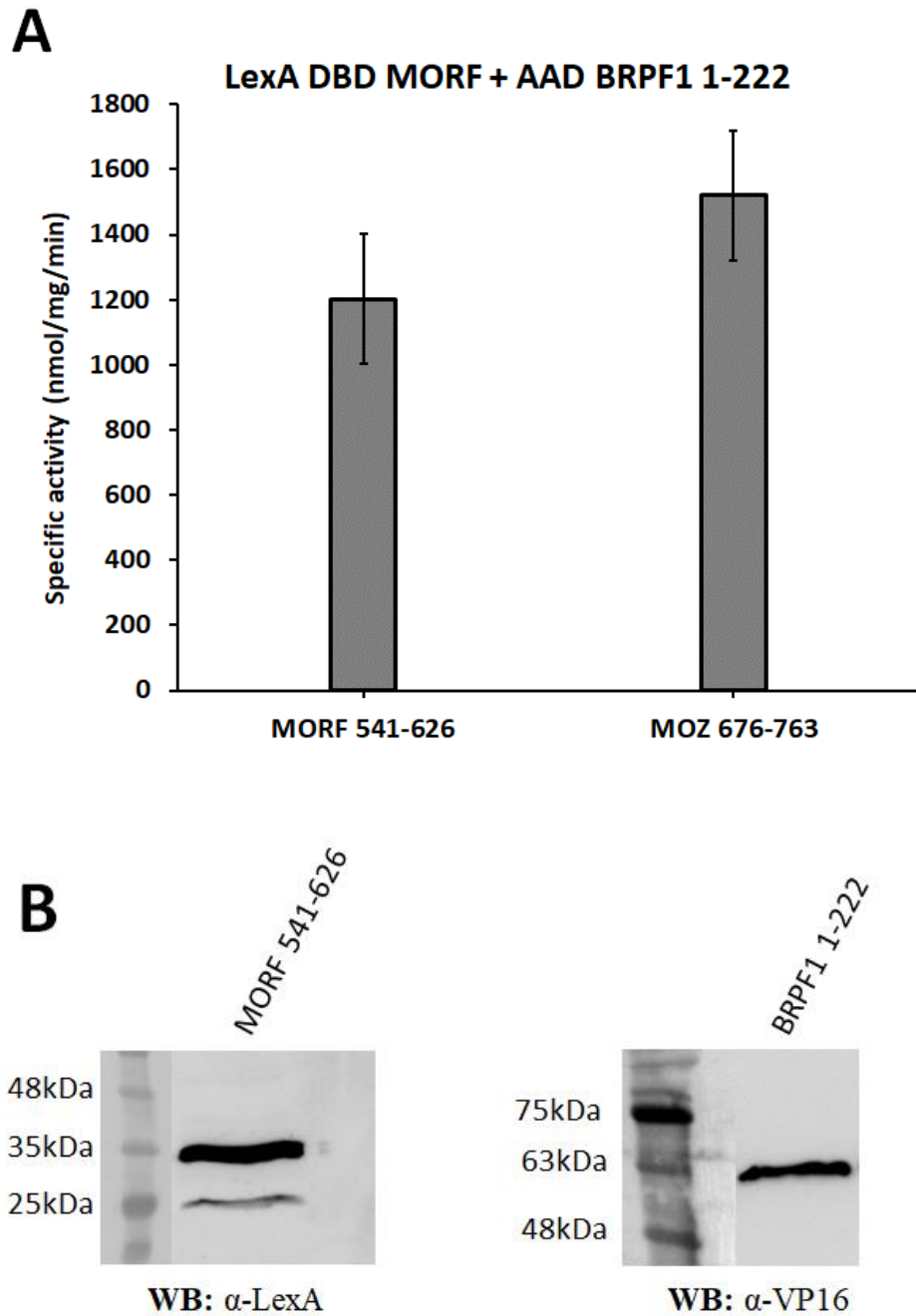


Figure 3. 13. β -galactosidase reporter assay from yeast two hybrid for MORF 541-626 and BRPF1 1-222 interaction.

A) Showing specific activity for the MORF 541-626 and BRPF1 1-222 interaction from the β -galactosidase reporter assay. MORF 541-626 (like its corresponding sequence in MOZ 676-763) shows an interaction with BRPF1 1-222. **B)** Western blot result confirming the expression for both MORF 541-626 and BRPF1 1-222.

3.8 *In vivo* confirmation of the interaction of BRPF1 with the MYST domain of MOZ

After defining the MOZ minimal sequence for interaction with BRPF1 in the yeast-two-hybrid assay, I have tried to confirm this interaction *in vivo*. The MOZ constructs used in this experiment were shorter versions of FLAG-MOZ ending just after the MYST HAT domain of MOZ, these were FLAG-MOZ 1-810 and FLAG-MOZ 1-779 (generated by Sian Deeves). These shorter MOZ constructs were each co-transfected with HA-BRPF1 full length into HEK293 cells for over expression of MOZ and BRPF1 together. Attempts to co-immunoprecipitate BRPF1 using a commercially available anti-FLAG antibody were successful.

HEK293 cells were co-transfected with FLAG MOZ 1-810 or FLAG MOZ 1-779 alone or in combination with HA-BRPF1 full length and after 48 hours post-transfection, cells were harvested and lysed in phosphate buffer (explained in Materials and Methods) to release soluble proteins. Western blots confirmed expression of FLAG-MOZ 1-810, FLAG-MOZ 1-779 and HA-BRPF1 proteins in the cell-free extracts (Figure 3.13 A). As expected the MOZ proteins were detected according to their size. FLAG MOZ proteins were then immunoprecipitated from the cell lysate using anti-FLAG antibody and tested for HA-BRPF1 interaction by western blotting with the anti-HA antibody. Using anti-HA antibody a protein band was detected in the co-IP lane (Figure 3.13 B) near 140kDa marker and also detected in co-IP membrane 1st and 3rd lane (Figure 3.13 B) only when both HA-BRPF1 and FLAG MOZ were co expressed, but not in the control lanes (Figure 3.13 B 2nd, 4th, 5th and 6th lane) which were transfected with either only MOZ constructs or only HA-BRPF1 or pCDNA3.1 empty. The identity of this protein was confirmed as BRPF1 by using anti-HA antibody

to detect the HA tag. This result confirms that the N-terminus of MOZ amino acid residue 1 to 779 is capable of binding to BRPF1 full length protein.

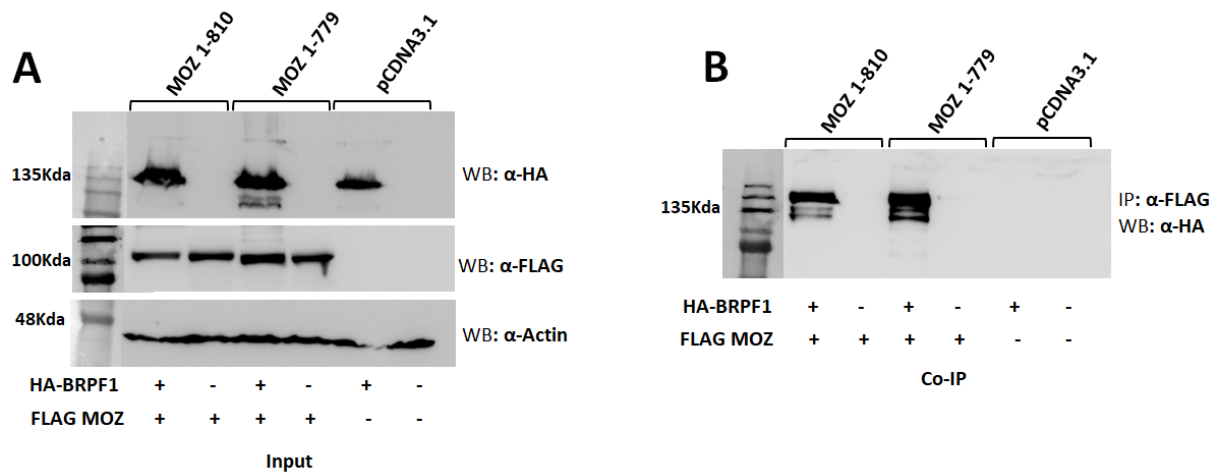


Figure 3. 14. Co-immunoprecipitation of shorter MOZ and BRPF1 complex *in vivo*.

A & B) Showing that shorter version of MOZ (1-810 and 1-779) can also pull BRPF1 down in co-immunoprecipitation which indicates that, the BRPF1 interaction site is present in between the MOZ residue 1 to 779. All the transfected proteins expression were also confirmed in the input.

3.9 Summary

Epigenetic regulation is driven by the chromatin structure and function. To read different chromatin modifications, there are several hundred chromatin readers encoded by the human genome. These chromatin regulatory proteins form complexes to sense DNA and histone modifications for chromatin based signalling (Taverna *et al.*, 2007 and Musselman *et al.*, 2012). The BRPF1/MOZ complex is one of those chromatin reader complexes. In this study, we tried to identify the minimal sequence of MOZ and BRPF1 involved in this MOZ/BRPF1 complex formation. Although there have been several studies investigating MOZ/BRPF1 interaction, so far they were performed in mammalian cells using co-immunoprecipitation (Ullah *et al.*, 2008; Shima *et al.*, 2014; Mishima *et al.*, 2011 and Lalonde *et al.*, 2013). In co-immunoprecipitated complexes, the interactions may be direct or indirect. So in this project we have used yeast-two-hybrid method to provide better insights into the sequences involved, as yeast lacks proteins with very close homology to BRPF1.

Herein this study started with confirming the MOZ/BRPF1 complex formation in a mammalian cell line, which was conducted by various researchers (Ullah *et al.*, 2008; Shima *et al.*, 2014; Mishima *et al.*, 2011). In co-immunoprecipitation we found both overexpressed full length MOZ and BRPF1 form a complex and can be co-immunoprecipitated from the cell lysate. After this co-immunoprecipitation confirmation, we found that MOZ physically associated with BRPF1, which extends previous findings (Doyon *et al.*, 2006) and this remains consistent with this immunofluorescence data, which was our next attempt to confirm the MOZ BRPF1 complex. This immunofluorescence data for co-localisation of MOZ and BRPF1 in mammalian cells indicate that BRPF1 required MOZ for proper chromatin localisation

and this result was consistent with previous works by Laue *et al.*, (2008) and Ullah *et al.*, (2008). However, it remains unclear whether this MOZ/BRPF1 complex formation is direct or indirect due to other chromatin regulatory protein such as ING5 (Doyon *et al.*, 2006).

Although HEK293 cells express MOZ endogenously, whenever HA-BRPF1 was transfected alone into HEK293 cells, it was detected in the cytoplasm. It is not very clear why HA-BRPF1 remains in the cytoplasm when endogenous MOZ is present. Perhaps HA-BRPF1 needs to be modified or the level of overexpressed protein (HA-BRPF1) is higher than endogenous MOZ and the majority of HA-BRPF1 stays in the cytoplasm until YFP-FLAG-MOZ is co-expressed exogenously.

Thus the yeast-two-hybrid reporter assay becomes a good alternative to map the essential region of MOZ and BRPF1 required for their complex formation. Initial sequence and structural analysis suggested that N-terminal BRPF1 up to BN2 domain and MOZ MYST domain is highly conserved in their respective protein families. Later our yeast-two-hybrid result revealed that BRPF1 BN1 and an adjacent small region of BN2 is sufficient to interact with MOZ MYST HAT domain (Figures 3.14 a). This finding was also reported by Ullah *et al.*, (2008); Laue *et al.*, (2008) and Mishima *et al.*, (2011) from their co-immunoprecipitation results. Our results showed that the N-terminal region of BRPF1 (59-190) interacts with the MOZ sequence (676-763) within the HAT domain and just before the adjacent sequences from the acidic region, but not with the N-terminal NEMM and DPF domain. This finding refutes the Shima *et al.*, (2014) finding, who suggested that MOZ 488-664 is responsible for BRPF1 recruitment. Also Ullah *et al.*, (2008) suggested MOZ 760-810 is the minimal sequence for BRPF1 interaction, which falls within the region we have mapped in

yeast-two-hybrid system for MOZ/BRPF1 complex interface. We also have confirmed our yeast-two-hybrid findings in mammalian cells as well. Our co-immunoprecipitation of N-terminus MOZ (up to and including the MYST domain) and full length data supported our yeast-two-hybrid findings, that the BRPF1 interaction site of MOZ is located within the MYST domain of MOZ.

Holbert *et al.*, (2007) reported secondary and tertiary structure of the MOZ MYST domain. That available PDB structural analysis of MOZ HAT domain (PDB 2RC4) revealed that there is a helix-turn-helix structure within the region MOZ 676-763 (Figure 3.14 B). It is this whole helix-turn-helix part that we determined was required for BRPF1 binding. Anything smaller than MOZ 676-763 disturbs the folding of the protein and disrupts the binding. This BRPF1 binding site of MOZ is situated very close to the active site of the HAT domain which facilitates incorporation acetyl-coA to the histone chain (Figure 3.14 B). This will provide better understanding of how complexes are formed between these proteins, whether there is competition for binding sites with other proteins, and it may also allow us to generate mutant versions of BRPF1 or MOZ that are defective for complex formation, enabling us to study their functions. However, in this helix-turn helix region of MOZ there is a loop like tertiary conformation, in which a major portion (amino acid 738-748) does not have any structure due to low electron density (Holbert *et al.*, 2007). Alignment of all the available PDB structures for other MYST family proteins revealed that this unstructured region of MOZ might consist of two β -strands, this region may play a crucial role in complex formation with BRPF1.

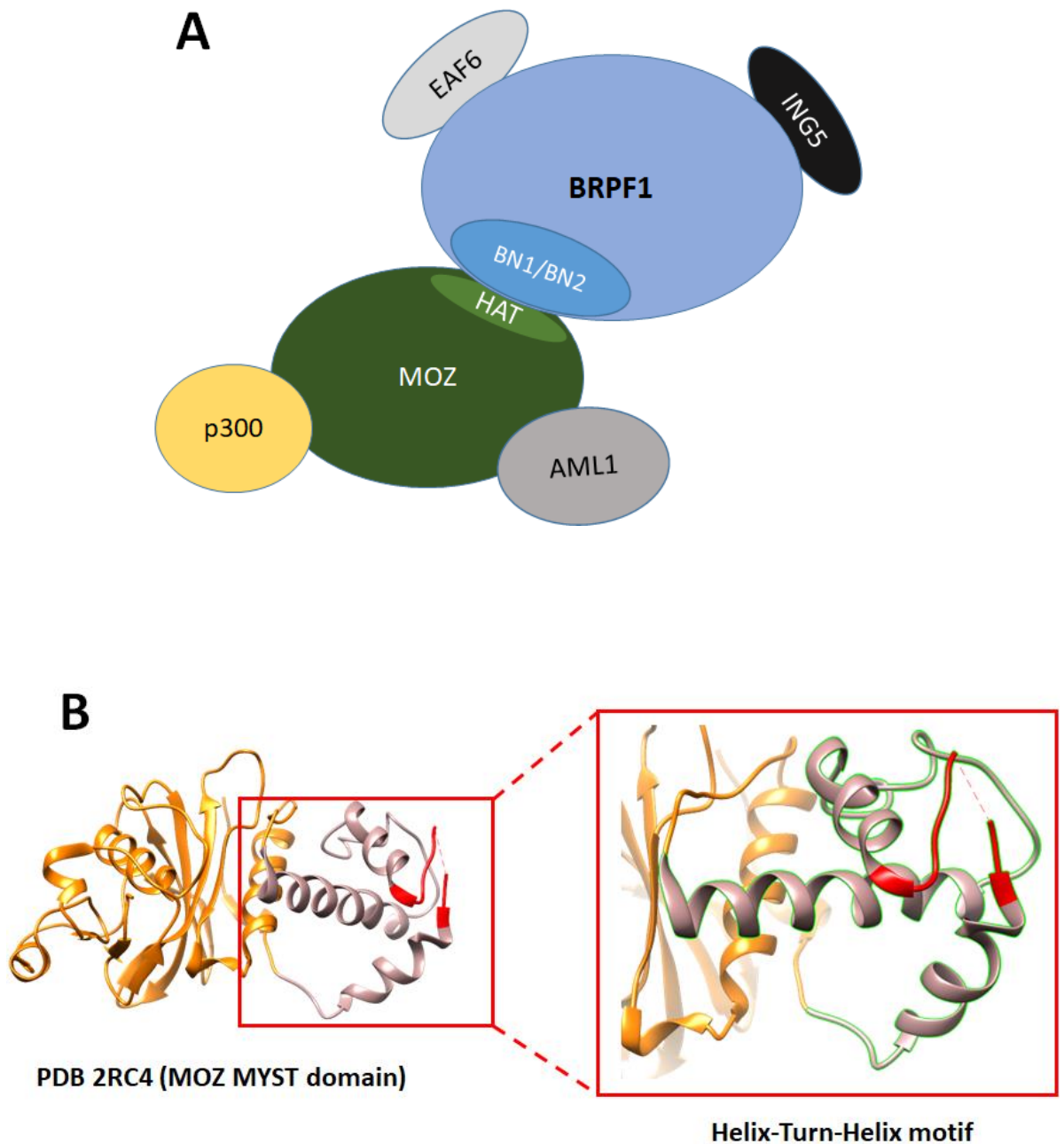


Figure 3. 15. Schematic for MOZ/BRPF1 complex and the tertiary structure of MOZ MYST domain and location of BRPF1 interacting site.

A) Schematic diagram of the MOZ/BRPF1 complex, where it is showing that MOZ MYST HAT domain interacts with BRPF1 BN1/BN2 domain to form the complex. **B)** PDB structure for MOZ MYST domain showing that MOZ MYST domain contains a helix-turn-helix motif, which is shown with pink ribbon and zoomed for better view. This helix-turn-helix motif is the responsible for the BRPF1 recruitment, but if that helix-turn-helix conformation get disturbed by shortening the sequence, BRPF1 interaction also gets disrupted.

After obtaining the minimal sequence of MOZ, which required for BRPF1 interaction, we have tested another MYST family homolog, MORF. Previously, Lalonde *et al.*, (2013); Doyon *et al.*, (2006) and Ullah *et al.*, (2008) reported on the BRPF1 and MORF interaction and we also found the same in our study. The yeast-two-hybrid result unveil that same N-terminal BRPF1 interacts with MORF 541-626, which is the corresponding region of MOZ 676-763. These findings also suggests that BRPF1 needs that same helix-turn-helix conformation to interact with MOZ or other MYST family proteins, which is sufficient to interact with BRPF1. Although, functional analysis of requirement of this Helix-turn-helix region of MOZ for interaction with BRPF1 is yet to be established, which will be explored in coming chapters.

Chapter 4: Effect of Mutations on MOZ/BRPF1 Complex

4.1. Introduction

The previous chapter demonstrated that the MOZ MYST domain contains a BRPF1 interaction site, which is located within the amino acid residues 676-763. Crystal structures of other MYST family domains including human MOZ, MOF, TIP60 and yeast ESA1P (Holbert *et al.*, 2007; Wu *et al.*, 2007; Yuan *et al.*, 2011; McCullough *et al.*, 2016) reveal that this region consists of three α -helices and a protruding 'finger' made up of two β -strands. In the published MOZ structure (Holbert *et al.*, 2007) electron density was poor in this region and so was undefined, but based on sequence similarity, we hypothesised this is also likely to form the anti-parallel β -sheet 'finger'. Changes in these amino acids may change the conformation of the protein and that leads to disruption of their interaction. Moreover changes of an amino acid from an aliphatic to a cyclic one can also alter the adaptor specificity (Verstak *et al.*, 2013 and Perez *et al.*, 2014). Incorporation of an extra secondary amine group, by switching from any non-cyclic amino acids to a cyclic amino acid, into the polypeptide chains of any protein, leads to a change in the folding of that protein and changing in folding will lead to change in function or specificity or interaction with other proteins (Verstak *et al.*, 2013). Therefore, we decided to undertake site-directed mutagenesis of the BRPF1 interaction region of MOZ.

As discussed briefly in the previous chapter, the BRPF1 N-terminus interaction with residues 676-763 of MOZ would potentially bring it into close proximity with the acetyl-CoA binding site as defined in the MOZ MYST domain structure (Holbert *et al.*, 2007). The interaction of BRPF1 has been proposed to activate the histone acetyltransferase activity of MOZ and guide its specificity toward acetylation of H3K14 in chromatin (Lalonde *et al.*, 2013). The isolated MYST HAT domain of MOZ

has been found to acetylate histones H3 and H4 at specific lysine residues such as H3K14ac, H4K4ac, H4K8ac, H4K12ac and H4K16ac (Kitabayashi *et al.*, 2001), and our group has shown that binding of the H3 N-terminal tail by the Double PHD Finger of MOZ greatly facilitates H3K9 and H3K14 acetylation, enhancing the rate of this reaction by around 8-fold (Dreveny *et al.*, 2014). Thus the fully functional acetyltransferase activity of MOZ relies not only on the MYST domain but other substrate interaction regions of the protein as well, and may be further supplemented by MOZ-interacting proteins such as BRPF1.

BRPF1 is a unique chromatin reader, which also possesses sequence motifs mediating interaction with MOZ and other MYST proteins (Yan *et al.*, 2016). The BRPF1 bromodomain can also recognise each of the histone PTMs acetylated by MOZ except H4K16ac and also it is involved in targeting the MOZ HAT complex (Champagne *et al.*, 2008). H4K16ac is very crucial for the formation of euchromatin to disrupt the formation of higher-order chromatin by inhibiting chromatin remodelling (Shogren-Knaak *et al.*, 2006). Thus, the epigenetic landscape indicating that BRPF1 recruits the MOZ HAT complex for acetylation of H4K16 to up-regulate gene expression. The BRPF1 and MOZ HAT complex also contains some other chromatin readers such as ING5 PHD fingers, MOZ DPF, the BRPF1 PWWP domain and BRPF1 double PHD finger, which bind H3K4me3, H3K14ac, H3K36me3 and unmodified H3, respectively (Vezzoli *et al.*, 2010; Liu *et al.*, 2012; Champagne *et al.*, 2008 and Qiu *et al.*, 2012). Thus, MOZ and BRPF1 recognise multiple PTMs on histone tails by their specific modules either individually or in a complex form to regulate the MOZ HAT activity based on the available epigenetic signals. Mutations within these modules of either BRPF1 or MOZ causes disruption in their ability to

recognise the PTMs of histone tails and that leads to developmental abnormalities (Yan *et al.*, 2016). Clinically there are many developmental disorders, such as intellectual disability and congenital abnormalities, which were analysed molecularly and it was pointed out that in those developmental disorders, either MOZ or BRPF1 was mutated and therefore their functions for acetylation of histone tails were also disrupted due to those mutations (You *et al.*, 2015; Yan *et al.*, 2016 and Mattioli *et al.*, 2017). However, our aim is not to check the effect of any mutations on the clinical function of MOZ or BRPF1, instead, we wanted to determine if the minimal sequence of MOZ we identified in our previous chapter, is essential and sufficient to recruit N terminal BRPF1 for complex formation. We have tried to replace some non-cyclic amino acids with the cyclic amino acids to incorporate an extra secondary amine group, which can have the effect to change the conformation of the folding in that helix-turn-helix region of MOZ.

4.2. Site-Directed Mutagenesis of the BRPF1 interaction region of MOZ

To further map the interaction interface of MOZ and BRPF1 we generated single amino acid substitution mutations in the construct LexA-MOZ 676-763. We selected residues in the three α -helices, as well as four residues within the unstructured loop region, which is predicted to form two β -strands by Phyre2 analysis (The Phyre2 web portal for protein modelling, prediction and analysis, Protein Homology/analogy Recognition Engine V 2.0) and one from the third α -helix. At all points the respective amino acids were replaced with proline. The only hetero cyclic amino acid is proline, which does not have a hydrogen on the α -amino group, therefore proline cannot

donate a hydrogen bond to stabilize any α -helix or any β strand while it is in a peptide bond. If proline is present in an α -helix, the helix will have a slight bend due to the lack of the hydrogen bond. Whenever proline is present within a polypeptide chain of any protein, it is often present at the end of α -helix or in turn or loops. All other amino acids exist exclusively in the trans-form of any polypeptide, but proline can exist in the cis-configuration and this cis and trans isomerisation plays a very crucial role in the folding of the protein (Baldwin et al., 2003). As described in Table 4.1, the following eight amino acids residues, alanine at 693, glutamine at 706, lysine at 718, phenylalanine at 739, aspartic acid at 742, glutamine at 743, valine at 745 and alanine at 758, were replaced with proline. These residues were replaced with proline to potentially disrupt the tertiary structure of the MOZ helix-turn-helix motif. Out of the eight point mutations three of them were located in the first and third α -helices, one of them at the turn before the second α -helix and the other four chosen were within the unstructured loop region of that helix-turn-helix motif (Figures 4.1A) assuming that region might be crucial for recruitment of BRPF1. According to the tertiary structure of the MOZ MYST HAT domain, all four mutations from the structured region were distributed in such a way so that we can cover all sides of that helix-turn helix region (Figures 4.1 B). This coverage allowed us to examine the whole BRPF1 interaction site of MOZ along with investigating the unstructured loop very precisely.

Name of the mutation	Wild type amino acid residue	Position at MOZ	Replaced by
A693P	Alanine	693	Proline
Q706P	Glutamine	706	Proline
K718P	Lysine	718	Proline
F739P	Phenylalanine	739	Proline
D742P	Aspartic acid	742	Proline
Q743P	Glutamine	743	Proline
V745P	Valine	745	Proline
A758P	Alanine	758	Proline

Table 4.1. List of selected point mutations on the MOZ 676-763.

The point mutations selected for incorporation into MOZ 676-763 by site-directed mutagenesis. The table is showing the position of the point mutations, original amino acid residue and replaced amino acid.

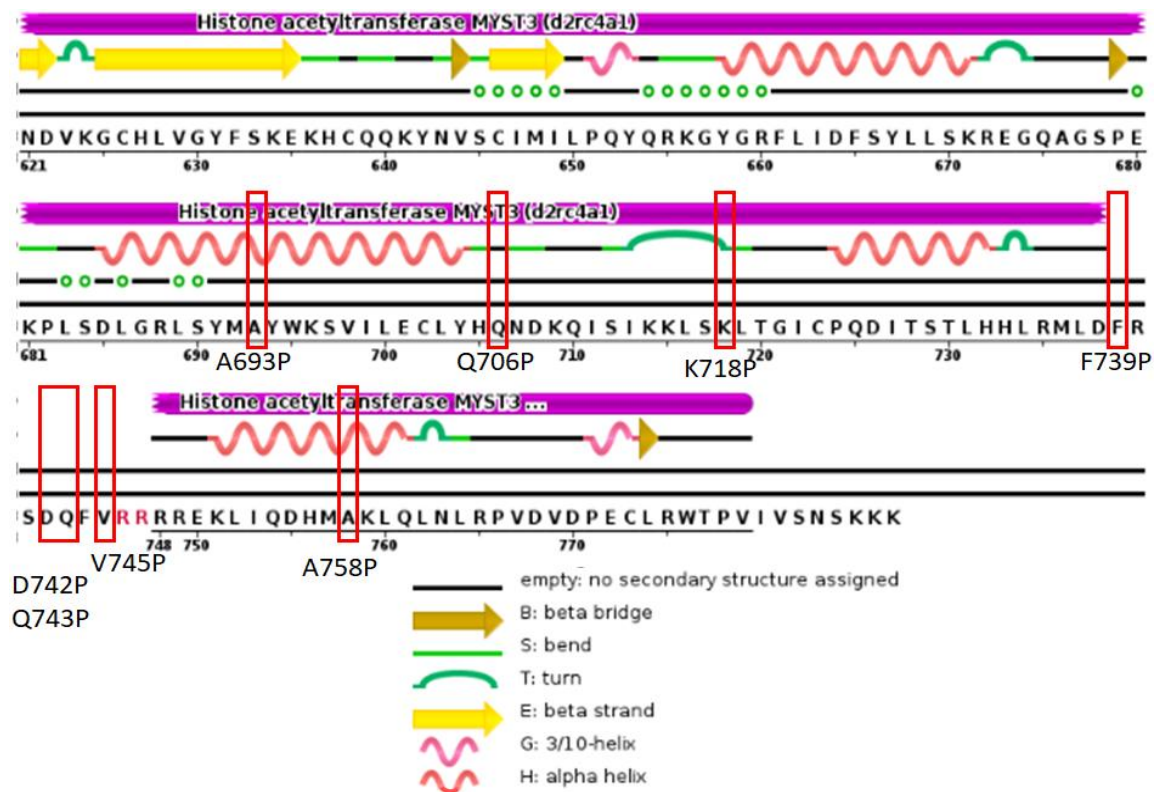
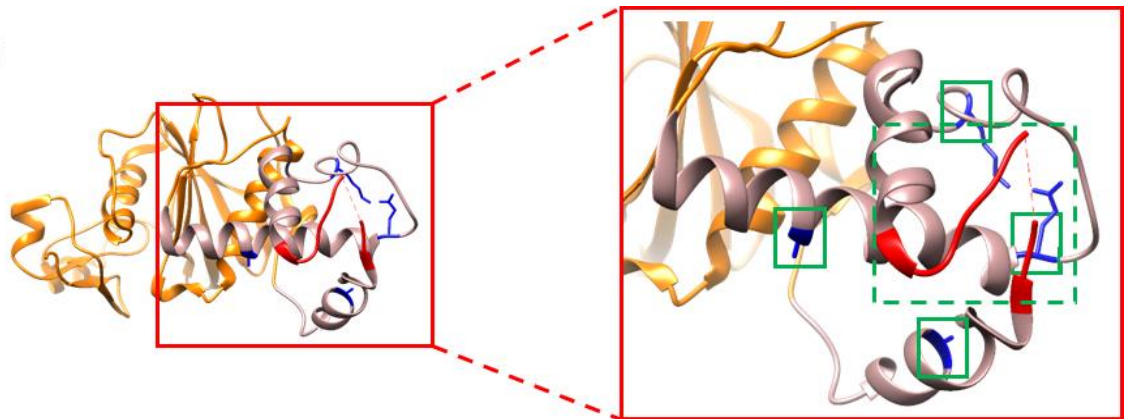
A**B**

Figure 4. 1. Positions of the point mutations on MOZ 676-763.

A) Secondary structure of the MOZ MYST HAT domain from the available PDB structure 2RC4 showing the position of the selected point mutations, marked with red boxes, on the helix-turn-helix motif of the MOZ MYST domain. **B)** Tertiary structure of the MOZ MYST domain showing the selected point mutations locations after folding into the functional protein. The zoomed image is showing the location of the mutations, marked with green boxes and out of the eight point mutations four are located within the unstructured loop region, which is marked with a dotted green box.

4.3. Incorporation of mutations using site-directed mutagenesis

There are several methods available to incorporate specific point mutations in any plasmid such as Kunkel's method, cassette mutagenesis, PCR site-directed mutagenesis or whole plasmid mutagenesis. However, we have chosen whole plasmid mutagenesis. The chosen point mutations were incorporated into the LexA-DBD MOZ 676-763 plasmid using PCR. Primers for the mutations were designed in such a way so that changes within the subjected codon responsible for the mutation remain minimal. As described in Material and Methods section, the PCR reaction was performed with a long extension time to synthesise the whole plasmid, which contained the desired point mutation. The original wild-type template plasmids were eliminated by treating with Dpn1, which digests methylated plasmids, leaving the newly synthesised plasmid unaffected. After Dpn1 treatment all eight newly synthesised plasmids were transformed into *E.coli* and plasmids from single colony were analysed by sequencing. All eight point mutations were confirmed from the sequencing data (Figures 4.2).

After confirming the presence of the mutations we confirmed the expression of the mutated LexA-DBD MOZ 676-763 constructs in yeast. Each of the mutated LexA-DBD MOZ 676-763 constructs were co-transformed separately into yeast with VP16-AAD BRPF1 1-222. The transformants were selected for growth on yeast selection media lacking leucine and tryptophan. After selection of double transformants, single colonies were selected and inoculated into 15 ml of selective media YNB glucose lacking tryptophan and leucine. Cell-free extracts were prepared followed by a western blot. Expression (and molecular weight) of all the LexA-DBD MOZ 676-763 mutants was confirmed using the anti-LexA antibody (Figures 4.3 A). This result also

confirmed that all the mutants are in frame and that these mutations have no effect on the stability on the proteins. VP16-AAD BRPF1 1-222 expression was again confirmed using the anti-VP16 antibody (Figures 4.3 B).

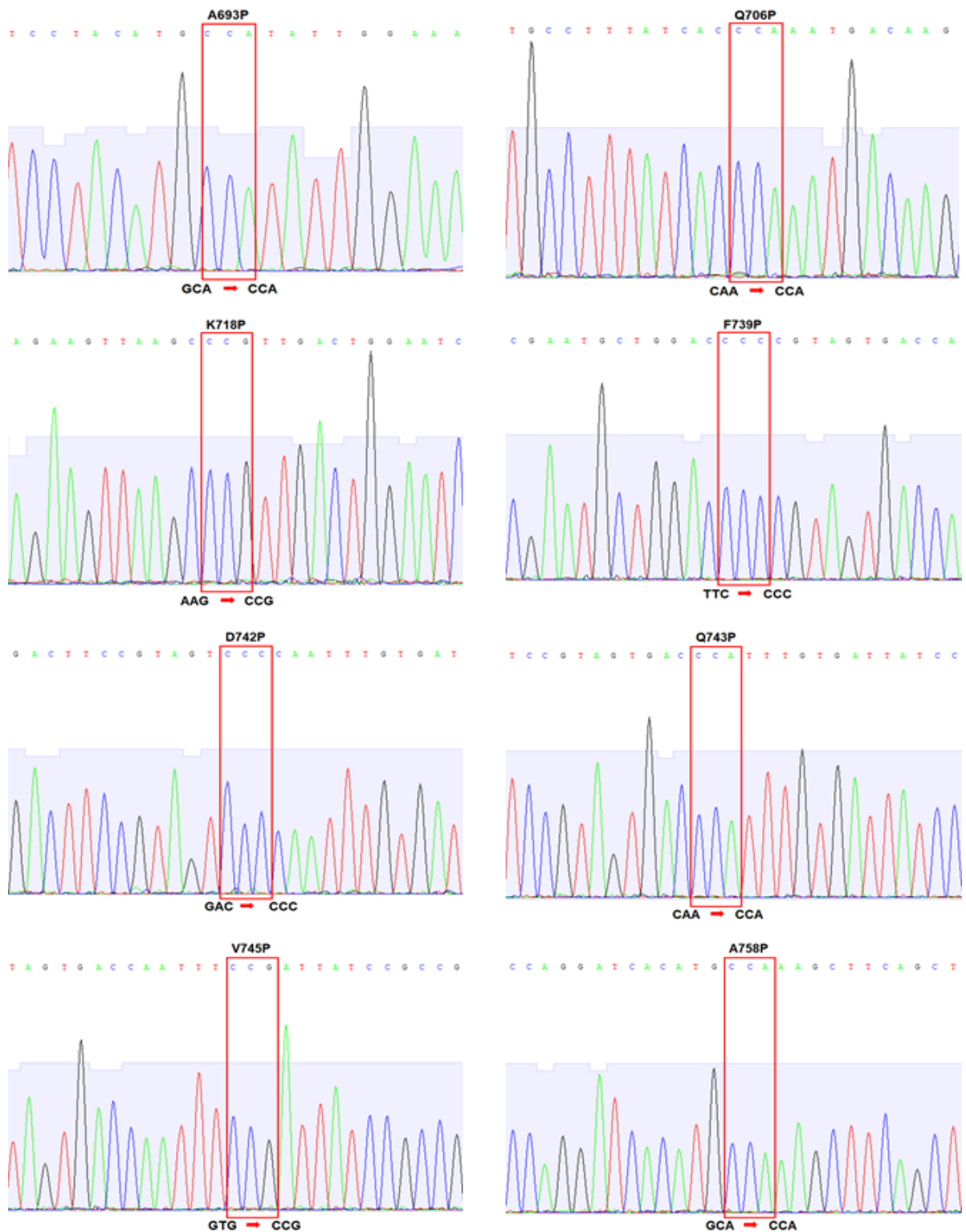


Figure 4. 2. Sequencing data to confirm the mutations.

Sequencing electropherograms are showing the confirmation of successful incorporation of all eight point mutations into MOZ 676-763 by site-directed mutagenesis. The changed codon

for the mutations marker with red box and changes of the codons are pointed at the bottom of each electropherograms for all eight point mutations.

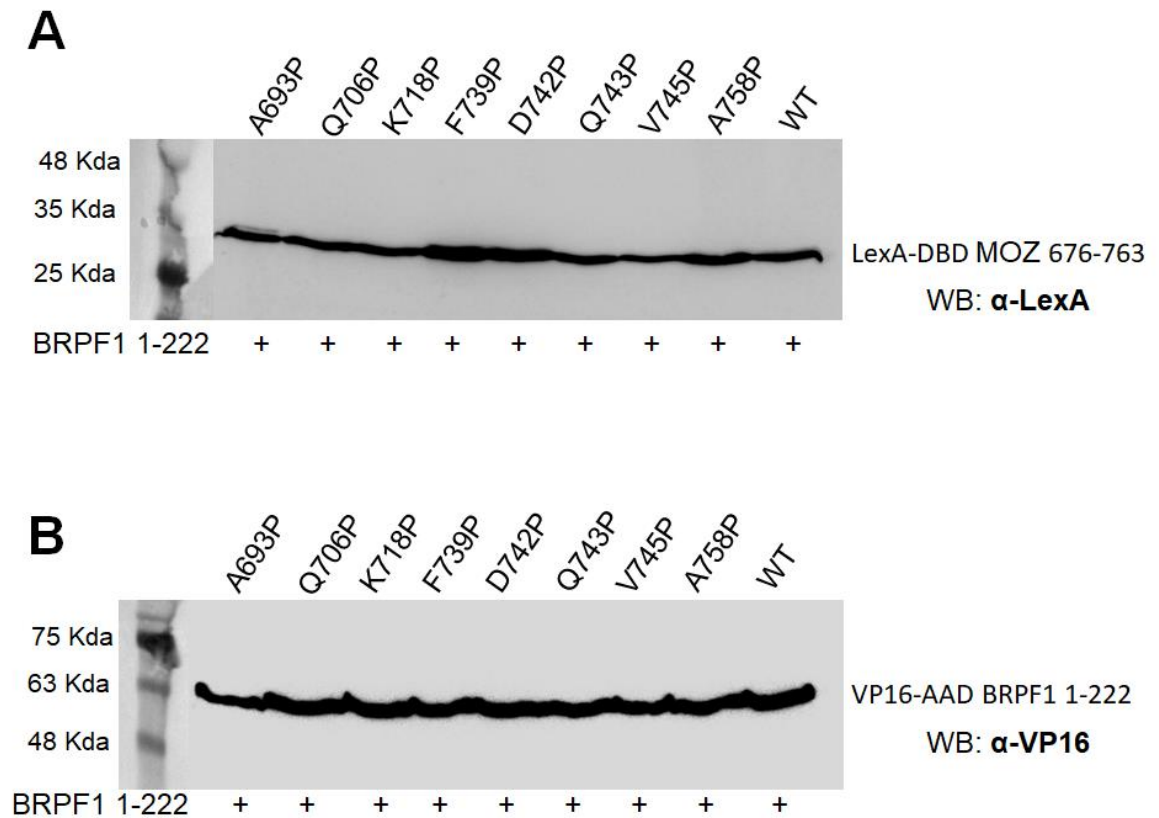


Figure 4. 3. Confirmation of expression of both BRPF1 1-222 and MOZ mutations.

A) Western blot result showing expression confirmation in yeast of all eight point mutations on LexA-DBD MOZ 676-763 along with wild-type MOZ 676-763 co-transformed with VP16-AAD BRPF1 1-222. **B)** Western blot confirming expression of the VP16-AAD BRPF1 1-222 fusion protein in the co-transformed yeast.

4.4. Disruption of MOZ/BRPF1 complex in yeast-two-hybrid assays

After successful incorporation of the desired point mutations into LexA-DBD MOZ 676-763 using site-directed mutagenesis, we performed yeast-two-hybrid reporter assay to check the effect of the mutations on MOZ/BRPF1 complex formations. All the mutants of LexA-DBD-MOZ 676-763 co-transformed separately with VP16-AAD BRPF1 1-222 into yeast reporter strain *S.cerevisiae* L40. The transformants were selected for double transformation on yeast selection media lacking leucine and tryptophan. Single colonies for each mutants were selected and inoculated into 15 ml of selective media YNB glucose lacking tryptophan and leucine. Cell-free extracts were prepared and β -galactosidase specific activities measured as described in previously.

The β -galactosidase assay results show that, out of eight point mutations, five of them completely disrupted the MOZ/BRPF1 interaction (Figures 4.4). As we can see from the β -galactosidase result, the A693P point mutation at the middle of the first α -helix of the helix-turn-helix motif of MOZ has no effect on the interaction, which suggested that this first α -helix does not take part in complex formation with BRPF1 directly. However, the mutation after the first α -helix, Q706P, completely disrupted the interaction. Likewise K718P, the point mutation at the turn before the second α -helix also completely disrupted the BRPF1 interaction. Out of four mutations from the unstructured loop part, three of them, F739P, Q743P, and V745P, similarly show complete disruption of the MOZ/BRPF1 complex formation. However, the other mutation D742P from that loop region and A758P from the third α -helix had approximately an eight fold reduction in specific activity in the β -galactosidase

reporter assay compared with wild-type but did not completely disrupt the interaction between MOZ and BRPF1.

Our yeast-two-hybrid result indicated that the unstructured loop region of the helix-turn-helix motif plays a crucial role in the interaction between BRPF1 and the MOZ MYST HAT domain. This site-directed mutagenesis study also suggested that, MOZ 676-763 is sufficient and essential for recruiting BRPF1. Importantly while we were investigating this MOZ/BRPF1 interface and trying to find the essential part for the interaction, Tao *et al.*, (2017) reported BRPF2 interacts with another MYST family protein HBO1 at the correspondence site. This group reported the crystal structure of the HBO1/BRPF2 complex, where it was reported that, BRPF2 binds to the HBO1 at the helix-turn-helix motif and more specifically at the anti-parallel β -strand region after the second α -helix which corresponds to the unstructured loop region of the MOZ MYST HAT domain. BRPF2's extended BN1 domain interacts with HBO1 double β -strands loop region of helix-turn-helix motif like a collar on the neck (Tao *et al.*, 2017).

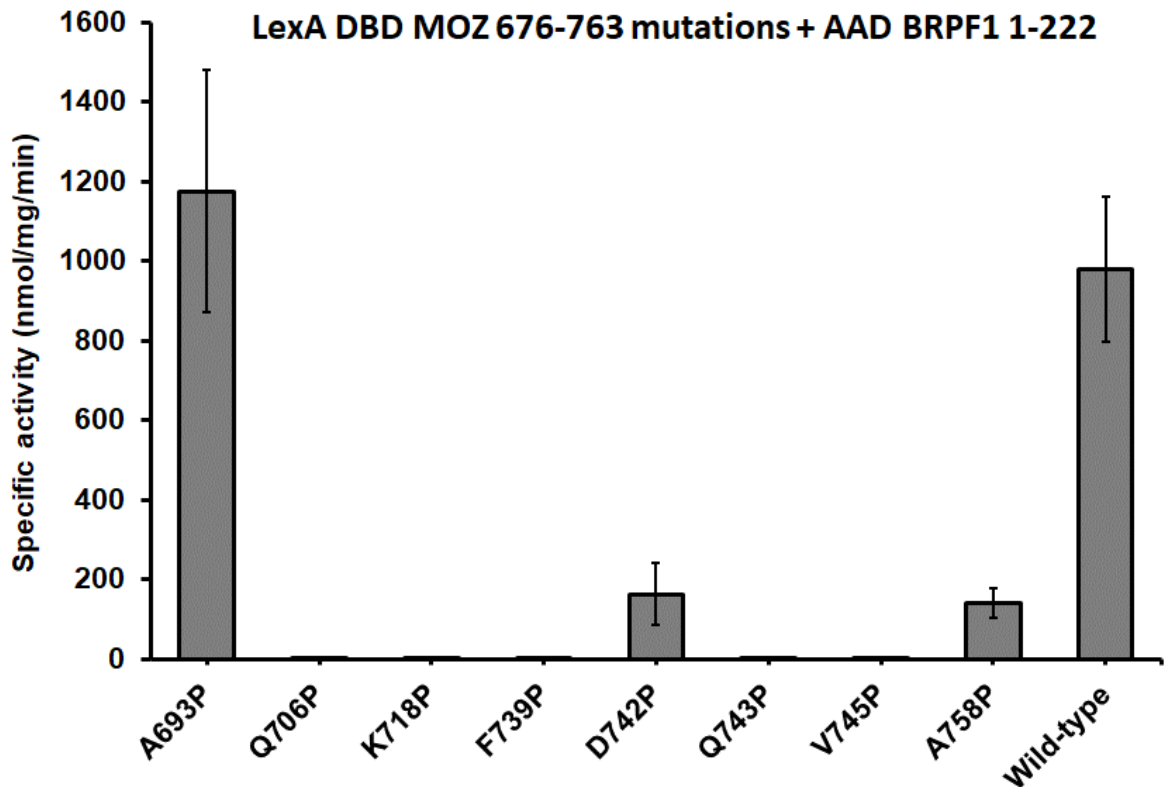


Figure 4. 4. β -galactosidase reporter assay from yeast two hybrid for all the LexA-DBD MOZ mutations vs VP16-AAD BRPF1 1-222.

β -galactosidase reporter assay result showing specific activity for all the MOZ 676-763 mutations along with wild-type MOZ 676-763 and BRPF1 1-222 interaction, where it indicates that first α -helix does not have any involvement in interactions with BRPF1 and the unstructured loop region of this helix-turn-helix region plays a crucial role in recruitment of BRPF1.

4.5. Attempts to disrupt MOZ/BRPF1 complexes in mammalian cells

The yeast two-hybrid results indicated that mutations within the helix-turn helix region of the MOZ MYST domain, and in particular the unstructured loop (MOZ 739-747) can disrupt the interaction between MOZ and BRPF1 subdomains, therefore we attempted to incorporate some of these mutations into mammalian expression plasmids to check the effect of the mutations in full length proteins using co-immunoprecipitation experiments.

Point mutations from the unstructured region (Q743P and V745P) were incorporated initially into the mammalian expression vectors FLAG MOZ 1-810 and FLAG MOZ 1-779 using site-directed PCR mutagenesis as before. After confirming the successful introduction of mutations into the respective expression vectors, they were co transfected with HA-BRPF1 full length into HEK293 cells. 48 hours post transfection, the soluble proteins were released by lysing the transfected cells using a phosphate lysis buffer. Input blots verified the expression of both the transfected proteins in the cell lysate (Figures 4.5 A). FLAG-MOZ proteins were then immunoprecipitated from the cell lysates with a FLAG antibody and tested for the HA-BRPF1 interaction by immuno-detection with the anti-HA antibody (Figures 4.5 B). Protein bands at ~135 kDa were detected using anti-HA antibody in co-IP membrane for both wild type and mutant FLAG-MOZ proteins (Figures 4.5 B). These bands were also detected on the input membrane at the same ~135 kDa size mark, which confirmed HA-BRPF1 full length. This result suggests that, even though the mutations in the MOZ 676-763 subdomain disrupt interaction with the N-terminus of BRPF1 in a yeast-two-hybrid assay, these mutations are not sufficient to disrupt the complexes involving full-length BRPF1 and N-terminal MOZ constructs in

mammalian cells. There are several possibilities to explain this; i) there are additional direct interactions occurring between other sequences in the overexpressed MOZ and BRPF1 proteins which are sufficient to allow the complexes to form in the mammalian cell experiment; ii) there is the possibility of third party involvement in mammalian cells which are absent from the yeast system e.g. ING5, MEAF6 which might stabilise the MOZ/BRPF1 complexes; iii) the mutations may disrupt the direct interaction of MOZ and BRPF1, but the proteins continue to be indirectly associated, e.g. via DNA or chromatin binding via their histone reader domains (e.g. MOZ DPF domain, and BRPF1 PZP, PWWP, bromodomains etc.); iv) lastly it is also possible that the yeast-two-hybrid interactions are artefactual, but given the evidence from other MYST/BRPF complexes, we believe this is the least likely explanation.

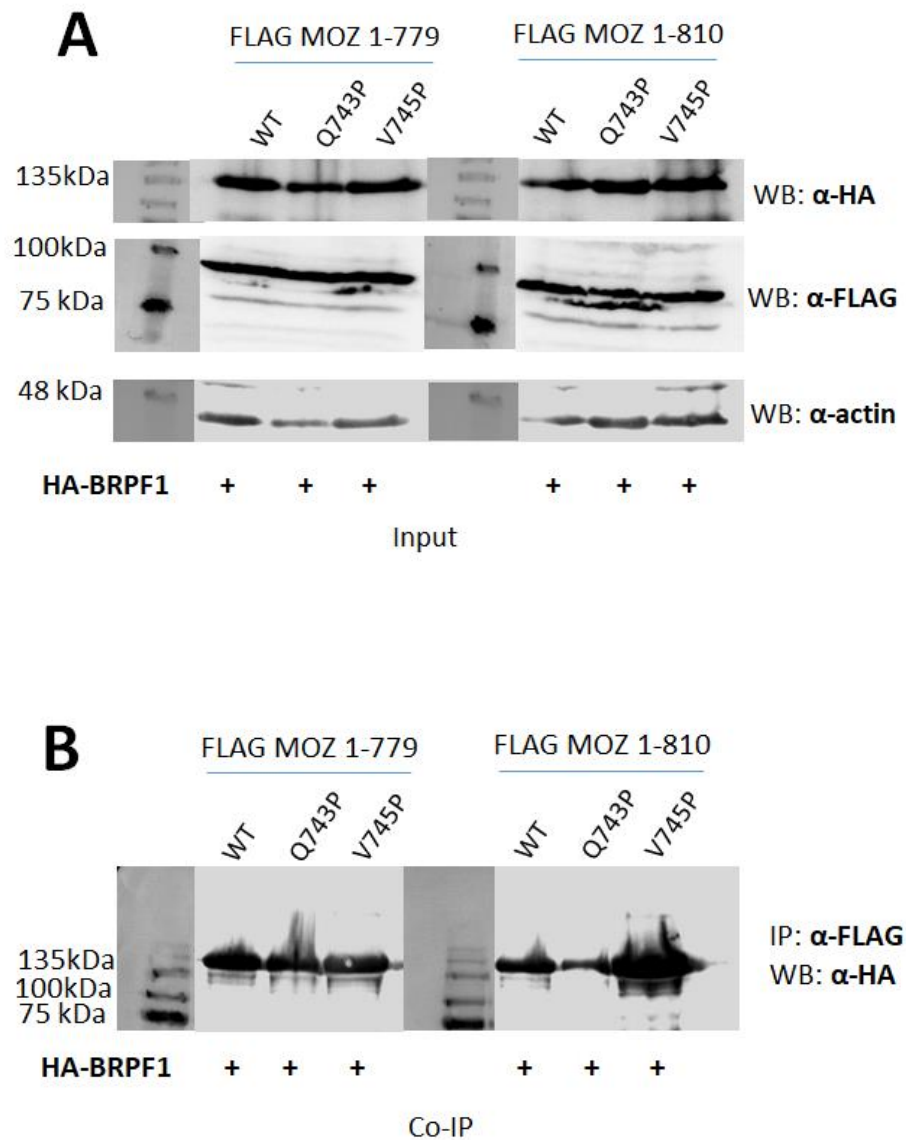


Figure 4. 5. Co-immunoprecipitation result for Q743P and V745P point mutation in FLAG-MOZ 1-779 and FLAG-MOZ 1-810.

Co-immunoprecipitation result showing the effect of point mutations *in vivo*. **A)** The confirmation of the expression of both HA-BRPF1 and FLAG-MOZ 1-779 or 1-810 (both wild type and mutant) in the cell lysate. **B)** Co-IP membrane showing that even though MOZ is mutated at the BRPF1 interacting site and HA-BRPF1 was still detected in the pulldown. This might be due to indirect interactions.

4.6. Targeting the loop ‘finger’ region (735-750) in full length MOZ

Single residue substitutions in FLAG-MOZ 1-810 or 1-779 were not successful in disrupting the co-precipitation of MOZ/BRPF1 complexes. Thus, we generated MOZ constructs in which we deleted the entire region encoding the unstructured loop /beta sheet finger by site-directed mutagenesis. This time we used the mammalian expression vector YFP-FLAG-MOZ full length in addition to FLAG-MOZ 1-810, to enable us to visualise any effect on the subcellular localisation of MOZ/BRPF1 complexes. The HA-BRPF1 full length protein was used in co-transfections as before.

Constructs containing the deletion of amino acid residues 735-750 (termed $\Delta\beta$) were confirmed by sequencing (Figures 4.6 A). HEK293 cells were co-transfected with HA-BRPF1 in conjunction with YFP-FLAG-MOZ FL, YFP-FLAG-MOZ FL $\Delta\beta$, FLAG-MOZ 1-810 or FLAG-MOZ 1-810 $\Delta\beta$ and 48 hours post transfection, cells were harvested and lysed to release the soluble proteins. Extracts were subjected to immunoprecipitation using the FLAG antibody. The detection of transfected proteins in the input lysates was confirmed by probing with the anti-HA antibody for HA BRPF1 and the anti-FLAG antibody for YFP-FLAG-MOZ FL ($\Delta\beta$ or wild type) and FLAG-MOZ 1-810 ($\Delta\beta$ or wild type). Interestingly, despite being sequence verified across the MOZ cDNA sequence, we were unable to detect FLAG-MOZ 1-810 $\Delta\beta$ in the input western blots (Figures 4.6 B). This suggests that the deletion of residues 735-750 may destabilise the FLAG MOZ 1-810 construct, or that there is an inadvertent mutation elsewhere in the expression plasmid. Given that we selected two independent clones of the construct, the latter seems unlikely.

The co-IP experiment using YFP FLAG MOZ FL or YFP FLAG MOZ FL $\Delta\beta$ showed that a strong interaction with HA BRPF1, was maintained in both cases, consistent

with the single point mutation data (Figures 4.6 C). Measurement of the band intensity obtained in co-IP using densitometry revealed a slightly reduced signal for the mutant compared to wild type (Figures 4.6 D). This result strengthens our previous hypothesis of either additional direct interactions between these proteins, or indirect contacts mediated by their binding to chromatin or other proteins.

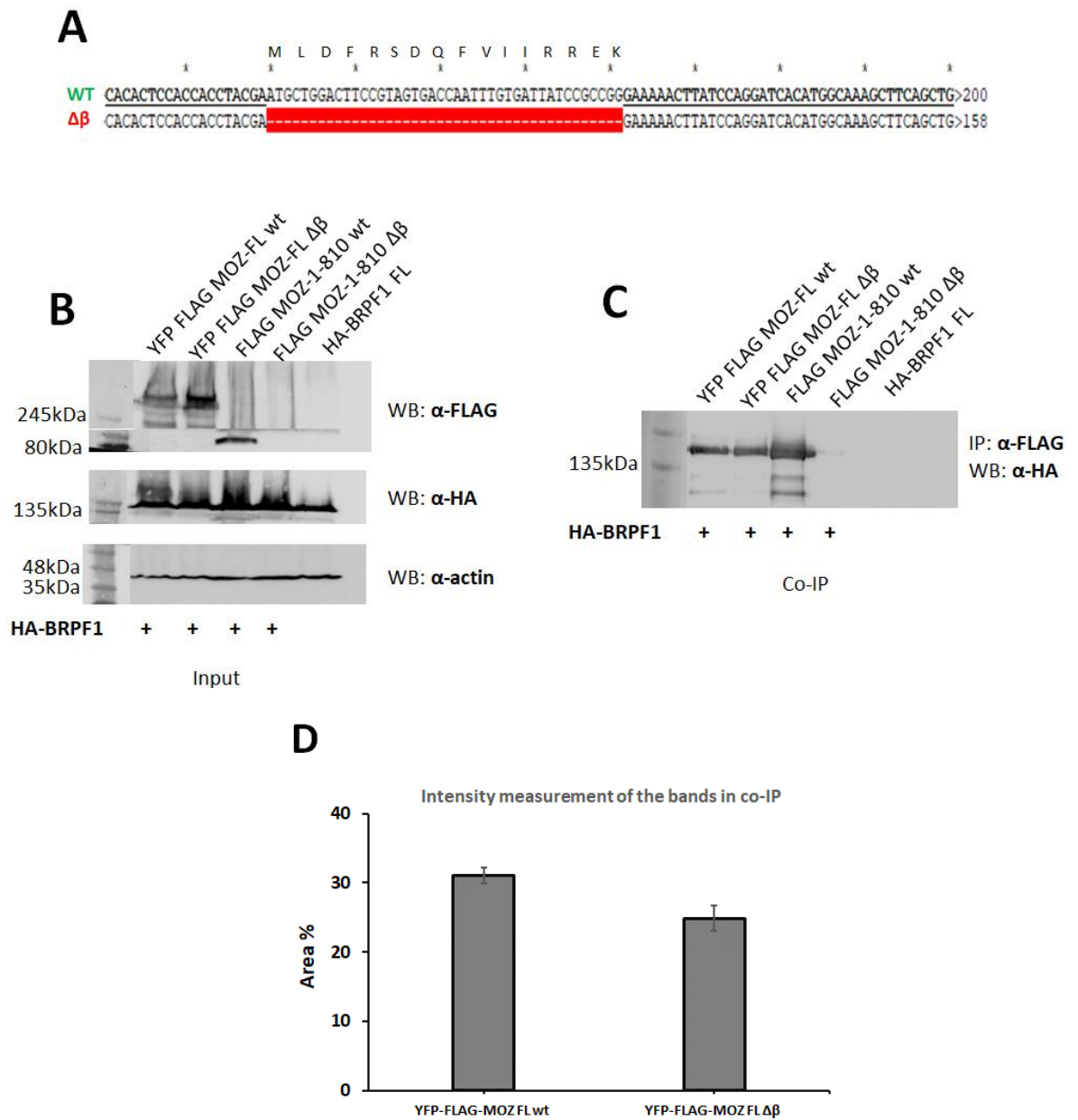


Figure 4. 6. Creating MOZ deleted double β -sheet ($\Delta\beta$) and co-immunoprecipitation result for deleted double β -sheet ($\Delta\beta$) on YFP-FLAG-MOZ full length and FLAG-MOZ 1-810.

A) Sequence alignment of wild type and mutant MOZ showing the successful deletion of the double β -sheet in the mutant plasmid. **B)** The confirmation of the expression of both HA-BRPF1 and FLAG-MOZ (both wild type and mutant) in the cell lysate, however FLAG-MOZ 1-810 $\Delta\beta$ failed to be detected by western blot. **C)** Co-IP membrane is showing even though MOZ is mutated at the BRPF1 interacting site and HA-BRPF1 was still detected. This might be due to indirect interaction but there was a reduction in the interaction. **D)** The bar chart confirming the reduction of interaction for YFP-FLAG-MOZ $\Delta\beta$ in comparison to wild type.

4.7. Co-localisation of BRPF1 and MOZ proteins in mammalian cell line

To assay MOZ/BRPF1 association intracellularly, we examined the subcellular localisation of exogenous YFP FLAG MOZ (full length wild type and mutant proteins) with HA BRPF1 full length proteins using confocal microscopy. In previous studies we have reported that MOZ is a nuclear protein that localises to nuclear speckles that are distinct from PML bodies or other known structures, as well as having a more diffuse nuclear distribution (Kindle *et al.*, 2005) YFP-tagged MOZ proteins display a similar nuclear distribution (Collins *et al.*, 2006). Other studies have reported that BRPF1 or HA-tagged BRPF1 shows a distribution in the cytoplasm (Laue *et al.*, 2008) whereas co- expression with MOZ, but not ING5 or MEAF6 can redirect BRPF1 protein to the nucleus.

HEK293 cells were seeded into eight well glass bottom μ -slide and transfected with YFP FLAG MOZ FL $\Delta\beta$ and HA-BRPF1. At 48 hours post-transfection, cells were immobilised, permeabilised and incubated with anti-HA antibody followed by an anti-rabbit Alexa594 conjugated secondary antibody. As YFP-FLAG MOZ fluoresces directly, antibody detection was not necessary to visualise this protein. The nucleus was visualised with DRAQ-5 stain which stains the DNA (purple). YFP-FLAG-MOZ FL (green) and HA-BRPF1 (red) were detected.

As shown in figure 4.7 D, YFP-FLAG-MOZ FL $\Delta\beta$ was present in the nucleus of transfected HEK293 cells. A non-uniform distribution pattern in the nucleus was observed, which is similar to the distribution we have seen before for wild type MOZ (Figures 4.7 A). Merging the image with the nuclear stain confirmed that YFP-FLAG MOZ FL $\Delta\beta$ was entirely nuclear. However, when HA-BRPF1 also co-transfected in

those HEK293 cells with YFP-FLAG MOZ FL $\Delta\beta$, the HA-BRPF1 was surprisingly found in the cytoplasm while YFP-FLAG MOZ FL $\Delta\beta$ was present in the nucleus. This subcellular localisation result confirms the mutation at the BRPF1 interaction site of MOZ disrupts the interaction, which we have also seen in the yeast two hybrid result and therefore, BRPF1 did not move into nucleus and remained in the cytoplasm due to disruption of the interaction. However, when MOZ wild type was co-expressed with BRPF1, both the proteins were found in the nucleus (Figures 4.7 C). This result also support out previous hypothesis about third party involvement for indirect interaction between MOZ and BRPF1 while both of them were released in cell lysate. As we have discussed before this indirect interaction could be due to interaction with histone tails for both MOZ as well as BRPF1. If we can somehow block the interaction between histones tails with MOZ or BRPF1 that could be beneficial in terms of disrupting the interaction between MOZ and BRPF1 in co-immunoprecipitation.

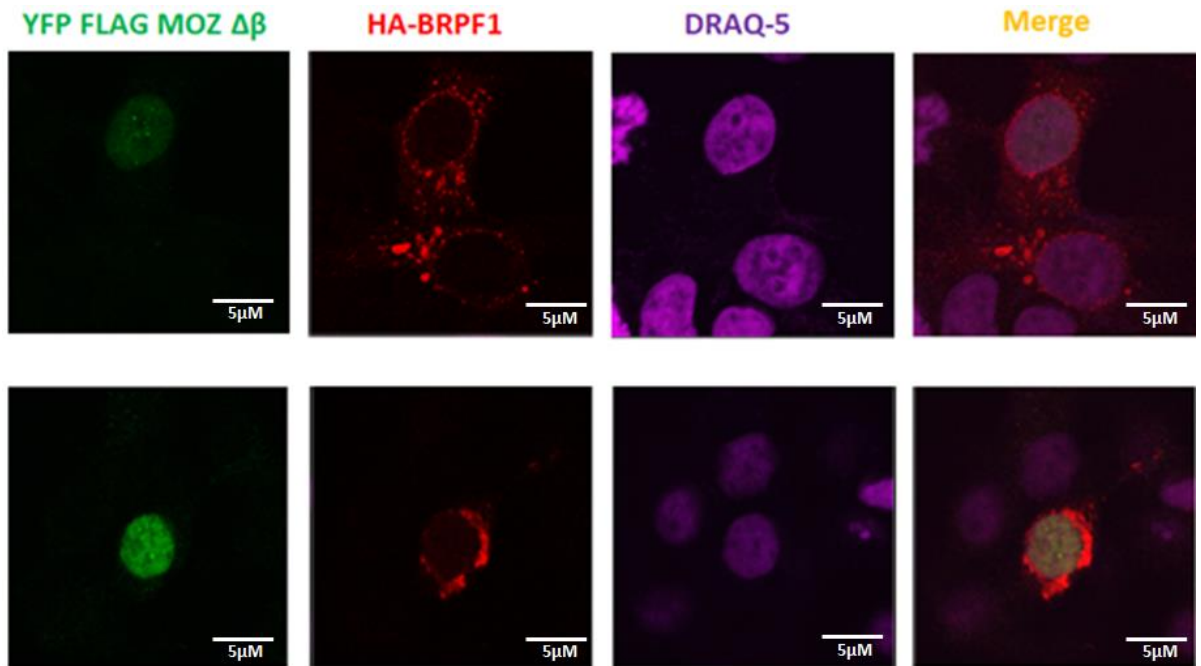


Figure 4. 7. Co-localisation of YFP-FLAG-MOZ FL $\Delta\beta$ and HA-BRPF1 in HEK293 cells.

Subcellular localisation of mutated MOZ and BRPF1 in HEK293 cells showing that BRPF1 remain in the cytoplasm when MOZ is mutated. Though mutated MOZ is present in the (similar to wild type) it failed to bring BRPF1 into nucleus due to loss of interaction.

4.8. BRPF1 bromodomain inhibitors partially disrupt MOZ/BRPF1 association

Recent advances achieved through genome sequencing studies indicate that chromatin regulators are frequently altered in tumours, and are therefore promising therapeutic targets to treat different tumour types. This has been demonstrated with the advance of HDAC inhibitors into clinical trials and the promise of bromodomain inhibitors in preclinical studies (Dawson *et al.*, 2009). Inhibitors of MOZ have recently been shown to block the advance of MOZ-TIF2 driven AML in mouse models (Baell *et al.*, 2018). Given the functional association of MOZ and BRPF1 in chromatin modification, the role of BRPF1 in targeting MOZ/MORF to chromatin (You *et al.*, 2015) and studies showing that BRPF1 is required for MOZ-TIF2 induced AML in animal models (Shima *et al.*, 2014), several groups have developed strategies to target BRPF1 function through small molecule inhibitors.

BRPF1 contains multiple domains that undergo direct interactions with histones or DNA in chromatin, including the PHD fingers domain (PZP), bromodomain and proline-tryptophan-tryptophan-proline (PWWP) domain. Small molecule inhibitors of bromodomain binding to acetylated lysines in H4 have been developed by several groups (Bennet *et al.*, 2015). In a collaboration with Dr Chung and colleagues at Glaxo Smith Kline, we examined the impact of a selective benzimidazole BRPF1 inhibitor GSK6835A and a poorly selective inhibitor GSK9311A on the interaction of BRPF1 with acetylated histone tails, and the effects on the assembly of MOZ/BRPF1 complexes in colP experiments.

The bromodomain is a conserved structural motif which recognises acetylated lysines, in particular in histones (Lubula *et al.*, 2014). It was also found that BRPF1

bromodomain is crucial along with PWWP domain for chromatin targeting (Laue *et al.*, 2008). The BRPF1 bromodomain binds more tightly to H4K12ac and H2AK5ac than any other histone tail modification, although some interaction with H3 via H3K14ac has been reported (Poplawski *et al.*, 2013; Lubula *et al.*, 2014). Recently the X-ray crystal structure for BRPF1 bromodomain in complex with H4K12ac and H2AK5ac histone peptides was revealed (Lubula *et al.*, 2014). There are several reports of selective inhibitors targeting BRPF1, including, 1,3 Dimethyl Benzimidazolones, which compete with acetylated lysine sidechains for occupancy of the bromodomain pocket (Demont *et al.*, 2014). Similarly, Palmer *et al.*, (2014) have used structure guided design to produce a series of N,N-dimethyl benzimidazolone compounds that inhibit the bromodomains of TIF1 α and BRPF1.

As part of this project we tested the two novel compounds provided by our pharmaceutical industry collaborator, GSK6835A and GSK9311A, which had been developed as having a high selectivity for BRPF1 bromodomain (Bamborough *et al.*, 2016). The development of specific inhibitors also provides valuable experimental tools to explore BRPF1 function and for testing in preclinical models.

4.8.1. Bacterial expression of BRPF1 bromodomain and large scale purification

The pET28 recombinant plasmid containing 6xhis tagged BRPF1 bromodomain (amino acid residue 622-738 is the boundary for BRPF1 bromodomain) received from our collaborator was transformed into *E.coli* BL21 and the positive clones were selected by using kanamycin selection LB agar plates. IPTG inductions were performed on mini-cultures (2ml) at 37°C to check for protein expression. As shown

in Figure 4.8A Coomassie Blue staining revealed the IPTG induction of a protein that migrated in the gel between the 10 kDa and 17kDa markers, consistent with the expected size (11kDa) of the recombinant bromodomain polypeptide. The clones were stored as frozen glycerol stocks at -80°C for further use.

After confirmation of expression from small scale induction, a glycerol stock from single clone was used to inoculate 1 litre LB (containing kanamycin 100µg/ml) for large scale expression and purification. After addition of IPTG the culture was divided into two and one culture allowed to grow at 18°C and the other at 37°C for overnight, to assess the effect of induction temperature. It was found that after adding IPTG, incubation at 18°C, helps with better BRPF1 bromodomain expression. The 6xHis-bromodomain was purified using nickel beads (Figures 4.8 B). The recombinant protein was eluted from the beads and this purified His-bromodomain protein was then used for further analysis.

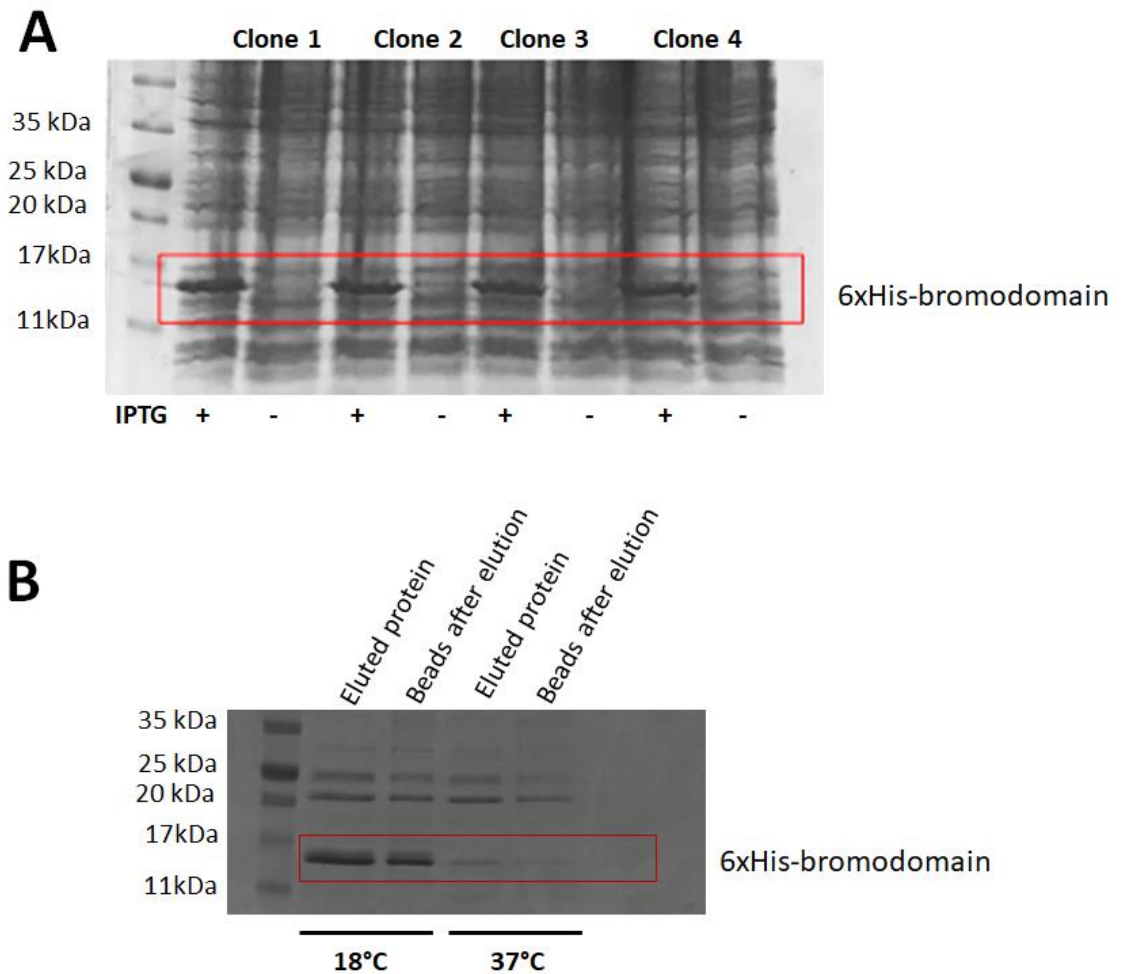


Figure 4. 8. Expression and purification of 6xHis-bromodomain in BL21 *E.coli*.

A) SDS-PAGE gel showing the successful induction of 6xHis-bromodomain in mini bacterial culture. **B)** From the mini-induction a clone was selected and large scale purification was performed. The SDS-PAGE gel picture is showing the purification of His-bromodomain using nickel beads. This is also showing that 18°C is the better temperature for expression of His-bromodomain, which might be due to better folding of the protein.

4.8.2. Detection of 6xHis-BRPF1 bromodomain and evaluating the specificity of bromodomain to histone peptide by western blotting

To confirm the identity of the BRPF1 His-bromodomain band, we performed a western blot. As no antibodies were available for this region of BRPF1, a penta-his specific antibody (Qiagen, cat# 34660) was used for detection. To determine the limit of detection different amounts of protein ranging from 1ug to 1ng were separated by SDS PAGE, followed by western blotting. To block the membrane generally we use 5% milk, however following the suppliers protocol, milk was substituted with 3% BSA. As shown in figure 4.9 A, the recombinant bromodomain protein was detected down to 250ng and anything below that was not detected. The detection for the His-bromodomain was linear according to the concentration.

After confirming the identity of the His-bromodomain of BRPF1, the specificity of the BRPF1 bromodomain binding to modified histone tails was assessed. As previously reported in the literature, the BRPF1 bromodomain is very specific for histone peptides, in particular acetylated H4K12 peptides and acetylated H2AK5. To confirm this is the case with our preparations of the BRPF1 bromodomain, binding assays with biotinylated histone tail peptides were performed as explained in material and methods chapter. In our experiment we tested biotinylated H4K12ac, unmodified H4, H4K5ac, H4K8ac and H4K16ac for BRPF1 bromodomain specificity. This results were consistent with the previous literature (Poplawski *et al.*, 2013) as the BRPF1 bromodomain showed strongest interaction with acetylated H4K12 (Figures 4.9 B). There was only very weak evidence if any, for the interaction of BRPF1 bromodomain with unmodified H4 or H4K5ac, H4K8ac and H4K16ac. The expression of BRPF1 bromodomain was confirmed in the first lane for input. This result helped us to select

the histone peptide containing H4K12Ac to assess the effect of the small molecule BRPF1 bromodomain inhibitor in our next step of this experiment.

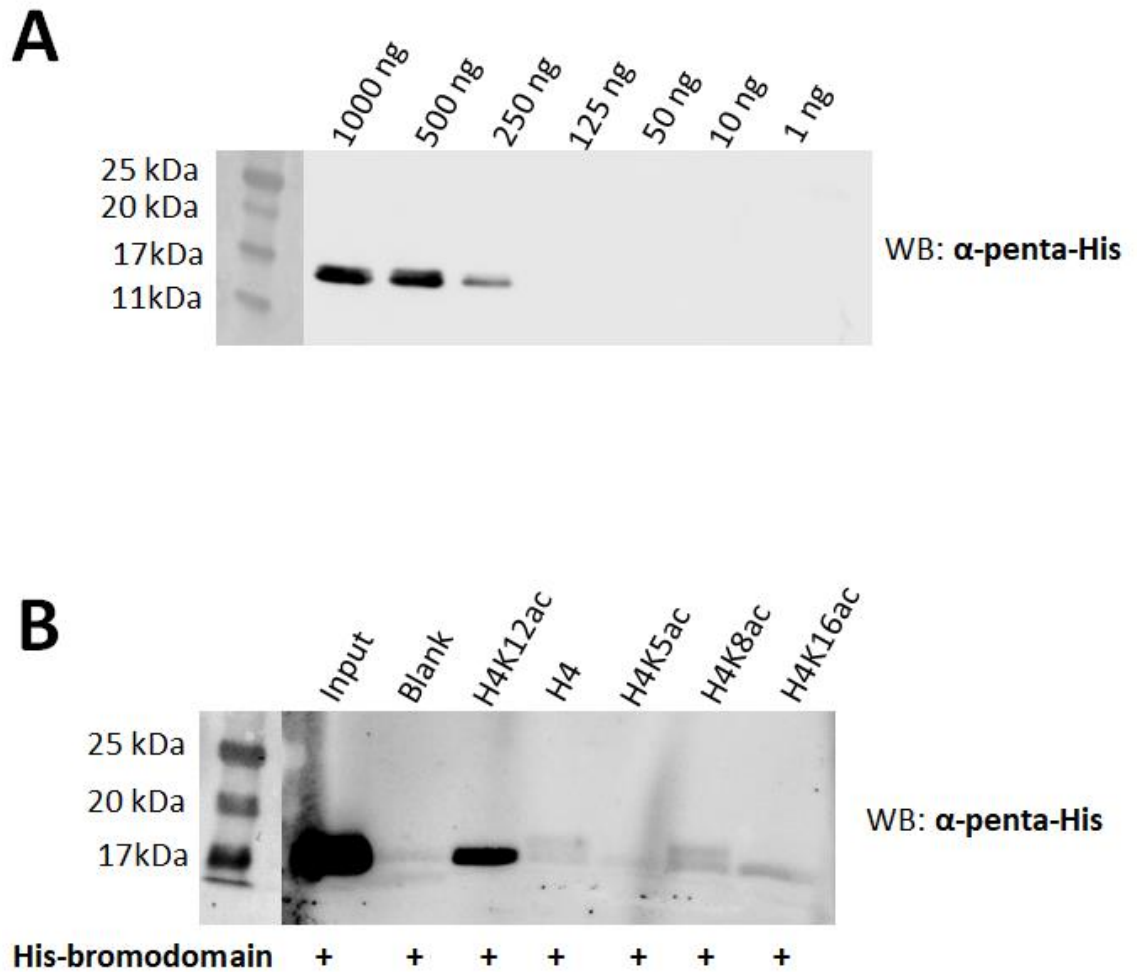


Figure 4. 9. Western blot result for detection and specificity assay of His-bromodomain.

A) His-bromodomain was subjected to western blot to check the minimum protein requirement for detection in western blot. This result shows that the minimum that can be detected 250 ng. **B)** The selectivity of BRPF1 bromodomain towards histone tails, it is specific for acetylated H4K12.

4.8.3. Effects of GSK6835A and GSK9311A on BRPF1 bromodomain binding to H4K12ac

Of the two small molecule compounds provided by our industrial partner, GSK6835A was developed to have high selectivity for the BRPF1 bromodomain compared to bromodomains of BRPF2 and BRPF3, whereas the alkylated analog GSK9311A was reported to show reduced BRPF1 inhibition (Bamborough *et al.*, 2016). After initial pilot assays *in vitro*, GSK6835A and GSK9311A were tested in binding assay at 10 μ M and 20 μ M, using vehicle (DMSO) as control. As shown in the figure 4.10, purified 6xhis-BRPF1 bromodomain was incubated at 4°C with the inhibitors respectively in binding buffer for an hour in a microfuge tube and in another microfuge tube biotinylated H4K12ac peptides were incubated with Dynabeads in binding buffer at 4°C for an hour. After one hour incubation the BRPF1 bromodomain and inhibitor mixture and the H4K12ac and Dynabeads mixture were merged together in one tube and incubated for overnight at 4°C. Next day the Dynabeads were processed and subjected to SDS-PAGE for western blot as described in Materials and Methods. In one sample bromodomain was incubated with vehicle DMSO to check whether DMSO has any effect on inhibition by itself.

As shown in Figure 4.11, GSK6835A was an effective inhibitor of BRPF1 bromodomain binding to H4K12ac, and this inhibition of H4K12Ac binding was dose-dependent, with complete inhibition observed at 20 μ M of GSK6835A. Some inhibition was also seen with GSK9311A at the highest dose. This result is consistent with the reported data that GSK6835A is a selective BRPF1 bromodomain inhibitor. A similar volume of vehicle DMSO did not have any apparent effect on interaction between BRPF1 bromodomain and H4K12ac. The input lane is to confirm the availability of the His-bromodomain. The positive control of this experiment was shown in the lane

H4K12ac only to confirm positive interaction between BRPF1 bromodomain and H4K12ac and the negative control of this experiment was shown in the lane marked as blank, where it was confirmed that in the absence of biotinylated H4K12ac, the Dynabeads alone could not bind with the BRPF1 bromodomain. Thus as GSK6835A is a selective BRPF1 bromodomain inhibitor, we assessed whether use of the inhibitor would help destabilise indirect association of MOZ and BRPF1 in the co-IP assays, as a consequence of chromatin association.

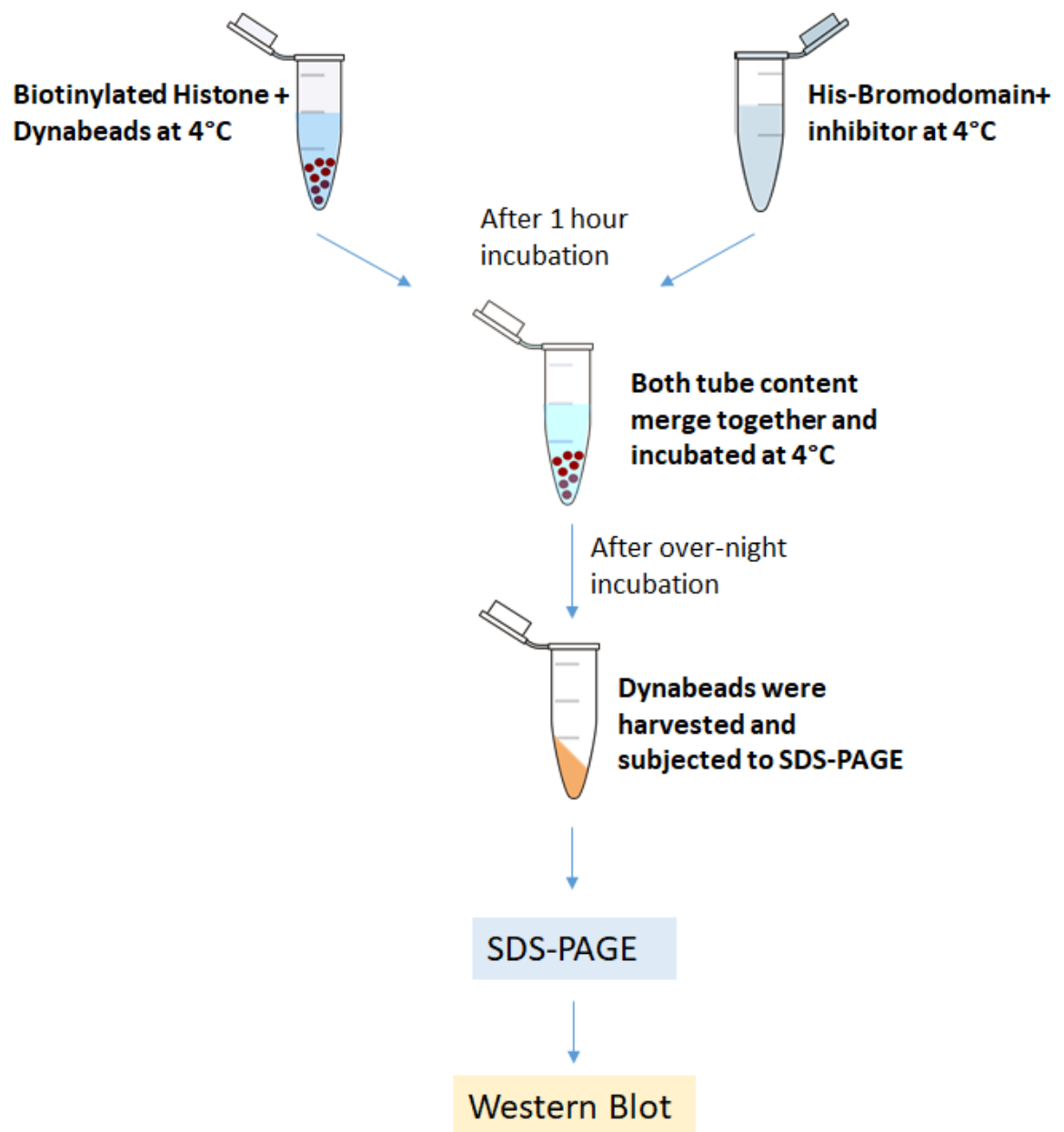


Figure 4. 10. Flow diagram for incubation stages for peptide binding and inhibition assay.

This flow diagram is showing that at first stage Dynabeads with histone peptides and BRPF1 bromodomain with inhibitors were incubated in separate tubes. Then after 1 hour incubation, both tube contents were merged together and incubated further overnight. Next day the Dynabeads were harvested and subjected to SDS-PAGE for western blotting.

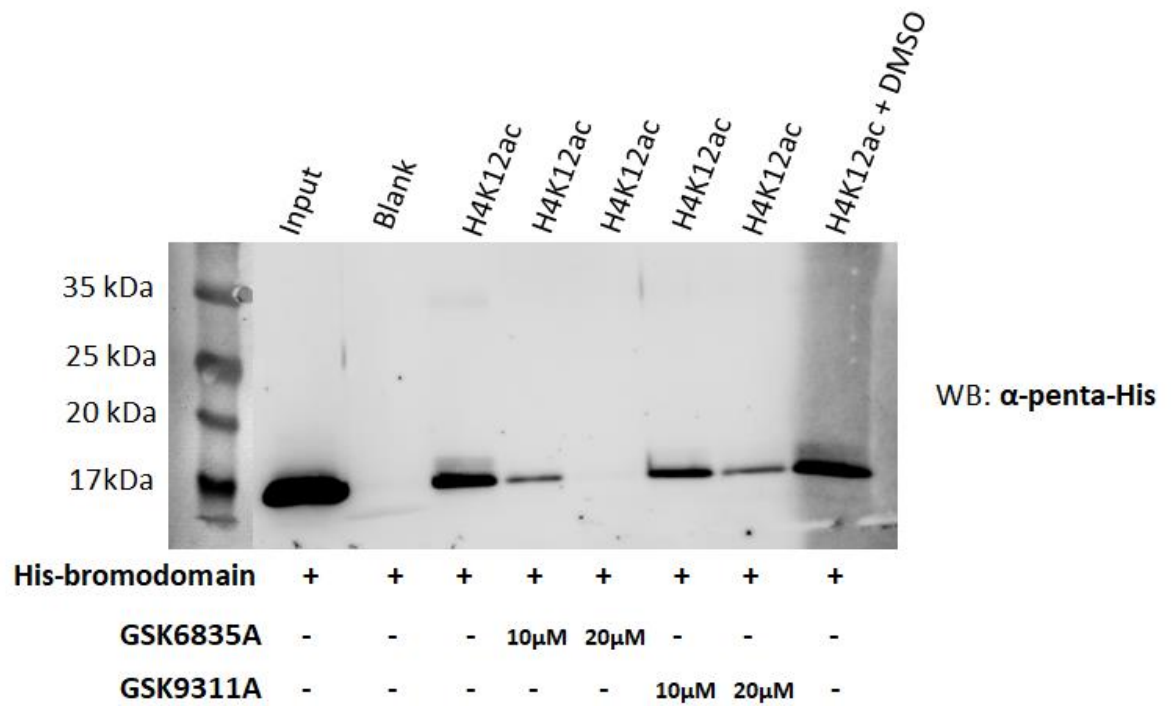


Figure 4. 11. Western blot result for BRPF1 bromodomain inhibition assay.

This western blot result is showing that among two inhibitor GSK6835A had better inhibitory effect on the BRPF1 bromodomain and H4K12ac interaction than GSK9311A. This inhibition is also dose dependent and it needs 20μM of inhibitor to inhibit the interaction. This is the best representation of three replicates of this experiment.

4.9. Effects of GSK6835A on MOZ/BRPF1 complex formation in mammalian cells

Having shown that GSK6835A can effectively block binding of BRPF1 to H4K12Ac peptides *in vitro*, we utilised this inhibitor in co-IP experiments. HEK293 cells were seeded in media containing 20 μ M of BRPF1 bromodomain inhibitor GSK6835A or control vehicle (DMSO) only. This was to ensure that both endogenous and transfected BRPF1 proteins would be exposed to the inhibitor during the experiment. The seeded cells were then co transfected with expression plasmids for full length HA-BRPF1 with full length wild type YFP-FLAG-MOZ, YFP-FLAG-MOZ full length $\Delta\beta$ or empty YFP vector control. 48 hours post-transfection the soluble proteins were extracted using a phosphate lysis buffer, followed by co immunoprecipitation using the anti-FLAG antibody.

Similar expression levels of wild type and mutant YFP-FLAG-MOZ proteins (top panel) or HA-BRPF1 (middle panel) in the inputs were confirmed as previously, with β -actin (lower panel) shown to indicate equal loading (Figures 4.12 A). The western blots indicated successful co-precipitation of HA-BRPF1 with YFP-FLAG-MOZ, which was moderately reduced with the YFP-FLAG-MOZ $\Delta\beta$ (Figures 4.12 B). Pre-inclusion of GSK6835A caused a further reduction in the amount of precipitated HA-BRPF1 with the WT and mutant MOZ proteins. Densitometry analysis revealed that the combination of inhibitor plus deletion of the BRPF1 binding sequence equated to around 50% reduction in the level of HA-BRPF1 protein detected compared to YFP-FLAG-MOZ with no inhibitor (Figure 4.12 C). No HA-BRPF1 was detected when co-transfected with empty pCDNA3.1 vector (Figure 4.12B, lane marked as 'HA-BRPF1 FL' only) and inclusion of GSK6835A had no effect on HA-BRPF1 expression levels (Figure 4.12A) This result supports the hypothesis that binding of BRPF1 to

acetylated H4 may contribute to the assembly of MOZ/BRPF1 complexes on chromatin. However, as other domains in MOZ (DPF) BRPF1 (PZP, PWWP) or their associated proteins may also contact histones and DNA, it may be difficult to completely disrupt these complexes without targeting multiple interactions. Alternatively, it remains possible that other regions of MOZ and BRPF1 mediate direct interactions of these proteins.

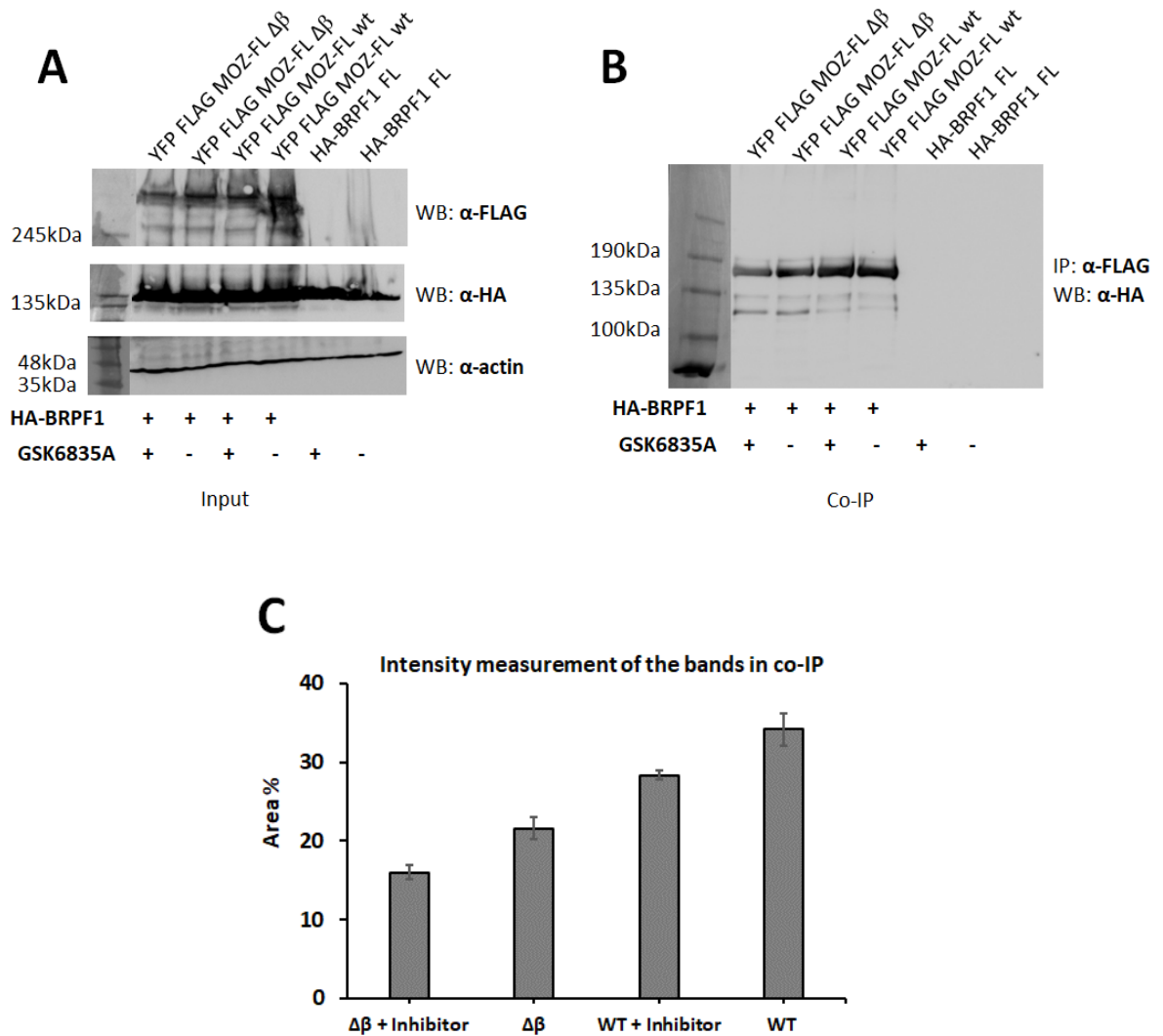


Figure 4. 12. Co-immunoprecipitation result for deleted double β -sheet ($\Delta\beta$) on YFP-FLAG-MOZ full length in presence of bromodomain inhibitor.

Co-immunoprecipitation result showing the effect of deletion of β -sheets on BRPF1 interaction site of MOZ *in vivo*. **A)** The confirmation of the expression of both HA-BRPF1 and YFP-FLAG-MOZ full length (both wild type and mutant) in the cell lysate. **B)** co-IP membrane is showing YFP-FLAG-MOZ $\Delta\beta$ had reduction of interaction with BRPF1 and in the presence of the bromodomain inhibitor the interaction reduced further. **C)** Bar chart is confirming the reduction of the band intensity for the MOZ/BRPF1 interaction when MOZ is mutated and cells were treated with the bromodomain inhibitor.

4.10. Summary

Previous work on MOZ/BRPF1 complex interface had summarised the minimal region required for MOZ to interact with BRPF1, which were MOZ 676-763 with BRPF1 59-222 as BRPF1 59-190 failed to express in yeast and as we discussed before it remain consistent with Ullah *et al.*, (2008) for BRPF1 mapping and partially consistent for MOZ mapping. Therefore it was important to show the minimal region of MOZ we have mapped in our previous chapter, is essential for the recruitment of BRPF1. As in this chapter we have explored the effect of various point mutation on the MOZ 676-763, the minimal region of BRPF1 recruitment, in our yeast-two-hybrid result we have seen that out of eight point mutations five of them had completely disrupted the interaction. The mutations, which disrupted the interactions, majority of them were located at the MOZ unstructured loop. This unstructured loop region has been predicted to consist a pair of β -sheet. During this study, Tao *et al.* (2017) published the crystal structure of BRPF2/HBO1 complex. According to their structure we can see that BRPF2 fits around the double β -sheet region of HBO1 like a collar on the shoulder (Figures 4.13 A). From our mutation studies we also found that exactly the same corresponding region of MOZ is important for interaction with BRPF1. As from structural analysis we already know that MOZ MYST domain and the HBO1 MYST domain are highly homologous and the essential region of interaction also very similar, therefore the interaction of MOZ and BRPF1 could be similar like HBO1/BRFP2 complex. In figure 4.13 B, we tried to superimpose MOZ MYST domain with the HBO1 MYST domain containing the BRPF2 fragment as a complex. This superimposed structure revealed that the MYST domains from both proteins aligned perfectly and small BRPF2 polypeptide fits nicely on both the MYST domain. The loop region of MOZ MYST domain at the helix-turn-helix motif has an

unstructured region (739-747) due to lack of electron density. If we can generate the structure of MOZ/BRPF1 complex with the minimal interacting region we may solve the structure for that unstructured loop of MOZ MYST domain. BRPF1 peptide (as it is a small interacting region only) may have stabilise the surrounding region for formation of the complex and that may lead to solve the structure. The region of the MYST domain of MOZ we have mapped and proved as the essential for interaction with BRPF1, is also found to be same correspondence sequence of interacting site of HBO1 with BRPF3 (Yan *et al.*, 2015). This similar correlation of interaction between MYST family proteins and BRPF family proteins suggesting that in BRPF proteins interacts with MYST proteins at similar region (Yan *et al.*, 2015).

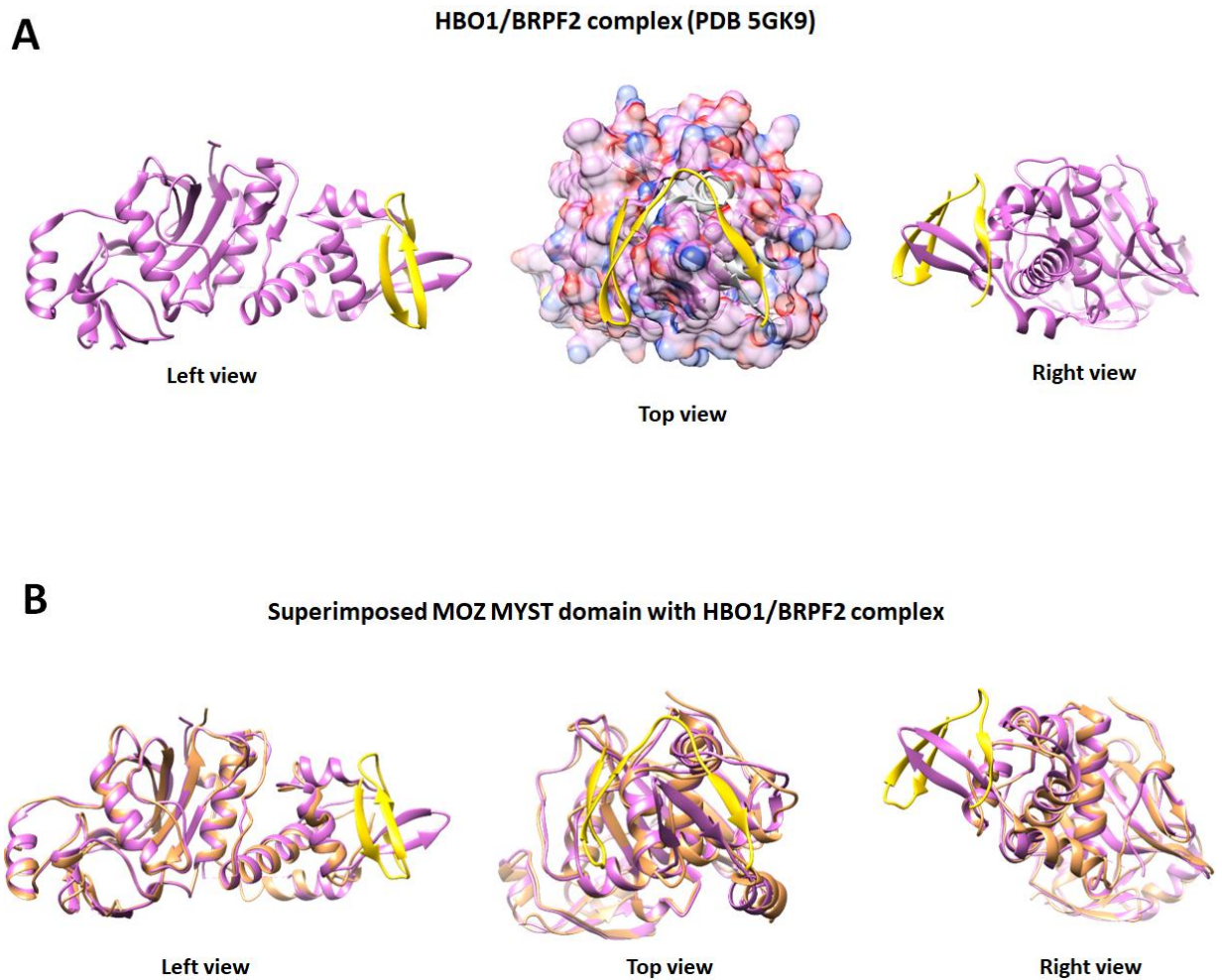


Figure 4. 13. Crystal structure of HBO1/BRPF2 complex.

A) Crystal structure of HBO1/BRPF2 complex showing how BRPF2 fits on the MYST domain of HBO1. BRPF2 peptide fits on HBO1 helix-turn-helix region as a collar on the shoulder.

B) MOZ MYST domain was superimposed with HBO1/BRPF2 complex structure and showing that BRPF2 peptide also nicely fits on the MOZ MYST domain in exactly similar way.

Although we have managed to disrupt the interaction between MOZ and BRPF1 in yeast-two-hybrid assay, in co-immunoprecipitation BRPF1 was detected in the lane for mutated MOZ, which suggested there was still an interaction between MOZ and BRPF1 even though the BRPF1 interaction site is deleted from MOZ. However in sub-cellular co-localisation study we clearly showed that when MOZ is mutated at the BRPF1 interaction site (YFP-FLAG-MOZ FL $\Delta\beta$), the interaction with BRPF1 is completely lost and BRPF1 remained in the cytoplasm, where MOZ was still found in the nucleus. The control study with wild type MOZ and wild type BRPF1 confirmed the interaction between them and BRPF1 moved to the nucleus from the cytoplasm. That means there is a clear interaction loss between them *in cellulo*, when MOZ is mutated. Now the question is why BRPF1 was still immunoprecipitated with mutated MOZ? The hypothesis we came up with, there are several domains of both MOZ and BRPF1 interacts with histone, which linked with chromatin and when both the proteins released in soluble lysate, both interacted with histone tails linked with chromatin. Due to this indirect interaction BRPF1 was still detected in co-IP even though the MOZ was mutated.

Therefore we decided to block one of the histone interacting site of any of these two proteins. Our industrial collaborator was also developing two specific inhibitors for BRPF1 bromodomain to block the interaction with histone tails. To assess the specificity of those inhibitors, we started to check the specificity of BRPF1 bromodomain towards histone tails. The outcome from the initial studies, suggested that BRPF1 bromodomain is very specific to the acetylated H4K12, which is consistent with previous works from (Poplawski *et al.*, 2013 and Lubula *et al.*, 2014). After finding out the specificity of BRPF1 bromodomain for the modified histone tails,

we next assessed the inhibitory effect of the newly developed BRPF1 bromodomain specific inhibitors (Bamborough *et al.*, 2016). From our peptide binding assay for the inhibitory effect, we found that GSK6835A inhibitor is more specific than GSK9311A. The result showed that GSK6835A had inhibited the interaction between BRPF1 bromodomain and histone H4K12ac. However, this inhibitory effect is dose dependent as we found that 20 μ M of GSK6835A was required to inhibit the bromodomain and H4K12ac binding completely.

After successfully blocking the BRPF1 bromodomain from binding with histone tails, we repeated the co-IP experiment but this time cells were treated with bromodomain inhibitor GSK6835A. As expected, there was a clear drop of band intensity from the western blot of co-IP for the mutated MOZ lane. There was almost 50% drop in the band density for the cells treated with bromodomain inhibitor and transfected with mutated MOZ. Positive control with wild type MOZ (+/- inhibitor) confirms that drop was due to lack of indirect interaction because of bromodomain blockage for histone bindings. However, there was not complete disappearance of the band for BRPF1 in co-IP, as BRPF1 still has other domains (PZP or PWWP), which interacts with histone tails (Vezzoli *et al.*, 2010; Liu *et al.*, 2012; Champagne *et al.*, 2008 and Qiu *et al.*, 2012) and also MOZ (MYST or DPF) interacts with histone tails (Kitabayashi *et al.*, 2001; Dreveny *et al.*, 2014). This co-IP findings supports the hypothesis that MOZ and BRPF1 has indirect interaction due to their interaction with histones and because of that, we still found BRPF1 immunoprecipitated even though MOZ was mutated.

Chapter 5: Functional Investigation of MOZ/BRPF1 Complexes

5.1. Introduction

The previous two Results chapters described the identification of MOZ 676-763, encoding a potential helix-turn-helix motif, as a BRPF1 interaction site. Site-directed mutagenesis studies of this sequence confirmed that, the entire helix-turn-helix motif is essential for BRPF1 recruitment. We have also seen that when MOZ and BRPF1 are released in cell lysates, they may have indirect association due to their interaction with different histone tails dependent on the presence of post translational modifications, namely specific lysine acetylations. In this section I will describe experiments to investigate how interaction with BRPF1 may impact on MOZ function. So far only MOZ MYST domain crystal structure is available but there is no structure available for the MOZ/BRPF1 complex like another MYST family protein HBO1 and BRPF2 has.

To facilitate functional studies for the MOZ and BRPF1 complex, we employed a KAT6A/MOZ deficient HEK293 cell line generated in the laboratory using CRISPR/Cas9 technique. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas9 is a cutting edge technology in recent molecular biology research for high-throughput genome editing in cells. The genome editing by CRISPR technique evolved as a bacterial survival mechanism to recognise and destroy viral genome permanently (Horvath *et al.*, 2010). In molecular biology, the Cas9 endonuclease target is guided by small homologous RNA sequence of the targeting locus and this RNA sequence is known as guide RNA. The guide RNA is the 20 nucleotide upstream sequence from the PAM (5' NGG) sequence. The Cas9 endonuclease complexes with the guide RNA to recognise its specific DNA target and creates a double stranded break (DSB) at 3-4 nucleotide upstream from the PAM

sequence (Mali *et al.*, 2013). The DNA repair mechanisms are activated due to this DSB and that generate indels via non-homologous end joining (NHEJ), which leads to inactivating mutations. Alternatively, a homologous template DNA with specific and desired modification can be used in DSB repair in homology directed repair (HDR), an approach that enables precise editing by introducing a point mutation or recombining sites (Shen *et al.*, 2014; Chiang *et al.*, 2016). The Cas9 endonuclease creates the double strand breaks by its two nuclease domain, HNH and RuvC. However, in gene targeting by Cas9, using both of these two nuclease domains, HNH and RuvC, can often lead to off target event (Hsu *et al.*, 2014). To reduce this off target event, one of the crucial amino acid residue of RuvC, D10 has been mutated and that leads to achieve modified Cas9 nickase enzyme, which creates only single strand break on the target DNA. Therefore, to create nick on the complementary DNA strand, it is essential to have another modified Cas9 nickase, which targets the opposite strand to generate a double strand break on the targeted gene (Hsu *et al.*, 2014). In this way the targeting becomes gene specific and off target events are reduced dramatically. In this chapter I describe preliminary rescue studies in which Wild type or mutant MOZ proteins are used to rescue gene expression defects in the CRISPR edited HEK293 cells and controls.

In addition to functional studies of MOZ/BRPF1 regulation of target gene expression, in this chapter I will also describe progress with regard to co-expression of MOZ and BRPF1 minimal interacting sequences towards obtaining the NMR or crystal structures of the MOZ/BRPF1 complex interface. While these studies were in progress last year, a report by Tao *et al.*, 2017 reported detailed mapping of the interface between MYST protein HBO1 and the BRPF2 protein. The report by Tao *et*

al., (2017) revealed that the BRPF2 N-terminal sequence encompassing the BN1/BN2 domain is required to bind HBO1, and fits on the helix turn helix region of the MYST like a collar, specifically enfolding the anti-parallel β -sheet forming the turn. The structure is stabilised by both hydrophobic and hydrophilic interactions. They also showed that, the N-terminus of BRPF2 is also potential to bind with free histones along binding with HBO1 and that can potentiate the HAT activity of HBO1. A stable crystal structure was obtained using the HBO1 MYST domain in complex with BRPF2 31-80, but not other fragments. Intriguingly the sequence 31-80 of BRPF2 is sufficient to enhance the H3K14 acetylation activity of HBO1, suggesting that a change in the structure of the MYST domain occurs upon binding BRPF2.

The BN1/BN2 region of BRPF2 was also found to have negatively charged patches that may be important for interaction with histone H3 and this potential histone binding site and HBO1 binding site are in close proximity. The structural and sequence similarities of MOZ and HBO1 as well as BRPF1 and BRPF2, suggest their functions and interactions might also have strong similarity. The BRPF2 binds with HBO1 around the β -sheet loop area, however this region of MOZ has no structure available due to lack of electron density (PDB 2RC4, Holbert *et al.*, 2007). The interaction with BRPF1 accessory proteins can affect the activity and substrate specificity of the HAT complex (Ullah *et al.*, 2008). At the outset of the project, the interaction regions of MOZ and BRPF1 were only partially mapped by other groups, and no structures were available for the MYST proteins in complex with their BRPF/JADE partners. Thus, having mapped the minimal interaction domains in the yeast two hybrid system and demonstrated the requirement for the MOZ minimal region in mammalian cells, we set out to co-express MOZ and BRPF1 minimal regions in *E.coli* to allow us to purify

them as a complex for crystallisation. This structural analysis of the MOZ/BRPF1 complex might provide us new insights into how the HAT activity of MOZ is controlled by BRPF1.

5.2. CRISPR edited MOZ-deficient HEK293 cell line

MOZ knockout HEK293 cells were generated by our group using CRISPR/Cas9 nickase technology (CJ Monteiro, HM Collins, J Whitchurch, B Mazumder, DM Heery unpublished data). For generating these MOZ knockout HEK293 cells, it started with designing the single guide RNA (sgRNA). This sgRNA is 20 nucleotides upstream of the PAM sequence (5' NGG), however the PAM sequence is not integrated within the sgRNA. In this case, the sgRNA sequences were designed to target exon 3 of MOZ gene or KAT6A and designed by using Zhang lab CRISPR target finder tool (crispr.mit.edu/nick). The complementary oligonucleotides were annealed together to generate a BbSI compatible overhang sequence on both side. The annealed sgRNA sequence was then ligated in BbSI restriction digested pX461 vector and transformed into *E. coli* DH5 α strain. The pX461 vector contains the endonuclease Cas9 encoding sequence along with a sequence for U6 RNA polymerase III promoter and a GFP encoding sequence. The clone was then confirmed by sequencing and this CRISPR/Cas9 vectors with targeted sgRNA sequences are now ready to be transfected into HEK293 cells.

For nickase targeting (Figures 5.1a), HEK293 cells were transfected with vectors expressing sgRNA pairs and as a non-targeting control cells were transfected with a single sgRNA. Single strand nicks are repaired to wild type and our laboratory has not observed any mutations using these controls. Transfection was performed by

using NeonTM Transfection System (Invitrogen). As mentioned before about GFP expression by pX461 vector, after 48 hours post-transfection GFP positive cells were sorted using Beckman Coulter MoFlo XDP cell sorter and 96 well plate culture was started with each well receiving a single GFP positive cell. After growing them for 3-4 weeks some cells were harvested from the cultures of targeted and control (single sgRNA) cells. The harvested cells were then subjected to genomic DNA extraction by using Sigma GenElute Mammalian Genomic DNA Miniprep Kit (G1N350) and followed by PCR reaction using primer flanking the region 285 and 983 of MOZ (KAT6A) gene (CJ Monteiro, personal communication). Following the PCR reaction, the PCR products were gel purified and sequenced to confirm the alterations in the KAT6A gene. The sequencing result confirmed that, targeted clones had indels in the KAT6A exon 3 whereas single guide transfected cells repaired the single nick and remained wild type. Two clones (G10 and E3) were selected for further analysis and all alleles were cloned and sequenced, confirming disruptive deletions in Exon 3. Western blots confirmed that in significant loss of MOZ expression in clones G10 and E3, whereas and the single guide control (AF10) showed wild type expression levels (Figures 5.1B). Moreover, the westerns also revealed a strong reduction in the H3K14 acetylation, consistent with the role of MOZ in targeting this histone modification (Figures 5.1B) These data were further confirmed by RNA sequencing (RNAseq) showing the loss of reads corresponding to the targeted region in clone G10, but not the AF10 control (Figures 5.1C). Analysis of the RNAseq data identified a set of differentially expressed genes (both up and downregulated) (Monteiro *et al.*, in preparation). We selected several of these genes from the list for analysis in rescue experiments, as described in the next section.

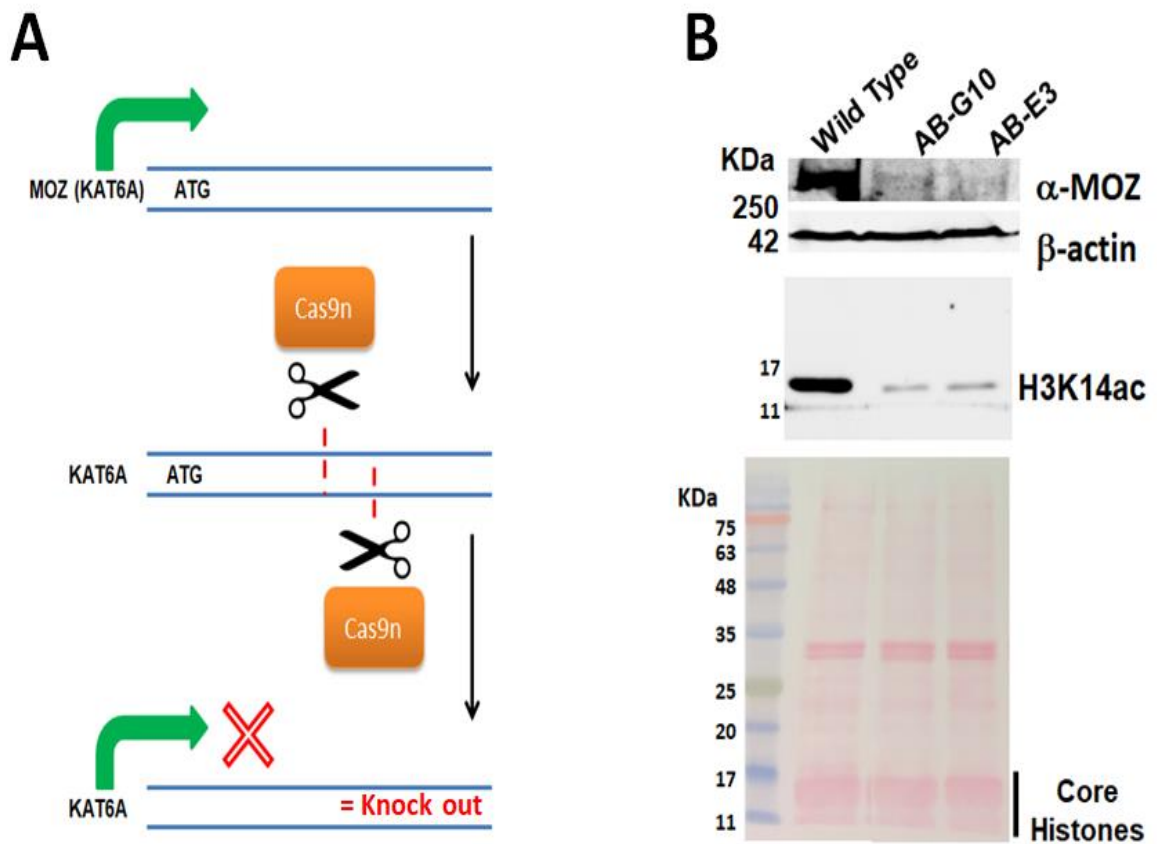


Figure 5. 1 (a). Schematic for CRISPR/Cas9 activity on MOZ gene (KAT6A).

A) Schematic diagram showing the steps for how CRISPR/Cas9 nickase technology work on KAT6A and knockout MOZ expression in mammalian cells. **B)** Western Blots showing the expression of MOZ proteins in the control (WT) and CRISPR edited HEK293 cells. Clone G10 was used in our study. The reduction of H3 K14 acetylation mediated by MOZ was confirmed and equal loading of the extracted histone preparations is shown for comparison. These data were generated by Dr Cintia Monteiro in our group and shown here with permission to justify why this clone was used.



Figure 5. 2 (b). Computer generated data showing deletion at Exon 3 of MOZ gene (KAT6A) after CRISPR/Cas9 targeting.

Schematic showing RNA Seq generated reads from HEK293 control cells treated with a single sgRNA (AF10) and HEK293 G10 (MOZ-deficient) cells transfected with sgRNA pair to target the ATG start codon in KAT6A Exon 3. The absence of sequences aligning to the targeted region is consistent with the deletions of this region in the allele sequencing. The data were visualised in the Integrated Genome Viewer and the figure was generated by Prof.David Heery.

5.3. MOZ Regulation of Gene Expression: Rescue Experiments in HEK293 G10 Cells

Based on the RNAseq data analysis, we selected several DEGs, those were strongly upregulated in HEK293 G10 clone compared to control cells, suggesting a role for MOZ in suppressing these genes. In haematopoietic stem cells (HSCs), MOZ has been demonstrated to suppress genes involved in senescence and differentiation, e.g. CDKN2A thus maintaining the stem-like properties of HSCs (Sheikh *et al.*, 2016). Similarly, BRPF1 knockout studies (Yan *et al.*, 2017) resulted in mostly upregulation of target genes, suggesting that MOZ/BRPF1 complexes may suppress expression of some gene networks. Thus, in this experiment we have tried to rescue the MOZ knockout phenotype by exogenous expression of MOZ. We wanted to check whether genes whose expression is altered expression is restored to the normal level of expression when MOZ is introduced into the HEK293 G10 cells. The selected genes are listed in the Table 5.1. In this experiment, a mutant MOZ protein deficient in BRPF1 binding was used for comparison to WT MOZ.

We have previously shown that a YFP-tagged version of MOZ functions similarly to wild type MOZ in that it shows similar nuclear localisation, activates AML1 dependent transcription in reporter assays, and acetylates H3 in a similar way to untagged MOZ protein (Kindle *et al.*, 2005; Collins *et al.*, 2006; Kindle *et al.*, 2010). The MOZ knockout HEK293 G10 cells and single guide control HEK293 AF10 cells were seeded for transfection. The cells were transfected in triplicate using calcium-phosphate method with wild type YFP-FLAG-MOZ FL (full length) or a YFP FLAG-MOZ FL $\Delta\beta$ mutant expression vector to disrupt BRPF1 interaction. After 48 hours post transfection YFP positive cells were sorted using Beckman Coulter MoFlo XDP cell sorter and 50,000 transfected cells were collected for each condition. After that

total RNA was extracted from the triplicate batches of sorted cells and reverse transcribed to make cDNA. The cDNA was used for RTqPCR analysis relative to GAPDH, and performed with specific primers for each of the selected target genes.

The qPCR data were found to be consistent with the previous RNAseq data, as it was shown that while the expression of SLC7AC, HMOX1, TIMP1, SPARC and RUNX3 was very low or undetectable in the control AF10 cell line, expression of these genes was strongly increased in MOZ-deficient HEK293 G10 cells (Figures 5.2). Interestingly, YFP-FLAG-MOZ was able to suppress the expression of SLC7AC, HMOX1, TIMP1, SPARC and RUNX3 and restoring them back to normal wild type level (Figures 5.2). This confirms the role of MOZ proteins in negative regulation of some genes, although it is not clear from these data if this is direct or indirect. Unexpectedly, YFP-FLAG-MOZ-FL $\Delta\beta$ was also able to suppress the five genes tested. However, we noted a somewhat reduced ability of, this construct to reach complete expression in several cases, suggesting that the BRPF1 recruitment site in the MYST domain is not essential for this function. However, a number of issues must be considered as follows;

- It is not known whether the selected DEGs are directly regulated by MOZ or BRPF1. It may also be the case that restoration of MOZ in G10 cells upregulates another MOZ target gene, which in turn suppresses the SLC7AC, HMOX1, TIMP1, SPARC and RUNX3 genes in a mechanism independent of BRPF1.
- It is unknown whether the depletion of MOZ in G10 cells destabilises the expression or localisation of endogenous BRPF1 complexes.

- Although HA-BRPF1 is located to the cytoplasm even when endogenous MOZ is present, we are unable to determine the subcellular location of endogenous BRPF1 due to absence of good quality antibodies.
- It is not known whether the YFP-FLAG-MOZ constructs can associate with endogenous BRPF1 proteins, which may already be assembled in other complexes.
- It is possible that YFP-FLAG-MOZ-FL $\Delta\beta$ and BRPF1 can be recruited to chromatin sites independently of their direct interaction, due to their chromatin reader domains.
- We did not have time to analyse whether YFP-MOZ rescues the global levels of H3K14 acetylation

However, the rescue experiments described show that it is feasible to use this approach to explore the functions of other MOZ domains e.g. the DPF domain, the MYST acetyltransferase activity etc. Thus, future work may focus on those areas.

List of chosen gene suppressed by KAT6A	Function of the chosen genes
SLC7AC	Sodium-independent cationic amino acid transporter.
SPARC	Encodes secreted proteins.
RUNX3	Transcription factor.
TIMP1	Metallopeptidase inhibitor.
HMOX1	Heme oxygenase functioning in heme metabolism.

Table 5.1. Showing list of selected genes repressed by MOZ and their functions in human.

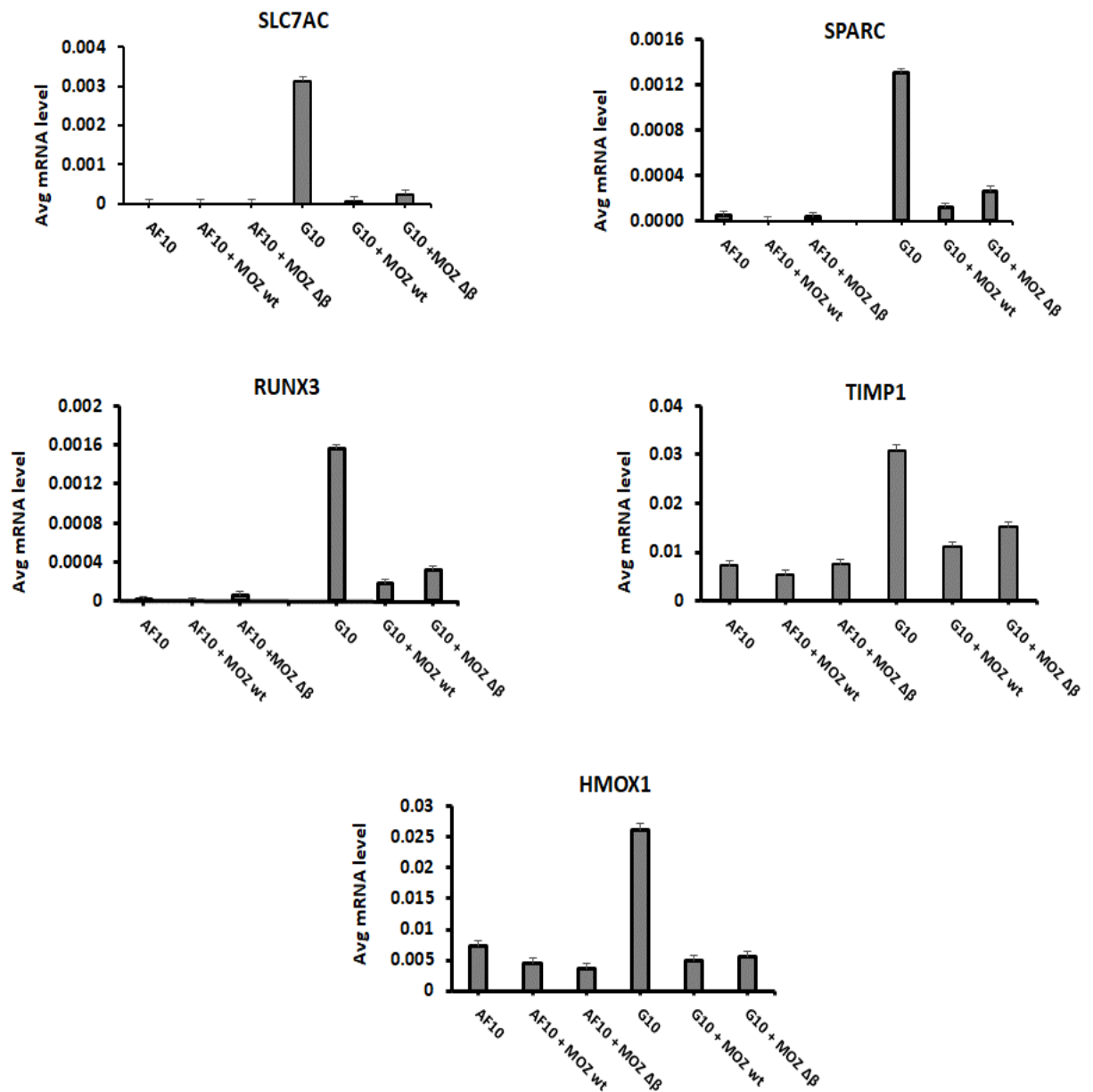


Figure 5. 3. Rescuing the gene expression by exogenous MOZ expression in MOZ knockout HEK293 cells.

RTqPCR result showing the mRNA level of selected genes. This result is also showing that when MOZ was knocked out using CRISPR/Cas9 these genes were over expressed and when they were transfected with MOZ wild type expressing vector the expressions for those gene came back to normal. It is also showing that mutant MOZ had slight less effectiveness in terms of bringing those expression back, which indicated that, BRPF1 might have a role on MOZ in gene expression control.

5.4. Co-expression of MOZ and BRPF1 minimal interacting sites in *E.coli*

At the outset of this study, no structures were available for the MOZ/BRPF1 complex or similar complexes of MYST proteins with BRPF or JADE family proteins. We therefore attempted to co-express the MOZ 676-763 region with the BRPF1 59-222 (BN1/BN2 domain) using expression in bacteria. Previous attempts to express a his-tagged construct consisting of BRPF1 59-190 were unsuccessful, possibly due to instability of the isolated BN1/BN2 region.

5.5.1. Gibson assembly cloning

The pACYC Duet-1 dual expression vector was selected for co-expression of MOZ and BRPF1 polypeptides. The cloning method used in this experiment was Gibson assembly method. For this process NEBuilder HiFi DNA assembly master mix was used, which contains mainly an exonuclease, DNA polymerase, dNTPs, ligase and ATP. In this process primers were designed with a 30 nucleotides overhang sequence along with our gene of interest on both side. As explained in figure 5.3, the inserts were generated by PCR with the overhang of the vector sequence. The vector was linearized by restriction digestion. Both linear vector and the newly generated inserts were gel purified in 1% agarose gel. 100ng of purified vector and 50ng of purified inserts and equal volumes (combining the total volume of insert and vector together) of HiFi master mix were mixed together and incubated for 50°C for 15 minutes. During this incubation T5 exonuclease removes extra 5' ends from the vector and creates 3' overhang and then the inserts with vector overhang fit in the vector. DNA polymerase repairs if there is any extra mismatch sequence due to

restriction digestion and at the end ligase enzyme ligate the inserts into linearized vector. After this 15 minutes incubation the whole reaction was transformed into E. coli DH5 α strain. Next day after obtaining the colonies single colony was selected and initially confirmed the cloning by colony PCR and followed by sequencing for final confirmation of the cloning. The selection for this vector transformation was by chloramphenicol selection.

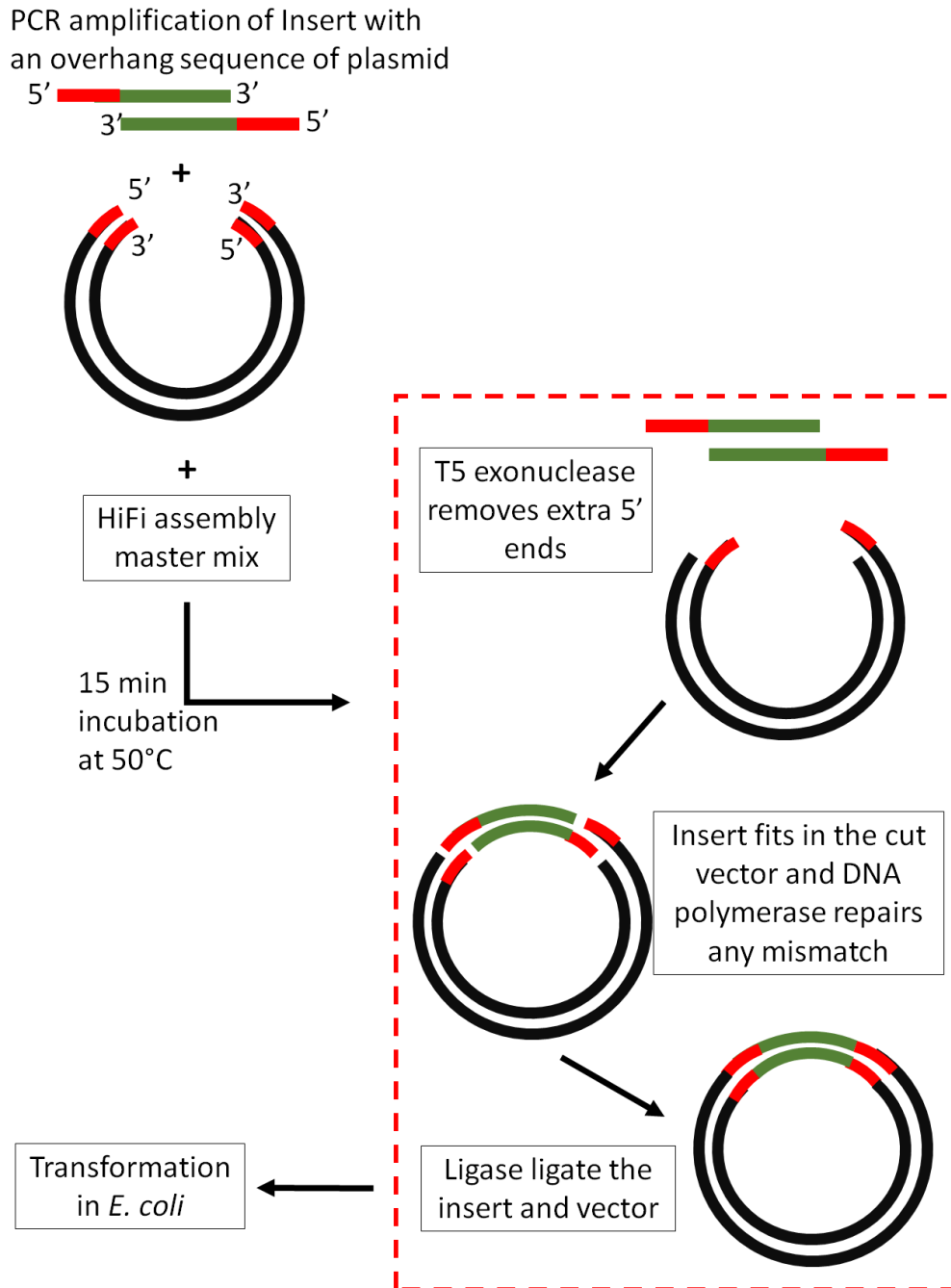


Figure 5. 4. Gibson assembly cloning steps in flow diagram.

Flow diagram showing the Gibson assembly steps where gel purified PCR generated insert with an overhang sequence of the target vector on both side incubated with linear vector and HiFi assembly master mix. During incubation T5 exonuclease removes extra 5' overhang sequence. The DNA polymerase repair if there any mismatch and at the end ligase enzyme ligate the insert with vector.

5.4.2. Cloning BRPF1 into first multiple cloning site of pACYC Duet-1

Initially it was decided that BRPF1 59-222 will be cloned into pACYC Duet-1 first multiple cloning site (MCS-1), which generates an in frame fusion with sequence for 6xHistidine tag for purification. Therefore as final product it will generate His-tagged BRPF1 59-222 fusion protein. The Gibson ready BRPF1 59-222 insert was generated by PCR and gel purified (Figures 5.4 C). The MCS 1 of the vector pACYC Duet-1 was digested with BamHI and HindIII restriction enzymes (Figures 5.4 B) and followed by gel purification. Next the insert and the linear vector were incubated with HiFi master mix at 50°C for 15 minutes followed by transforming into *E. coli* DH5 α . Selected clones were confirmed by colony PCR (Figures 5.4 D) and sequencing. After this confirmation this pACYC Duet-1 with BRPF1 59-222 was transformed into *E. coli* BL21 DE3 strain to assess expression of the BRPF1 polypeptide. Several clones were selected and expression was induced in mini scale cultures with 1mM IPTG. SDS-PAGE analysis by Coomassie brilliant blue staining had confirmed the induction of the expression of His-BRPF1 59-222 (Figures 5.4 E).

While pursuing this project for the crystallisation of MOZ/BRPF1 complex, Tao *et al.*, (2017) published the crystal structure of HBO1/BRPF2 complex. Their boundary of the interaction sites fall exactly within the regions we mapped for the MOZ/BRPF1 complex. One difference to our study is that they only used the HBO1 full MYST domain in their experiments, whereas our data shows that the helix turn helix subdomain is both necessary and sufficient for BRPF1 recruitment. Interestingly, the Tao *et al.* study further defined the minimal BRPF2 sequence required to bind HBO1 as residing within a 50 residue sequence (31-80), and this was used in their crystal structure. Based on their results we also attempted to clone the corresponding

sequence of BRPF1 (90-139). This smaller BRPF1 sequence BRPF1 90-139 was cloned into pACYC Duet-1 in similar way how BRPF1 59-222 was cloned. After confirming positive cloning the BRPF1 90-139 containing pACYC Duet-1 was transformed into *E. coli* BL21 DE3 and induced the expression with IPTG. This detected an induced band at around 20kda, higher than the expected size for the construct. However, sequence analysis confirmed the correct insert was present, therefore we assumed that this band corresponded to His-BRPF1 90-139 (Figures 5.5 C), so we proceeded to the next step of cloning MOZ 676-763 in MCS-2 of pACYC Duet-1.

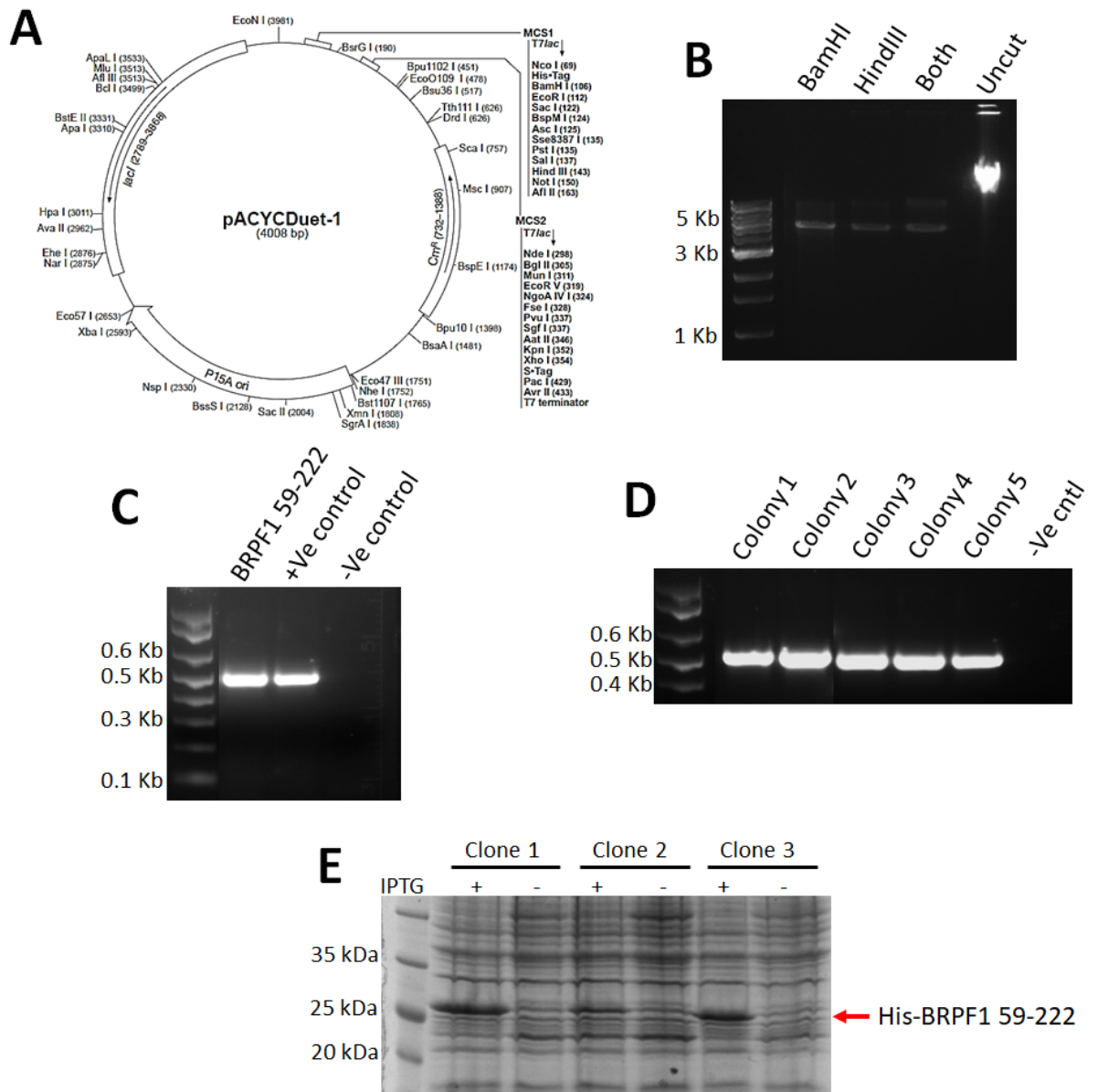


Figure 5. 5. Cloning of BRPF1 59-222 into pACYC Duet-1 vector at MCS-1 and mini scale induction.

A) Plasmid map for pACYC Duet-1 showing the position of both multiple cloning site and their tags. **B)** Agarose gel image showing the digestion of the pACYC Duet-1 vector with BamHI and HindIII restriction enzymes. **C)** Agarose gel showing the generation of BRPF1 59-222 insert by PCR. **D)** Confirmation of cloning showing in 1% agarose gel by colony PCR for BRPF1 59-222. **E)** SDS-PAGE gel analysis showing the induction of His-BRPF1 59-222 in presence of IPTG. It was appeared at ~25 kDa.

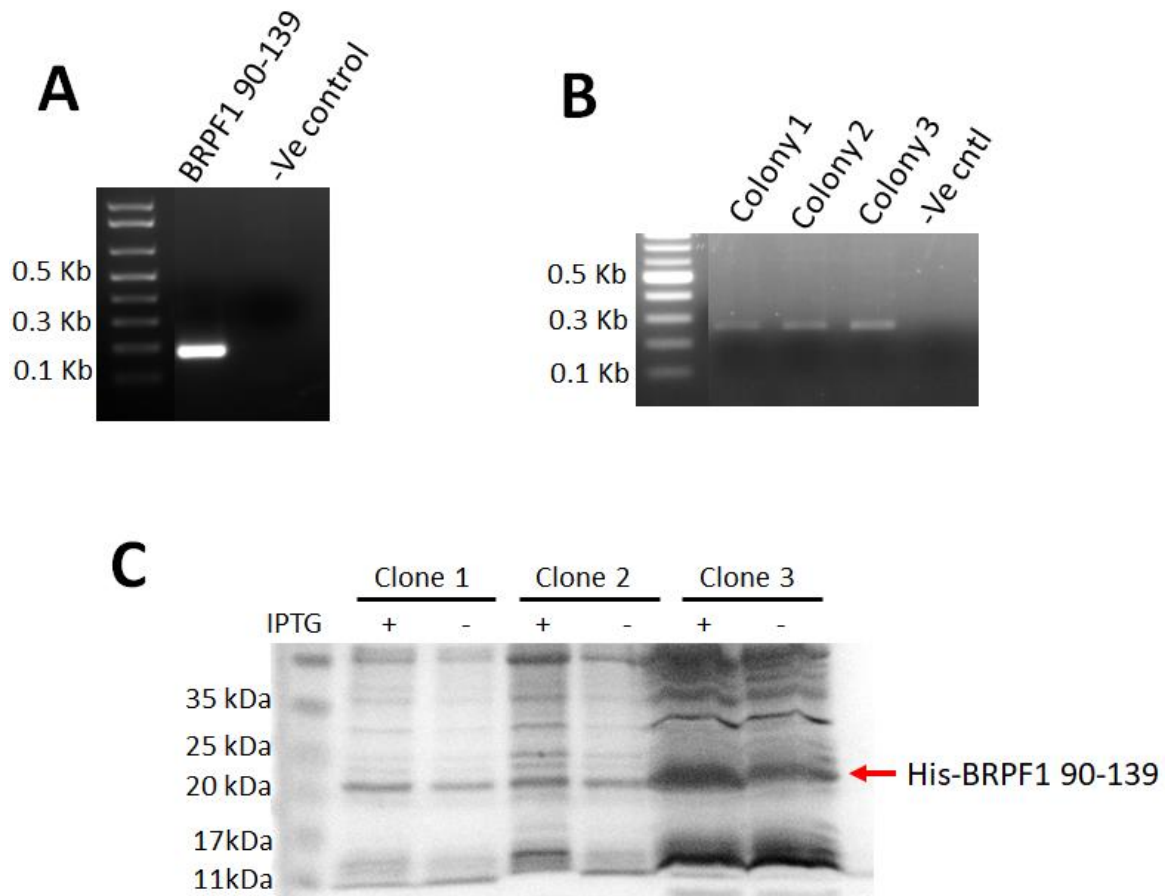


Figure 5. 6. Cloning of BRPF1 90-139 into pACYC Duet-1 vector at MCS-1 and mini scale induction.

A) Agarose gel showing the generation of BRPF1 90-139 insert by PCR. **D)** Confirmation of cloning showing in 1% agarose gel by colony PCR for BRPF1 90-139. **E)** SDS-PAGE gel analysis showing the induction of His-BRPF1 90-139 in presence of IPTG. It was appeared at ~20 kDa.

5.4.3. Cloning of MOZ 676-763 into MCS-2 of pACYC Duet-1 containing 6xHis-BRPF1 90-139

The MOZ mapping results from our previous chapters confirmed that MOZ 676-763 is the minimal sequence required for interaction with BRPF1. Therefore, MOZ 676-763 sequence was attempted to clone into MCS-2 of pACYC Duet-1 vector, which already containing the BRPF1 90-139. The MCS-2 is equipped with a sequence to express S-tag. The vector was digested with NdeI and XhoI restriction enzymes (Figures 5.6 A) and similarly like BRPF1 cloning for MCS-1, MOZ 676-763 was also cloned into the linear vector using HiFi assembly master mix. The clone was confirmed by colony PCR (Figures 5.6 C) and followed by sequencing confirmation.

After confirming positive clone of MOZ 676-763 into MCS-2 of pACYC Duet-1 containing BRPF1 90-139, the vector was transformed into *E. coli* BL21 DE3 for expression. Some clones were selected and induced with 1mM IPTG for mini scale induction of the expression. SDS-PAGE gel analysis with Coomassie brilliant blue stain confirmed the induction of the BRPF1 polypeptide running at 20kDa as before, in addition to a strongly induced band at the expected size (11kDa) for the MOZ peptide (Figures 5.6). We next attempted to further purify the BRPF1 protein using the 6x his tag to determine if it co-purified with the co-expressed MOZ fragment.

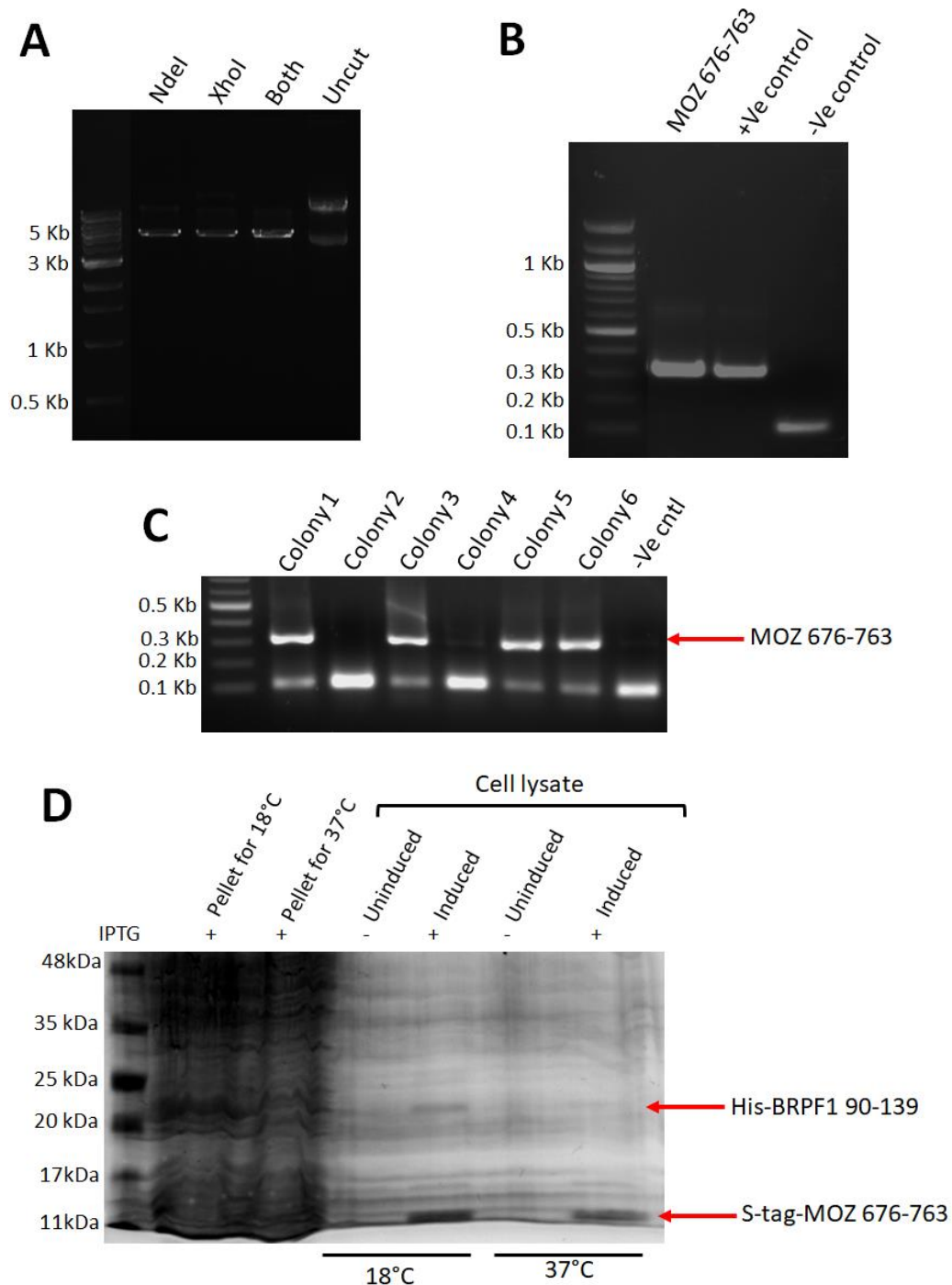


Figure 5. 7. Cloning of MOZ 676-763 into pACYC Duet-1 vector containing BRPF1 90-139 and mini scale induction of both the protein.

A) Agarose gel image showing the digestion of the pACYC Duet-1 vector with NdeI and XhoI restriction enzymes. **B)** Agarose gel showing the generation of MOZ 676-763 insert by PCR. **C)** Confirmation of cloning showing in 1% agarose gel by colony PCR. **D)** SDS-PAGE gel analysis showing the induction of both MOZ 676-763 and BRPF1 90-139 in presence of IPTG. It is also showing that 18°C is better for the expression of both the proteins.

5.5. Purification of His-BRPF1 90-139 complexes

To attempt to confirm expression and association of the MOZ-S tag and 6xHis-BRPF1 polypeptides purification was performed to capture the 6xHis-BRPF1 90-139 using cobalt beads and determine if the MOZ 676-763 would be co-purified indicating complex formation. After confirmation of expression from small scale induction, a glycerol stock from single clone was used to inoculate 200ml LB (containing chloramphenicol 100µg/ml) for medium scale expression and purification. After addition of IPTG the culture was divided into two and one culture allowed to grow at 18°C and the other at 37°C for overnight, to assess the effect of induction temperature. Then the His-BRPF1 90-139 along with MOZ 676-763 was purified using cobalt beads (Figures 5.7 A). It was found that after adding 1mM IPTG incubation at 18°C helps better expression of both the proteins. The recombinant protein complex was eluted from the beads. This purified protein complex was then subjected to SDS-PAGE for western blot.

The western blot was performed using anti-histidine pentapeptide antibody to confirm the expression of His-BRPF1 90-139 as we are purifying the complex with this protein. As shown in figure 5.7 B a 20kDa band was detected for the eluted protein only after IPTG induction, which was consistent with the inductions, although larger than the predicted size of 6xhis BRPF1 90-139. This band was also detected on beads used with induced, but not uninduced lysates, indicated that the elution was not complete and there were still some proteins left on the beads. A longer elution might help to elute the protein completely. An attempt to detect S-tagged MOZ 676-763, unfortunately failed, thus it is not clear if we have successfully expressed the MOZ polypeptide in these assays. The antibody for the S-tag which was available

was mentioned by the manufacturer for the detection for immune-flow cytometry but there was no mention about western blot.

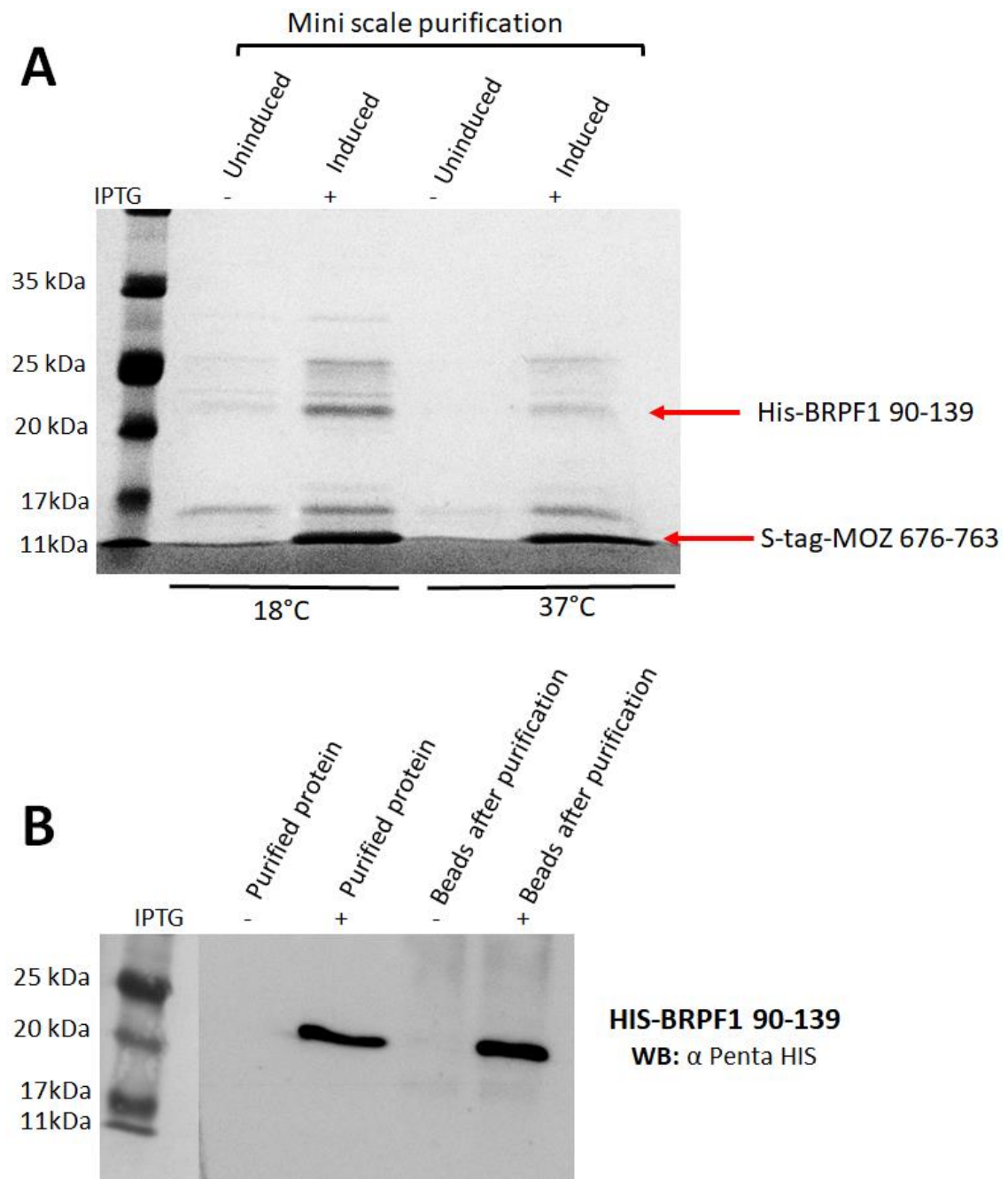


Figure 5. 8. Purification of MOZ 676-763 and BRPF1 90-139 by His-tag purification and western blot confirmation of His-BRPF1 90-139.

A) SDS-PAGE gel analysis by coomassie brilliant blue staining showing the His-tag purification of the MOZ676-763 and BRPF1 90-139 complex by using cobalt beads. This result also indicated that 18°C temperature is better for the expression of the complex. **B)** Western blot result showing the confirmation of the His-BRPF1 90-139 in the purified protein. It is also showing that there is still some amount of protein was not eluted and remain on the beads.

5.6. Summary

In order to explore the functional consequences of disrupting the MOZ/BRPF1 association in cells, we performed rescue experiments utilising a CRISPR generated MOZ deficient line of HEK293 cells (G10) and a control line AF10 that was targeted with a single sgRNA, that will not induce indels when used with CAS9 nickase. The AF10 clone was validated as lacking any change in sequence or expression of KAT6A/MOZ. Previous work on these and other CRISPR/Cas9 edited MOZ knockout cells from our group (Monteiro *et al.*, unpublished) had shown that MOZ depletion leads to reductions in global H3K9 and H3K14 acetylation, hampers cell proliferation and has a significant impact on cell transcriptomes. Using RNAseq profiling we identified genes that are differentially expressed in MOZ-deficient HEK293 G10 cells, These included many transcription factors or other transcriptional regulators which are likely to have downstream effects on other genes, and thus DEGs identified in our studies may be direct or indirect MOZ targets. Due to the proposed role of BRPF1 in gene repression (Yan *et al.*, 2017) we focussed on candidate genes that were upregulated in the G10 clone compared to AF10, and confirmed this by RTqPCR. Rescue experiments using YFP-FLAG-MOZ were performed which was found to restore repression of these genes, at least at the level of transcription, and these results are consistent with similar data assessing MOZ regulated genes in K562 erythroleukemia cells (Monteiro *et al.*, in preparation). We noted a small difference between wild type MOZ and the BRPF1 binding-deficient mutant MOZ in terms of the degree of ability to restore the expression patterns of five candidate genes. This indicates that there might be a role for BRPF1 in complex with MOZ to control their expression, at least in part.

The five MOZ target genes selected for analysis were primarily chosen for their robust expression in response to MOZ depletion in the HEK293 G10 clone. SPARC encodes secreted protein acidic and cysteine rich and has been implicated as having raised expression in AML and other tumour types. SLC7A3 encodes a sodium-independent cationic amino acid transporter associated with autism spectrum disorders. Runx 3 is a member of the Runx family of transcription factors (also termed AML1, 2, 3). These proteins are known to associate directly with MOZ, which acts as a cofactor, and regulated Runx expression is important in T-cell function and development. TIMP1 is a metalloproteinase inhibitor implicated as a prognostic factor in leukaemia and erythroid cell development, and HMOX1 is a heme oxygenase functioning in heme metabolism, erythroid development and neural function. However, further studies will be required to also look at rescue of genes that are activated by MOZ, such as the HOX cluster, which is known to be regulated by MOZ.

Efforts to co-purify bacterially expressed constructs corresponding to MOZ 676-763 and BRPF1 90-139 for crystallisation studies, were successful in detecting 6xHis tagged BRPF1 90-139, which corresponds to the minimal BRPF2 sequence needed for interaction with HBO1 (Tao *et al.*, 2017). However, the BRPF1 construct in our study was shown to run at higher than the expected molecular size. As the sequence of the construct was confirmed, we propose that this is likely due to the very high ratio of acidic amino acid residue presence in BRPF1 90-139 sequence with the total amino acid number of protein polypeptide fragment. The amino acid content of this polypeptide is almost 25% glutamic acid (E) and aspartic acid (D) residues, in comparison to 10% acidic content of the MOZ construct, which runs at the expected size. It is known that high content of acidic residues can polypeptides appear at

higher than expected apparent molecular weight in SDS-PAGE gels (Raceffa *et al.*, 1992 and Alves *et al.*, 2004). The comparative acidic amino acid content of the constructs is highlighted in Figure 5.8.

An induced band at the expected size of 11kDa for MOZ-676-763 was clearly visible in the *E.coli* cell-free extracts post-induction, which co-purified with the His-BRPF1 polypeptide. However, it remains to be confirmed with western detection of the S-tag that this is tag of the MOZ polypeptide. Once this is confirmed future work can optimise the expression and purification conditions to achieve sufficient quantities to attempt to crystallise the complex for structure determination.

BRPF1 90-139

90- MSYAQAQRMV E V D LHGRVHRISIF D N L D
 VVS E D E E AP E E AP E NGSNK E NT -139

MOZ 676-763

676- AGSP E KPLS D LGRLSYMAYWKSVIL E CLYHQ
 N D KQISIKKLSKLTGICPQ D ITSTLHHLRML D
 FR S D QFVIIR R E KLIQ D HMAKLQLN -763

Figure 5.8. Comparison of amino acid chains for MOZ 676-763 and BRPF1 90-139.

A comparison of the ratio of acidic amino acids such as glutamic acids and aspartic acids in MOZ 676-763 and BRPF1 90-139 sequences. The acidic residues were marked with red and underlined. It was seen that BRPF1 90-139 has very high ratio of acidic amino acids with its total amino acid chain, whereas MOZ 676-763 has very low ratio with its polypeptide chain.

Chapter 6: Concluding Remarks

MOZ is a histone acetyltransferase, which functions as an essential coactivator for the genes that are involved in haematopoietic and developmental process. For being functionally active, MOZ often found as a catalytic subunit in several stable and nuclear multi-subunit complexes (Kitabayashi *et al.*, 2001; Ullah *et al.*, 2008; Mishima *et al.*, 2011). After being identified with various interacting partner, it was found that MOZ/BRPF1 complex is one of the important complexes for stimulating the acetyltransferase domain activity in acetylation of nucleosomal histones (Ullah *et al.*, 2008). BRPF1 is a bromodomain containing protein, which plays crucial role in epigenetic regulation of gene expression and developmental process as well. In this complex BRPF1 acts as a bridging protein for other subunits attachment to the complex. Therefore, the interface of the MOZ/BRPF1 complex is very crucial subject to explore. As we discussed before different groups had mapped the interface and there is a clear contradiction with the mapping reported so far. All the previous mapping was performed by using co-immunoprecipitation but there could be several indirect interaction in co-immunoprecipitation. Therefore we were needed to map the interface using some direct interaction method such as yeast-two-hybrid.

To map the MOZ/BRPF1 complex interface, we started with confirmation of the interaction of MOZ and BRPF1 *in cellulo*. Sub-cellular localisation confirmed the complex formation very clear and straight forward way. We saw that MOZ is nuclear protein, whereas BRPF1 is cytoplasmic when it expressed alone. During co-expression of MOZ and BRPF1 together, BRPF1 was found in the nucleus along with MOZ. This findings suggested that MOZ form a complex with BRPF1 and moved BRPF1 from cytoplasm to nucleus, but BRPF1 without interacting with MOZ remain in cytoplasm. This result of co-localisation of MOZ and BRPF1 *in vivo* also confirmed

by Laue *et al.*, (2008). In our preliminary result from yeast-two-hybrid reporter assay we identified N-terminal BRPF1 (BRPF1 59-222) containing the BN1 domain with adjacent small part of BN2 domain is sufficient to interact with the MYST domain of MOZ. This result remain consistence with Ullah *et al.* (2008). Further mapping of MOZ MYST domain revealed that minimal BRPF1 interacting site is situated at the helix-turn-helix motif of MYST domain, which was MOZ 676-763. For interacting with BRPF1 MOZ required this whole helix-turn-helix motif and any further shortening disturb the interaction. Co-immunoprecipitation confirmation of the interaction also suggested that N-terminal MOZ (up to MYST domain) MOZ 1-779 contains the BRPF1 interacting site.

After identifying the minimal region of MOZ for interacting with BRPF1, we were investigating whether that region is essential or not for recruiting BRPF1. To achieve that we incorporated eight point mutations within that helix-turn-helix region. A series of β -galactosidase reporter assay from yeast-two-hybrid system revealed that out of eight point mutations five of them completely disrupted the interaction with BRPF1. Among these five point mutations majority of them located at the loop region after second β -sheet of the helix-turn-helix motif. Though Holbert *et al.*, (2007) reported the MYST domain crystal structure of MOZ but this loop region of the helix-turn-helix motif does not have any structure due to lack of electron density. However, our site-directed mutagenesis study in yeast-two-hybrid system suggested that this unstructured loop region is very important for MOZ/BRPF1 interaction. Our overall aim also fulfilled from this result, which is the MOZ 676-763 is essential for BRPF1 recruitment. Interestingly, during this study of site-directed mutagenesis, Tao *et al.*, (2017) reported the crystal structure of HBO1/BRPF2 complex. HBO1 MYST domain

and BRPF2 both are highly similar according to their sequence and our findings for the MOZ/BRPF1 interface mapping and correspondence sequence for HBO1/BRPF2 complex interface reported by Tao *et al.*, (2017) are very consistence.

Mutation at the BRPF1 interacting site had disrupted the interaction between MOZ and BRPF1 in yeast but now we need to check how this mutations affect *in vivo*. The initial co-immunoprecipitation result failed to show any disruption of the interaction when mutated MOZ and wild type full length BRPF1 co-expressed. Then we further deleted the whole BRPF1 interacting loop region (735-750) of MOZ and it still failed to stop the interaction. However, when sub-cellular localisation of both the proteins was checked, it was found that mutation (deletion of 735-750) at the BRPF1 interaction site of the MOZ had shown disruption of the interaction and due to that disruption BRPF1 remain in the cytoplasm even though MOZ (mutated) was present in the nucleus. This result suggested that when both MOZ and BRPF1 released in a cell free lysate, there is an indirect interaction takes place due to their interaction with histone tails. Both the proteins consist with histone interacting domains such as DPF domain of MOZ or PZP, bromo or PWWP domain of BRPF1 and their interaction with histone attached with chromatin providing this indirect interaction and if we could stop this interaction with histones we may see the disruption of the interaction between MOZ and BRPF1 due to mutation.

In collaboration with an industrial partner we have explored for specific BRPF1 bromodomain inhibitor to inhibit the interaction between histone tails and bromodomain. During this study we reported that BRPF1 has histone specificity towards histone H4K12ac and the novel inhibitor GSK6835A can completely inhibit the interaction between BRPF1 bromodomain and histone H4K12ac. Taking this

result in account, we treated our cells with this novel GSK6835A inhibitor before and after the transfection. There were a clear drop in intensity of the band detected in co-IP for the mutation. This findings supporting the fact that there are indirect interaction as well present and that is why BRPF1 was still detected even though MOZ was mutated. If we could block all the histone interacting domains from both the proteins that could lead to a complete disruption of interaction in co-IP due to mutation of BRPF1 interacting site of MOZ.

After identifying the minimal region for both interaction proteins from MOZ/BRPF1 complex, we also manage to co-express MOZ 676-763 and BRPF1 90-139 (this boundary was taken from the correspondence sequence reported by Tao *et al.*, 2017) in a dual expression vector as a complex form. We also have managed to purify the complex from the medium scale bacterial culture using His-tag purification.

Apart from interface mapping we had explored the functional part of the MOZ. MOZ knockout cell line was made using CRISPR/Cas9 technology. From the RNA sequencing data we found several gene expressions are regulated by MOZ. Some of those gene were chosen, to whom MOZ works as suppressor. We tried to rescue the MOZ knockout HEK293 cells by exogenous expression of MOZ as well as mutated MOZ (at BRPF1 interacting site). The result showed that exogenous MOZ can suppress the expression of those selected genes' expression. It was noticed that mutated MOZ was slightly less effective in terms of the rescue the gene expression. This findings suggest that there may be a role for BRPF1 on the gene expression controlling activity of MOZ. MOZ needs the complex formation with BRPF1 to control the gene expression properly.

Appendix

Primer Name	Sequences	Restriction site	Purpose
BRPF1 1F	aaaactcgagatgggggtggactttgatgtg	XhoI	Mapping
BRPF1 222R	aaaaccgcggtcagtagtcctcctcgctcatgtc	sacII	Mapping
BRPF1 59F	aaaactcgagcgcaagcacaagaaaaagggg	XhoI	Mapping
BRPF1 117R	aaaaccgcggccaatccaggtgtcaaagatgct	sacII	Mapping
BRPF1 190R	aaaaccgcggccaatcaggggtgtcctgttccag	sacII	Mapping
BRPF1 117F	aaaactcgaggatgtggtgtcagaggatgag	XhoI	Mapping
BRPF1 190F	aaaactcgaggatgccccaccccgccaact	XhoI	Mapping
BRPF1 90F	aaaactcgagatgagctatgcacaggcc	XhoI	Crystallisation
BRPF1 139R	aaaaccgcggccaagtgttctccttggtgct	sacII	Crystallisation

Primer name	Sequences	Restriction site	Purpose
MOZ 478F	aaaaccgcggactgagaaagatatggaa	sacII	Mapping
MOZ 498F	aaaaccgcgggtgactgggtcccctgat	sacII	Mapping
MOZ 561F	aaaaccgcggggttggtccatcctcct	sacII	Mapping
MOZ 586F	aaaaccgcggagtaccatttattgtcaa	sacII	Mapping
MOZ 603F	aaaaccgcggcacaaaaccctctattac	sacII	Mapping
MOZ 675R	aaaaggatcctcattggccttcacgcttga	BamHI	Mapping
MOZ 676F	aaaaccgcgggcagggctccagagaaa	sacII	Mapping
MOZ 763R	aaaaggatcctcaaagctttgcattcagctg	BamHI	Mapping
MOZ 806R	aaaaggatcctcattcctggcactgtggctc	BamHI	Mapping
MOZ 764F	aaaaccgcggttgcgacctgtagatgta	sacII	Mapping
MOZ 807F	aaaaccgcggagagaattagagatcagt	sacII	Mapping
MOZ 779R	aaaaggatcctcagactggagtccagcgcaa	BamHI	Mapping
MOZ 707F	aaaaccgcggaatgacaagcagatcagc	sacII	Mapping
MOZ 723F	aaaaccgcggtgccctcaagacatcact	sacII	Mapping
MOZ 736F	aaaaccgcggatgctggacttccgtagt	sacII	Mapping
MOZ 750F	aaaaccgcgggaaaaacttatccaggat	sacII	Mapping
MOZ 707R	aaaaggatcctcaattttgtgtataaaggca	BamHI	Mapping

MOZ 721R	aaaaggatcctcatccagtcaacttgcttaa	BamHI	Mapping
MOZ 735R	aaaaggatcctcatcgtaggtggtggagtgt	BamHI	Mapping
MOZ 749R	aaaaggatcctcaccggcggataatcacaaa	BamHI	Mapping
MOZ 676F	aaaacatatggcagggctccagagaaa	NdeI	Crystallisation
MOZ 763R	aaaactcgagtcaaagcttgcattcagctg	XhoI	Crystallisation
A693P Fwd	T CGT CTT TCC TAC ATG CCA TAT TGG AAA AGT GTA A	N/A	Mutation
A693P Rvs	T TAC ACT TTT CCA ATA TGG CAT GTA GGA AAG ACG A	N/A	Mutation
Q706P Fwd	G GAG TGC CTT TAT CAC CCA AAT GAC AAG CAG ATC A	N/A	Mutation
Q706P Rvs	T GAT CTG CTT GTC ATT TGG GTG ATA AAG GCA CTC C	N/A	Mutation
K718P Fwd	C ATT AAG AAG TTA AGC CCG TTG ACT GGA ATC TGC C	N/A	Mutation
K718P Rvs	G GCA GAT TCC AGT CAA CGG GCT TAA CTT CTT AAT G	N/A	Mutation
F739P Fwd	C CTA CGA ATG CTG GAC CCC CGT AGT GAC CAA TTT G	N/A	Mutation
F739P Rvs	C AAA TTG GTC ACT ACG GGG GTC CAG CAT TCG TAG G	N/A	Mutation
D742P Fwd	G CTG GAC TTC CGT AGT CCC CAA TTT GTG ATT ATC C	N/A	Mutation
D742P Rvs	G GAT AAT CAC AAA TTG GGG ACT ACG GAA GTC CAG C	N/A	Mutation
Q743P Fwd	G GAC TTC CGT AGT GAC CCA TTT GTG ATT ATC CGC C	N/A	Mutation
Q743P Rvs	G GCG GAT AAT CAC AAA TGG GTC ACT ACG GAA GTC C	N/A	Mutation
V745P Fwd	C CGT AGT GAC CAA TTT CCG ATT ATC CGC CGG GAA A	N/A	Mutation
V745P Rvs	T TTC CCG GCG GAT AAT CGG AAA TTG GTC ACT ACG G	N/A	Mutation

A758P Fwd	T ATC CAG GAT CAC ATG C CA AAG CTT CAG CTG AAT	N/A	Mutation
A758P Rvs	ATT CAG CTG AAG CTT TGG CAT GTG ATC CTG GAT A	N/A	Mutation
MOZ Δβ Fw	ACA CTC CAC CAC CTA CGA GAA AAA CTT ATC CAG GAT	N/A	Mutation
MOZ Δβ Rv	TCG TAG GTG GTG GAG TGT 'ATC CTG GAT AAG TTT TTC	N/A	Mutation
MORF 541F	aaaaccgcgggcagggctcctggaaag	sacII	Mapping
MORF 626R	aaaaggatcctcatttcagctttccatgtg	BamHI	Mapping

References

- Aguilera, A. 2002. The connection between transcription and genomic instability. *EMBO J.* 21:195-201.
- Allis, C.D., S.L. Berger, J. Cote, S. Dent, T. Jenuwien, T. Kouzarides, L. Pillus, D. Reinberg, Y. Shi, R. Shiekhatar, A. Shilatifard, J. Workman, and Y. Zhang. 2007. New nomenclature for chromatin-modifying enzymes. *Cell.* 131:633-6.
- Alves, V. S., Pimenta, D. C., Sattlegger, E., & Castilho, B. A. (2004). Biophysical characterization of Gir2, a highly acidic protein of *Saccharomyces cerevisiae* with anomalous electrophoretic behavior. *Biochemical and biophysical research communications*, 314(1), 229-234
- Baell, Jonathan B, David J Leaver, Stefan J Hermans, Gemma L Kelly, Margs S Brennan, Natalie L Downer, Nghi Nguyen, et al. 2018. "Inhibitors of Histone Acetyltransferases KAT6A/B Induce Senescence and Arrest Tumour Growth." *Nature* 560 (7717): 253–57.
- Baldwin, T. & Lapointe, M. (2003). The Chemistry of Amino Acids. The Biology Project [Online], Available: <http://www.biology.arizona.edu/biochemistry/>
- Bamborough, Paul, Heather A. Barnett, Isabelle Becher, Mark J. Bird, Chun Wa Chung, Peter D. Craggs, Emmanuel H. Demont, et al. 2016. "GSK6853, a Chemical Probe for Inhibition of the BRPF1 Bromodomain." *ACS Medicinal Chemistry Letters* 7 (6): 552–57
- Bannister, A.J., and T. Kouzarides. 2005. Reversing histone methylation. *Nature.* 436:1103-6.
- Bannister, A.J., and T. Kouzarides. 2011. Regulation of chromatin by histone modifications. *Cell Research.* 21:381–395.

- Bedford, M.T., and S. Richard. 2005. Arginine methylation an emerging regulator of protein function. *Mol Cell*. 18:263-72.
- Berndsen, C.E., B.N. Albaugh, S. Tan, and J.M. Denu. 2007. Catalytic mechanism of a MYST family histone acetyltransferase. *Biochemistry*. 46:623-9.
- Borrow, J., V.P. Stanton, Jr., J.M. Andresen, R. Becher, F.G. Behm, R.S. Chaganti, C.I. Civin, C. Disteché, I. Dube, A.M. Frischau, D. Horsman, F. Mitelman, S. Volinia, A.E. Watmore, and D.E. Housman. 1996. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat Genet*. 14:33-41.
- Champagne KS, Saksouk N, Pena PV, Johnson K, Ullah M, Yang XJ, Cote J, Kutateladze TG. The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. *Proteins*. 2008; 72:1371–6.
- Chiang, T.W., et al., CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing. *Sci Rep*, 2016. 6: p. 24356.
- Chrivia, J.C., R.P. Kwok, N. Lamb, M. Hagiwara, M.R. Montminy, and R.H. Goodman. 1993. Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature*. 365:855-9.
- Choudhary, C., C. Kumar, F. Gnad, M.L. Nielsen, M. Rehman, T.C. Walther, J.V. Olsen, and M. Mann. 2009. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. 325:834-40.
- Clapier, C.R., and B.R. Cairns. 2009. The biology of chromatin remodelling complexes. *Annu Rev Biochem*. 78:273-304.

- Collins, H.M., K.B. Kindle, S. Matsuda, C. Ryan, P.J. Troke, E. Kalkhoven, and D.M. Heery. 2006. MOZ-TIF2 alters cofactor recruitment and histone modification at the RARbeta2 promoter: differential effects of MOZ fusion proteins on CBP- and MOZ-dependent activators. *J Biol Chem.* 281:17124-33.
- Core, L.J., and J.T. Lis. 2008. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science.* 319:1791-2.
- D. Parchaliuk, 1999 Online Yeast-2 Hybrid Method.
- Dawson, M.A., A.J. Bannister, B. Gottgens, S.D. Foster, T. Bartke, A.R. Green, and T. Kouzarides. 2009. JAK2 phosphorylates histone H3Y41 and excludes HP1alpha from chromatin. *Nature.* 461:819-22.
- Demont, E. H., Bamborough, P., Chung, C., Craggs, P. D., Fallon, D., Gordon, L. J., Watson, R. J. (2014). 1,3-Dimethyl Benzimidazolones Are Potent, Selective Inhibitors of the BRPF1 Bromodomain. *ACS Medicinal Chemistry Letters*, 5, 1190–1195.
- Deguchi, K., P.M. Ayton, M. Carapeti, J.L. Kutok, C.S. Snyder, I.R. Williams, N.C. Cross, C.K. Glass, M.L. Cleary, and D.G. Gilliland. 2003. MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. *Cancer Cell.* 3:259-71.
- Dilworth, F.J., K.J. Seaver, A.L. Fishburn, S.L. Htet, and S.J. Tapscott. 2004. *In vitro* transcription system delineates the distinct roles of the co-activators pCAF and p300 during MyoD/E47-dependent transactivation. *Proc Natl Acad Sci U S A.* 101:11593-8.

- Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., T.J. Richmond. 2004. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*. 306:1571-3.
- Doyon, Y., C. Cayrou, M. Ullah, A.J. Landry, V. Cote, W. Selleck, W.S. Lane, S. Tan, X.J. Yang, and J. Cote. 2006. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol Cell*. 21:51-64.
- Dreveny, I., Deeves, S. E., Fulton, J., Yue, B., Messmer, M., Bhattacharya, A., Collin, H.M., Heery, D. M. 2014. The double PHD finger domain of MOZ/MYST3 induces α -helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. *Nucleic Acids Research*, 42(2), 822–835
- Eckner, R., M.E. Ewen, D. Newsome, M. Gerdes, J.A. DeCaprio, J.B. Lawrence, and D.M. Livingston. 1994. Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes Dev*. 8:869-84.
- Ferreira, H., J. Somers, R. Webster, A. Flaus, and T. Owen-Hughes. 2007. Histone tails and the H3 alphaN helix regulate nucleosome mobility and stability. *Mol Cell Biol*. 27:4037-48.
- Fuda, N.J., M.B. Ardehali, and J.T. Lis. 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*. 461:186-92.
- Grossman, S.R., M. Perez, A.L. Kung, M. Joseph, C. Mansur, Z.X. Xiao, S. Kumar, P.M. Howley, and D.M. Livingston. 1998. p300/MDM2 complexes participate in MDM2-mediated p53 degradation. *Mol Cell*. 2:405-15.

- Guelman, S., T. Suganuma, L. Florens, S.K. Swanson, C.L. Kiesecker, T. Kusch, S. Anderson, J.R. Yates, 3rd, M.P. Washburn, S.M. Abmayr, and J.L. Workman. 2006. Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in *Drosophila*. *Mol Cell Biol*. 26:871-82.
- Hahn, S. 2004. Structure and Mechanism of RNA Polymerase II Transcription Machinery. *Nature Structural & Molecular Biology*. 11:394-403.
- Hassa, P.O., S.S. Haenni, M. Elser, and M.O. Hottiger. 2006. Nuclear ADPribosylation reactions in mammalian cells: where are we today and where are we going? *Microbiol Mol Biol Rev*. 70:789-829.
- Hassan, Y.I., and J. Zemleni. 2008. A novel, enigmatic histone modification: biotinylation of histones by holocarboxylase synthetase. *Nutr Rev*. 66:721-5.
- Heery, D.M., and P.M. Fischer. 2007. Pharmacological targeting of lysine acetyltransferases in human disease: a progress report. *Drug Discov Today*. 12:88-99.
- Holbert, M.A., T. Sikorski, J. Carten, D. Snowflack, S. Hodawadekar, and R. Marmorstein. 2007. The human monocytic leukemia zinc finger histone acetyltransferase domain contains DNA-binding activity implicated in chromatin targeting. *J Biol Chem*. 282:36603-13.
- Horvath, P. and R. Barrangou, CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 2010. 327(5962): p. 167-70.
- Hsu, P.D., E.S. Lander, and F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 2014. 157(6): p. 1262-78.

- Imhof, A., X.J. Yang, V.V. Ogryzko, Y. Nakatani, A.P. Wolffe, and H. Ge. 1997. Acetylation of general transcription factors by histone acetyltransferases. *Curr Biol.* 7:689-92.
- Illingworth, R.S., and A.P. Bird. 2009. CpG islands--'a rough guide'. *FEBS Lett.* 583:1713-20
- Kindle, K.B., H.M. Collins, and D.M. Heery. 2010. MOZ-TIF2-mediated destruction of CBP/p300 is blocked by calpain inhibitor 2. *Leukemia.* 24:1359-61.
- Kindle, K.B., P.J. Troke, H.M. Collins, S. Matsuda, D. Bossi, C. Bellodi, E. Kalkhoven, P. Salomoni, P.G. Pelicci, S. Minucci, and D.M. Heery. 2005. MOZ-TIF2 inhibits transcription by nuclear receptors and p53 by impairment of CBP function. *Mol Cell Biol.* 25:988-1002.
- Kitabayashi, I., Y. Aikawa, L.A. Nguyen, A. Yokoyama, and M. Ohki. 2001a. Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. *EMBO J.* 20:7184-96.
- Kitabayashi, I., Y. Aikawa, A. Yokoyama, F. Hosoda, M. Nagai, N. Kakazu, T. Abe, and M. Ohki. 2001b. Fusion of MOZ and p300 histone acetyltransferases in acute monocytic leukemia with a t(8;22)(p11;q13) chromosome translocation. *Leukemia.* 15:89-94.
- Klein, B.J., Muthurajan, U.M., Lalonde, M.E., Gibson, M.D., Andrews, F.H., Hepler, M., Machida, S., Yan, K., Kurumizaka, H., Poirier, M.G., et al. (2016). Bivalent interaction of the PZP domain of BRPF1 with the nucleosome impacts chromatin dynamics and acetylation. *Nucleic Acids Res.* 44, 472–484.
- Kornberg, R.D. 2005. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci.* 30:235-9.

- Kornberg, R.D. 2007. The molecular basis of eucaryotic transcription. *Cell Death Differ.* 14:1989-97.
- Komander, D. 2009. The emerging complexity of protein ubiquitination. *Biochem Soc Trans.* 37:937-53.
- Kouzarides, T. 2007. Chromatin modifications and their function. *Cell.* 128:693-705.
- Lalonde, M. E., Avvakumov, N., Glass, K. C., Joncas, F. H., Saksouk, N., Holliday, M., Côté, J. 2013. Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. *Genes and Development*, 27(18), 2009–2024.
- Laue, K., Daujat, S., Crump, J. G., Plaster, N., Roehl, H. H., Kimmel, C. B., Hammerschmidt, M. 2008. The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity. *Development*, 135(11), 1935–1946
- Le Douarin, B., Heery, D.M., Gaudon, C., Baur, E., and Losson, L. Yeast Two-Hybrid Screening for Proteins that Interact with Nuclear Hormone Receptors. *Methods in Molecular Biology*, vol. 176, 227-248
- Lee, K.K., and J.L. Workman. 2007. Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol.* 8:284-95.
- Li, B., M. Carey, and J.L. Workman. 2007. The role of chromatin during transcription. *Cell.* 128:707-19.
- Liu, L., Qin, S., Zhang, J., Ji, P., Shi, Y., and Wu, J. (2012). Solution structure of an atypical PHD finger in BRPF2 and its interaction with DNA. *J. Struct. Biol.* 180, 165–173.

- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 18;389(6648):251-60.
- Luger, K. 2003. Structure and dynamic behavior of nucleosomes. *Curr Opin Genet Dev*. 13:127-35.
- Lubula, M. Y., Eckenroth, B. E., Carlson, S., Poplawski, A., Chruszcz, M., & Glass, K. C. 2014. Structural insights into recognition of acetylated histone ligands by the BRPF1 bromodomain. *FEBS Letters*, 588(21), 3844–3854.
- Mali, P., *et al.*, RNA-guided human genome engineering via Cas9. *Science*, 2013. 339(6121): p. 823-6.
- Martin, C., and Y. Zhang. 2005. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol*. 6:838-49.
- Mattioli, F., Schaefer, E., Magee, A., Mark, P., Mancini, G. M., Dieterich, K., Von Allmen, G., Alders, M., Coutton, C., van Slegtenhorst, M., Vieville, G., Engelen, M., Cobben, J. M., Juusola, J., Pujol, A., Mandel, J. L., ... Piton, A. (2016). Mutations in Histone Acetylase Modifier BRPF1 Cause an Autosomal-Dominant Form of Intellectual Disability with Associated Ptosis. *American journal of human genetics*, 100(1), 105-116.
- McCullough, C.E., Song, S., Shin, M.H., Johnson, F.B., Marmorstein, R. (2016) Structural and Functional Role of Acetyltransferase hMOF K274 Autoacetylation. *J.Biol.Chem*. 291: 18190-18198
- Metzger, E., A. Imhof, D. Patel, P. Kahl, K. Hoffmeyer, N. Friedrichs, J.M. Muller, H. Greschik, J. Kirfel, S. Ji, N. Kunowska, C. Beisenherz-Huss, T. Gunther, R. Buettner, and R. Schule. 2010. Phosphorylation of histone H3T6 by PKCβ(I) controls demethylation at histone H3K4. *Nature*. 464:792-6.

- Meulmeester, E., and F. Melchior. 2008. Cell biology: SUMO. *Nature*. 452:709-11.
- Mizuguchi, G., X. Shen, J. Landry, W.H. Wu, S. Sen, and C. Wu. 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodelling complex. *Science*. 303:343-8.
- Musselman, C.A., Lalonde, M.E., Coˆte´, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* 19, 1218–1227.
- Nathan, D., K. Ingvarsdottir, D.E. Sterner, G.R. Bylebyl, M. Dokmanovic, J.A. Dorsey, K.A. Whelan, M. Krsmanovic, W.S. Lane, P.B. Meluh, E.S. Johnson, and S.L. Berger. 2006. Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Genes Dev.* 20:966-76.
- Nelson, C.J., H. Santos-Rosa, and T. Kouzarides. 2006. Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell*. 126:905-16.
- Palmer, W. S., Poncet-Montange, G., Liu, G., Petrocchi, A., Reyna, N., Subramanian, G., Andersen, J. N. 2015. Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *Journal of Medicinal Chemistry*, 150610165235000
- Park, Y.J., J.V. Chodaparambil, Y. Bao, S.J. McBryant, and K. Luger. 2005. Nucleosome assembly protein 1 exchanges histone H2A-H2B dimers and assists nucleosome sliding. *J Biol Chem*. 280:1817-25.
- Park, Y.J., and K. Luger. 2008. Histone chaperones in nucleosome eviction and histone exchange. *Curr Opin Struct Biol*. 18:282-9.

- Perez, R. B., Tischler, A., Auton, M., & Whitten, S. T. 2014. Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins. *Proteins*, 82(12), 3373-84.
- Peterson, C.L., and M.A. Laniel. 2004. Histones and histone modifications. *Curr Biol*. 14:R546-51.
- Poplawski, A., Hu, K., Lee, W., Natesan, S., Peng, D., Carlson, S., Glass, K. C. 2014. Molecular insights into the recognition of N-terminal histone modifications by the BRPF1 bromodomain. *Journal of Molecular Biology*, 426(8), 1661–1676.
- Poux, A.N., and R. Marmorstein. 2003. Molecular basis for Gcn5/PCAF histone acetyltransferase selectivity for histone and nonhistone substrates. *Biochemistry*. 42:14366-74.
- Puri, P.L., V. Sartorelli, X.J. Yang, Y. Hamamori, V.V. Ogryzko, B.H. Howard, L. Kedes, J.Y. Wang, A. Graessmann, Y. Nakatani, and M. Levrero. 1997. Differential roles of p300 and PCAF acetyltransferases in muscle differentiation. *Mol Cell*. 1:35-45.
- Qian, C., and M.M. Zhou. 2006. SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol Life Sci*. 63:2755-63.
- Qiu Y, Liu L, Zhao C, Han C, Li F, Zhang J, Wang Y, Li G, Mei Y, Wu M, Wu J, Shi Y. Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev*. 2012; 26:1376–91.
- Raceffa, P., Jancsó, A., & Mabuchi, K. (1992). Modification of acidic residues normalizes sodium dodecyl sulfate-polyacrylamide gel electrophoresis of

caldesmon and other proteins that migrate anomalously. Archives of biochemistry and biophysics, 297(1), 46-51.

- Reinke, H., P.D. Gregory, and W. Horz. 2001. A transient histone hyperacetylation signal marks nucleosomes for remodeling at the PHO8 promoter in vivo. Mol Cell. 7:529-38
- Routh, A., S. Sandin, and D. Rhodes. 2008. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proc Natl Acad Sci U S A. 105:8872-7.
- Ryan, C.M., J.C. Harries, K.B. Kindle, H.M. Collins, and D.M. Heery. 2006. Functional interaction of CREB binding protein (CBP) with nuclear transport proteins and modulation by HDAC inhibitors. Cell Cycle. 5:2146-52.
- Sabari, Benjamin R, Robert G Roeder, C David, and Allis Correspondence. 2015. "Intracellular Crotonyl-CoA Stimulates Transcription through P300-Catalyzed Histone Crotonylation." Molecular Cell 58: 203–15.
- Santos-Rosa, H., A. Kirmizis, C. Nelson, T. Bartke, N. Saksouk, J. Cote, and T. Kouzarides. 2009. Histone H3 tail clipping regulates gene expression. Nat Struct Mol Biol. 16:17-22.
- Sheikh, B. N., Yang, Y., Schreuder, J., Nilsson, S. K., Bilardi, R., Carotta, S., McRae, H. M., Metcalf, D., Voss, A. K., & Thomas, T. 2016. MOZ (KAT6A) is essential for the maintenance of classically defined adult hematopoietic stem cells. Blood, 128(19), 2307-2318.
- Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science. 2006; 311:844–7.

- Sheppard, H. M., Harries, J. C., Hussain, S., Bevan, C., & Heery, D. M. (2001). Analysis of the Steroid Receptor Coactivator 1 (SRC1)-CREB Binding Protein Interaction Interface and Its Importance for the Function of SRC1. *Mol Cell Biol.* 1(1), 39–50.
- Shahbazian, M.D., and M. Grunstein. 2007. Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem.* 76:75-100.
- Shaw, P.E. 2007. Peptidyl-prolyl cis/trans isomerases and transcription: is there a twist in the tail? *EMBO Rep.* 8:40-5.
- Shen, B., et al., Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods*, 2014. 11(4): p. 399-402
- Shima, H., Yamagata, K., Aikawa, Y., Shino, M., Koseki, H., Shimada, H., & Kitabayashi, I. 2014. Bromodomain-PHD finger protein 1 is critical for leukemogenesis associated with MOZ-TIF2 fusion. *International Journal of Hematology*, 99(1), 21–31
- Smith, K.T., and J.L. Workman. 2009. Introducing the acetylome. *Nat Biotechnol.* 27:917-9.
- Sowa, M.E., E.J. Bennett, S.P. Gygi, and J.W. Harper. 2009. Defining the human deubiquitinating enzyme interaction landscape. *Cell.* 138:389-403.
- Svejstrup, J.Q. 2010. The interface between transcription and mechanisms maintaining genome integrity. *Trends Biochem Sci.* 35:333-8.
- Takagi, Y., and R.D. Kornberg. 2006. Mediator as a general transcription factor. *J Biol Chem.* 281:80-9.
- Tan, Minjia, Hao Luo, Sangkyu Lee, Fulai Jin, Jeong Soo Yang, Emilie Montellier, Thierry Buchou, et al. 2011. "Identification of 67 Histone Marks and

Histone Lysine Crotonylation as a New Type of Histone Modification.” *Cell* 146: 1016–28.

- Tao, Y., Zhong, C., Zhu, J., Xu, S., Ding, J. (2017). Structural and mechanistic insights into regulation of HBO1 histone acetyltransferase activity by BRPF2. *Nucleic Acids Res.* 45: 5707-5719
- Taverna, S.D., H. Li, A.J. Ruthenburg, C.D. Allis, and D.J. Patel. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol.* 14:1025-40.
- Thomson, J.P., P.J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A.R. Kerr, A. Deaton, R. Andrews, K.D. James, D.J. Turner, R. Illingworth, and A. Bird. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature.* 464:1082-6.
- Trievel, R.C., J.R. Rojas, D.E. Sterner, R.N. Venkataramani, L. Wang, J. Zhou, C.D. Allis, S.L. Berger, and R. Marmorstein. 1999. Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. *Proc Natl Acad Sci U S A.* 96:8931-6.
- Ullah, M., N. Pelletier, L. Xiao, S.P. Zhao, K. Wang, C. Degerny, S. Tahmasebi, C. Cayrou, Y. Doyon, S.L. Goh, N. Champagne, J. Cote, and X.J. Yang. 2008. Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Mol Cell Biol.* 28:6828-43.
- Verstak, B., Arnot, C. J., & Gay, N. J. 2013. An alanine-to-proline mutation in the BB-loop of TLR3 Toll/IL-1R domain switches signalling adaptor specificity from TRIF to MyD88. *Journal of immunology (Baltimore, Md.: 1950)*, 191(12), 6101-9.

- Verdone, L., E. Agricola, M. Caserta, and E. Di Mauro. 2006. Histone acetylation in gene regulation. *Brief Funct Genomic Proteomic*. 5:209-21.
- Vezzoli, A., Nicolas B., Mark D. A., Stefan M. V. F., Clara M. S., Brynn T. K., Brian J. P. H., Berthold G., and Mark B. 2010. "Molecular Basis of Histone H3K36me3 Recognition by the PWWP Domain of Brpf1." *Nature Structural & Molecular Biology* 17 (April): 617.
- Woodcock, C.L., and R.P. Ghosh. 2010. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol*. 2:a000596.
- Wu, H., Min, J., Dombrovski, L., Loppnau, P., Weigelt, J., Sundstrom, M., Arrowsmith, C.H., Edwards, A.M., Bochkarev, A., Plotnikov, A.N. (To be published). The Crystal Structure of acetyltransferase domain of Human HIV-1 Tat interacting protein in complex with acetylcoenzyme A.
- <http://www.merckmillipore.com/GB/en/life-science-research/genomic-analysis/Epigenetics-and-Nuclear-Function/Histone-Modification/73yb.qB.U9cAAAF0BtA1ITAF,nav>
- Yang, X.J. 2004. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res*. 32:959-76.
- Yan, Y., S. Harper, D.W. Speicher, and R. Marmorstein. 2002. The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. *Nat Struct Biol*. 9:862-9.
- Yan, K., You, L., Degerny, C., Ghorbani, M., Liu, X., Chen, L., Li, L., Miao, D., Yang, X. J. 2015. The Chromatin Regulator BRPF3 Preferentially Activates the HBO1 Acetyltransferase but Is Dispensable for Mouse Development and Survival. *The Journal of biological chemistry*, 291(6), 2647-63.

- Yan, K., Rousseau, J., Littlejohn, R. O., Kiss, C., Lehman, A., Rosenfeld, J. A., Stumpel, C., Stegmann, A., Robak, L., Scaglia, F., Nguyen, T., Fu, H., Ajeawung, N. F., Camurri, M. V., Li, L., Gardham, A., Panis, B., Almannai, M., Sacoto, M., Baskin, B., Ruivenkamp, C., Xia, F., Bi, W., DDD Study, CAUSES Study, Cho, M. T., Potjer, T. P., Santen, G., Parker, M. J., Canham, N., McKinnon, M., Potocki, L., MacKenzie, J. J., Roeder, E. R., Campeau, P. M., ... Yang, X. J. (2016). Mutations in the Chromatin Regulator Gene BRPF1 Cause Syndromic Intellectual Disability and Deficient Histone Acetylation. *American journal of human genetics*, 100(1), 91-104.
- You, L., Yan, K., Zhou, J., Zhao, H., Bertos, N. R., Park, M., Yang, X.-J. 2015. The Lysine Acetyltransferase Activator Brpf1 Governs Dentate Gyrus Development through Neural Stem Cells and Progenitors. *PLOS Genetics*, 11(3)
- Yang, X.J. 2015. MOZ and MORF acetyltransferases: Molecular interaction, animal development and human disease. *Molecular Cell Research*, 1853(8), 1818–1826
- Yang, X.J., and M. Ullah. 2007. MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene*. 26:5408-19.
- Yuan, H., Rossetto, D., Mellert, H., Dang, W., Srinivasan, M., Johnson, J., Hodawadekar, S., Ding, E.C., Speicher, K., Abshiru, N., Perry, R., Wu, J., Yang, C., Zheng, Y.G., Speicher, D.W., Thibault, P., Verreault, A., Johnson, F.B., Berger, S.L., Sternglanz, R., McMahon, S.B., Cote, J., Marmorstein, R. MYST protein acetyltransferase activity requires active site lysine autoacetylation (2011) *Embo J*. 31: 58-70

- Yuan, L.W., and A. Giordano. 2002. Acetyltransferase machinery conserved in p300/CBP-family proteins. *Oncogene*. 21:2253-60.
- Zhovmer, A., V. Oksenysh, and F. Coin. 2010. Two sides of the same coin: TFIIH complexes in transcription and DNA repair. *ScientificWorldJournal*. 10:633- 43.